WILEY-BLACKWELL

RMetS

Thermal Physics of the Atmosphere

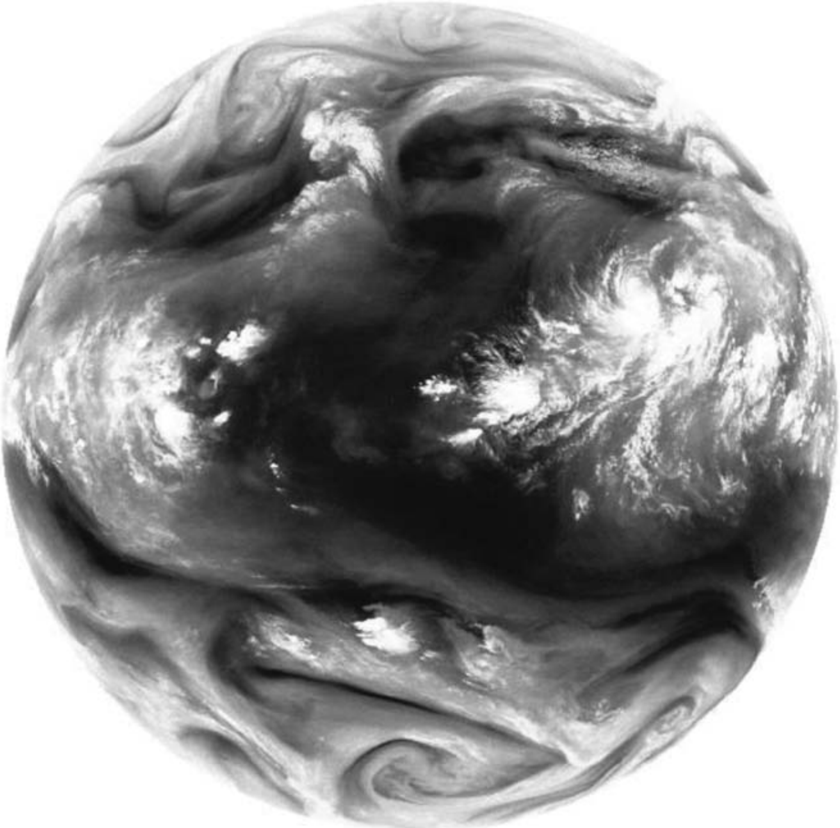
Maarten H. P. Ambaum

Advancing Weather and Climate Science

Thermal Physics
of the Atmosphere

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1



Thermal Physics of the Atmosphere

Maarten H. P. Ambaum
Department of Meteorology
University of Reading

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2010
© 2010 by John Wiley & Sons, Ltd

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Other Editorial Offices:
9600 Garsington Road, Oxford, OX4 2DQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloguing-in-Publication Data

Ambaum, Maarten H. P.
Thermal physics of the atmosphere / Maarten H.P. Ambaum.
p. cm.
ISBN 978-0-470-74515-1 (cloth)
1. Atmospheric thermodynamics. I. Title.
QC880.4.T5A438 2010
551.5'22–dc22
92010005172

ISBN: 978-0 470-74515-1

A catalogue record for this book is available from the British Library.

Set in 10/12 pt Charter by Thomson Digital, Noida, India
Printed in Great Britain by Antony Rowe Ltd., Chippenham, Wilts

First printing 2010

Contents

Preface

This book grew out of the notes for a masters level course in atmospheric physics at the Department of Meteorology at the University of Reading. The course and the book have evolved in parallel with the desire to provide a more explicit and straightforward link between classical thermodynamics, as taught in university physics departments, and thermal physics, as presented in the atmospheric science literature.

In this book we cover the key thermal physics aspects of atmospheric science from a mainstream thermodynamics point of view. It could be argued that there is no such thing as mainstream thermodynamics. Indeed, not many branches of physics invite such heated debates as thermodynamics. Here we ignore those often metaphysical debates and build on classical techniques of thermodynamics as described in many excellent texts. In particular, the text by C. J. Adkins[1] has influenced my view on thermodynamics and its imprint can be found throughout this book.

Atmospheric science is characterized by the inclusion of gravity and vertical structure, the importance of phase transitions in mixtures of air and water, the peculiar physics of droplet formation, the transfer of radiation through deep, inhomogeneous layers of atmosphere, and the lack of thermodynamic equilibrium in the atmosphere as a whole. This pushes the subject well beyond the traditional boundaries of thermodynamics. Hence the use of the words 'thermal physics' in the title, with a nod to the classical text by Kittel and Kroemer.[2]

The book is aimed at graduate, or advanced undergraduate level students. Students must be thoroughly at ease with the use of calculus, in particular partial differentiation, and, in the case of Chapter 10, vector calculus and fluid mechanics. There are two introductory chapters on ideal gases and the laws of thermodynamics but they do not present a comprehensive discussion of the fundamental laws of thermodynamics. There are many excellent textbooks available that cover these subjects in more detail.

The first five chapters contain the fundamentals of atmospheric thermal physics, presented largely as a linear narrative. The final five chapters may be

---

[1] Adkins, C. J. (1983) *Equilibrium Thermodynamics*, 3rd edn. Cambridge University Press, Cambridge.

[2] Kittel, C. & Kroemer, H. (1980) *Thermal Physics*, 2nd edn. W. H. Freeman, New York.

viewed as individual applications, although in each of them additional physics is introduced. Black arrows are used to indicate key equations throughout.

In our case, this text provides the background material for a one trimester masters-level course, but it contains more material than could be covered in such a period. Just by historical precedent, we normally exclude sections 2.4–2.6, 3.4, 3.7, 4.5, 4.6, 5.4, 7.4, 9.7–9.9, and Chapters 8 and 10. The problems at the end of each chapter often invite extensions and elaborations on theoretical ideas in the main text. The problems can be open-ended and may require numerical solution. In our own tutorials we add many hands-on exercises to help students get to grips with the basics and get a feeling for the typical values and ranges of variables.

The depth of the subject of atmospheric thermal physics is only revealed when we start to discover the many links that it provides between apparently distant fields of enquiry. To this end, the book contains a selection of subjects that goes beyond the traditional utilitarian set that is normally associated with atmospheric science. I hope that the reader looks kindly on this, perhaps occasionally flawed, attempt to convey this deep beauty.

I received substantial help through discussions, proofreading, and course development from Suzanne Gray, David Grimes, Giles Harrison, Keith Shine, and Remi Tailleux. Many students and colleagues have contributed by asking questions and suggesting improvements in the material. I thank all of them.

M. AMBAUM
Reading, September 2009

Useful data

Universal constants:

Boltzmann constant	$k_B = 1.381 \times 10^{-23}\,\mathrm{J\,K^{-1}}$.
Avogadro number	$N_A = 6.022 \times 10^{23}$
Universal gas constant	$R^\star = 8.314\,\mathrm{J\,mol^{-1}\,K^{-1}}$
Planck constant	$h = 6.626 \times 10^{-34}\,\mathrm{J\,s}$
Speed of light	$c = 2.998 \times 10^{8}\,\mathrm{m\,s^{-1}}$
Stefan–Boltzmann constant	$\sigma = 5.670 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$
Permittivity of vacuum	$\epsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$
Elementary charge	$e = 1.602 \times 10^{-19}\,\mathrm{C}$

Dry air properties:

effective molar mass	$\mu_d = 29.0\,\mathrm{g\,mol^{-1}}$
specific gas constant	$R, R_d = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$
specific heat at constant pressure	$c_p, c_{pd} = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$

Water properties (temperature dependent properties determined at 0 °C):

molar mass	$\mu_v = 18.015\,\mathrm{g\,mol^{-1}}$
specific gas constant	$R_v = 461.5\,\mathrm{J\,kg^{-1}\,K^{-1}}$
specific heat at constant pressure (vapour)	$c_{pv} = 1859\,\mathrm{J\,kg^{-1}\,K^{-1}}$
specific heat at constant pressure (liquid)	$c_{pl} = 4218\,\mathrm{J\,kg^{-1}\,K^{-1}}$
specific heat at constant pressure (solid)	$c_{pl} = 2050\,\mathrm{J\,kg^{-1}\,K^{-1}}$
enthalpy of vaporization	$L = 2.501 \times 10^{6}\,\mathrm{J\,kg^{-1}}$
enthalpy of fusion	$L_f = 0.334 \times 10^{6}\,\mathrm{J\,kg^{-1}}$
enthalpy of sublimation	$L + L_f = 2.835 \times 10^{6}\,\mathrm{J\,kg^{-1}}$
surface tension	$\gamma = 75.6 \times 10^{-3}\,\mathrm{N\,m^{-1}}$

Other useful data:

Standard pressure	$= 101325\,\mathrm{Pa}$
Standard temperature	$= 15^\circ\,\mathrm{C} = 288.15\,\mathrm{K}$
Acceleration of gravity (Earth surface)	$g_0 = 9.81\,\mathrm{m\,s^{-2}}$
Earth: mean radius	$R_E = 6.371 \times 10^{6}\,\mathrm{m}$
Earth: mean orbital radius	$r_E = 149.5 \times 10^{9}\,\mathrm{m}$
Earth: orbital eccentricity	$e = 0.0167$
Sun: radius	$R_S = 696 \times 10^{6}\,\mathrm{m}$

1

Ideal gases

In this chapter we introduce the concept of an ideal gas, a gas of non-interacting molecules. An ideal gas is an accurate model of dilute gases such as the atmosphere.

We further introduce the notion of macroscopic variables, amongst them such familiar ones as temperature or pressure. These macroscopic variables must be related to some property of the microscopic state of the molecules that make up the substance. For example, for the systems we consider here, temperature is related to the mean kinetic energy of the molecules. The linking of the macroscopic and microscopic worlds is the subject of statistical mechanics. In this chapter we give an elementary application of it to ideal gases.

1.1 THERMODYNAMIC VARIABLES

Consider a volume of gas. A useful mental picture is that of a gas in a closed cylinder with a piston, similar to the driving cylinder of a steam engine, see Figure 1.1. In this way we can control certain properties of the gas, such as its volume or temperature, and perform experiments on it. Such experiments are normally thought experiments, although in principle they can be performed in the laboratory.

At the macroscopic level, the gas has some familiar properties:

- volume V (units: m^3)
- mass M (units: kg)
- density $\rho = M/V$ (units: $\mathrm{kg\,m^{-3}}$)
- temperature T (units: K, Kelvin)
- pressure p (units: $\mathrm{Pa} = \mathrm{N\,m^{-2}}$, Pascal).

The gas is made up of molecules with individual mass M_1, so the total mass of gas is

$$M = NM_1, \tag{1.1}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

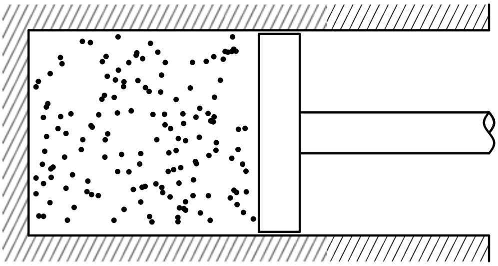


FIGURE 1.1 Gas in a cylinder with piston.

with N the number of molecules. The number of molecules N is often expressed as a multiple of the Avogadro number N_A,

$$N_A = 6.022 \times 10^{23}. \tag{1.2}$$

The Avogadro number is defined as the number of molecules in 12 g of carbon-12. The number of molecules is then defined as a multiple n of N_A

$$N = n N_A, \tag{1.3}$$

where n is the number of *moles*. With this definition of the mol, the mass of the gas can be written as

$$M = n \mu \tag{1.4}$$

with $\mu = N_A M_1$ the *molar mass.* So the molar mass of carbon-12 is by definition precisely 12 g mol^{-1}.

The temperature can be defined as ‘that property which can be measured with a thermometer’. This definition sounds circular but it can be shown to be a perfectly valid definition. The SI unit[3] for temperature is the Kelvin (K). Temperature is often denoted in degrees Celsius, °C with $T(^\circ\mathrm{C}) = T(\mathrm{K}) - 273.15$, or in degrees Fahrenheit, °F with $T(^\circ\mathrm{F}) = 1.8\,T(^\circ\mathrm{C}) + 32$, see Figure 1.2. Temperature can never be lower than 0 K, or absolute zero; the temperature in Kelvin is also called the *absolute temperature*.

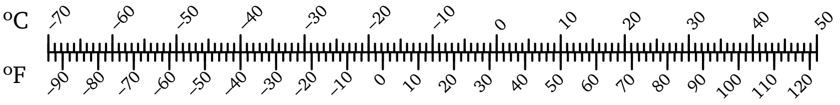


FIGURE 1.2 Nomogram for Celsius–Fahrenheit conversion.

---

[3] SI stands for *Système International d’Unités*, the internationally agreed system of units for physical quantities.

Figure 1.3 illustrates the typical mean temperatures encountered through the depth of the Earth's atmosphere. This figure uses the logarithm of pressure as a vertical coordinate because this is approximately proportional to the geometric height in the atmosphere.

Pressure is the force a gas exerts on its bounding walls per unit area. This does not mean that gas only has a pressure defined at the bounding walls: the internal pressure of a gas can in principle be measured by inserting some probe and measuring the force per unit area on the probe. There are several units of pressure in use, each with its specific area of application. The SI unit for pressure is the Pascal (Pa) which is equivalent to one Newton per square metre. In atmospheric applications we normally use the hectoPascal (hPa; by definition, $1\,\mathrm{hPa} = 100\,\mathrm{Pa}$) or millibar (mbar; with $1\,\mathrm{mbar} = 1\,\mathrm{hPa}$).

Pressure and temperature do not correspond to a property of individual molecules. They are bulk properties that can only be defined as a statistical property of a large number of molecules. This will be discussed in the next section.

There are several other macroscopic variables that can be used to describe the state of a simple gas; these are known as *thermodynamic variables*. If we know all the relevant thermodynamic variables, we know the full thermodynamic state of the gas. All these variables are interrelated and it turns out that for a simple substance (a substance with a fixed composition, such as dry air) we only need two thermodynamic variables to describe the whole thermodynamic state.[4]

For more complex systems we need more variables. For example, in a mixture of varying composition we need to know the concentrations of the constituents. Moist air is such a mixture. The number of water molecules in the air is highly variable and these variations need to be taken into account. For sea water, on the other hand, we need to know the salinity – the quantity of dissolved salts – because it has important consequences for the density. Finally, for cloud drops we need to know the surface area as well as the amount of dissolved solute, both of which have profound consequences for the thermodynamics of the drops.

Thermodynamic variables are either:

- *extensive*, proportional to the mass of the system
- *intensive*, independent of the mass of the system.

Volume and mass are extensive variables, temperature and density are intensive variables. For most variables it is obvious whether they are extensive or intensive.

---

[4] The number $\mathcal{N}$ of thermodynamic variables required to define the state of any system is given by the *Gibbs' phase rule*,

$$\mathcal{N} = 2 + C - P,$$

with C the number of independent constituents and P the number of coexisting phases (gas, liquid, solid) in the system.

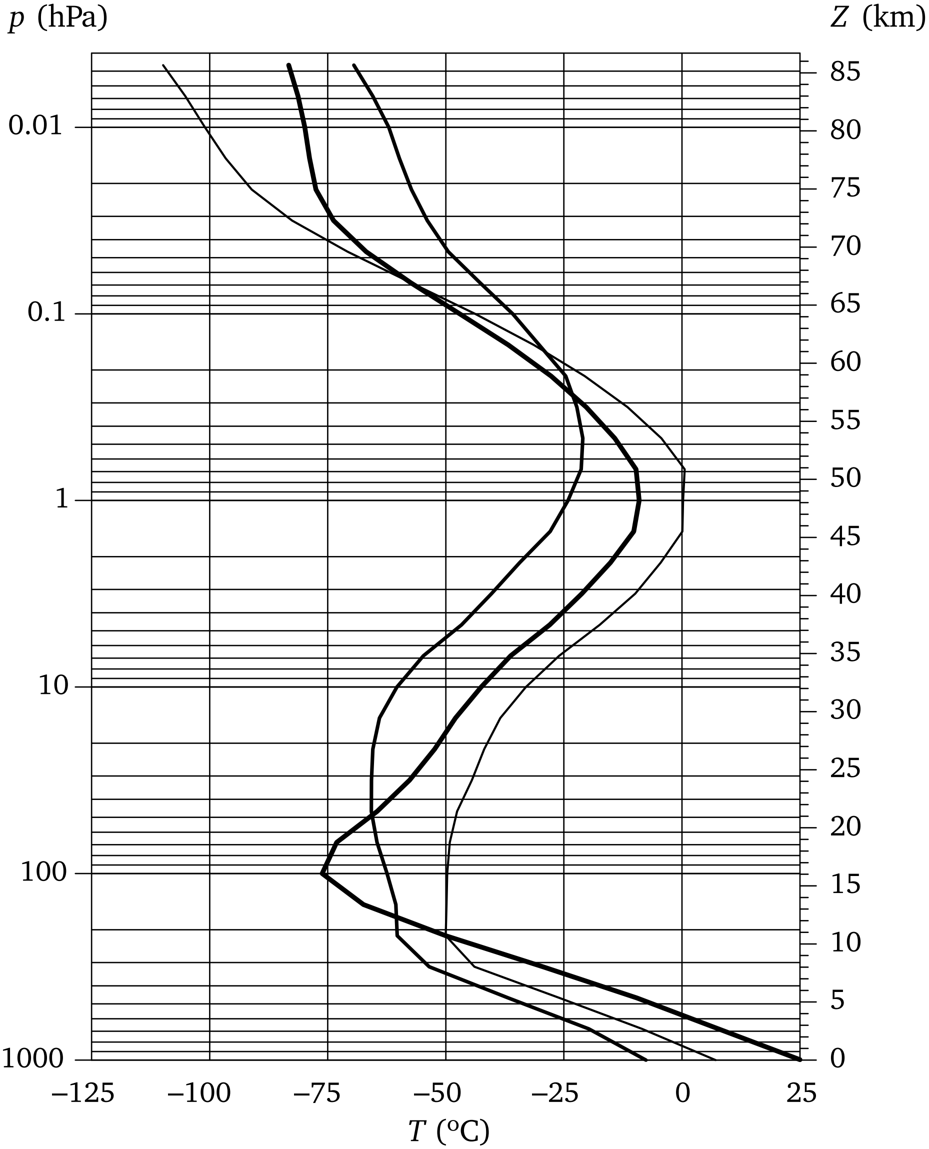


Figure 1.3 Temperature, in °C, as a function of height. Tropical annual mean (thick line), extratropical winter mean (medium line) and extratropical summer mean (thin line). The tropics here correspond to the latitudes between the tropics of Cancer and Capricorn; the extratropics correspond to the latitudes beyond 45° in either hemisphere at the corresponding season. Based on data from Randel, W. et al. (2004) *Journal of Climate* **17**, 986–1003.

Going up in altitude, the temperature first decreases (troposphere), increases (stratosphere), and then decreases (mesosphere). The mesosphere ends at about 90 km altitude, above which the temperature starts to increase again (thermosphere). These atmospheric layers are separated by the tropopause, stratopause, and mesopause, respectively. The temperature increase in the stratosphere is due to the photo-dissociation of ozone, which

Extensive variables can be divided by the mass of the system to become intensive; such new variables are then called *specific* variables. Specific and extensive variables are usually denoted by the same letter, but with the specific variable written in lower case and its extensive equivalent in upper. For example, the volume V of a system divided by the mass M of the system becomes the specific volume v with $v = V/M$. Note that

$$v = 1/\rho, \tag{1.5}$$

where ρ is the density. Later we will come across other extensive variables. For example, the entropy S of a system is an extensive variable, so we can define $s = S/M$ as the specific entropy. Although temperature T is an intensive variable it is normally denoted by an upper case letter, a convention we adopt here as well.

We have ignored internal variations in the volume of gas or material under consideration. For example, we assume there is no internal macroscopic motion of the gas, which would be associated with pressure variations and internal kinetic energy. Clearly this is not the case for the atmosphere as a whole. The pressure and density vary enormously through the atmosphere, usually most dramatically in the vertical: at 10 km height the pressure is about a quarter of its surface value. Indeed, such variations are the source of all atmospheric motion.

We assume that we can define the intensive thermodynamic variables locally and that they have their usual equilibrium thermodynamic relations. We then say that the gas is in *local thermodynamic equilibrium*. Local thermodynamic equilibrium is valid if there is a large separation between the spatial and temporal scales of macroscopic variations and those of microscopic variations. The spatial scale of macroscopic variations needs to be much larger than the *mean free path* of molecules, the mean distance a molecule travels between collisions with other molecules. The temporal scale of macroscopic variations needs to be much larger than the mean time between molecular collisions. Near the Earth's surface the mean free path in the atmosphere is

---

FIGURE 1.3 (*continued*) absorbs the solar energy in the UV part of the spectrum (wavelengths shorter than about 320 nm). Indeed, the ozone itself is formed by photodissociation of molecular oxygen, which occurs at wavelengths shorter than 240 nm. The maximum ozone concentration ('the ozone-layer') is at about 25 km altitude.

It is of note that the temperature in the troposphere is at its maximum in the tropics, while in the stratosphere it is at a maximum in the summer hemisphere and a minimum in the winter hemisphere. This latitudinal temperature gradient is reversed in the mesosphere. Note also that the tropopause is coldest and highest in the tropics.

The thermosphere (outside this plot) is heated by absorption of UV radiation and subsequent ionization of the molecular constituents, thus forming the ionosphere. At these altitudes the density is so low that energy does not get thermalized effectively and local thermodynamic equilibrium is not fully attained. The thermosphere gives way to space in the exosphere.

about 0.1 μm (about 30 times the average molecular distance) with typical molecular velocities of several hundreds of metres per second, so local thermodynamic equilibrium is satisfied. It turns out that above about 100 km height, local thermodynamic equilibrium breaks down.

A small volume of gas in the atmosphere, for which the internal motion can be ignored and which has well-defined density, temperature, and so on, is called an *air parcel*. Because an air parcel is, by definition, in local thermodynamic equilibrium, its thermodynamic variables satisfy all the relationships that are found in equilibrium systems. At the level of an air parcel we need not worry about non-equilibrium effects.

1.2 MICROSCOPIC VIEWPOINT

From the microscopic viewpoint, temperature is defined as the average kinetic energy of the molecules,

$$\blacktriangleright \quad \frac{3}{2}k_B T = \left\langle \frac{1}{2} M_1 \mathscr{U}^2 \right\rangle + \left\langle \frac{1}{2} M_1 \mathscr{V}^2 \right\rangle + \left\langle \frac{1}{2} M_1 \mathscr{W}^2 \right\rangle, \tag{1.6}$$

with $(\mathscr{U}, \mathscr{V}, \mathscr{W})$ the three-dimensional vector velocity of the molecule. The brackets $\langle \ldots \rangle$ denote the average, a time average for a single molecule, the average over all molecules, or the average over an ensemble of gases in the same macroscopic state. A key assumption of statistical mechanics is that all these averages lead to the same result. The constant k_B is the Boltzmann constant,

$$k_B = 1.381 \times 10^{-23}\,\mathrm{J\,K^{-1}}. \tag{1.7}$$

In statistical physics as well as macroscopic thermodynamics, energy is the fundamental quantity. Temperature is a derived quantity which has been given its own units because it is measured with a thermometer. The Boltzmann constant is merely a proportionality constant between energy and absolute temperature. The fundamental point is that statistical mechanics can be formulated such that the microscopic definition of temperature in terms of the mean kinetic energy of the molecules corresponds to the thermodynamic definition of temperature.

The factor 3/2 in the microscopic definition of temperature reflects a classic result in the mechanics of systems with many components, namely that each *degree of freedom* contains, on average, the same energy. A degree of freedom is an independent variable in which the system can vary. A single molecule carries three translational degrees of freedom: motion in the x, y, and z-directions. There can also be internal degrees of freedom corresponding to rotations and vibrations of the molecule. The *equipartition theorem* states that each accessible degree of freedom[5] carries on average the same energy,

---

[5] Not all available degrees are necessarily accessible. Quantization of energy levels implies that there is a minimum energy required to excite any degree of freedom.

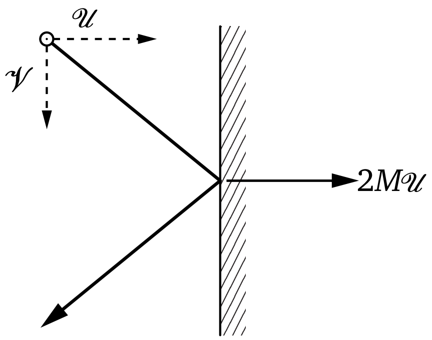


FIGURE 1.4 Transfer of momentum by a molecule colliding with the wall. The total momentum transfer is twice the momentum in the x-direction of the molecule.

and this energy equals $k_B T/2$. Adding the average kinetic energies in the three spatial directions then gives the result of Eq. 1.6.

Pressure is the result of many collisions of individual molecules against the walls of a vessel or a probe. If a molecule approaches the wall with a velocity $\mathscr{U}$ and elastically collides with the wall, then the molecule's momentum in the direction of the wall changes by $2M_1\mathscr{U}$, from $M_1\mathscr{U}$ to $-M_1\mathscr{U}$. This momentum is transferred to the wall. By Newton's laws, the amount of momentum transferred per unit time is the force on the wall, see Figure 1.4. For an interior point we can define the local pressure as the momentum flux density through some imaginary surface in the interior of the fluid.

So how many molecules collide with the wall? Let the number density of molecules, that is the number of molecules per unit volume, be denoted with $\tilde{n}$. We can now write the number density of molecules with x-velocities between $\mathscr{U}$ and $\mathscr{U} + \mathrm{d}\mathscr{U}$ as $\tilde{n}_{\mathscr{U}}$, which is related to the total number density $\tilde{n}$ by

$$\tilde{n} = \int \tilde{n}_{\mathscr{U}} \,\mathrm{d}\mathscr{U}. \tag{1.8}$$

Over a time δt, those molecules with positive velocity between $\mathscr{U}$ and $\mathscr{U}+\mathrm{d}\mathscr{U}$ that are located within a distance $\mathscr{U}\,\delta t$ of the wall will collide with the wall. Therefore, the number of such molecules that have collided with the wall will be $\tilde{n}_{\mathscr{U}}\mathscr{U}\,\delta t\,A$, with A the area of the wall. To get the momentum transfer per unit time, simply multiply this number by the momentum transfer per molecule, $2M_1\mathscr{U}$, and divide by the time taken, δt. This is the force $F_{\mathscr{U}}$ exerted on the wall by molecules with positive velocities between $\mathscr{U}$ and $\mathscr{U} + \mathrm{d}\mathscr{U}$,

$$F_{\mathscr{U}} = 2\tilde{n}_{\mathscr{U}} M_1 \mathscr{U}^2 A. \tag{1.9}$$

To find the total pressure we need to divide by A and integrate the force over all positive velocities, that is to say those $\mathscr{U} > 0$, because molecules with negative velocities will not collide with the wall and thus will not contribute to the pressure,

$$p = \int_{\mathscr{U}>0} 2\tilde{n}_{\mathscr{U}} M_1 \mathscr{U}^2 \, \mathrm{d}\mathscr{U}. \tag{1.10}$$

By symmetry, there will be an equal number of molecules with positive and negative $\mathscr{U}$. We can therefore integrate over all velocities $\mathscr{U}$, positive and negative, and divide the result by two. The expression for the pressure then becomes

$$p = \tilde{n} \, \langle M_1 \mathscr{U}^2 \rangle, \tag{1.11}$$

with $\tilde{n}$ the total number density. The equipartition theorem states that $\langle M_1 \mathscr{U}^2 \rangle = k_B T$ so that the pressure satisfies

▶ $$pV = N k_B T, \tag{1.12}$$

where we have substituted $\tilde{n} = N/V$. This is the *ideal gas law.*

By writing the total number of molecules N as nN_A, the ideal gas law can be written

$$pV = nR^\star T, \tag{1.13}$$

where $R^\star$ is called the *universal gas constant,*

$$R^\star = N_A k_B = 8.314 \, \mathrm{J\,mol^{-1}\,K^{-1}}. \tag{1.14}$$

Before the microscopic definitions of temperature and pressure were known, it was already hypothesized by Avogadro (and later confirmed to be true) that the constant $R^\star$ is the same for all types of gases.

Another form of the ideal gas law follows by dividing by the mass $M = n\mu$ of the gas to find

▶ $$pv = RT, \quad \text{or} \quad p = \rho RT, \tag{1.15}$$

where R is the so-called *specific gas constant*,

$$R = R^\star/\mu. \tag{1.16}$$

This is the form of the ideal gas law that is normally used in atmospheric science. Confusingly, the convention is to use a capital R for the specific gas constant even though it is a specific quantity. Note also that in most general physics literature the letter R stands for the universal gas constant; it should be clear from the context which is meant. This is one of those instances where the convention used in atmospheric science literature is not particularly helpful. Furthermore, the ideal gas law in the form of Eq. 1.13 is more general and more useful.

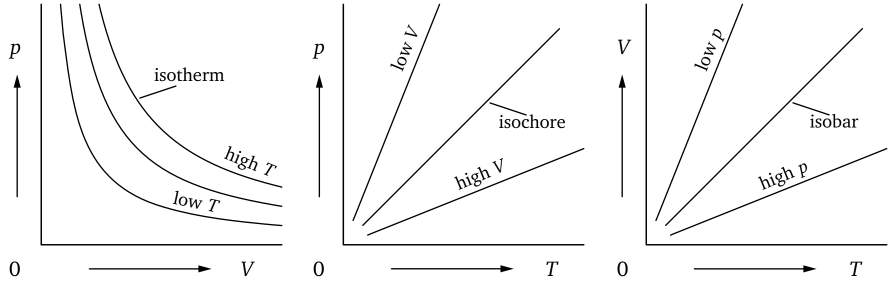


FIGURE 1.5 The left panel illustrates Boyle's law and the middle panel Gay-Lussac's law. The right panel illustrates that, for an ideal gas at fixed pressure, the volume of a gas is proportional to its temperature; this is sometimes known as *Charles's law*.

The ideal gas law encompasses:

- *Boyle's law:* at constant temperature, the product of pressure and volume is constant
- *Gay-Lussac's law:* at constant volume, the pressure of a gas is proportional to its temperature.

Figure 1.5 illustrates these laws in diagrams. These laws were originally determined experimentally. They are only strictly valid for ideal gases.

In deriving the ideal gas law, we have not considered subtleties such as inelastic collisions, where energy transfer between the gas and the wall occurs, or the consideration that the wall is not a mathematical flat plane but made up of molecules. These complications do not alter the basic result.

We have also not considered interactions between the molecules and interactions at a distance between the molecules and the wall. This *does* make a difference and it defines the difference between real gases and ideal gases. Ideal gases are made up of non-interacting molecules, vanishingly small molecules that are unaware of the presence of any other molecules.

We assume that molecules in an ideal gas do not interact with each other and also that the molecules are in thermal equilibrium. Strictly speaking these assumptions are inconsistent, as a gas can only achieve thermal equilibrium through many collisions between the molecules. The colliding molecules distribute the energy amongst all the accessible degrees of freedom and thus achieve equipartition. This process of energy distribution is called *thermalization*. A gas is in local thermodynamic equilibrium if all the available energy is thermalized. If collisions are rare, energy cannot be thermalized effectively and the gas cannot achieve local thermodynamic equilibrium. This occurs at high altitudes in the atmosphere (higher than, say, 100 km) where the energy input from radiation is not thermalized due to the low number of collisions.

The ideal gas law is an example of an *equation of state.* Real gases are not ideal and will therefore have a different relationship between pressure, density and temperature. For example, the equation of state for real gases is more accurately described by *van der Waals' equation*,

$$\left(p + a\left(\frac{n}{V}\right)^2\right)(V - nb) = nR^\star T, \tag{1.17}$$

with a and b constants that depend on the particular gas. The term $-nb$ represents the reduction in available volume due to the finite effective size of the molecules. The term $a(n/V)^2$ is related to the average interaction energy between molecules (which is why the term is quadratic in the number density) and it contributes as an effective pressure; the relation between pressure and energy density will be explained in Section 2.2, and a derivation of van der Waals' equation will be presented in Section 3.4. Van der Waals' equation is more accurate for gases at high densities, and approximately describes such important processes as phase transitions.

Densities of gases in the atmosphere are such that the ideal gas law gives essentially the same results as van der Waals' equation, so we stick to the much simpler ideal gas law. Indeed, for low densities van der Waals' equation reduces to the ideal gas law,

$$\text{Eq. 1.17} \rightarrow pV = nR^\star T \qquad \text{if } n/V \rightarrow 0. \tag{1.18}$$

Conversely, for large pressures,

$$\text{Eq. 1.17} \rightarrow V = nb \qquad \text{if } p \rightarrow \infty. \tag{1.19}$$

So the high pressure, incompressible liquid is also a limiting case of van der Waals' equation.

In general, an equation of state is some relationship between the variables of the system,

$$f(p, N, V, T, a_1, a_2, a_3, \dots) = 0, \tag{1.20}$$

where $a_1, a_2 \dots$ represent any other variables that influence the state of the system, such as humidity in air, or salinity in water. The equation of state is dependent on the precise nature of the system and results derived using an equation of state are therefore only valid for that particular system. We will see that many of the results in this book are derived without reference to an equation of state and will therefore be valid for any substance.

1.3 IDEAL GAS MIXTURES

Ideal gases are defined as gases where the molecules themselves have negligible volume and have negligible interactions with each other. So if we mix several ideal gases at the same temperature in a single volume they do not 'feel' each other's presence. This means that each constituent gas contributes independently to the pressure. The contribution of each constituent to the total pressure is called the *partial pressure*. So n_i mols of constituent i will have a partial pressure p_i equal to:

$$p_i = \frac{n_i R^\star T}{V}. \tag{1.21}$$

The fact that these partial pressures independently make up the total pressure of the mixture is called *Dalton's law:*

$$p = \sum\nolimits_i p_i. \tag{1.22}$$

Dalton's law is only true for ideal gases. For real gases, partial pressures cannot be easily defined.

From Dalton's law it follows that

$$p = \sum\nolimits_i \frac{n_i R^\star T}{V} = \sum\nolimits_i \frac{M}{V} \frac{n_i \mu_i}{M \mu_i} R^\star T = \rho R T, \tag{1.23}$$

where

$$R = \frac{R^\star}{\mu_e}, \tag{1.24}$$

with the effective molar mass μ_e of the mixture defined as

$$\frac{1}{\mu_e} = \sum\nolimits_i \frac{c_i}{\mu_i}, \tag{1.25}$$

and $c_i = n_i \mu_i / M$ the mass fraction of constituent i – that is, the fraction constituent i contributes to the total mass. Another way of interpreting this result is that the specific gas constant R of the mixture is the weighted mean of the specific gas constant of the pure constituents $R_i = R^\star / \mu_i$, that is,

▶ $$R = \sum\nolimits_i c_i R_i. \tag{1.26}$$

Table 1.1 lists the main constituents of air. The dry air constituents are well mixed and long-lived which means that the bulk composition of dry air is fixed throughout the atmosphere up to very high altitudes. There is a long-term upward trend in CO_2 concentration over time due to human activity but this has only a minor effect on the ideal gas properties of dry air, see Figure 1.6. However, the CO_2 trend does have a profound effect on the radiative properties of the air and it is the main agent of human-induced climate change.

Table 1.1 Main constituents of air.

Constituent	*Molar mass, μ (g mol^{-1})*	*Mass fraction, c*
N_2	28.02	0.755
O_2	32.00	0.231
Ar	39.93	0.013
CO_2	44.01	~ 0.0006
H_2O	18.02	0 to ~ 0.03

Using the numbers in Table 1.1 it is straightforward to verify that the effective molar mass μ_d for dry air is

$$\mu_d = 29.0\,\mathrm{g\,mol^{-1}} \tag{1.27}$$

and the specific gas constant R for dry air is

▶ $$R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}. \tag{1.28}$$

This value of R can be used in the ideal gas law, Eq. 1.15, which then relates the pressure, density, and temperature for dry air. Figure 1.7 is a nomogram expressing this relation graphically.

Now we can also calculate the effect of water vapour on the ideal gas law. The amount of water vapour in the air can be quantified by its mass

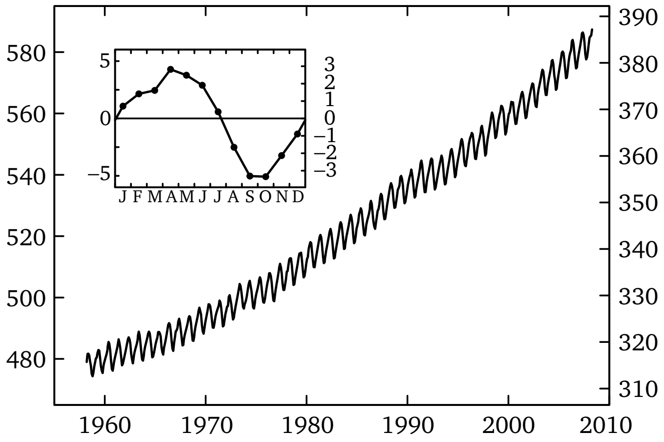


Figure 1.6 Mass fraction (left axis, in parts per million) and molar mixing ratio (right axis, in parts per million) of CO_2 as measured at Mauna Loa Observatory, Hawaii. Inset: mean annual cycle of CO_2, scales as main graph. This annual cycle is dominated by vegetation growth and resulting CO_2 capture in the Northern Hemisphere summer. Data from Tans, P. P. *Trends in Atmospheric Carbon Dioxide - Mauna Loa*. Available from: www.esrl.noaa.gov/gmd/ccgg/trends/. See also Keeling, C. D. et al. (1976) *Tellus* **28**, 538–551.

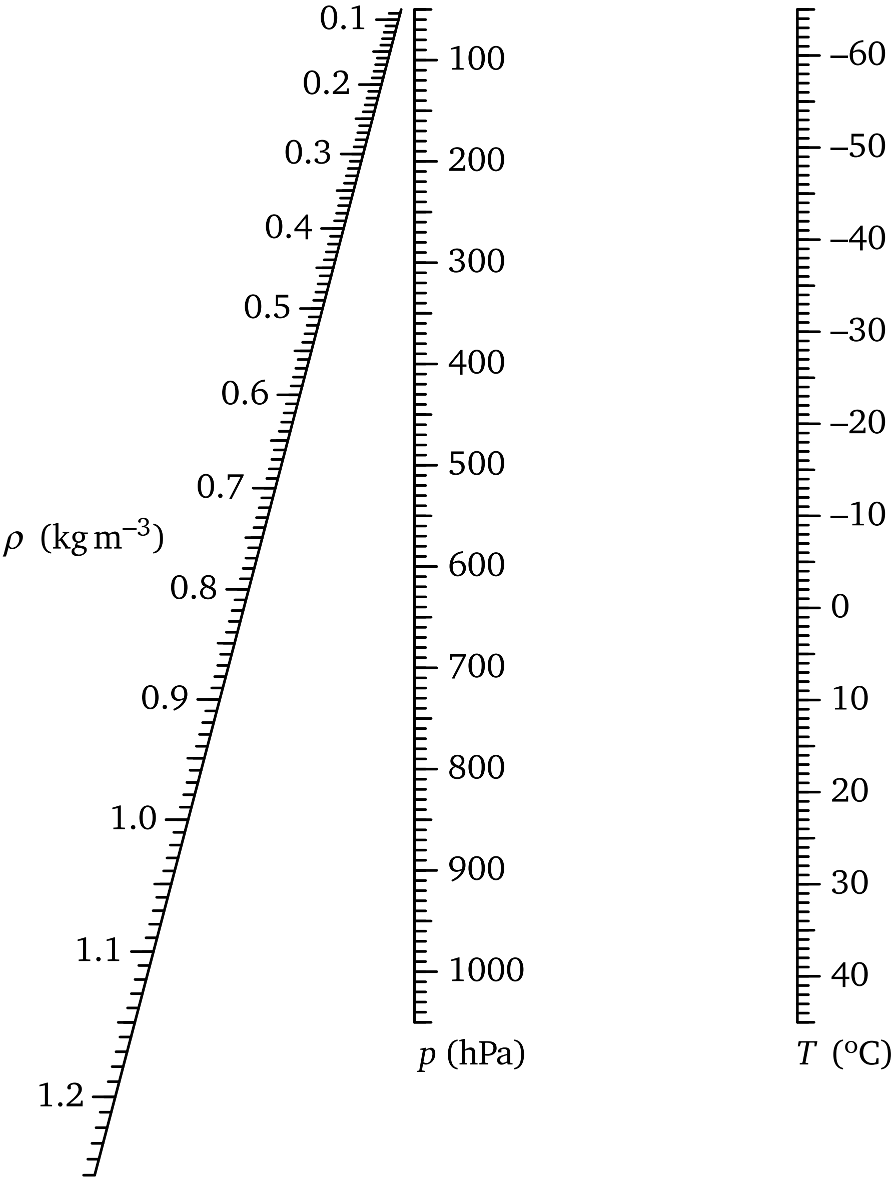


FIGURE 1.7 Nomogram expressing the ideal gas law for dry air,

$$p = \rho RT \quad (\text{with } R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}).$$

The nomogram works by laying any straight index line over the figure so that it crosses all three scales. The values for ρ, p, and T where the index line then crosses the graduated scales satisfy the ideal gas law for dry air.

fraction. For water vapour this mass fraction is called the *specific humidity*, usually denoted q. This means that the mass fraction of dry air is $1-q$. From Eq. 1.26 it then follows that the specific gas constant for the moist air is

$$(1-q)\frac{R^\star}{\mu_d}+q\frac{R^\star}{\mu_v}=R\left(1+\left(\frac{\mu_d}{\mu_v}-1\right)q\right), \tag{1.29}$$

with μ_d and μ_v the molar masses for dry air and water, respectively, and R the specific gas constant for dry air, Eq. 1.28. It is common usage to absorb the factor with the dependency on specific humidity in a redefined temperature, the *virtual temperature*, T_v. So the ideal gas law for moist air is

$$\blacktriangleright \qquad pv=RT_v \quad \text{or} \quad p=\rho R T_v, \tag{1.30}$$

where the virtual temperature T_v is defined as

$$\blacktriangleright \qquad T_v=\left(1+\left(\frac{\mu_d}{\mu_v}-1\right)q\right)T=(1+0.61\,q)\,T, \tag{1.31}$$

μ_d/μ_v being equal to 1.61. So it turns out that moist air can be treated as dry air as long as we replace the temperature in the equation of state by the virtual temperature.

The virtual temperature increases with specific humidity. Over humid tropical areas the specific humidity can be about $30\,\mathrm{g\,kg^{-1}}$, leading to a virtual temperature that is about 2% above the temperature (about 6°C). Over colder areas the specific humidity is always much less, so here the virtual temperature and the temperature are always less than 1°C apart.

PROBLEMS

1.1. How many molecules are there in a litre of air at standard pressure and temperature? What is the volume, in litres, of one mole of ideal gas at standard pressure and temperature?

1.2. Use the microscopic definition of temperature, Eq. 1.6, to find a typical root-mean-square molecular velocity for air.

1.3. (i) Calculate the density of dry air at a temperature of 15°C and a pressure of 1000 hPa (or use the nomogram in Fig 1.7). (ii) Calculate the density if we now assume that this air has a specific humidity of $3\,\mathrm{g\,kg^{-1}}$. (iii) Calculate the density of air at a virtual temperature of 15°C and a pressure of 1000 hPa.

1.4. The proportions by mass of the major constituents of the atmosphere on Mars are given in the following table. Calculate the effective molar mass and the specific gas constant for the Martian atmosphere.

	CO_2	N_2	Ar
Mars	0.96	0.02	0.02

1.5. Saturn's largest moon, Titan, has an atmosphere mainly made up of nitrogen (N_2). The temperature at Titan's surface is about −94°C and the pressure is about 1460 hPa. What is the density at the surface? The main varying constituent on Titan is methane (which has a molar mass of $16\,\mathrm{g\,mol^{-1}}$). What is the virtual temperature, with respect to methane, at the surface if the number concentration of methane is 1.5%?

2

The first and second laws

In this chapter we will discuss the thermodynamic version of the law of conservation of energy and the law relating heat to entropy. These laws are called the first law and the second law of thermodynamics, respectively, and they are part of the axiomatic framework of thermodynamics.

To make the axiomatic framework of thermodynamics complete we also need the third law, stating how thermodynamic properties behave near absolute zero. The third law does not play an important role in atmospheric applications and we will not further expand on it. We also need the zeroth law, which states that if system A is in thermal equilibrium with system B and system B is in thermal equilibrium with system C then system A has to be in thermal equilibrium with system C. The zeroth law makes physical sense and is usually taken for granted, although in a strictly axiomatic development it needs to be stated explicitly.

The second law has a particularly profound interpretation in that it describes the irreversibility of macroscopic systems and thus provides the direction of time in macroscopic systems. It also provides the link between the macroscopic world of thermodynamics and the microscopic world of molecular dynamics, in that entropy measures the microscopic uncertainty of a macroscopic system.

2.1 WORK

From classical mechanics we know that if we exert a force $\boldsymbol{F}$ on a body in order to displace it over a distance $\Delta\boldsymbol{r}$, we put in an energy ΔW equal to the inner product of the force and the displacement: $\Delta W = \boldsymbol{F}\cdot\Delta\boldsymbol{r}$. This energy of displacement with a force is called *work* (units J, Joule). The force may be a function of location so that it is better to take infinitesimal displacements $\mathrm{d}\boldsymbol{r}$ for which infinitesimal amounts of work $\mathrm{d}W$ are required:

$$\mathrm{d}W = \boldsymbol{F}\cdot\mathrm{d}\boldsymbol{r}. \qquad (2.1)$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum
ISBN: 978-0-470-74515-1

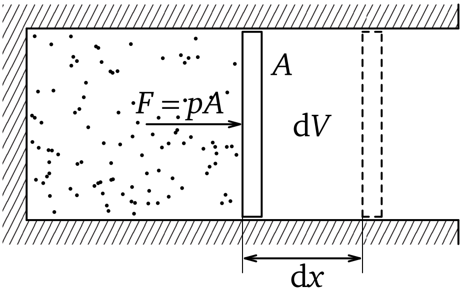


Figure 2.1 Work done during expansion of a gas.

So for a finite displacement we find that the total work W required is

$$W = \int_{r_0}^{r_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}, \tag{2.2}$$

where $\boldsymbol{F}$ can be a function of $\boldsymbol{r}$.

A similar set up occurs in thermodynamics: if we put a gas in a cylinder with a piston then the gas exerts a pressure p on the outer walls, see Figure 2.1. Moving the piston therefore requires a force of magnitude $F = pA$ (positive when compressing the gas). We assume here that the movement of the piston is frictionless so no work is required to overcome the friction. The work done on the gas in compressing the piston over a distance $\mathrm{d}x$ is therefore equal to $pA\,\mathrm{d}x$. This can be written as the work $\mathrm{d}W$ put into the gas,

▶ $$\mathrm{d}W = -p\,\mathrm{d}V, \tag{2.3}$$

with $\mathrm{d}V = -A\,\mathrm{d}x$, the change in volume of the gas; note that an expanding gas ($\mathrm{d}V > 0$) performs pressure work *on* its surroundings and the work input, $\mathrm{d}W$, is negative. It is again a good idea to use infinitesimal displacements; Boyle's law provides an example of how pressure can be a function of the volume of the gas, just as in Eq. 2.1 the force can be a function of x. Dividing by the mass M of the system we find the specific form of the above equation,

$$\mathrm{d}w = -p\,\mathrm{d}v. \tag{2.4}$$

This equality is only true for systems without friction or other dissipative processes. If there is friction, some of the work $\mathrm{d}W$ required to compress a volume of gas is used to overcome the friction and will therefore be larger than $-p\,\mathrm{d}V$; similarly, for expansion, the work delivered by the gas is partly used to overcome friction and will therefore produce less external work. In either case we have

▶ $$\mathrm{d}W \geq -p\,\mathrm{d}V. \tag{2.5}$$

This is the generalization of Eq. 2.3 for systems with friction or dissipation; for frictionless, non-dissipative systems, the above equation reduces to Eq. 2.3. Below we will discuss this result in the context of the second law of thermodynamics.

The above definition of work done on a gas under pressure is not only true for gases in a cylinder: an amount of gas (or any substance) in any shape will experience a certain pressure p on its boundary. Assume its surface A is made up of different area elements A_i, which are displaced outward by a distance $\mathrm{d}n_i$. The total work performed on the gas equals the sum of work performed on each individual area element,

$$\mathrm{d}W = -\sum_i pA_i \,\mathrm{d}n_i. \tag{2.6}$$

But on displacing the area elements by $\mathrm{d}n_i$, the volume of the gas changes by

$$\mathrm{d}V = \sum_i A_i \,\mathrm{d}n_i. \tag{2.7}$$

Combining the last two equations we get Eq. 2.3 again.

As an example of pressure work, take an ideal gas at initial volume V_0 and pressure p_0. If we change the volume to V_1 while keeping the temperature constant, the total work done would be

$$W = p_0 V_0 \ln (V_0/V_1). \tag{2.8}$$

This situation can be achieved by keeping the gas in thermal contact with a heat reservoir, such as a big block of metal, at a given temperature, and changing the volume slowly so as to allow thermal equilibration between the gas and the thermal reservoir. Isothermally compressing a litre ($10^{-3}\,\mathrm{m}^3$) of gas at atmospheric pressure to half its volume would require about 70 J of energy; enough to heat a cup of tea up by about 0.15°C or to power a 60 W lightbulb for about a second.

Change in volume against pressure is not the only form of work a thermodynamic system can perform. For example, a body of liquid could change its surface area by an amount $\mathrm{d}A$ (for example, by changing its shape while conserving its volume or by evaporating liquid into the gas phase). However, a liquid–gas interface behaves like an elastic membrane that is trying to contract. The tensile strength of the membrane is measured by the surface tension γ. The amount of work $\mathrm{d}W$ required to change the surface area $\mathrm{d}A$ against surface tension γ is

$$\mathrm{d}W = \gamma \,\mathrm{d}A. \tag{2.9}$$

A droplet will tend to minimize the energy associated with its surface tension, just as a particle will tend to minimize its potential energy. The result of this is that a droplet at rest will tend to have a minimum surface area given its volume; this is achieved by assuming a spherical shape. The thermodynamic effects of surface tension on water droplets are further described in Chapter 7.

Another example is a charged droplet that has an electric surface potential of V_e (units V, Volt) which is charged by an amount $\mathrm{d}Q_e$ (units C, Coulomb). The electrostatic energy required for this charging process requires an energy $\mathrm{d}W$,

$$\mathrm{d}W = V_e\,\mathrm{d}Q_e. \tag{2.10}$$

In general, the work $\mathrm{d}W$ required to produce a particular infinitesimal change can be written as

$$\mathrm{d}W = y\,\mathrm{d}X, \tag{2.11}$$

where, inspired by Eq. 2.1, y is called the *generalized force* and $\mathrm{d}X$ is called the *generalized displacement*.

Because the infinitesimal variation in work is a compound term, see Eq. 2.3, it follows that work is not a state variable; that is, it cannot be determined by making a single measurement of the properties of the gas. This can be readily understood from a mathematical perspective. Consider a gas for which the states are plotted on a pV diagram, see Figure 2.2. The temperature can be deduced from the equation of state. Any state will correspond to a single point on this diagram. Any transformation of the system – a change in volume, temperature, pressure, etc. – will move the point around on the diagram. So on the pV diagram we can think of a process transforming the system from A to B, but we can choose different routes: we could go via C or via D, or any other route from A to B. The total work W required to perform such a process is

$$W = -\int_{A\to B} p\,\mathrm{d}V, \tag{2.12}$$

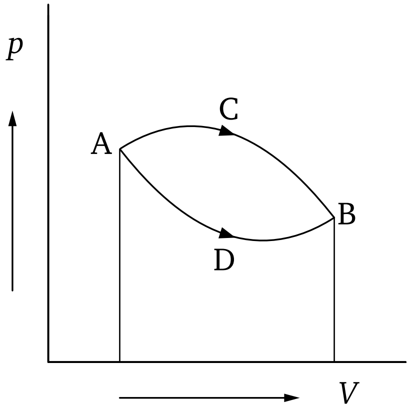


FIGURE 2.2 *pV* diagram for a system changing its state from A to B via ACB or ADB.

which is the area under the graph on the pV diagram. The area depends on the route taken: curve ACB has a larger area, and more negative work is required (the gas performs more work), than curve ADB. The path ACB performs the change in volume $V_B - V_A$ on average at a larger pressure than the path ADB, although the pressures at their joint endpoints, A and B, are the same. Going from A to B, the changes in state variables p, V, T, etc., are well defined, while the work W depends on the path taken and is therefore not uniquely determined by the system's location on the pV diagram. We have to conclude that work is not a state variable.

2.2 ENERGY CONSERVATION: THE FIRST LAW

If we put energy into a gas under pressure by compressing it ($\mathrm{d}V < 0$) where does the energy $\mathrm{d}W = -p\,\mathrm{d}V$ go? If we extract work from a gas by expanding it ($\mathrm{d}V > 0$) where does the energy come from? The energy is stored in all the internal degrees of freedom, see Section 1.2. The thermodynamic expression for the sum total of this stored energy is called *internal energy U* (units J). Combining the equipartition theorem (all degrees of freedom store on average an equal amount of energy) and the microscopic definition of temperature (temperature is proportional to the energy in the translational degrees of freedom) it follows that, like temperature, the internal energy is a state variable.[6]

If we bring a hot body into contact with a cold body we know from experiments that the hot body will cool down and the cold body will warm up. If the volume of the bodies does not change much in this process, this change of temperature of each body has to be associated with a change of their internal energy. In fact, conservation of energy dictates that the cool body gains the same amount of energy as the warm body loses. The energy that flows between bodies of different temperatures in contact is called *heat Q* (units J).[7]

So if, besides work, we also allow a system to exchange heat $\mathrm{d}Q$ with its environment the law of conservation of energy becomes

$$\blacktriangleright \qquad \mathrm{d}U = \mathrm{d}Q + \mathrm{d}W. \tag{2.13}$$

This is the first law of thermodynamics. It states that the internal energy of a system can increase by putting in heat or by performing work on it. We

---

[6] Although our argument appears only valid for substances with a fixed number of degrees of freedom, or a number that only depends on temperature, it turns out to be true for any substance; the effective number of degrees of freedom can be a complex function of the state variables. Essentially, we assume that the energy in all the internal degrees of freedom for a particular state does not depend on the way we arrived at this state.

[7] In a more formal treatment of thermodynamics, temperature is *defined* to be the quantity that indicates the direction of spontaneous heat flow; so heat will flow by definition from high to low temperatures. See Baierlein, R. (1999) *Thermal Physics*. Cambridge University Press, Cambridge.

can divide this equation by the mass of the system to rewrite all extensive variables as specific variables,

$$\mathrm{d}u = \mathrm{d}q + \mathrm{d}w. \tag{2.14}$$

We will usually use this specific form of the first law, but the transformation between the two forms is trivial.

The above form of the first law is valid for systems that only exchange energy with their environment through heat or work. In this form it excludes matter exchange. However, in most relevant cases matter exchange can be dealt with using the concept of enthalpy, see Section 3.5. In its most general form, the first law also includes terms related to molecule number changes, and the associated generalized force is called the *chemical potential*. Chapter 8 provides the necessary background.

The first law is very general; it is the thermodynamic version of the law of conservation of energy. However, the real meat of the the first law is the realization that heat is a form of energy. This realization was a major step in the development of thermodynamics.

Let us now briefly return to our example of isothermal compression of a litre of air. The isothermal constraint implies that the internal energy of the gas does not change: $\mathrm{d}U = 0$. The first law then states that a volume change of $\mathrm{d}V$ requires a heat input in the gas of $\mathrm{d}Q = p\,\mathrm{d}V$. In the absence of such heat input, if, for example, the gas is thermally isolated, the internal energy would change by an amount $\mathrm{d}U = -p\,\mathrm{d}V$, leading to a change in temperature.

For an insulated system (that is, no heat exchange, $\mathrm{d}Q = 0$) which can perform pressure work, the first law states that

$$\mathrm{d}U = -p\,\mathrm{d}V. \tag{2.15}$$

This leads to the interpretation of pressure as a volumetric energy density. If a system changes volume while there are no other sources of internal energy, then the volumetric energy density U/V changes as

$$\mathrm{d}\left(\frac{U}{V}\right) = -\left(\frac{U}{V}\right)\frac{\mathrm{d}V}{V} - p\frac{\mathrm{d}V}{V}. \tag{2.16}$$

The first term on the right-hand side is a purely geometric effect that describes how the energy density changes if the given energy is redistributed over a different volume; this term is present for any quantity for which a change in volumetric density is computed, for example molecule number density. The second term on the right-hand side is a physical effect that results from the energy that is produced by performing work on the system. The two effects can be interpreted as purely geometric contributions if we interpret p as an additional volumetric energy density,

$$\mathrm{d}\left(\frac{U}{V}\right) = -\left(\frac{U}{V} + p\right)\frac{\mathrm{d}V}{V}. \tag{2.17}$$

For many physical phenomena it is useful to reverse this argument: if some physical process implies some additional volumetric energy density then this is manifested as an effective pressure. For example, let us consider the contribution of the surface tension to the internal energy of a spherical water drop of volume $V = (4/3)\pi r^3$. The volumetric energy density then changes as (see Eq. 2.9)

$$\mathrm{d}\left(\frac{U}{V}\right) = -\left(\frac{U}{V}\right)\frac{\mathrm{d}V}{V} + \gamma\frac{\mathrm{d}A}{V} = -\left(\frac{U}{V} - \frac{2\gamma}{r}\right)\frac{\mathrm{d}V}{V}, \tag{2.18}$$

where it is taken that for a spherical drop $2\,\mathrm{d}V = r\,\mathrm{d}A$. The surface tension then corresponds to an effective negative pressure (the tension tries to compress the drop) of magnitude $2\gamma/r$, the so-called *capillary pressure.* For a water drop in dynamic equilibrium, the pressure inside and outside the drop has to be equal. However, because the surface tension acts as an additional negative pressure inside the drop, the actual pressure inside the drop has to be larger by $2\gamma/r$ compared to its environment.

2.3 ENTROPY AND THE SECOND LAW

From the first law we know that the change in internal energy of a system equals the sum of the heat flow into the system and the work done on the system, $\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W$. Because internal energy is a state variable, but work is not, we need to conclude that heat cannot be a state variable either. It can be shown that, just like work W, heat input Q can also be written as a compound variable:

▶ $$\mathrm{d}Q = T\,\mathrm{d}S, \tag{2.19}$$

with S the *entropy*. Entropy *is* a state variable; the combination $T\,\mathrm{d}S$ is not. The arguments leading to this important result are beautiful and somewhat convoluted. There are several ways of demonstrating this important result but the classical arguments are based on a careful analysis of thermodynamic *cycles,* where a system undergoes several well-defined transformations to return to its original state; most general textbooks on thermodynamics will provide a version of these arguments.

We can illustrate the definition of entropy in the case of an ideal monoatomic gas. The only internal degrees of freedom for such a gas are the translational degrees of freedom. This means that the internal energy follows from the definition of temperature, Eq. 1.6,

$$U = \frac{3}{2} N k_B T. \tag{2.20}$$

The differential form of the first law for a monoatomic ideal gas therefore can be written

$$\mathrm{d}Q = \mathrm{d}U - \mathrm{d}W = \frac{3}{2} N k_B \,\mathrm{d}T + p\,\mathrm{d}V. \tag{2.21}$$

From the ideal gas law, $pV = Nk_B T$, it then follows that

$$\mathrm{d}S = \frac{\mathrm{d}Q}{T} = \frac{3}{2} N k_B \frac{\mathrm{d}T}{T} + N k_B \frac{\mathrm{d}V}{V}. \tag{2.22}$$

The above equation can be integrated (we introduced an *integrating factor* $1/T$ to achieve this) and we find that for an ideal monoatomic gas the entropy S equals

$$S = \frac{3}{2} N k_B \ln\left(T/T_0\right) + N k_B \ln\left(V/V_0\right), \tag{2.23}$$

with T_0 and V_0 integration constants. So entropy S *is* a state function; we have an expression of S in terms of other state variables.

We now have two forms of the first law of thermodynamics for simple systems:

$$\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W, \tag{2.24a}$$
$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V. \tag{2.24b}$$

Subtracting these equations, we find that

$$T\,\mathrm{d}S - \mathrm{d}Q = p\,\mathrm{d}V + \mathrm{d}W. \tag{2.25}$$

In general, the right-hand side is non-negative, Eq. 2.5, and will only be zero for non-dissipative systems.[8] If we allow for friction or other dissipative processes, we have

$$\blacktriangleright \qquad dQ \leq T\,\mathrm{d}S. \tag{2.26}$$

---

[8] These arguments are often a source of confusion: the $\mathrm{d}W$ and the $\mathrm{d}Q$ represent the work and heat put into the full system, e.g. a cylinder with gas. They are, in a sense, external to the working fluid, the gas. The $-p\,\mathrm{d}V$ and $T\,\mathrm{d}S$ represent the way this work and heat input change the equilibrium thermodynamic properties of the working fluid and are thus internal to the working fluid.

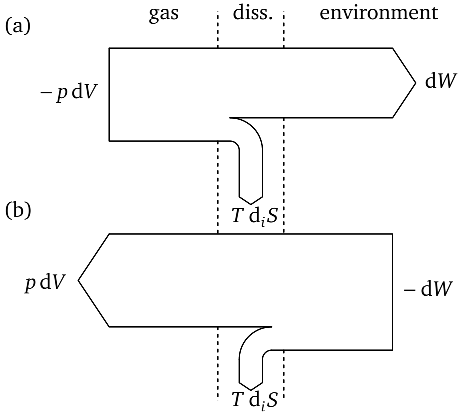


FIGURE 2.3 Compression (a) and expansion (b) of a system with friction. In both cases we find $\mathrm{d}W + p\,\mathrm{d}V = T\,\mathrm{d}_iS > 0$ with $\mathrm{d}W$ the work done on the environment and $p\,\mathrm{d}V$ the work done by the gas.

The difference $T\,\mathrm{d}S - \mathrm{d}Q$ is called *uncompensated heat* and it is always positive. For thermally isolated systems ($\mathrm{d}Q = 0$) we find that

$$\mathrm{d}S \geq 0, \tag{2.27}$$

that is, for thermally isolated systems the entropy will never decrease.

In general, we can write the entropy change $\mathrm{d}S$ as the sum of a reversible (external) change d_eS, due to heat flow (and possibly matter flow), and an irreversible change d_iS, due to uncompensated heat:[9]

$$▶ \quad dS = \mathrm{d}_eS + \mathrm{d}_iS = \frac{\mathrm{d}Q}{T} + \mathrm{d}_iS. \tag{2.28}$$

The irreversible entropy change is always positive. This is the second law of thermodynamics.

Now we will consider a couple of further examples of the use of these equations. Firstly, consider again the system of a gas in a cylinder, which is kept at constant temperature but which now has a piston moving under friction, see Figure 2.3. The work put into the gas equals $-p\,\mathrm{d}V$, but because the piston moves under friction we need to put more work into the system of gas and cylinder. To keep the gas at constant temperature we need to flux heat between the gas and its environment such that $\mathrm{d}Q = p\,\mathrm{d}V$, from which it follows that

$$\mathrm{d}_eS = \frac{p\,\mathrm{d}V}{T}. \tag{2.29}$$

---

[9] See also Kondepudi, D. & Prigogine, I. (1998) *Modern Thermodynamics*. J. Wiley & Sons, Chichester.

We can now use Eqs. 2.25 and 2.28 to deduce that

$$\mathrm{d}_i S = \frac{\mathrm{d}W + p\,\mathrm{d}V}{T}, \tag{2.30}$$

which is positive if the cylinder moves under friction, Eq. 2.5. In fact, we could write the work put into the system as $\mathrm{d}W = -p_{\mathrm{piston}}\,\mathrm{d}V$, where on compression ($\mathrm{d}V < 0$), the piston pressure p_{piston} is larger than the gas pressure p to overcome the friction. Under those circumstances we have

$$\mathrm{d}_i S = \frac{(p - p_{\mathrm{piston}})\,\mathrm{d}V}{T} > 0. \tag{2.31}$$

For an expanding gas, $\mathrm{d}V > 0$, in a cylinder with friction we would find $p_{\mathrm{piston}} < p$, so that again $\mathrm{d}_i S > 0$.

Another instructive example is the spontaneous flow of heat between two bodies initially at different temperatures. Suppose we have two bodies at different temperatures T_1 and T_2 with $T_1 > T_2$ and that these are briefly put into contact so that a small amount of heat $\mathrm{d}Q$ is exchanged between them, see Figure 2.4. For simplicity, we assume that any volume changes are insignificant in this process.

Heat will flow spontaneously from the hot body to the cold body. The hot body therefore changes its entropy by an amount $\mathrm{d}S_1 = -\mathrm{d}Q/T_1$ and the cool body changes its entropy by an amount $\mathrm{d}S_2 = \mathrm{d}Q/T_2$. The total change of entropy of the two bodies together is therefore (remember that entropy is an extensive quantity)

$$\mathrm{d}S = \mathrm{d}S_1 + \mathrm{d}S_2 = \mathrm{d}Q\left(\frac{1}{T_2} - \frac{1}{T_1}\right) > 0. \tag{2.32}$$

Each subsystem changes its entropy according to $\mathrm{d}S = \mathrm{d}Q/T$, which would qualify as a reversible entropy change, but the system *as a whole* does not exchange heat with its environment and the entropy change of the whole system is therefore irreversible. The second law states that the irreversible entropy change must be positive. For our system this means that we must have $T_1 > T_2$ if $\mathrm{d}Q > 0$. In other words, the heat will not *spontaneously* flow from lower to higher temperatures, something we are well aware of from everyday experience.

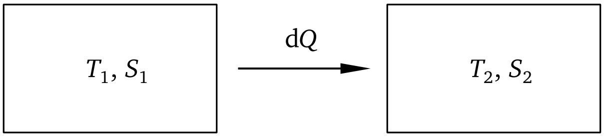


FIGURE 2.4 Exchange of heat between two bodies at different temperatures.

The fact that heat will spontaneously flow from higher to lower temperatures is made precise in another formulation of the second law, called *Clausius' postulate:*

> *A closed system cannot have as sole result the transfer of heat from a body at lower temperature to a body at higher temperature.*

Clausius' postulate is valid for any thermodynamic system, not just a heat conduction system. There are several equivalent formulations of the second law of thermodynamics. The textbook by Fermi[10] gives an overview of these different formulations and why they must be considered equivalent. Clausius' postulate is probably the most intuitive of them.

An important class of processes are those reversible ($\mathrm{d}_i S = 0$) processes for which the system does not exchange heat with the environment, that is $\mathrm{d}Q = T\,\mathrm{d}_e S = 0$. Such processes are called *isentropic* processes or *adiabatic* processes. Many flows in the atmosphere (and nearly all flows in the interior oceans) are, to a large extent, adiabatic. Any fluid parcel undergoing adiabatic transformations conserves its specific entropy. Processes that do change the specific entropy of a parcel are called *diabatic.*

2.4 BOLTZMANN ENTROPY

The microscopic world of molecules is thought to be reversible. If we could reverse the direction of all molecules at once, the world would run backwards. Planets would rotate the other way around, mixed paint would un-mix, candles would go out with the flame retreating into the matchstick, and our thoughts would run backwards – whatever that may mean. Clearly, this is not consistent with experience. The macroscopic world is *not* reversible; it satisfies the second law of thermodynamics, Eq. 2.27. Entropy is Nature's arrow of time.[11]

How can the second law be made consistent with microscopic reversibility? Consider an ideal gas in a thermally isolated cylinder with a wall in the middle, such that the gas is initially located on one side of this wall, see Figure 2.5a. We will now consider what happens if we remove this separating wall.

When the separating wall is removed, the gas will expand to fill the whole cylinder, see Figure 2.5b. The gas expands into the vacuum and therefore will not encounter any inhibition to its expansion; it does not perform any work. Because the cylinder is thermally isolated it too will not exchange any heat in the process of expansion. We conclude from the first law of thermodynamics that the internal energy of the gas does not change on expansion into a vacuum.

---

[10] Fermi, E. (1956) *Thermodynamics*. Dover, New York.

[11] A popular account is in Ruelle, D. (1991) *Chance and Chaos*. Princeton University Press, Princeton.

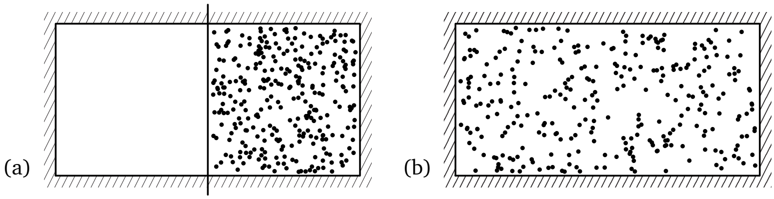


FIGURE 2.5 Expansion of a gas into vacuum. In panel (a) the gas is confined to one side of the thermally isolated cylinder by a separating wall; in panel (b) the separating wall is removed and the gas expands to fill the whole cylinder.

From the first law in the form of Eq. 2.24b we can now calculate the entropy change during expansion into a vacuum as[12]

$$\Delta S = \int_{V_0}^{V_1} \frac{p\,\mathrm{d}V}{T}, \tag{2.33}$$

with V_0 the initial volume and V_1 the final volume. The ideal gas law states that $pV = Nk_BT$, with N the number of molecules and k_B Boltzmann's constant. With the ideal gas law we can perform the integration explicitly and deduce that for an ideal gas expanding into a vacuum

$$\Delta S = k_B \ln{(V_1/V_0)^N}. \tag{2.34}$$

Because there is no heat flow to or from the system, this entropy change must be irreversible; it is the result of the gas occupying a larger volume and because the occupied volume increases ($V_1 > V_0$) we find that the entropy change is positive.

The statement of the second law – that the irreversible entropy change for any system always has to be positive – corresponds in the present setup to $V_1 > V_0$. In other words, the gas will not *spontaneously* occupy a smaller volume, something we are well aware of from everyday experience. If the second law stopped working we would all die of asphyxiation.

Suppose a gas is confined to the cylinder of volume V_1. What is the probability that all N molecules spontaneously reside in sub-volume V_0? If the molecules move in a random fashion through volume V_1 then the chance of finding a particular molecule in sub-volume V_0 is V_0/V_1. If the molecules move independently, then the chance of finding all N molecules at the same time in sub-volume V_0 is $(V_0/V_1)^N$. For realistic values of N this number is unimaginably small, but more importantly, we see that the entropy change in Eq. 2.34 is proportional to the logarithm of the relative probability of the two states.

---

[12] Although the expansion is not a reversible process, we can in principle link the initial and final state with a reversible process and thus calculate the change in state variables between the initial and final states.

Boltzmann realized that this observation is more general: he posited that the entropy of a given system at equilibrium would equal

$$\blacktriangleright \qquad S = k_B \ln \mathscr{W}, \tag{2.35}$$

where $\mathscr{W}$ is the total number of microscopic states that are consistent with the macroscopic constraints on the system. For example, let us say that the gas in the cylinder of volume V_1 has $\mathscr{W}_1$ microscopic states. We need not be concerned here with the details of how to get the precise number of microstates.[13] It now readily follows that, all other things being equal, the number of states corresponding to having all particles in volume V_0 will be $\mathscr{W}_0 = \mathscr{W}_1(V_0/V_1)^N$; each molecule now has access to a fraction V_0/V_1 of the original states, namely those states where the molecule resided in volume V_0. According to the Boltzmann formula, the difference in entropy between those states equals

$$\Delta S = k_B(\ln \mathscr{W}_1 - \ln \mathscr{W}_0) = k_B \ln (V_1/V_0)^N, \tag{2.36}$$

which is the same as Eq. 2.34.

The Boltzmann definition implies that entropy is an extensive quantity. Suppose we have two systems, A and B, with $\mathscr{W}_A$ and $\mathscr{W}_B$ microstates respectively. For each single state of system A, system B can be in $\mathscr{W}_B$ different states. That means that the total number of states of the combined system (A, B) is $\mathscr{W}_A \times \mathscr{W}_B$. The Boltzmann definition of entropy for the combined system then gives

$$S_{(A,B)} = k_B \ln (\mathscr{W}_A \mathscr{W}_B) = k_B \ln \mathscr{W}_A + k_B \ln \mathscr{W}_B = S_A + S_B, \tag{2.37}$$

proving that the Boltzmann S is extensive. In fact, the logarithm in the Boltzmann definition of entropy is the only possible function of the number of microstates for which S is extensive, see Problem 2.1.

The entropy is a measure of the probability of a particular state. Suppose there are $\mathscr{W}$ microstates consistent with the macroscopic constraints. Now consider some macroscopic state i that is consistent with the original macroscopic constraints (for example, defined by some local density field for a given mean density) and suppose there are $\mathscr{W}_i$ microstates that correspond to this particular macrostate. If all microstates are equally probable, then the probability P_i of a particular macrostate i will be $P_i = \mathscr{W}_i/\mathscr{W}$. But according to Boltzmann's entropy definition this probability can therefore be written as

$$\blacktriangleright \qquad P_i = P_0 \exp (S_i/k_B), \tag{2.38}$$

---

[13] This problem is solved in quantum mechanics, where the phase space available to molecules is made up of elementary volumes with a size equal to Planck's constant, h.

where S_i is the entropy of the chosen macrostate and P_0 is fixed to ensure that the sum of probabilities over all possible macrostates add up to 1,

$$P_0 \sum_i \exp(S_i/k_B) = 1. \tag{2.39}$$

So we find that the probability of a particular macrostate is an increasing function of its entropy. Typically, the probability distribution is very sharply peaked around the observed macrostate and any out-of-equilibrium system (such as illustrated in Figs 2.4 and 2.5) will evolve towards a state of maximum entropy because this state is vastly more probable. In this sense the maximum entropy state defines the observed macroscopic state. It also explains why observed macroscopic states are stable: according to the second law, a system can only increase its entropy if left to its own devices. But then a state of maximum entropy cannot evolve any further.

The fact that Boltzmann's microscopic definition of entropy corresponds to the independently derived thermodynamic definition of entropy of Section 2.3 is one of the great triumphs in the history of physics. We will not develop the many beautiful and deep results associated with entropy much further as this is not central to the rest of this book. However, we will return to the equivalence of the microscopic and macroscopic entropies in Section 4.6.

2.5 ENTROPY AND PROBABILITY: A MACROSCOPIC EXAMPLE

Here we present an explicit example of how entropy can be used to distinguish different states of a system and how maximization of entropy selects the thermodynamically stable state. This example illustrates the various connections between entropy, probability, and fluctuations in the thermodynamic state variables.

First consider a system that cannot change its volume, so it cannot exchange work with its environment. A block of metal or a gas in a solid container will be a useful mental picture. Each system has a fixed *heat capacity* C; that is, for any heat input dQ the system changes its temperature by

$$dT = dQ/C. \tag{2.40}$$

We can use this relationship to calculate the entropy change in the system when it receives a finite amount of heat ΔQ. For an infinitesimal transfer of heat, the entropy of the system changes as

$$dS = dQ/T = C\,dT/T. \tag{2.41}$$

This can be integrated to find the total change in entropy when the system changes its temperature from T_a to T_b,

► $$\Delta S_{a\to b} = C \ln(T_b/T_a). \tag{2.42}$$

We find the same expression for entropy change for an ideal gas at constant volume, see Eq. 2.23. Indeed, the heat capacity of a monoatomic ideal gas at constant volume is $3Nk_B/2$.

The initial and final temperatures are related by

$$T_b = T_a + \Delta Q_{a\to b}/C, \tag{2.43}$$

with $\Delta Q_{a\to b}$ the total amount of heat transferred to the system to change its temperature from T_a to T_b. The heat transfer can be either positive, $T_b > T_a$, or negative, $T_b < T_a$. Using this relationship between T_a and T_b, we can write the expression for the entropy change as

$$\Delta S_{a\to b} = C \ln\left(1 + \frac{\Delta Q_{a\to b}}{CT_a}\right). \tag{2.44}$$

For small heat transfers $\Delta Q_{a\to b} \ll C/T_a$ we can use the Taylor expansion $\ln(1+\epsilon) \approx \epsilon$ to find the familiar expression

$$\Delta S_{a\to b} = \Delta Q_{a\to b}/T_a. \tag{2.45}$$

For small heat transfers, or large heat capacity, the temperature of the system will not vary by much and the heat transfer occurs at a nearly constant temperature.

The change of entropy of the system is only dependent on its initial and final temperatures; it does not matter what temperatures the system had in between. Suppose we change the system from temperature T_a to T_c, but first change it from T_a to T_b and then from T_b to T_c. The total entropy change in the combined transformation is

$$\begin{aligned}\Delta S_{a\to b\to c} &= \Delta S_{a\to b} + \Delta S_{b\to c} \\ &= C\ln(T_b/T_a) + C\ln(T_c/T_b) = C\ln(T_c/T_a).\end{aligned} \tag{2.46}$$

This is an important property: we can construct any number of ways a system can achieve equilibrium, but the resulting entropy change is only dependent on the initial and final state. It demonstrates how, for this simple system, the entropy is a function of state.

Now consider two such systems with temperatures T_1 and T_2, respectively. If we bring the two systems into thermal contact, heat will start to flow from the high temperature system to the low temperature system. Because the systems have the same heat capacity, the final temperature for both systems will be the average T_a of the two initial temperatures,

$$T_a = (T_1 + T_2)/2. \tag{2.47}$$

The total entropy change in the two systems is

$$\Delta S = \Delta S_1 + \Delta S_2 = C\ln(T_a/T_1) + C\ln(T_a/T_2). \tag{2.48}$$

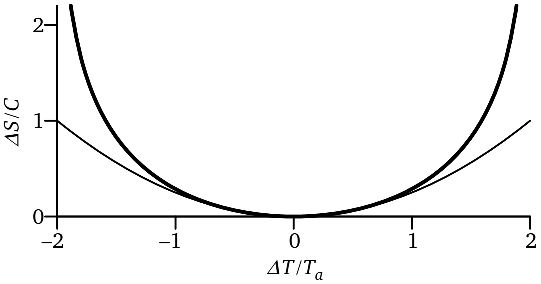


FIGURE 2.6 Entropy gain, ΔS, in units of heat capacity, C, associated with removing a temperature difference ΔT between two systems with an average temperature T_a. Thick line: exact result of Eq. 2.50; thin line: quadratic approximation of Eq. 2.51.

The latter expression can be most easily analyzed if we write

$$T_1 = T_a + \Delta T/2, \qquad T_2 = T_a - \Delta T/2, \tag{2.49}$$

where $\Delta T = T_1 - T_2$. Note that ΔT can be either positive or negative. The total entropy change, on putting the two systems into thermal contact, can then be written

$$\Delta S = C \ln \frac{T_a^2}{T_a^2 - \Delta T^2/4}. \tag{2.50}$$

Figure 2.6 shows this total entropy change as a function of ΔT. As can be seen, the entropy change is always positive; the heat conduction system satisfies the second law of thermodynamics.

For small differences in the initial temperature, the total entropy change can be approximated with the usual Taylor expansion to

$$\Delta S = \frac{C\Delta T^2}{4T_a^2}, \qquad \text{if } \Delta T^2 \ll T_a^2. \tag{2.51}$$

Now using the relationship between entropy and the probability of a particular state, Eq. 2.38, we can write the ratio of the probability of finding the two systems with a temperature difference ΔT to the probability of finding them with no temperature difference as

$$\frac{P(\Delta T)}{P(0)} = \exp\left(-\frac{C\Delta T^2}{4k_b T_a^2}\right). \tag{2.52}$$

This is a Gaussian distribution, with the variance of the temperature fluctuations given by

$$\langle \Delta T^2 \rangle = 2k_B T_a^2/C. \tag{2.53}$$

So if the heat capacity of the systems increases, the temperature fluctuations decrease, as expected. The heat capacity of any system increases with the number of molecules N in the system. For a monoatomic gas, we have $C = 3Nk_B/2$, so the relative size of the temperature fluctuations is

$$\langle \Delta T^2 \rangle / T_a^2 = 4/3N. \tag{2.54}$$

For macroscopic systems, spontaneous macroscopic fluctuations are vanishingly small.

A useful way of rewriting the above fluctuation equation is

$$\frac{C\langle \Delta T^2 \rangle}{4T_a} = \frac{1}{2} k_B T_a. \tag{2.55}$$

The left-hand side is the average energy required to produce a temperature fluctuation of the size ΔT (this energy is $T_a \Delta S$, see Eq. 2.51). We see that this energy is $k_B T_a/2$, which is precisely what would be expected from the equipartition theorem: any accessible degree of freedom in the system contains, on average, $k_B T_a/2$ of energy.

Now consider a large system made up of many such subsystems. If we put heat into the system – perhaps in the sub-systems representing the boundary of the large system – then the entropy of the large system will increase reversibly. The system will be out of equilibrium, however, because the temperature of the subsystems will now be different. If we allow this heat to be redistributed equally amongst all subsystems, the entropy continues to increase. We can generalize Eq. 2.50 to demonstrate that the entropy will increase when the temperature difference between two subsystems decreases through heat exchange. The final equilibrium state must have constant temperature throughout. It will also be the maximum entropy state (amongst all possible redistributions of the heat throughout the subsystems) because any redistribution of heat that increases the temperature difference between two subsystems would reduce the entropy.

The maximum entropy state thus defines the final equilibrium state of the large system when there are no internal constraints that keep the heat from redistributing over the subsystems.

We can impose extra macroscopic constraints on the system that would prevent the heat from being redistributed equally amongst the subsystems. But the large system would then again evolve to a maximum entropy state *consistent with the externally imposed constraints.* The entropy of this constrained system is necessarily lower than the entropy of the unconstrained system. In this sense, entropy measures how many system configurations are consistent with the macroscopic constraints.[14]

---

[14] The configuration freedom of a system can be quantified by its *information entropy,* as defined by Claude Shannon. Edwin Jaynes showed how Shannon's information entropy maps onto the thermodynamic entropy we use here. See Shannon, C. E. (1948) *Bell System Technical Journal* **27**, 379–423, 623–656, Jaynes, E. T. (1957) *Physical Review* **4**, 620–630, Jaynes, E. T. (1965) *American Journal of Physics* **33**, 391–398. See also Problem 4.8.

2.6 ENTROPY AND PROBABILITY: A STOCHASTIC EXAMPLE

A stochastic model for molecules in a cylinder illustrates some microscopic aspects of the second law. Consider a system of molecules in a cylinder that is divided into two equal subvolumes, say ‘left’ and ‘right’. The molecules can move freely between the two subvolumes without preference for direction. We assume that the chance P_{lr} that some molecule moves from left to right per unit time is the same as the chance P_{rl} that some molecule moves from right to left. This is microscopic reversibility,

$$P_{lr} = P_{rl}. \tag{2.56}$$

If at some point there are N_l molecules to the left and $N_r = N - N_l$ molecules to the right then the chance that any molecule moves from left to right is $N_l P_{lr}$ and the chance that any molecule moves from right to left is $N_r P_{rl}$. These chances have a ratio N_l/N_r and if $N_l > N_r$ then there is a bigger chance that a molecule moves from left to right, and vice versa. In equations, we can write the evolution of N_l as

$$\frac{\mathrm{d}N_l}{\mathrm{d}t} = N_r P_{rl} - N_l P_{lr}, \tag{2.57}$$

where the right-hand side is the sum of a gain term and a loss term. We now set $P_{lr} = P_{rl}$ and $N_l + N_r = N$ to find

$$\frac{\mathrm{d}N_l}{\mathrm{d}t} = (N - 2N_l)\, P_{lr}, \tag{2.58}$$

which has the solution

$$N_l(t) = \frac{N}{2} + e^{-2P_{lr}t}\left(N_l(0) - \frac{N}{2}\right). \tag{2.59}$$

So N_l decays exponentially to its equilibrium value $N/2$. The same is true for N_r. This regression to the mean is the second law as applied to this simple system.

We can also analyze this system in terms of its entropy. It is a standard result from combinatorics that the number of states with N_l molecules in the left subvolume is

$$\mathscr{W}(N_l) = \frac{N!}{N_l!\,(N - N_l)!}. \tag{2.60}$$

The situation is equivalent to considering the number of states with N_r molecules in the right volume because $\mathscr{W}(N_r) = \mathscr{W}(N - N_l) = \mathscr{W}(N_l)$. The maximum of $\mathscr{W}$ is achieved for $N_l = N_r = N/2$.

The total number of possible microstates is 2^N because each molecule can be in either the left or right subvolume. This means that the probability of finding N_l molecules in the left subvolume is $P(N_l) = \mathscr{W}(N_l)2^{-N}$. The entropy of the state follows from the Boltzmann definition $S(N_l) = k_B \ln \mathscr{W}(N_l)$. Because of Eq. 2.59, the entropy will on average relax exponentially to its equilibrium value $k_B \ln \mathscr{W}(N/2)$, which is a maximum.

We can quantify this further by using Stirling's approximation for the factorial of a number,

$$n! \approx \sqrt{2\pi n}\,(n/e)^n, \tag{2.61}$$

which is accurate for $n > 10$. Using Stirling's approximation we can write

$$\mathscr{W}(N_l) = \frac{1}{\sqrt{2\pi N}} \frac{N^{N+1}}{N_l^{N_l+1/2}(N-N_l)^{N-N_l+1/2}}. \tag{2.62}$$

We can now use a Taylor expansion around $N_l = N/2$ in terms of the small number $(N_l - N/2)/N$ to find an approximation for the logarithm of $\mathscr{W}$. After some algebra we find

$$\ln \mathscr{W}(N_l) = \ln \mathscr{W}(N/2) - \frac{(N_l - N/2)^2}{N/2} + \ldots \tag{2.63}$$

where the dots represent small terms of higher order in $(N_l - N/2)/N$ and $\mathscr{W}(N/2) = 2^{N+1}/\sqrt{2\pi N}$. Ignoring the small, higher order terms, this can be rewritten as

$$\mathscr{W}(N_l) = \mathscr{W}(N/2) \exp\left(-\frac{1}{2}\left(\frac{N_l - N/2}{\sigma}\right)^2\right), \tag{2.64}$$

with the width of the distribution

$$\sigma = \sqrt{N/4}. \tag{2.65}$$

This means that the number of possible microstates becomes much smaller than its maximum value at $N_l = N/2$ if N_l deviates more than σ from this value. The relative width of the distribution is measured by

$$\sigma/(N/2) = 1/\sqrt{N}. \tag{2.66}$$

This result is analogous to the result for thermal fluctuations in the conduction system, see Eq. 2.54. For macroscopic systems this number is very small: for $N = N_A$, Avogadro's number, the relative width of the distribution is about one part in 10^{12}. There is a very small probability indeed that any observed value of N_l is measurably different from its mean $N/2$.

The entropy can now be approximated by

$$\frac{S(N_l)}{k_B} = \frac{S(N/2)}{k_B} - \frac{1}{2}\left(\frac{N_l - N/2}{\sigma}\right)^2. \tag{2.67}$$

So the entropy is maximum for $N_l = N/2$ and it decays rapidly away for other values of N_l.

The narrowness of the peak of the distribution means that macroscopic variables are well defined and usually limited by the accuracy of the measuring device rather than the microscopic variations in molecule densities. Even though the system does explore the whole phase space of microstates, the macroscopic variables in practice do not deviate from their mean values. Because the maximum entropy state is so much more probable than any other state, the thermodynamic equilibrium state is a very accurate description of the observed state of the system.

PROBLEMS

2.1. Boltzmann definition of entropy. Suppose the entropy S is only a function of the number of microstates $\mathcal{W}$ so that $S = g(\mathcal{W})$. Because S is extensive we need to have

$$g(\mathcal{W}_1\mathcal{W}_2) = g(\mathcal{W}_1) + g(\mathcal{W}_2).$$

Show that the only possibility for g is that $g(\mathcal{W}) \propto \ln \mathcal{W}$. Hint: start by differentiating the above formula twice, first with respect to $\mathcal{W}_1$ and then with respect to $\mathcal{W}_2$ to find a differential equation for g with variable $x = \mathcal{W}_1\mathcal{W}_2$.

2.2. Stirling engine. Atmospheric motion gets its energy from thermodynamic cycles. Here we examine a version of a thermodynamic cycle which forms the basis of the *Stirling engine*. The Stirling engine is a simple, high efficiency engine used in small-scale applications where energy needs to be generated at relatively low excess temperatures (for example from waste heat in an office or factory).

Figure 2.7 shows a schematic of a Stirling engine and the associated thermodynamic cycle of the working fluid, which consists of two isothermal ($dT = 0$) and two isochoric ($dV = 0$) transformations. Assume the working fluid is an ideal gas of n moles.

(i) Show that the total work output L of the cycle ABCDA equals

$$L = nR^\star (T_1 - T_0) \ln (V_1/V_0).$$

(ii) The heating and cooling branches A and C can be performed with no net energy input because heat extracted from the working fluid on the cooling stroke can be stored and re-injected in the working fluid on the

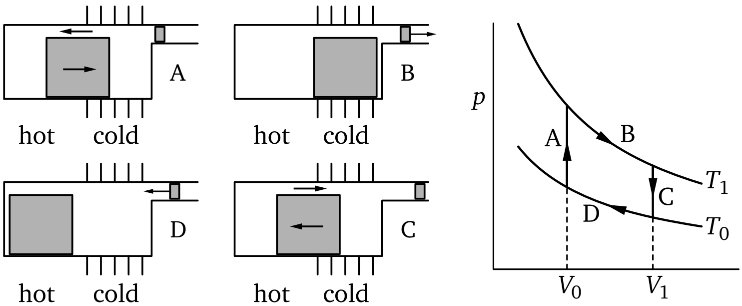


FIGURE 2.7 Left panel: Schematic of four stages in the cycle of a Stirling engine. The engine has a hot and a cold side. The *power* piston (small grey block) moves a flywheel (not-depicted) which in turn moves the *displacer* piston (large grey block). The displacer piston moves the working fluid isochorically (at constant volume, so no work is performed or extracted) between the hot and the cold side (parts A and C of the cycle). When the working fluid is on the hot side it expands and drives the power piston (part B of the cycle). When the working fluid is on the cold side it is compressed by the power piston (part D of the cycle). Right panel: thermodynamic transformations of the working fluid on a pV diagram.

heating stroke. In a real Stirling engine this is accomplished by pushing the working fluid through a *regenerator* in the displacement part of the cycle. A regenerator is often some porous material which stores the heat extracted from the working fluid and which offers little friction to the motion of the fluid (two incompatible features in practice). Show that the total heat input Q during the cycle ABCDA equals

$$Q = nR^\star T_1 \ln (V_1/V_0).$$

(iii) The efficiency η of an engine is the ratio of power output to power input,

$$\eta = L/Q.$$

Show that for the Stirling engine the efficiency equals

$$\blacktriangleright \qquad \eta = 1 - T_0/T_1. \tag{2.68}$$

This equation for the efficiency turns out to be valid for any reversible and frictionless engine, irrespective of the engine design or substance of the working fluid. For real cycles, the efficiency is always less, which is an expression of the second law of thermodynamics.

(iv) Estimate the maximum efficiencies of the climate system and a hurricane.

3

General applications

Now that we have reviewed the first and second laws of thermodynamics we can start to look at some of their basic consequences. This will provide us with relationships that will be used throughout the rest of the book and that are also applicable outside the field of atmospheric physics.

In this chapter we only consider *simple substances.* Simple substances have only one constituent, or are accurately described as if made up of one constituent, as opposed to compound substances where variations in composition would contribute to variations in internal energy through their chemical potentials. Dry air is a good example of a simple substance: the bulk composition of dry air is constant up to great heights. Sea water would be a simple substance if we discarded any variation in the salinity. However, variation in salinity is an important driver of the ocean currents, so any approximation of sea water as a simple substance has limited applicability.

The developments in this chapter are largely of general validity. We often apply these results specifically to an ideal gas. But the more general approach will be advantageous, for example when considering the more complex situation of a vapour–dry air mixture in Chapter 5.

3.1 THERMODYNAMIC POTENTIALS

First we introduce the four *thermodynamic potentials* for simple substances: the internal energy, the enthalpy, the Helmholtz free energy, and the Gibbs function. These thermodynamic potentials are energy-like quantities, which can be defined as transformed versions of the internal energy. Each of them has particular applications. For example, enthalpy plays a key role in the thermodynamics of flow processes, as discussed in Section 3.5. The four thermodynamic potentials lead to the four Maxwell relations, general and non-obvious differential relationships between thermodynamic variables.

The notation for partial differentiation used here and throughout the rest of this book is reviewed in Appendix A.

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

Internal energy

In differential form, the first law of thermodynamics for any simple substance is written as

$$\mathrm{d}u = T\,\mathrm{d}s - p\,\mathrm{d}v. \tag{3.1}$$

The form with specific variables is used here, but all the following derivations can be equally performed using extensive variables. This formulation of the first law is completely general; it does not depend on the particular substance or what conditions the substance is in. One way to read the first law is that the internal energy for any substance is a natural function of the specific entropy s and the specific volume v, so $u = u(s, v)$. The first law then defines the partial derivatives of u with respect to each of the variables,

$$T = \left(\frac{\partial u}{\partial s}\right)_v, \quad p = -\left(\frac{\partial u}{\partial v}\right)_s. \tag{3.2}$$

These relationships are always true, whether the system is a gas in a cylinder, a cup of water, or a stone.

For compound substances, such as sea water (in which salinity plays a role), a chemical mixture, or a system where various phases occur simultaneously the first law is still valid but only in its extensive form, that is $\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V$, and where there are no other fluxes into or out of the system (such as salinity fluxes, in the case of sea water). As long as the system only interacts with its environment through pressure work or heat exchange, the first law in extensive form is the appropriate expression of conservation of total energy in the system.

For repeated derivatives, the order of differentiation is not important. So on differentiating the specific internal energy with respect to s and then with respect to v we get the same result if we perform these differentiations in reverse order. The result is

▶ $$\left(\frac{\partial T}{\partial v}\right)_s = -\left(\frac{\partial p}{\partial s}\right)_v. \tag{3.3}$$

This relationship is called a *Maxwell relation*. It can also be written in extensive variables. The Maxwell relation is general because it is derived from the general form of the first law; it does not depend on the particular substance.

For simple substances, there are four Maxwell relations in all. We will derive these next. For mixtures or systems where the internal energy depends on other processes, there are more Maxwell relations corresponding to double differentiation, including the additional pairs of generalized forces and displacements.

Enthalpy

The specific enthalpy h is defined as

$$\blacktriangleright \qquad h = u + pv. \tag{3.4}$$

The specific enthalpy h is an intensive variable, while enthalpy $H = U + pV$ is an extensive variable. The differential form of the specific enthalpy follows from $\mathrm{d}h = \mathrm{d}u + p\,\mathrm{d}v + \mathrm{d}p\,v$ and the first law in differential form, $\mathrm{d}u = T\,\mathrm{d}s - p\,\mathrm{d}v$. We find that

$$\mathrm{d}h = T\,\mathrm{d}s + v\,\mathrm{d}p. \tag{3.5}$$

Enthalpy is a natural function of entropy and pressure. In mathematical terms, enthalpy is a Legendre transformation of the internal energy, where a function of variables v and s is transformed to a new function of variables p and s.[15] There are several transformed versions of the internal energy (three in all for a substance made up of a single component; for substances made up of several components more transformed versions can be defined); these transformed versions are called *thermodynamic potentials*.

Enthalpy plays a role when there is flow of substance or when considering open systems. Section 3.5 discusses this use of enthalpy. However, like all the thermodynamic potentials, it is initially perhaps easiest to consider enthalpy just as a transformed version of internal energy with some useful properties.

The first derivatives of enthalpy give

$$T = \left(\frac{\partial h}{\partial s}\right)_p, \quad v = \left(\frac{\partial h}{\partial p}\right)_s, \tag{3.6}$$

and the double derivatives give a second Maxwell relation

$$\blacktriangleright \qquad \left(\frac{\partial T}{\partial p}\right)_s = \left(\frac{\partial v}{\partial s}\right)_p, \tag{3.7}$$

which, as before, is generally valid for any simple substance.

---

[15] Legendre transforms occur in several areas of physics, thermodynamic potentials being one of them. Another notable example is the transformation between the Lagrangian and the Hamiltonian of a system in mechanics. The Legendre transform of a function f of variable x is a function g of variable z where z is defined through $z = \mathrm{d}f/\mathrm{d}x$ and g is defined through $f + g = xz$. More explicitly, the definition of z leads to an implicit relation $x = x(z)$ and the Legendre transform g is then $g(z) = x(z)z - f(x(z))$. For functions with curvature of a single sign, the transform is defined everywhere and is invertible: the Legendre transform of g is f, so it is its own inverse.

Helmholtz free energy

The Helmholtz free energy, or simply *free energy,* is another thermodynamic potential. The specific free energy f (in the literature the letter a is also used) is defined as

$$\blacktriangleright \qquad f = u - Ts. \tag{3.8}$$

Specific free energy is an intensive variable; the free energy $F = U - TS$ is an extensive variable. Its differential form again follows from the first law in differential form, analogous to the derivation of the differential of the enthalpy. We find

$$\mathrm{d}f = -s\,\mathrm{d}T - p\,\mathrm{d}v. \tag{3.9}$$

The free energy has as natural variables the temperature and the volume. Free energy plays a central role in statistical mechanics[16] where it is natural to consider systems with a given temperature and volume, so that their free energy is fixed. In atmospheric science free energy is perhaps used less often. We find that

$$s = -\left(\frac{\partial f}{\partial T}\right)_v, \quad p = -\left(\frac{\partial f}{\partial v}\right)_T. \tag{3.10}$$

The corresponding Maxwell relation is:

$$\blacktriangleright \qquad \left(\frac{\partial s}{\partial v}\right)_T = \left(\frac{\partial p}{\partial T}\right)_v. \tag{3.11}$$

Gibbs function

The Gibbs function is the fourth thermodynamic potential for a simple substance. The specific Gibbs function (or *Gibbs free energy,* or sometimes just *the thermodynamic potential*) g is defined as

$$\blacktriangleright \qquad g = u - Ts + pv. \tag{3.12}$$

---

[16] A typical technique of statistical mechanics is to maximize the entropy S of a system (a measure of how probable a particular state is according to the Boltzmann definition of S) for a fixed total energy U. We then introduce a Lagrange multiplier β so that we maximize $S - \beta U$. This can be interpreted as the (negative) free energy if β is interpreted as the inverse temperature. In this way, the microscopic world of statistical mechanics is linked to the macroscopic world of thermodynamics; see also Section 4.6.

The specific Gibbs function is an intensive variable; the Gibbs function $G = U - TS + pV$ is an extensive variable. The differential form follows again by using the differential form of the first law,

$$\mathrm{d}g = -s\,\mathrm{d}T + v\,\mathrm{d}p. \tag{3.13}$$

The Gibbs function has as natural variables the temperature and the pressure. The Gibbs function is unique amongst the thermodynamic potentials in that both its natural variables are intensive. The Gibbs function is used when describing phase transitions; it turns out that for a substance where two phases are in contact, such as water vapour over a water surface, the specific Gibbs function for the two phases is the same; see Sections 3.6 and 5.1. Chapter 8 describes the role of the Gibbs function in describing substances with varying composition: the Gibbs function describes the energetic effects of changing the composition of a substance.

For the first derivatives of the Gibbs function we find

$$s = -\left(\frac{\partial g}{\partial T}\right)_p, \quad v = \left(\frac{\partial g}{\partial p}\right)_T. \tag{3.14}$$

The corresponding Maxwell relation is

► $$\left(\frac{\partial s}{\partial p}\right)_T = -\left(\frac{\partial v}{\partial T}\right)_p. \tag{3.15}$$

The four thermodynamic potentials and their Maxwell relations can be summarized as:

Internal energy u

$du = T\,\mathrm{d}s - p\,\mathrm{d}v$

$$\left(\frac{\partial T}{\partial v}\right)_s = -\left(\frac{\partial p}{\partial s}\right)_v$$

Enthalpy h

$dh = T\,\mathrm{d}s + v\,\mathrm{d}p$

$$\left(\frac{\partial T}{\partial p}\right)_s = \left(\frac{\partial v}{\partial s}\right)_p$$

Helmholtz free energy f

$df = -s\,\mathrm{d}T - p\,\mathrm{d}v$

$$\left(\frac{\partial s}{\partial v}\right)_T = \left(\frac{\partial p}{\partial T}\right)_v$$

Gibbs function g

$dg = -s\,\mathrm{d}T + v\,\mathrm{d}p$

$$\left(\frac{\partial s}{\partial p}\right)_T = -\left(\frac{\partial v}{\partial T}\right)_p$$

Although the thermodynamic potentials can be thought of as merely transformed versions of the internal energy of a system, they do find a more physical justification when considering interactions between the system and its environment. This is formalized in the analysis of the *exergy* of a system, see Appendix B.

3.2 HEAT CAPACITY

We know from experiments that if we put heat into a substance its temperature will usually increase. In fact, this observation is the source of the confusion of heat and temperature in informal language. Moreover, with the Boltzmann constant any temperature can be expressed as an energy, so heat and temperature can both be expressed as energies. The fundamental difference is that heat represents the microscopic transfer of internal energy from one substance to another while temperature is a measure of how much energy is stored as molecular kinetic energy. Temperature measures a state; heat measures a change of state.

Let us consider a gas: the way we put heat into a gas influences how much its temperature will increase. For example, if we allow the gas to expand while we put heat into it, some of the heat energy will be transformed to mechanical work (by $p\,\mathrm{d}V$) and therefore less energy will be available to warm up the gas compared to the situation where we would keep the volume constant.

The amount of heat input δQ per unit change in temperature δT is called the *heat capacity* and it is denoted by a C (units $\mathrm{J\,K^{-1}}$). We can also divide by the mass of the system to get the specific heat capacity c (units $\mathrm{J\,kg^{-1}\,K^{-1}}$). In equations,

$$C = \frac{\delta Q}{\delta T} = T\frac{\mathrm{d}S}{\mathrm{d}T}, \quad c = \frac{\delta q}{\delta T} = T\frac{\mathrm{d}s}{\mathrm{d}T}, \tag{3.16}$$

where the limit of small δQ and δq is implied. These equations are not meant to imply a differentiation of Q or q with respect to T, as heat input is not a state function; entropy is a state function so it can be differentiated with respect to T.

As noted above, we need to make clear under what constraints the heat is being put into the system. For example, we expect the heat capacity at constant volume to be smaller than the heat capacity at constant pressure. The specific heat capacities at constant volume c_v and at constant pressure c_p are defined as

► $$c_v = T\left(\frac{\partial s}{\partial T}\right)_v, \quad c_p = T\left(\frac{\partial s}{\partial T}\right)_p. \tag{3.17}$$

These two heat capacities are necessarily related: if we heat up a substance at a constant pressure (while allowing it to expand) then its temperature will have changed by $\delta T = \delta q/c_p$ and its specific volume will have changed by $\delta v = (\partial v/\partial T)_p\,\delta T$ or

$$\delta v = \left(\frac{\partial v}{\partial T}\right)_p \frac{\delta q}{c_p}. \tag{3.18}$$

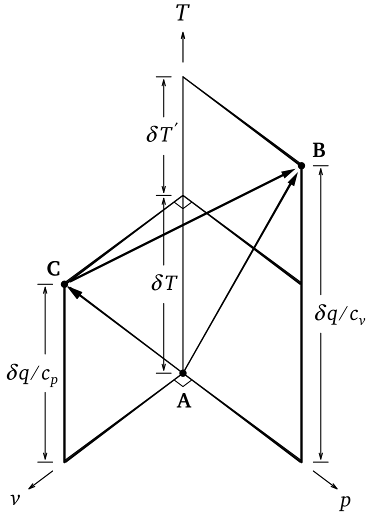


FIGURE 3.1 Illustration on a *pvT* diagram of the relationship between heat capacities at constant volume and at constant pressure. Process A→B is isovolumetric, process A→C is isobaric, and process C→B is isentropic.

We can now compress the substance isentropically (no further heat is added to or removed from the substance) until it has returned to its original volume. This would lead to a further temperature change $\delta T'$, which is related to δv by

$$\delta T' = -\left(\frac{\partial T}{\partial v}\right)_s \delta v. \tag{3.19}$$

At the end of this process we have added heat δq and we are back at the original (specific) volume. This is illustrated in Figure 3.1: going straight from A to B in the *pvT* diagram is the same as going from A to B via C. The total temperature change (going from A to B in the diagram) is therefore by definition equal to $\delta q/c_v$ and it is equal to $\delta T + \delta T'$. So we find that

$$\frac{\delta q}{c_v} = \delta T + \delta T' = \frac{\delta q}{c_p}\left(1 - \left(\frac{\partial v}{\partial T}\right)_p \left(\frac{\partial T}{\partial v}\right)_s\right). \tag{3.20}$$

Now, multiplying by $c_v c_p/\delta q$, substituting $c_v = T(\partial s/\partial T)_v$, and using the first Maxwell relation, Eq. 3.3, we find that

$$c_p = c_v + T\left(\frac{\partial v}{\partial T}\right)_p \left(\frac{\partial p}{\partial s}\right)_v \left(\frac{\partial s}{\partial T}\right)_v. \tag{3.21}$$

Combining the two partial derivatives at constant volume we arrive at

$$c_p = c_v + T\left(\frac{\partial v}{\partial T}\right)_p \left(\frac{\partial p}{\partial T}\right)_v. \tag{3.22}$$

This relationship is valid for any substance. An analogous relationship can be found for the extensive heat capacities C_V and C_p.

A quicker but less intuitive derivation follows from considering s as a function of T and v. Its first order variations then satisfy

$$\mathrm{d}s = \left(\frac{\partial s}{\partial T}\right)_v \mathrm{d}T + \left(\frac{\partial s}{\partial v}\right)_T \mathrm{d}v. \tag{3.23}$$

This leads to

$$\left(\frac{\partial s}{\partial T}\right)_p = \left(\frac{\partial s}{\partial T}\right)_v + \left(\frac{\partial s}{\partial v}\right)_T \left(\frac{\partial v}{\partial T}\right)_p, \tag{3.24}$$

which is equivalent to Eq. 3.22, using the third Maxwell relation, Eq. 3.11.

In the next section we will consider these results as applied to ideal gases. For liquids or solids, these results can be simplified: under normal circumstances (away from the critical point and for pressure variations that are typical for the atmosphere) the volume changes of a liquid or solid are so small as to contribute very little to the exchange of work with its environment. The first law for a liquid or a solid can therefore be approximated as $\mathrm{d}u = \mathrm{d}q$. Given that the difference in c_p and c_v is due to the exchange of work with the environment in the isobaric case, the two heat capacities are virtually the same for liquids or solids. We therefore find for liquids or solids that

$$c_p \approx c_v = c \tag{3.25}$$

and the internal energy and enthalpy for liquids or solids can be approximated as

$$u \approx h = u_0 + cT. \tag{3.26}$$

See Problem 3.6 for an application to the specific entropy for liquids or solids. The fact that work is not very important to the internal energy budget of liquids or solids does not mean that volume changes are dynamically irrelevant; all motions in liquids are ultimately driven by volume changes of the liquid – think of the convective motions in a pan of water heated from below. All that Eq. 3.26 says is that such volume changes are mainly driven by temperature changes, not by work performed by or on the liquid.

The definition of the heat capacities, Eq. 3.17, can be used to rewrite the entropy difference between two states at different temperatures but at fixed pressure:

$$S(T) - S(T_0) = \int_{T_0}^{T} \frac{C_p}{T}\,\mathrm{d}T. \tag{3.27}$$

The quantity $C_p\,\mathrm{d}T$ is determined experimentally by monitoring the heat input required to change the temperature by a certain amount. This equation can then be numerically integrated to find the entropy difference between two states. In fact, this integration can be done from absolute zero (where entropy is zero, according to the third law), through the melting point and boiling point of a substance to find the entropy of a gas. Such experiments have been used to verify the *Sackur–Tetrode equation*, which gives the absolute value of the entropy of an ideal gas from quantum mechanical arguments.

3.3 PROPERTIES OF IDEAL GASES

For an ideal gas we have $pv = RT$, so that the relationship between the two heat capacities becomes

$$\blacktriangleright \qquad c_p = c_v + R. \tag{3.28}$$

It is a beautiful result from statistical mechanics that for ideal gases we have[17]

$$c_v = \begin{cases} 3R/2 & \text{for a monoatomic ideal gas,} \\ 5R/2 & \text{for a diatomic ideal gas.} \end{cases} \tag{3.29}$$

The corresponding equations for c_p can be found from $c_p = c_v + R$. Using the fact that dry air is predominantly diatomic and that $R = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$ we find that for dry air

$$c_v = 717\,\mathrm{J\,kg^{-1}\,K^{-1}}, \tag{3.30}$$

$$c_p = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}. \tag{3.31}$$

In general, the heat capacity of an ideal gas at constant volume is given by

$$c_v = fR/2, \tag{3.32}$$

with f the number of degrees of freedom per molecule. This is the equipartition theorem, which states that, at equilibrium, all accessible degrees of freedom store an amount of energy equal to $k_B T/2$ per molecule or, equivalently, $RT/2$ per unit mass; see also Sections 1.2, 2.5, and 4.6. A monoatomic molecule has three degrees of freedom associated with kinetic energy in the

---

[17] A monoatomic molecule is made of one atom and a diatomic molecule is made of two atoms.

x, y and z directions. A diatomic molecule would have an additional three degrees of freedom associated with the rotation about three axes. However, rotation around the lengthwise axis does not contribute to the energy because the molecule is symmetric for such rotations, which prevents coupling between the translational modes and this rotational mode. So we are left with three translational modes and two rotational modes, $f = 5$. See Problem 3.2 for an application to water vapour.

All this assumes that the kinetic energy does not couple to the vibrational energy of the molecule: in quantum theory, there is a discrete minimum energy required to excite the first vibrational mode above its ground state.[18] It turns out that for atmospheric gases at typical temperatures, this energy is larger than $k_BT/2$; temperatures of thousands of Kelvins are typically required to significantly excite the vibrational modes. From a classical point of view, the relative excitation of the vibrational modes is expected to be proportional to the Boltzmann factor $\exp(-\Delta E/k_BT)$ with ΔE the typical excitation energy of a vibrational mode (see Section 4.6). So with increased temperature the vibrational degrees of freedom become more occupied. We therefore expect that the heat capacity will increase with temperature. However, in the range of temperatures encountered in the atmosphere, the heat capacity of dry air and most other ideal gases can be considered constant.

The first law in differential form $\mathrm{d}u = T\,\mathrm{d}s - p\,\mathrm{d}v$ and the definition of enthalpy in differential form $\mathrm{d}h = T\,\mathrm{d}s + v\,\mathrm{d}p$ imply that

$$c_v = \left(\frac{\partial u}{\partial T}\right)_v, \quad c_p = \left(\frac{\partial h}{\partial T}\right)_p. \tag{3.33}$$

For an ideal gas, the specific heat capacity at constant volume is constant. This implies that $u = u_0 + c_vT + f(v)$ where f is a function of the specific volume only. We now show that for an ideal gas this function vanishes. From the first law in differential form it follows that

$$\left(\frac{\partial u}{\partial v}\right)_T = T\left(\frac{\partial s}{\partial v}\right)_T - p = T\left(\frac{\partial p}{\partial T}\right)_v - p, \tag{3.34}$$

where the third Maxwell relation, Eq. 3.11, is used. This is a general relationship for the variation of the internal energy with volume. Substituting the ideal gas law in the right-hand side of this expression, we find $(\partial u/\partial v)_T = 0$. In other words, the internal energy of an ideal gas is a function of temperature only; this is also known as *Joule's law*. In fact, an ideal gas can be *defined* as a gas which satisfies Boyle's law and Joule's law. We now find for an ideal

---

[18] The vibrational levels of a quantum oscillator are separated by an amount of $\hbar\sqrt{(k/M)}$ with $\hbar$ Planck's constant, k the stiffness of the molecular bond (the spring constant in classical mechanics) and M the reduced mass of the constituent atoms. For larger atomic masses the distance between the vibrational modes reduces and will therefore be more easily accessible at room temperature.

gas that

▶ $$u = c_v T, \tag{3.35}$$

where the integration constant u_0 has been set to zero as it has no physical consequences. A similar argument shows that, for an ideal gas, enthalpy is a function of temperature only and we find that

$$h = c_p T, \tag{3.36}$$

a result which also follows from the definition of enthalpy $h = u + pv$ applied to an ideal gas.

We are now in a position to calculate the specific entropy for an ideal gas. From the fourth Maxwell relation, Eq. 3.15, we have

$$\left(\frac{\partial s}{\partial p}\right)_T = -\left(\frac{\partial v}{\partial T}\right)_p = -\frac{R}{p}, \tag{3.37}$$

where the ideal gas law $pv = RT$ is used to calculate $(\partial v/\partial T)_p$. We also have

$$\left(\frac{\partial s}{\partial T}\right)_p = \frac{c_p}{T}. \tag{3.38}$$

So for an ideal gas we have

$$ds = c_p \frac{\mathrm{d}T}{T} - R\frac{\mathrm{d}p}{p}. \tag{3.39}$$

Because R and c_p are constant for an ideal gas we can integrate this expression to find

▶ $$s = c_p \ln{(T/T_0)} - R \ln{(p/p_0)}. \tag{3.40}$$

The integration constants T_0 and p_0 are arbitrary here as they change the specific entropy by an additive constant. The entropy constant for an ideal gas can be fixed using quantum mechanics (the Sackur–Tetrode equation); in this book we will not encounter situations where the entropy constant plays a role.

The ideal gas law can now be used to express the specific entropy as a function of any pairing of variables; the expression for entropy of a monoatomic gas in terms of temperature and volume has already been derived in Section 2.3. The specific entropy of an ideal gas as a function of pressure and density is

$$s = c_v \ln{(p/p_0)} - c_p \ln{(\rho/\rho_0)}. \tag{3.41}$$

This particular form is often applied to well-mixed gases.

When an ideal gas is *well mixed* the specific entropy will be constant throughout the gas because specific entropy is a particle tracer (we assume any diabatic processes such as heat diffusion are weak). This may seem counterintuitive but the situation is analogous to stirring paint: suppose we started off with red and white paint and mix them – the final result is a homogeneous pink paint. In such situations Eq. 3.41 is valid throughout the gas, with constant values of s, ρ_0 and p_0. A well-mixed, adiabatic, ideal gas will therefore have the equation of state

$$p = k\rho^{c_p/c_v} \tag{3.42}$$

with k a constant, dependent on the specific entropy and reference pressure. The ratio c_p/c_v is often denoted γ. For a diatomic ideal gas (such as air, approximately) we have

$$\gamma = \frac{c_p}{c_v} = \frac{7}{5}. \tag{3.43}$$

Monoatomic gases have $\gamma = 5/3$.

The speed of sound $\mathscr{C}$ in a compressible fluid is given by

$$\mathscr{C}^2 = \left(\frac{\partial p}{\partial \rho}\right)_s. \tag{3.44}$$

This partial derivative can be easily found by taking the derivative of Eq. 3.42 with respect to density. We find, using the ideal gas law,

$$\mathscr{C} = \sqrt{\gamma RT}. \tag{3.45}$$

So the speed of sound in an ideal gas is a function of temperature only. For typical atmospheric values of temperature close to the surface we find values of $\mathscr{C}$ between about 330 and 350 $\mathrm{m\,s^{-1}}$. This leads to the well-known trick for estimating the distance to a lightning flash: divide the number of seconds between the flash and the rumble of thunder by three to get the distance in kilometres.

3.4 VAN DER WAALS' GASES

An instructive example of the use of thermodynamic potentials is the derivation of van der Waals' equation of state, Eq. 1.17. We start from the Helmholtz free energy, F_{ideal}, for a volume of ideal gas and make two modifications to it, representing the finite size of real molecules and the average attractive force between real molecules. We then use

$$p = -\left(\frac{\partial F}{\partial V}\right)_T \tag{3.46}$$

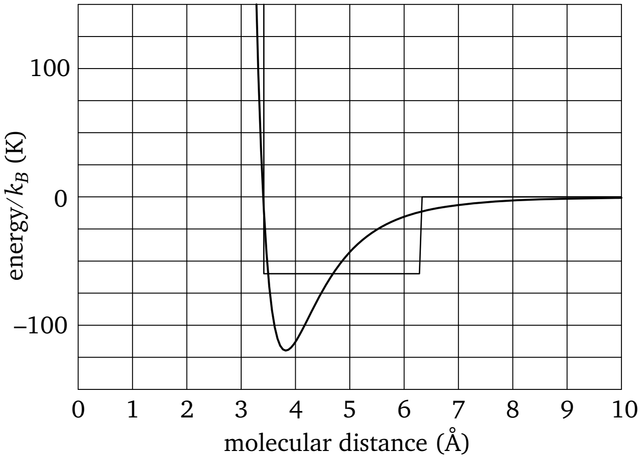


FIGURE 3.2 Interaction potential between two argon atoms. The interaction energy is expressed as a temperature (in Kelvin), and the energy in Joules can be found by multiplying by k_B. The potential minimum thus corresponds to 1.7×10^{-21}J. The horizontal distance is in units of $1\,\text{Å} = 10^{-10}$m. The interaction potential is fairly similar to that for N_2 and O_2 molecules. The thin line is the approximating square well potential, which is discussed in the text.

to calculate the pressure for this non-ideal gas. We can explicitly write down an expression for the free energy of an ideal gas, F_{ideal}, but presently we only require that

$$p_{\text{ideal}}(V,T) = -\left(\frac{\partial F_{\text{ideal}}}{\partial V}\right)_T = \frac{nR^{\star}T}{V}, \tag{3.47}$$

which follows from the ideal gas law.

The ideal gas is made up of non-interacting point particles. In a real gas molecules will have a finite effective size, sometimes called the *hard core.* The hard core is a result of the repulsive force at short distance between two molecules. This repulsive force is equivalent to a strong increase in the interaction potential when two molecules come close to each other, see Figure 3.2. The repulsive force becomes so large at small distances that the molecules effectively cannot come closer than some finite distance, r_c. Each molecule therefore has less volume available than the volume V of the gas. The excluded volume is the total volume of the hard cores of the molecules and can be written as V_cN, with V_c the effective volume of the core and N the number of molecules. The effective core volume is $V_c = (2/3)\pi r_c^3$.[19] This

---

[19] The effective core volume is half the size of a single core because the first molecule in a box does not experience any excluded volume, the second molecule is excluded from $(4/3)\pi r_c^3$, the third molecule is excluded from $2\,(4/3)\pi r_c^3$, etc. So the Nth molecule is excluded from $(N-1)\,(4/3)\pi r_c^3$ of volume. For large N, the average excluded volume per molecule is therefore $(2/3)\pi r_c^3$.

excluded volume is customarily written as bn, with n the number of mols and $b = V_c N_A$ a molecule size parameter which is specific for each gas.

In a real gas, two molecules will attract each other on average. This attractive force, the so-called *van der Waals force*, is due the induced electric dipole moments in the molecules. It corresponds to a negative interaction potential energy compared to the non-interacting ideal gas. Let us assume that only molecules within a certain distance of each other, molecules in the *potential well*, interact with each other and they do so with an average interaction potential energy $-E_{\text{well}}$, see Figure 3.2. The total potential energy difference ΔE_p between the real gas and the ideal gas can then be written as

$$\begin{aligned}\Delta E_p &= -E_{\text{well}} \times \text{molecules per potential well} \times N \\ &= -E_{\text{well}}\, V_{\text{well}} \frac{N}{V} N, \end{aligned} \tag{3.48}$$

where V_{well} is the effective volume of the potential well. Customarily, this total interaction potential energy is written as

$$\Delta E_p = -an^2/V, \tag{3.49}$$

with n the number of mols and $a = E_{\text{well}} V_{\text{well}} N_A^2$, a molecule interaction parameter, which is specific for each gas. The size of the interaction potential energy decreases with density; at very small densities it can be ignored and we are back at the ideal gas situation. The approximation of this interaction potential using the finite well can be fairly straightforwardly generalized to realistic interaction potentials;[20] the resulting equation is the same.

We can now include the effects of the finite volume and the interaction potential in the free energy of the ideal gas to find an approximation of the free energy for a more realistic gas:

$$F(V, T) = F_{\text{ideal}}(V - nb, T) - an^2/V. \tag{3.50}$$

It now follows that for a gas with this free energy the pressure is given by

$$p = -\left(\frac{\partial F}{\partial V}\right)_T = \frac{nR^\star T}{V - bn} - a\left(\frac{n}{V}\right)^2, \tag{3.51}$$

which is van der Waals' equation.

For typical atmospheric parameters the excluded volume and the interaction potential can be ignored: for N_2 we have $a = 0.1370\,\text{Pa}\,\text{m}^6\,\text{mol}^{-2}$ and $b = 0.0387 \times 10^{-3}\text{m}^3\,\text{mol}^{-1}$, which at standard atmospheric values corresponds to a correction in the pressure of 0.8 parts per thousand and a correction in the volume of 1.6 parts per thousand. If we use the equation of state to calculate the air density at standard pressure and temperature, then

---

[20] See, for example, Kittel, C. & Kroemer, H. (1980). *Thermal physics*, 2nd edn. W. H. Freeman, New York.

the ideal gas law underestimates the density of air by $1\,\mathrm{g\,m^{-3}}$ compared to van der Waals' equation. We can safely use the ideal gas law for all practical applications.

The above derivation uses an interaction potential based on a constant average density throughout the gas, the so-called *mean field method.* This ignores correlations between the positions of individual molecules (the attractive force makes molecules preferentially cluster together on average). Such correlations can be taken into account in a systematic way to find more accurate approximations to the free energy. Typically, the pressure or the free energy is written as a power series in the density. Such an expansion is called a *virial expansion*. The virial expansion for the pressure has the form

$$p = \rho RT \left(1 + B_2(T)\rho + B_3(T)\rho^2 + \ldots\right) \tag{3.52}$$

where the B_i for $i = 2, 3\ldots$ are virial coefficients, which are functions of the temperature. For small densities the equation of state reduces to that of an ideal gas. For a van der Waals gas, the second and third virial coefficients are given by

$$B_2 = \frac{b}{\mu} - \frac{a}{\mu^2 RT}, \quad B_3 = \frac{b^2}{\mu^2}, \tag{3.53}$$

with μ the molar mass of the substance, R its specific gas constant, and a and b its van der Waals coefficients, see Problem 3.9.

At high densities the free energy cannot be determined as a modification of the free energy of the ideal gas. For extremely high densities the free energy is dominated by entropy variations due to the possible configurations of densely packed molecules. Here an ordered state provides most freedom of motion to the molecules and will therefore correspond to the largest entropy: the molecules undergo an entropy driven phase transition to a crystalline state.

3.5 OPEN SYSTEMS: ENTHALPY FLUX

The first law expresses conservation of energy for closed systems, systems that interact with their environment through heat and work exchange but not through mass exchange. An open system allows mass exchange. Figure 3.3 shows a schematic of an open system. For simplicity the open system has one inlet and one outlet. We allow an amount of mass δM_1 to flow in through the inlet and an amount δM_2 to flow out of the outlet. The system gains an amount of energy equal to $U_1 = \delta M_1\, u_1$, with u_1 the specific internal energy of the matter flowing into the inlet port. But to make this flow enter the system an amount of work equal to $p_1\, \mathrm{d}V_1 = p_1\, \delta M_1\, v_1$ has to be performed on the system, with p_1 the pressure at the inlet. The total energy gain of the system is therefore $\delta M_1\,(u_1 + p_1 v_1) = \delta M_1\, h_1$ with h_1 the specific enthalpy at the inlet. Similarly, the total energy loss is $\delta M_2\, u_2$ plus the work $p_2\, \mathrm{d}V_2$ done by the system to expel the fluid. The total energy loss of the system is

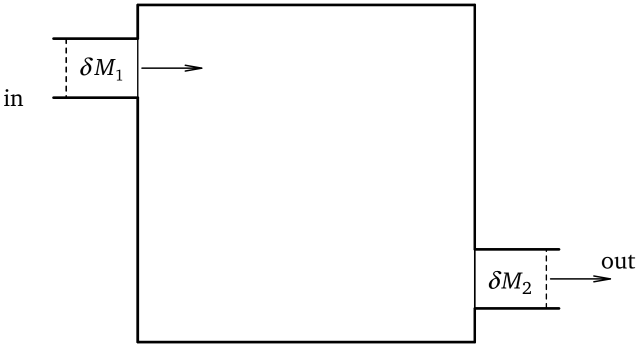


Figure 3.3 Open system with one inlet and one outlet port.

therefore $\delta M_2\, h_2$, with h_2 the specific enthalpy at the outlet. The total energy change δU of the open system is

$$\delta U = \delta M_1\, h_1 - \delta M_2\, h_2. \tag{3.54}$$

So when matter flows into and out of a system, its energy change is due to the flux of specific enthalpy into and out of the system, not the flux of specific internal energy. This is the key application of enthalpy. The difference between specific enthalpy and specific internal energy is pv, which is called the *flow work*. Any matter flow contributes internal energy as well as flow work to the energy budget.

The above argument can be generalized to a system with several inlets and outlets. If the mass flux δM_i through port i of the system is considered positive when directed outward, then energy change due to the enthalpy flux through any number of ports is

$$\delta U = -\sum_i \delta M_i\, h_i = -\sum_i \rho_i\, \mathscr{U}_i h_i A_i\, \delta t, \tag{3.55}$$

with $\rho_i, A_i, \mathscr{U}_i$ the density, port cross-sectional area, and outward flow velocity at port i, and δt the time over which the energy change is considered. More generally we can choose our system V to be some fixed volume in three-dimensional space with boundary area A and express the rate of change of energy due to the enthalpy flux through A. If we denote $\boldsymbol{U} = (\mathscr{U}, \mathscr{V}, \mathscr{W})$ the three-dimensional flow velocity and $\hat{\boldsymbol{n}}$ the unit vector normal to the surface A (by convention $\hat{\boldsymbol{n}}$ points outward), then we can see that the rate of internal energy change $\mathrm{d}U/\mathrm{d}t$ of the volume due to the enthalpy flux has to be

$$\frac{\mathrm{d}U}{\mathrm{d}t} = -\int_A (\rho \boldsymbol{U} h) \cdot \hat{\boldsymbol{n}}\, \mathrm{d}A. \tag{3.56}$$

Because the volume choice is arbitrary, we can use Gauss' theorem to transform this into a local change of specific internal energy due to the enthalpy flux,

$$\frac{\partial(\rho u)}{\partial t} = -\nabla \cdot (\rho \boldsymbol{U} h). \tag{3.57}$$

The local change of volumetric energy density (ρu) is due to a divergence of the enthalpy flux $(\rho \boldsymbol{U} h)$.[21]

3.6 LATENT HEAT

Consider the situation where the open system is made up of an open container with a liquid and the associated vapour above it. The two subsystems, liquid and vapour, can exchange matter through evaporation and condensation, and so can each individually be considered as an open system exchanging enthalpy. Now, some mass δM evaporates from the liquid to the vapour. According to our analysis this process transfers some enthalpy from the liquid to the vapour. The vapour gains an enthalpy of $\delta M\, h_v$, while the liquid loses an enthalpy of $\delta M\, h_l$. Because in general $h_v \neq h_l$ (normally $h_v > h_l$), the enthalpy budget for such an evaporation is not closed. Indeed, it costs energy to evaporate a liquid to a vapour. This energy is associated with the energy barrier molecules have to overcome to escape the intermolecular attractive forces in the liquid. The energy required is $\delta M\, (h_v - h_l)$, which is proportional to the specific enthalpy difference between the vapour and the liquid. This enthalpy difference defines

$$\blacktriangleright \qquad L = h_v - h_l \tag{3.58}$$

where L is called the *enthalpy of vaporization* (units $\mathrm{J\,kg^{-1}}$). In atmospheric science the enthalpy of vaporization is usually called the *latent heat of evaporation.* The energy associated with evaporation or condensation processes is called *latent heat.*

If we evaporate a mass of liquid at a fixed temperature T then all the heat we put into the liquid δQ is used for evaporation. The enthalpy of evaporation can then be written as an entropy difference between the vapour and liquid states,

$$\delta M\, (h_v - h_l) = \delta Q = \delta M\, T\, (s_v - s_l). \tag{3.59}$$

So the evaporation of the liquid corresponds to an increase in entropy. This equation is only valid at equilibrium; that is, if the evaporation occurs at a fixed temperature, T. We can now rewrite this equation as

$$h_l - T\, s_l = h_v - T\, s_v. \tag{3.60}$$

---

[21] The expression $\rho \boldsymbol{U} h$ is really a flux density, that is, a flux per unit area. Here, and throughout the rest of this book, we will use the term 'flux' to mean either 'total flux' or 'flux density', depending on the context.

We therefore find that at equilibrium the specific Gibbs functions $g = h - Ts$ for the liquid and vapour phases are the same,

$$g_l = g_v. \tag{3.61}$$

This important aspect of phase transitions is covered in more detail in Section 5.1.

Note that at equilibrium, the entropy change due to evaporation is reversible. To achieve evaporation at a fixed temperature, we need to put in energy δQ, as in Eq. 3.59, to overcome the enthalpy barrier between the vapour and the liquid. We can now extract the same amount of energy at the same temperature to to reverse the process through condensation.

At equilibrium the latent heat of evaporation is a function of temperature only because at equilibrium the vapour pressure is a function of temperature, see Section 5.1. However, to a very good approximation this is also true in more general cases. From Eq. 3.36 we see that if we consider the vapour an ideal gas, its enthalpy h_v is a function only of temperature. From Eq. 3.26 we see that this is also true to a good approximation for the enthalpy h_l of a liquid. From the definition of L in Eq. 3.58 it then follows that for both the vapour and the liquid

$$\frac{\mathrm{d}h_{v,l}}{\mathrm{d}T} = \left(\frac{\partial h_{v,l}}{\partial T}\right)_p = T\left(\frac{\partial s_{v,l}}{\partial T}\right)_p = c_{p\,v,l}. \tag{3.62}$$

Combining these results with the definition of the latent heat of evaporation, Eq. 3.58, we find that to a good approximation

► $$\frac{\mathrm{d}L}{\mathrm{d}T} = c_{pv} - c_{pl}. \tag{3.63}$$

This equation is sometimes called *Kirchhoff's equation.* As for all substances $c_{pl} > c_{pv}$, we find that the latent heat of evaporation is a decreasing function of temperature. This is expected because at the critical temperature, the temperature above which vapour and liquid become indistinguishable, the latent heat of evaporation vanishes.

So we find that the specific latent heat of evaporation can be approximated as

$$L = L_0 - (c_{pl} - c_{pv})(T - T_0), \tag{3.64}$$

where it is assumed that both heat capacities are constant over the temperature range of interest, and where L_0 is the value of the latent heat at some reference temperature T_0. Figure 3.4 illustrates how well this approximation works for water when compared to measured values. The approximation with a reference temperature of 25°C is better than one part in a thousand for temperatures between 0°C and 50°C and better than one part in two hundred for

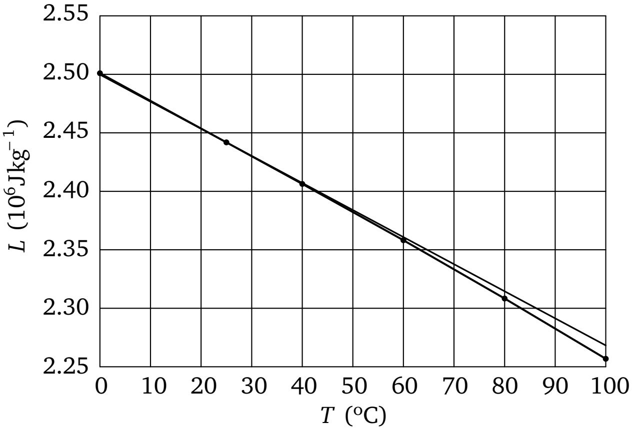


FIGURE 3.4 Points: measured latent heat of evaporation for water. Thin line: linear approximation of Eq. 3.64 with $T_0 = 25°\text{C}$. The other parameters in the linear approximation are $L_0 = 2.44 \times 10^6\,\text{J}\,\text{kg}^{-1}$, $c_{pl} = 4180\,\text{J}\,\text{kg}^{-1}\,\text{K}^{-1}$, and $c_{pv} = 1865\,\text{J}\,\text{kg}^{-1}\,\text{K}^{-1}$. Data from Marsh, K. N., ed. (1987) *Recommended reference materials for the realization of physicochemical properties* Blackwell, Oxford.

temperatures up to 100°C. A typical value often used for quick calculations is

$$L = 2.5 \times 10^6\,\text{J}\,\text{kg}^{-1}, \tag{3.65}$$

which is very close to the value of the latent heat of evaporation at 0°C.

The above discussion of evaporation from liquid to vapour can be applied to the melting from solid to liquid as well; we thus find the so-called *enthalpy of fusion*. For water at the melting point of 0°C at standard pressure the enthalpy of fusion is $L_m = 0.33 \times 10^6\,\text{J}\,\text{kg}^{-1}$. Similarly, we find an *enthalpy of sublimation* for the transition from solid to vapour; for water (ice) at 0°C it is $L_{\text{ice}} = 2.826 \times 10^6\,\text{J}\,\text{kg}^{-1}$.

3.7 TURBULENT ENERGY FLUXES

Apart from local sources, the budget of internal energy for any volume of air is determined by the enthalpy fluxes into and out of the volume. The enthalpy flux $\boldsymbol{F}_h$ is given by

$$\boldsymbol{F}_h = \rho \boldsymbol{U} h. \tag{3.66}$$

It is often convenient to decompose the flux into a time average contribution and a fluctuation from the time average – a so-called *Reynolds decomposition*. So we write

$$h = \overline{h} + h', \qquad \rho\boldsymbol{U} = \overline{\rho\boldsymbol{U}} + (\rho\boldsymbol{U})', \tag{3.67}$$

with the overbar denoting the time average, and the prime denoting the fluctuation. This decomposition has the property that the time average of the fluctuation is zero,

$$\overline{h'} = 0, \qquad \overline{(\rho \boldsymbol{U})'} = 0. \tag{3.68}$$

Using this property, we can write the enthalpy flux as

$$\rho \boldsymbol{U} h = \overline{\rho \boldsymbol{U}}\,\overline{h} + \overline{(\rho \boldsymbol{U})' h'}. \tag{3.69}$$

The other terms in the expanded product vanish. The first term is called the mean flux, the second term is called the turbulent flux.

In the planetary boundary layer the turbulent flux dominates. In the turbulent flux, the term associated with the vertical transport of mass and enthalpy dominates. In this case we find a vertical enthalpy flux F_h of

$$F_h = \overline{(\rho w)' h'}. \tag{3.70}$$

A capital H is often used to denote the vertical enthalpy flux, but we will not use that notation here to avoid confusion with the enthalpy proper.

For dry air we can use the ideal gas expression, $h' = c_p T'$. Also, the density is often replaced by its mean so that only the fluctuating vertical wind contributes to the fluctuating vertical mass transport; that is, $(\rho w)' = \overline{\rho} w'$. The resulting expression for the vertical enthalpy flux is[22]

▶ $$F_h = \overline{\rho} c_p \overline{w' T'}. \tag{3.71}$$

In this form, the dry-air enthalpy flux is called the *sensible heat flux,* to distinguish it from any contribution to the enthalpy flux associated with the transport of moisture, as discussed below.

Rapidly sampled measurements of vertical wind and temperature allow us to estimate the sensible heat flux near the surface. Theoretically, we need to sample at the highest frequency that occurs in the turbulent flow field to find the sensible heat flux. In practice, we can only sample at a lower frequency and therefore cannot measure the full sensible heat flux.

Because of the latent heat contributions of phase changes, the turbulent flux of water vapour is also of importance in assessing the energy budget of a volume of air. The vapour mass flux $\boldsymbol{F}_v$ can be written as

$$\boldsymbol{F}_v = \boldsymbol{U} \rho_v = \rho \boldsymbol{U} q, \tag{3.72}$$

---

[22] The choice to use the mean density $\overline{\rho}$ is consistent with the Boussinesq approximation that is often used in studies of turbulent fluxes. Without this approximation, the general result for ideal gases is

$$\overline{(\rho w)' h'} = \overline{\rho} c_p \overline{w' T'} + c_p \overline{w}\, \overline{\rho' T'} + c_p \overline{w' \rho' T'}.$$

The second term is often ignored on setting $\overline{w} = 0$ and the third term vanishes when the ideal gas law is linearized around the mean.

with ρ_v the vapour density, ρ the total density, and $q = \rho_v/\rho$ the specific humidity of the air. This vapour flux can be associated with an energy flux if we multiply the vapour flux by the latent heat of evaporation, L. The resulting energy flux is called the *latent heat flux*. Using the same approximations as those leading to Eq. 3.71, we find for the turbulent vertical latent heat flux F_l that

$$\blacktriangleright \qquad F_l = \overline{\rho} L \, \overline{w'q'}. \qquad (3.73)$$

Latent heat flux is really a water vapour flux, apart from a proportionality constant which changes its dimension into an energy flux. However, when there is a divergence of the latent heat flux (that is, there is condensation or evaporation of water vapour), there is a local contribution to the heating field, as expected from a real heat flux.

In Section 5.4 we will show that the variations in specific enthalpy of moist unsaturated air can be written as variations in the so-called *moist static energy*[23]

$$h_v^\star = c_p T + Lq. \qquad (3.74)$$

As we can see, the sensible heat flux and the latent heat flux are essentially the two main contributions to the total enthalpy flux in moist air; that is,

$$\blacktriangleright \qquad \overline{\rho}\,\overline{w'h_v^{\star\prime}} = \overline{\rho} c_p \overline{w'T'} + \overline{\rho} L \, \overline{w'q'} = F_h + F_l. \qquad (3.75)$$

The sensible and latent heat fluxes are contained in the total enthalpy flux for moist air. To measure the total enthalpy flux we need to know both the sensible and the latent heat fluxes, but for evaluating energy budgets it makes little sense to consider them separately.

Over humid surfaces, such as the sea, much of the total enthalpy flux is in the form of latent heat. Over dry surfaces much of the total enthalpy flux is in the form of sensible heat. The ratio of the sensible and latent heat fluxes is called the *Bowen ratio* β,

$$\blacktriangleright \qquad \beta = F_h/F_l. \qquad (3.76)$$

When the latent heat flux is small we often use the evaporative fraction, EF, which is the fractional contribution of the latent heat flux to the total enthalpy flux,

$$\mathrm{EF} = \frac{F_l}{F_h + F_l}. \qquad (3.77)$$

---

[23] The specific heat capacity in this equation is $c_p = (1-q)\,c_{pd} + qc_{pl}$, with c_{pd} and c_{pl} the specific heat capacities for dry air and water, respectively.

The Bowen ratio (or evaporative fraction) is a diagnostic variable describing properties of vertical heat exchange. However, the original use of the Bowen ratio was to estimate the evaporation from surfaces by first estimating the Bowen ratio using energy constraints and then measuring the sensible heat flux.[24]

3.8 POTENTIAL TEMPERATURE

In atmospheric science it is common usage to map the specific entropy of a parcel onto a temperature scale called the *potential temperature*. It is defined as follows:

> *The potential temperature θ is the temperature a fluid parcel would have if brought at constant entropy to a reference pressure p_0.*

From this definition it follows that the potential temperature is a function of the specific entropy of the fluid parcel (p_0 is a fixed parameter). Like the specific entropy, the potential temperature of a fluid parcel is a conserved quantity for adiabatic (isentropic) transformations.

So by definition, the specific entropies of the parcel at its initial pressure and at its reference pressure are the same and the temperature at the reference pressure defines the potential temperature, see Figure 3.5. Considering entropy a function of pressure p and temperature T, potential temperature is implicitly defined by

$$\blacktriangleright \qquad s(p, T) = s(p_0, \theta). \qquad (3.78)$$

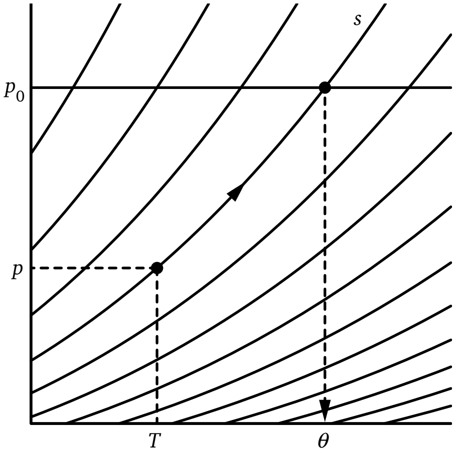


FIGURE 3.5 Illustration on a pT diagram of the definition of potential temperature. For a fixed reference pressure p_0, the isolines of specific entropy s uniquely map any (p, T) point onto a (p_0, θ) point.

---

[24] See Lewis, J. M. (1995) *Bull. Am. Met. Soc.* **76**, 2433–2443.

If we know the specific entropy as a function of pressure and temperature we can invert this equation to find the potential temperature.

Remembering that p_0 is a fixed reference pressure, infinitesimal variations of Eq. 3.78 satisfy

$$\mathrm{d}s = \left(\frac{\partial s}{\partial T}\right)_{p_0} \mathrm{d}\theta. \tag{3.79}$$

This can be rewritten as

$$\blacktriangleright \qquad \mathrm{d}s = c_{p_0} \frac{\mathrm{d}\theta}{\theta}, \tag{3.80}$$

where $c_{p_0} = \theta\,(\partial s/\partial T)_{p_0}$ is the specific heat capacity at the reference pressure. We can write out this equation in terms of variations of pressure and temperature (using the fourth Maxwell relation, Eq. 3.15):

$$-pv\alpha_p \frac{\mathrm{d}p}{p} + c_p \frac{\mathrm{d}T}{T} = c_{p_0} \frac{\mathrm{d}\theta}{\theta}, \tag{3.81}$$

with the isobaric expansion coefficient α_p defined as

$$\alpha_p = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_p. \tag{3.82}$$

If the isobaric expansion coefficient and the isobaric heat capacity are known functions of temperature and pressure, then this equation can be integrated to calculate the potential temperature.

Let us apply the definition of the potential temperature to an ideal gas. For an ideal gas the specific entropy is given in Eq. 3.40. The potential temperature for an ideal gas therefore satisfies

$$c_p \ln(T/T_0) - R\ln(p/p_0) = c_p \ln(\theta/T_0) - R\ln(p_0/p_0) \tag{3.83}$$

and we find for an ideal gas that

$$\blacktriangleright \qquad \theta = T\left(\frac{p_0}{p}\right)^{R/c_p}. \tag{3.84}$$

In the atmosphere, p_0 is usually taken to be 1000 hPa or standard pressure, 1013.25 hPa. In any calculation that leads to measurable results, the constant p_0 will never occur. In the formula for potential temperature of an ideal gas we use the ratio R/c_p; this ratio is often denoted by the greek letter κ. For a diatomic ideal gas (such as air, approximately) we have, from Eqs. 3.28 and 3.29

$$\kappa = \frac{R}{c_p} = \frac{2}{7}. \tag{3.85}$$

A monoatomic gas has $\kappa = 2/5$. Combining the definitions of specific entropy and potential temperature for an ideal gas we can verify that

▶ $$s = c_p \ln(\theta/T_0), \tag{3.86}$$

with T_0 a constant temperature that only changes the specific entropy by an additive constant. From this equation immediately follows the differential relation, Eq. 3.80, as applied to an ideal gas.

We can combine the definition of potential temperature for an ideal gas with the ideal gas law to find

$$\frac{p}{p_0} = \frac{\rho T}{\rho_0 \theta}, \tag{3.87}$$

with $p_0 = \rho_0 R\theta$. From this we can derive the following relationships that are valid for adiabatic transformations of an ideal gas, sometimes called the *Poisson equations*:

▶ $$\frac{p}{p_0} = \left(\frac{T}{\theta}\right)^{c_p/R} = \left(\frac{\rho}{\rho_0}\right)^{c_p/c_v}. \tag{3.88}$$

So these relationships quantify how pressure and temperature change on adiabatic compression and expansion. For example, an adiabatic compression of air by a factor of 2 leads to an increase in pressure by a factor of 2.6 and an increase in temperature by a factor of 1.3. Anyone who has ever used a bicycle pump will be aware of these effects.

Although in atmospheric science it is common usage to map specific entropy onto potential temperature, this is not without its drawbacks. Because of its widespread use in atmospheric science we will not avoid it in this text. But there are no sound physical arguments why we should use potential temperature instead of specific entropy in any of the applications encountered here.

Perhaps part of the attraction of using potential temperature is its apparent measurability: it is the temperature a fluid parcel would have if brought adiabatically to a reference pressure. This is a useful thought experiment but no one has ever used it as a practical method to measure the potential temperature of a fluid parcel.

The notion of potential temperature may distract from the more fundamental notions of entropy and heat. Derivations using potential temperature are usually less intuitive than those using entropy. If carelessly applied, they give less general results: potential temperature is often read as the ideal gas potential temperature but, for example, a moist parcel is not an ideal gas. We are then tempted to introduce such notions as equivalent potential temperature or wet-bulb potential temperature (see Chapter 6), which in fact are not potential temperatures in the sense of Eq. 3.78. In fact, a proper definition of potential temperature for a moist parcel requires the precise definition of how the parcel was brought reversibly to the reference pressure. In other words, there is no unique local definition of potential temperature. We can easily avoid any convoluted definitions and derivations by sticking to the more fundamental notion of entropy.

There is also a more philosophical drawback. Like specific entropy, potential temperature is an intensive quantity; it does not scale with the size of the system. However, specific entropy is based on the extensive total entropy. Entropies can be added; subsystems can be joined together and we can calculate the total entropy of the compound system. No such notion exists for potential temperature; temperatures cannot be added together. It appears somewhat illogical to replace a variable based on an extensive quantity, such as entropy, with a fundamentally intensive quantity such as temperature.

PROBLEMS

3.1. Legendre transforms. Consider a function f of x: $f = f(x)$. Now define a new variable p by

$$p = \frac{\mathrm{d}f}{\mathrm{d}x}$$

Show that the function g, defined as

$$g = xp - f$$

can be seen as a natural function of p (show that $\mathrm{d}g = x\,\mathrm{d}p$). The function g is called the *Legendre transform* of function f. In the same way, show that the function $\tilde{f}$ defined as

$$\tilde{f} = xp - g$$

is a natural function of x. Show that $\tilde{f}$ is the same as the original f; in other words, show that the Legendre transform is its own inverse. Successive Legendre transforms of the internal energy define the different thermodynamic potentials.

3.2. A water molecule H_2O is made up of two hydrogen atoms attached to a central oxygen atom at an angle of about 105 degrees $({}_{\mathrm{H}}\;{}^{\mathrm{O}}\;{}_{\mathrm{H}})$. Suppose that at room temperature, the vibrational modes are not excited. What is the expected specific heat capacity at constant pressure c_p for water vapour? (The observed value is $1.87 \times 10^3\,\mathrm{J\,kg^{-1}\,K^{-1}}$.)

3.3. Show that the specific heat capacity at constant volume for moist air can be written as

$$c_{v(\mathrm{moist})} = (1 + q\,(a_v - 1))\,c_{v(\mathrm{dry})},$$

with q the specific humidity and $a_v = c_{v(\mathrm{vapour})}/c_{v(\mathrm{dry})} \approx 1.96$, the ratio of the heat capacities at constant volume for water vapour and dry air. Demonstrate the validity of the analogous equation for $c_{p(\mathrm{moist})}$ with, instead of a_v, an isobaric heat capacity ratio $a_p \approx (20/21)a_v$.

3.4. The isobaric thermal expansivity and the isothermal compressibility are defined as

$$\alpha_p = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_p, \quad \beta_T = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T.$$

Show that the relationship between c_p and c_v can be rewritten as

$$c_p = c_v + vT\alpha_p^2/\beta_T.$$

Hint: Use the reciprocity relation for partial derivatives, see Appendix A. Find relevant tabulated values for water to show that the isobaric and isovolumetric heat capacities of water are typically less than 1% apart.

3.5. Show that the isentropic compressibility β_s and the isothermal compressibility β_T are related through

$$\beta_T = \beta_s + vT\alpha_p^2/c_p \;\text{ and hence }\; \beta_T/\beta_s = c_p/c_v.$$

Hint: Write the differential of v as a function of p and T and hence calculate $(\partial v/\partial p)_s$. What is the physical reason why β_T is larger than β_s?

3.6. Show that the specific entropy of a solid or liquid with specific heat capacity c can be approximated as

$$s = c\ln(T/T_0).$$

3.7. Use the equation for the speed of sound in an ideal gas, Eq. 3.45, to explain why the pitch of your voice goes up after breathing in helium gas. Show that this equation predicts a pitch change of close to one octave plus a perfect fifth.

3.8. Show that the speed of sound in humid air $\mathscr{C}_h$ can be approximated by

$$\mathscr{C}_h = \mathscr{C}\,(1 + 0.26\,q),$$

with $\mathscr{C}$ the speed of sound in dry air at the given temperature and q the specific humidity. Hint: See Problem 3.3. In light of the above equation, can we measure temperature in the field by measuring the speed of sound between two points?

3.9. Show that van der Waals' equation, Eq. 3.51, can be rewritten as

$$p = \rho R T \left(\frac{1}{1 - b\rho/\mu} - \frac{a\rho}{\mu^2 R T} \right),$$

with μ the molar mass of the substance, R its specific gas constant, and a and b its van der Waals coefficients. Hence show that the second and third virial coefficients for a van der Waals gas are given by Eq. 3.53.

3.10. Concorde used to cruise at a pressure of around 100 hPa, where the typical air temperature is around $-50°$C. If the cabin pressure in Concorde is 850 hPa, what would the consequences be of adiabatic compression of the outside air to provide ventilation?

4

The atmosphere under gravity

The gravitational potential energy of an air parcel varies considerably over the depth of the atmosphere when compared to variations in internal energy. This means that the gravitational potential energy has to be part of the thermodynamic energy budget when considering vertical structure in the atmosphere.

Hydrostatic balance is the first manifestation of the gravitational potential energy in fluid systems. It reflects the fact that the pressure at any point in a static fluid is determined by the weight of the fluid above it. In this chapter we give a general thermodynamic derivation of hydrostatic balance.

When the air is in motion, hydrostatic balance is only valid as an approximation. In dynamic terms, hydrostatic balance requires that the vertical acceleration of air parcels is much smaller than the acceleration of gravity. In thermodynamic terms, hydrostatic balance requires that the kinetic energy variations are much smaller than potential energy or internal energy variations. These conditions are essentially equivalent and are usually met in the atmosphere. Only in strong updrafts or downdrafts is an assumption of hydrostatic balance expected to be unwarranted.

The rest of the chapter is devoted to exploring what the assumption of hydrostatic balance implies for the vertical structure of the atmosphere.

The chapter concludes with a discussion of the vertical structure of an isothermal atmosphere in the context of statistical mechanics. Some key concepts in statistical mechanics are introduced, illustrating the way statistical mechanics can be used to link the microscopic world of molecules to the macroscopic world of thermodynamics.

4.1 GEOPOTENTIAL

The acceleration due to gravity is a vector $\boldsymbol{g}$, which can be written as the gradient of a *geopotential* ϕ as

$$\blacktriangleright \qquad \boldsymbol{g} = -g\hat{\boldsymbol{k}} = -\nabla\phi, \tag{4.1}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

with $\hat{\boldsymbol{k}}$ the unit vector pointing vertically upward and g the magnitude of the acceleration. The geopotential is in fact used to *define* the local vertical: vertically upward means parallel to the gradient of the geopotential. Because the vertical direction (spanned by the z-coordinate) is defined to be parallel to the gradient of ϕ, the magnitude of the gravitational acceleration can always be written as

$$g = \frac{\mathrm{d}\phi}{\mathrm{d}z}. \tag{4.2}$$

For simplicity, cartesian geometries with fixed g are often used. In this case

$$\phi = gz, \tag{4.3}$$

which is a good approximation near the Earth's surface. The energy required to lift a parcel of mass M against gravity over some small distance $\mathrm{d}\boldsymbol{r}$ equals $-M\boldsymbol{g} \cdot \mathrm{d}\boldsymbol{r}$. So to move a parcel from some initial location $\boldsymbol{r}_0$ to some final location $\boldsymbol{r}_1$ requires energy

$$-\int_{\boldsymbol{r}_0}^{\boldsymbol{r}_1} M\boldsymbol{g} \cdot \mathrm{d}\boldsymbol{r} = \int_{\boldsymbol{r}_0}^{\boldsymbol{r}_1} M\nabla\phi \cdot \mathrm{d}\boldsymbol{r} = M(\phi(\boldsymbol{r}_1) - \phi(\boldsymbol{r}_0)). \tag{4.4}$$

This demonstrates that geopotential is the same as the specific gravitational potential energy.

Around a spherical gravitating body, such as the Earth, the geopotential is (see Problem 4.1)

► $$\phi = g_0 \frac{az}{a+z}, \tag{4.5}$$

with a the radius of the body, z the height above the surface of the body, and g_0 the acceleration of gravity at the surface of the body. For the Earth, $a = 6371$ km and $g_0 = 9.8\,\mathrm{m\,s^{-2}}$. For small z it is clear that $\phi = g_0 z$.

Geopotential is often translated into *geopotential height* Z, which is defined as

► $$Z = \phi/g_0. \tag{4.6}$$

So geopotential height has the units of height but it is in fact a geopotential. The relationship between Z and z is illustrated in Fig 4.1. The differentials of Z and z are related by

$$g_0\,\mathrm{d}Z = g\,\mathrm{d}z. \tag{4.7}$$

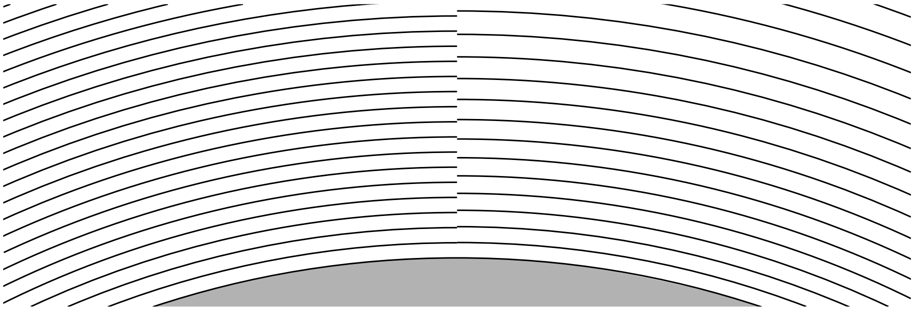


FIGURE 4.1 Isolines of geometric height z, left, and geopotential height Z, right, around the Earth's surface, both drawn every 100 (geopotential) km. Near the surface the geometric and geopotential heights are nearly the same.

For a spherical gravitating mass the geometric height and the geopotential height transform into each other by

$$Z = \frac{az}{a+z} \quad \text{and} \quad z = \frac{aZ}{a-Z}. \tag{4.8}$$

Near the Earth's surface the geopotential height is very similar to the geometric height, with the difference increasing for increasing height. The geopotential height is smaller than the geometric height at any point in the atmosphere. Geometric height can have any value larger than zero while geopotential height has a maximum value of a.

The atmosphere is usually considered in a coordinate frame rotating with the Earth. In this frame of reference, a fixed centrifugal acceleration will add to the purely gravitational acceleration. This can be taken into account by defining an effective gravity and an effective geopotential that include this centrifugal acceleration, see Problem 4.2. The local vertical in the rotating frame of reference is now defined by the gradient of the effective geopotential. From now on, 'effective' is omitted and we shall simply refer to geopotential or gravitational acceleration.

The main effect of rotation is that the geopotential surfaces are not spheres but oblate spheroids. The Earth's surface is, apart from orographic features, a geopotential surface. It is an oblate spheroid with a polar radius about 1/300 smaller than its equatorial radius. The transformations in Eq. 4.8 are therefore only approximate for a rotating planet.

4.2 HYDROSTATIC BALANCE

In a fluid at rest the net force on any fluid parcel has to vanish. So in a gravitational field the gravitational force has to be balanced by an internal force due to a decrease of pressure with height. This is called *hydrostatic balance.*

We can use thermodynamics to calculate the hydrostatic pressure decrease with height.[25] Think of the geopotential field as made up of thin layers of a given geopotential. By symmetry, in hydrostatic balance all thermodynamic variables are functions of the geopotential alone. We now examine the energy budget when moving a body of fluid from one geopotential to a nearby geopotential without putting any heat in. If it was possible to generate energy through such a process then the atmosphere would be unstable: any small perturbation of the atmosphere could release internal and potential energy to create kinetic energy.

We can now view the initial geopotential layer (layer 1) and the final geopotential layer (layer 2) as open systems with matter flowing from one to the other. From Section 3.5 we know that the change in internal energy δU of the whole system on moving the fluid from layer 1 to layer 2 has to equal the enthalpy gain of layer 2 minus the enthalpy loss of layer 1. The enthalpy change includes the flow work that is required to move the fluid from layer 1 to layer 2. Because we add no heat to the system, the enthalpy change can only be due to a pressure change, that is, $\mathrm{d}H = V\,\mathrm{d}p$. We thus find

$$\delta U = H_2 - H_1 = \int_{\phi_1}^{\phi_2} V \frac{\mathrm{d}p}{\mathrm{d}\phi}\,\mathrm{d}\phi. \tag{4.9}$$

The change in potential energy, δP, is

$$\delta P = M\,(\phi_2 - \phi_1) = \int_{\phi_1}^{\phi_2} M\,\mathrm{d}\phi, \tag{4.10}$$

with M the mass of the displaced fluid. At equilibrium, energy variations must vanish because otherwise we could generate kinetic energy by some spontaneous rearrangement of the vertical profile. We thus have

$$0 = \delta U + \delta P = \int_{\phi_1}^{\phi_2} \left(V \frac{\mathrm{d}p}{\mathrm{d}\phi} + M \right) \mathrm{d}\phi. \tag{4.11}$$

Because the two geopotential surfaces were chosen randomly, the integrand has to vanish. We therefore find, after dividing the integrand by V,

► $$\frac{\mathrm{d}p}{\mathrm{d}\phi} = -\rho. \tag{4.12}$$

Writing out the geopotential variations in terms of the gravitational acceleration,

► $$\frac{\mathrm{d}p}{\mathrm{d}z} = -\rho g \quad \text{or} \quad \frac{\mathrm{d}p}{\mathrm{d}Z} = -\rho g_0, \tag{4.13}$$

---

[25] Ambaum, M. H. P. (2008) *Proc. Roy. Soc. A* **464**, 943–950.

where Eqs. 4.2 and 4.7 have been used. This equation is called the *hydrostatic equation.* It is fully general for any fluid system at rest and was first derived (essentially in its integral form, but using a method based on dynamical arguments) by Newton in his *Principia.* The above thermodynamic derivation makes no reference to the geometry of the fluid and is therefore more general than some shorter derivations that are incomplete due to the implicit use of a flat geometry.

We have assumed that the fluid in our thought experiment is at rest; that is, there is no kinetic energy. Indeed, if the fluid is in motion, hydrostatic balance is not strictly valid any longer. However, the atmosphere is usually very close to hydrostatic balance, despite the motions in the atmosphere. From a dynamical point of view, as long as vertical accelerations are substantially smaller than g, the assumption of hydrostatic balance provides an accurate representation of the atmosphere. From a thermodynamic point of view, as long as vertical variations in internal or gravitational potential energy are substantially larger than variations in kinetic energy, hydrostatic balance is accurate. This depends on the vertical scale considered. For example, the variation of specific gravitational potential energy over the depth of the troposphere is about $10^5\,\mathrm{J\,kg^{-1}}$. The specific kinetic energy is $\mathcal{V}^2/2$, with $\mathcal{V}$ the flow speed. The variations in specific kinetic energy in the atmosphere are much smaller than this variation in gravitational potential energy. However, the variation of specific gravitational potential energy over one metre depth is about $10\,\mathrm{J\,kg^{-1}}$. This is comparable to observed variations in specific kinetic energies, especially in areas of substantial updrafts. So for such scales, hydrostatic balance is not accurate anymore.

For an ideal gas the hydrostatic relation becomes

$$\frac{1}{p}\frac{\mathrm{d}p}{\mathrm{d}z} = -\frac{g}{RT} \quad \text{or} \quad \frac{1}{p}\frac{\mathrm{d}p}{\mathrm{d}\phi} = -\frac{1}{RT}. \tag{4.14}$$

This means that if the temperature as a function of height is known then the pressure as a function of height can be found. Often, the above equation is reversed:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\ln p} = -RT. \tag{4.15}$$

This equation gives a relationship between geopotential and the (logarithm of) pressure. This equation is useful because radiosondes normally measure both pressure and temperature; the above equation is then used to calculate the geopotential height of the radiosonde.[26]

---

[26] For moist air, we need to replace the temperature in these equations by the virtual temperature, see Section 1.3.

An interesting application of hydrostatic balance is to assume a constant temperature T_0. The hydrostatic equation then integrates to

$$p = p_0\, e^{-Z/H} \quad \text{with} \quad H = \frac{RT_0}{g_0} \tag{4.16}$$

with p_0 the pressure at $Z = 0$. This equation is strictly only valid for geopotential height Z because g varies with height. A typical value for H, the so-called *scale height*, for the atmosphere is 8 km. Constant temperature is obviously not the most accurate approximation in the atmosphere. Nevertheless, the pressure still decays with height close to exponentially, with a scale height fairly close to 8 km.

Although the precise height–pressure relationship depends on the temperature profile it is useful to be aware of approximate values. For illustration, Table 4.1 shows typical values of height, temperature, pressure, and density for the 1976 US Standard Atmosphere, a fictional atmosphere with ‘typical’ mid-latitude values of the temperature profile.

Another application is to assume a fixed *lapse rate*. Lapse rate, usually denoted Γ, is the negative derivative of temperature with height or geopotential

Table 4.1 Selected values from the US Standard Atmosphere, 1976. Here, the 1013.25 hPa level provides the reference values for potential temperature θ and specific entropy s.

p (hPa)	z (km)	T (°C)	ρ (kg m^{-3})	θ (K)	s (J kg^{-1} K^{-1})
1013.25	0.00	15.00	1.225	288.15	0.00
1000	0.11	14.3	1.21	289	1
925	0.76	10.1	1.14	291	9
850	1.5	5.5	1.06	293	17
700	3.0	−4.6	0.91	298	35
600	4.2	−12.3	0.80	303	50
500	5.6	−21.2	0.69	308	68
400	7.2	−31.7	0.58	315	89
300	9.2	−44.6	0.46	324	117
250	10.4	−52.4	0.39	329	134
200	11.8	−56.5	0.32	344	179
150	13.6	−56.5	0.24	374	262
100	16.2	−56.5	0.16	420	378
70	18.5	−56.5	0.11	465	481
50	20.6	−55.9	0.080	513	580
30	23.9	−52.6	0.047	603	741
20	26.6	−50.0	0.031	684	870
10	31.2	−45.5	0.015	852	1089
5	36.0	−33.9	0.0073	1091	1338
2	42.7	−15.3	0.0027	1527	1676
1	48.2	−2.5	0.0013	1954	1924

height,

$$-\frac{\mathrm{d}T}{\mathrm{d}Z} = \Gamma. \tag{4.17}$$

So the temperature profile is

$$T = T_0 - \Gamma Z, \tag{4.18}$$

with T_0 the temperature at geopotential height $Z = 0$. This can be substituted in the hydrostatic equation and integrated with boundary condition $p(Z = 0) = p_0$. The result is

► $$p = p_0 \left(\frac{T_0 - \Gamma Z}{T_0} \right)^{g_0/R\Gamma}. \tag{4.19}$$

This equation can be related to the result for an isothermal atmosphere by rewriting as

$$p = p_0 \left(1 - \frac{R\Gamma}{g_0} \frac{Z}{H} \right)^{g_0/R\Gamma} \quad \text{with} \;\; H = \frac{RT_0}{g_0}. \tag{4.20}$$

Using a well-known result in calculus,

$$\lim_{q \to 0} (1 - qx)^{1/q} = e^{-x}, \tag{4.21}$$

we see that the Eqs 4.16 and 4.19 become the same when the lapse rate Γ is much smaller than $g_0/R = 34\,\mathrm{K\,km^{-1}}$.

The US Standard Atmosphere is, in the troposphere, defined to have a constant lapse rate in geopotential height of $6.5\,\mathrm{K\,km^{-1}}$. When cruising aircraft report height they actually refer to a pressure level in the US Standard Atmosphere, where Eq. 4.19 is used to calculate the geopotential height from the pressure. Equations 4.8 and 4.19 can be used to check the tropospheric values of Table 4.1. Figure 4.2 shows a graphical comparison of some vertical pressure profiles.

Integrating the hydrostatic equation between two height levels, we find that the mass between the two levels is proportional to the pressure difference. The mass M per unit area between two height levels is

$$M = \int_{z_0}^{z_1} \rho \,\mathrm{d}z = \int_{p_1}^{p_0} \frac{\mathrm{d}p}{g} \approx \frac{p_0 - p_1}{g}, \tag{4.22}$$

where we have assumed that g is constant between the two levels. In other words, equal pressure differences correspond to equal masses.

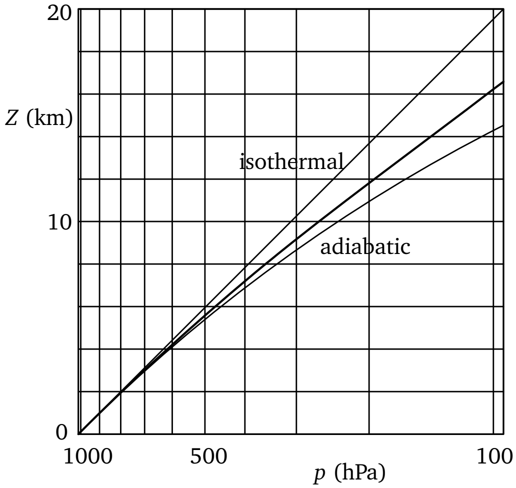


FIGURE 4.2 Comparison of the US Standard Atmosphere, 1976 (thick line) with an isothermal profile and a dry adiabatic profile with constant lapse rate of g_0/c_p (thin lines). Following Eq. 4.15, the slope of each line in this graph is proportional to the temperature.

Integrating the hydrostatic equation between two pressure levels, we find a useful expression for the geopotential difference between the two levels,

$$\phi_1 - \phi_0 = g_0\,(Z_1 - Z_0) = -\int_{p_0}^{p_1} v\,\mathrm{d}p. \tag{4.23}$$

With the ideal gas law this equation becomes the *hypsometric equation,*

$$\blacktriangleright \quad Z_1 - Z_0 = \int_{p_1}^{p_0} \frac{RT}{g_0 p}\,\mathrm{d}p \tag{4.24}$$

which relates the geopotential difference between two pressure levels to an average temperature of the layer. The *geopotential thickness*, indicated in geopotential metres or decametres, is a commonly used diagnostic to characterize the temperature of a layer, see Figure 4.3. Remember that in humid conditions, such as the tropical troposphere, these equations need to be written in terms of virtual temperature.

4.3 ADIABATIC LAPSE RATE

The hydrostatic equation can be used to relate the specific entropy profile to the temperature profile. Considering specific entropy a function of p and T we find

$$\frac{\mathrm{d}s}{\mathrm{d}z} = \left(\frac{\partial s}{\partial p}\right)_T \frac{\mathrm{d}p}{\mathrm{d}z} + \left(\frac{\partial s}{\partial T}\right)_p \frac{\mathrm{d}T}{\mathrm{d}z}. \tag{4.25}$$

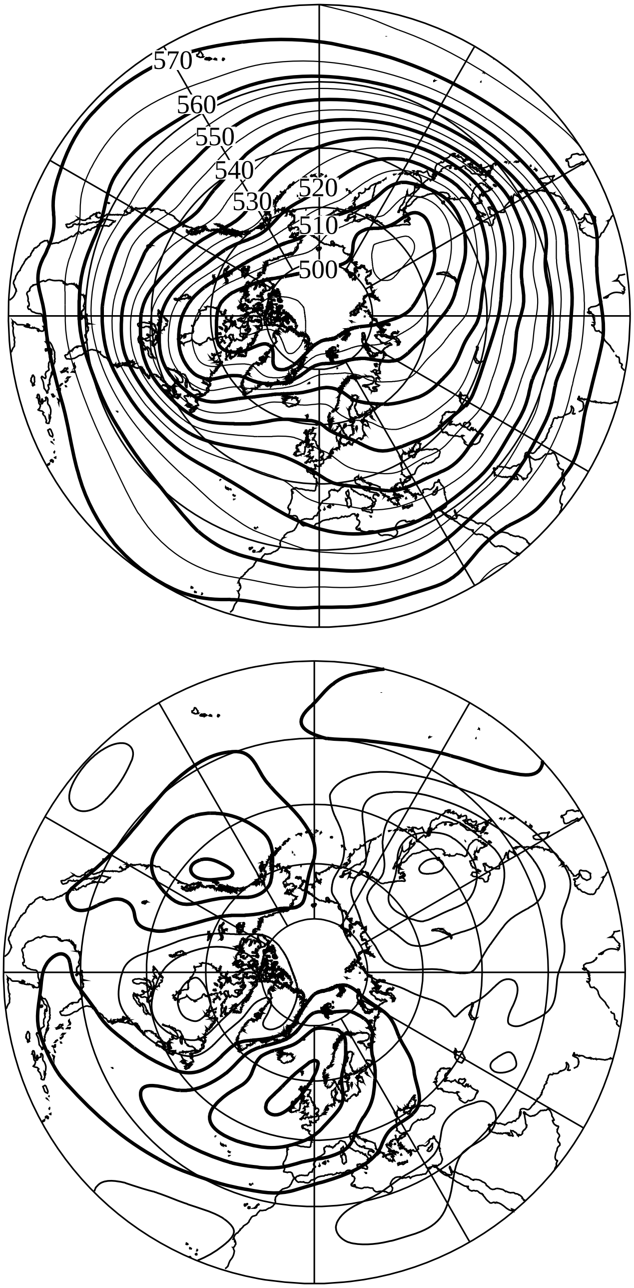


FIGURE 4.3 Top panel: Mean wintertime geopotential thickness of the 1000–500 hPa layer in decametres. Bottom panel: Same as top panel but with the longitudinal mean removed to emphasize the deviations from the dominant pole-to-equator thickness gradient. Thin contours represent negative values. Contour values are . . . , −75, −25, 25, 75, . . . m. Data from NCEP Reanalyses, Kalnay et al. (1996). *Bull. Amer. Met. Soc.* **77**, 437–471.

Substituting hydrostatic balance and using a Maxwell relation this can be rewritten as

$$\frac{\mathrm{d}s}{\mathrm{d}z} = g\,\alpha_p + \frac{c_p}{T}\frac{\mathrm{d}T}{\mathrm{d}z} \tag{4.26}$$

with the isobaric expansion coefficient α_p and isobaric specific heat capacity c_p defined as

$$\alpha_p = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_p, \quad c_p = T\left(\frac{\partial s}{\partial T}\right)_p. \tag{4.27}$$

Rearranging gives

► $$\frac{\mathrm{d}s}{\mathrm{d}z} = \frac{c_p}{T}\left(\Gamma_a + \frac{\mathrm{d}T}{\mathrm{d}z}\right), \tag{4.28}$$

where the *adiabatic lapse rate, Γ_a,* is defined as

► $$\Gamma_a = \frac{g\alpha_p T}{c_p}. \tag{4.29}$$

For an ideal gas $\alpha_p = 1/T$ and the adiabatic lapse rate becomes

► $$\Gamma_d = \frac{g}{c_p}. \tag{4.30}$$

This lapse rate is usually called the *dry adiabatic lapse rate,* hence the subscript '*d*'. These equations are also valid for lapse rates in geopotential height if g is replaced by g_0. The dry adiabatic lapse rate for the Earth's atmosphere is about $10\,\mathrm{K\,km^{-1}}$. It is typical for the observed temperature profile when there is vigorous turbulence, such as in the planetary boundary layer: here air parcels, carrying their specific entropy, get stirred and the specific entropy becomes on average constant in the vertical. The dry adiabatic lapse rate is in practice an upper bound on observed lapse rates because larger lapse rates are gravitationally unstable.

It is worth emphasizing that Eq. 4.29 for the adiabatic lapse rate is valid for any substance for which entropy is a function of temperature and pressure only. Section 6.1 demonstrates its application in finding the adiabatic lapse rate for moist saturated air, see also Problem 4.4.

The adiabatic lapse rate also describes how the temperature of a parcel changes with height if it moves adiabatically in the vertical. By definition, an air parcel moving adiabatically conserves its specific entropy. The Lagrangian derivative of the temperature, that is the time derivative of the temperature following the air parcel, can then be written as

$$\frac{\mathrm{D}T}{\mathrm{D}t} = \left(\frac{\partial T}{\partial s}\right)_p \frac{\mathrm{D}s}{\mathrm{D}t} + \left(\frac{\partial T}{\partial p}\right)_s \frac{\mathrm{D}p}{\mathrm{D}t} = \left(\frac{\partial T}{\partial p}\right)_s \frac{\mathrm{D}p}{\mathrm{D}t}, \tag{4.31}$$

because $\mathrm{D}s/\mathrm{D}t = 0$ for adiabatic motion. Considering pressure a function of height only, that is, ignoring variations of pressure in the horizontal and in time, we can write

$$\frac{\mathrm{D}p}{\mathrm{D}t} = \frac{\mathrm{d}p}{\mathrm{d}z}\frac{\mathrm{D}z}{\mathrm{D}t} = \frac{\mathrm{d}p}{\mathrm{d}z}\mathcal{W} \tag{4.32}$$

with $\mathcal{W}$ the vertical velocity. Substituting this in the above equation we find (see Problem 4.3),

$$\frac{\mathrm{D}T}{\mathrm{D}t} = \left(\frac{\partial T}{\partial p}\right)_s \frac{\mathrm{d}p}{\mathrm{d}z}\mathcal{W} = -\Gamma_a \mathcal{W}. \tag{4.33}$$

So with these assumptions, we find that the temperature change with height follows the adiabatic lapse rate. The assumptions on the pressure fields can be relaxed if we use pressure coordinates: here the vertical coordinate is pressure rather than geometric height. The vertical velocity in pressure coordinates is $\mathrm{D}p/\mathrm{D}t$ and the dry adiabatic lapse rate in pressure coordinates is $(\partial T/\partial p)_s$.

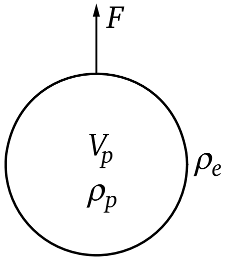
The dry adiabatic lapse rate explains the *Föhn effect*, where air descends from a mountain adiabatically; the temperature of that air will increase according to the dry adiabatic lapse rate. Because the dry adiabatic lapse rate is usually larger than the environmental lapse rate, the descending air will become warmer than the environment. We would still need to explain why the descending parcel had a higher entropy than the environment from which it originated on the upstream side of the mountain. Here condensation processes play a role: the rising air parcel is dried at relatively high temperature by condensation and subsequent precipitation on the upstream side of the mountain.

4.4 BUOYANCY

Vertical stability (or *parcel stability*) of the atmosphere depends on the weight of a parcel relative to its environment. If it is warmer and lighter, it will experience an upward buoyancy force in accordance with Archimedes' law; if it is cooler and heavier, it will experience a downward buoyancy force. According to Archimedes' law the upward buoyancy force F on a fluid parcel is

$$F = g(\rho_e - \rho_p)V_p \tag{4.34}$$

with ρ_p and ρ_e the densities of the parcel and environment, respectively, and V_p the volume of the parcel. The specific upward force (the force per unit mass, that is, acceleration) on the parcel is called the *buoyancy*, b



► $$b = \frac{F}{\rho_p V_p} = g\frac{\rho_e - \rho_p}{\rho_p} = g\frac{v_p - v_e}{v_e}. \tag{4.35}$$

So the buoyancy is due to the relative difference in specific volumes of the parcel and its environment. For an ideal gas, the buoyancy can be written as

$$b = g\frac{T_p - T_e}{T_e} = g\frac{\theta_p - \theta_e}{\theta_e}, \tag{4.36}$$

where it is used the fact that under hydrostatic balance the parcel pressure and the environmental pressure are the same. This equation formalizes the familiar notion that warm air parcels will rise, $b > 0$.

For a moist atmosphere, the temperatures in the expression for buoyancy need to be replaced by virtual temperatures, and the potential temperatures need to be replaced by the *virtual potential temperatures, θ_v*, defined as

► $$\theta_v = T_v(p_0/p)^{R/c_p}, \tag{4.37}$$

where the R and the c_p are the usual dry air values.

For small differences between the environment and the parcel, we can derive a linear approximation to the buoyancy:

$$\begin{aligned} b &= g\frac{v_p - v_e}{v_e} && \text{by Eq. 4.35} \\ &= \frac{g}{v_e}\left(\frac{\partial v}{\partial s}\right)_p (s_p - s_e) && \text{linearization} \\ &= -\frac{\mathrm{d}p}{\mathrm{d}z}\left(\frac{\partial T}{\partial p}\right)_s (s_p - s_e) && \text{by Eqs 3.7 and 4.13} \\ &= \Gamma_a(s_p - s_e) && \text{by Eq. 4.86} \end{aligned} \tag{4.38}$$

This linearized buoyancy can also be written in terms of potential temperature:

$$b = c_{p_0}\Gamma_a\frac{\theta_p - \theta_e}{\theta_e}. \tag{4.39}$$

This expression for buoyancy generalizes the ideal gas result, Eq. 4.36, to any simple substance although it is only valid in the linear regime. Note that moist unsaturated air is not necessarily a simple substance: the buoyancy can also be a function of the humidity contrast.

If a parcel of specific entropy s_p is lifted over a (small) distance δz, then the environment specific entropy s_e and parcel specific entropy s_p are related by

$$s_e = s_p + \frac{\mathrm{d}s_e}{\mathrm{d}z}\,\delta z, \tag{4.40}$$

because s_p is the entropy of the environment at the departure point. Substituting this in the linearized buoyancy we find

$$b = -\Gamma_a\frac{\mathrm{d}s_e}{\mathrm{d}z}\,\delta z. \tag{4.41}$$

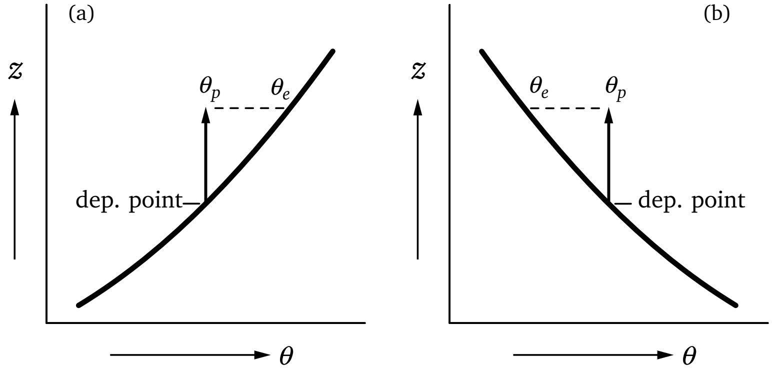


FIGURE 4.4 Illustration of a stable profile (a) and unstable profile (b) of potential temperature versus height (solid line). When lifting a parcel in case (a) the parcel potential temperature will be lower than that of the environment, and therefore the parcel will sink. In case (b) the parcel potential temperature will be larger than that of the environment, and the parcel will rise.

This equation is general for small displacements in any simple hydrostatic fluid.

In an environment where the specific entropy increases with height, $\mathrm{d}s_e/\mathrm{d}z > 0$, the buoyancy force on a displaced parcel has the opposite sign to the displacement, $b\,\delta z < 0$. This means that the fluid is stable for (small) parcel displacements: any fluctuation in vertical location is counteracted by a restoring buoyancy force. In an environment where the specific entropy decreases with height, $\mathrm{d}s_e/\mathrm{d}z < 0$, the buoyancy force on a displaced parcel has the same sign as its displacement, $b\,\delta z > 0$. This means that the fluid is unstable: any fluctuation in vertical location of the parcel will be enhanced by its buoyancy force. An equivalent argument is illustrated in Figure 4.4. An environment with constant specific entropy (so its temperature lapse rate is the dry adiabatic lapse rate) has neutral stability. A well-mixed boundary layer is often neutrally stable.

We can give a more dynamic interpretation of parcel stability. According to Newton's second law, the acceleration of the fluid parcel satisfies

$$\frac{\mathrm{d}^2\,\delta z}{\mathrm{d}t^2} = b = -\Gamma_a \frac{\mathrm{d}s_e}{\mathrm{d}z}\,\delta z. \tag{4.42}$$

This equation can be written as

$$\frac{\mathrm{d}^2\,\delta z}{\mathrm{d}t^2} + N^2\,\delta z = 0. \tag{4.43}$$

This is an oscillation equation for δz with solutions proportional to $\cos(Nt + \phi_0)$ where the buoyancy frequency N, also called the *Brunt–Väisälä frequency*, is given by

$$\blacktriangleright \quad N^2 = \Gamma_a \frac{ds}{dz} = \frac{c_p}{T}\Gamma_a \left(\Gamma_a + \frac{dT}{dz}\right), \tag{4.44}$$

where the subscript *e* has been omitted and the vertical profile of entropy has been written in terms of the temperature profile. Because variations in specific entropy can be mapped onto variations in potential temperature, $ds = c_{p_0} d\theta/\theta$, the Brunt–Väisälä frequency can also be written as

$$N^2 = \frac{c_{p_0}\Gamma_a}{\theta}\frac{d\theta}{dz}. \tag{4.45}$$

For ideal gases this gives

$$\blacktriangleright \quad N^2 = \frac{g}{\theta}\frac{d\theta}{dz}. \tag{4.46}$$

This equation is also the appropriate expression for moist unsaturated air if we replace the potential temperature by the virtual potential temperature.

For stable cases we find a real buoyancy frequency (oscillations of δz), for unstable cases we find an imaginary buoyancy frequency (exponential growth of δz), and in neutral cases, a vanishing buoyancy frequency (steady δz). For a typical tropospheric lapse rate of 6.5 $\mathrm{K\,km^{-1}}$ the buoyancy frequency is about $N = 10^{-2}\,\mathrm{s^{-1}}$, corresponding to a buoyancy period of about 10 minutes.

The instability of profiles with specific entropy decreasing with height means that such profiles are not normally observed, at least not for periods longer than the time scale associated with the parcel instability. For example, the time scale associated with a 11 $\mathrm{K\,km^{-1}}$ profile is about 3 minutes. Observed profiles will therefore in practice be stable or neutral. From Eq. 4.28 it then follows that the adiabatic lapse rate is an upper bound on the observed lapse rates.

The Brunt–Väisälä frequency should be interpreted as a measure of stability instead of the precise oscillation frequency of a displaced parcel. The buoyancy *b* does not equal the acceleration of the parcel as it underestimates the inertia of the parcel. The parcel attains a so-called *added mass* because any moving parcel also has to displace, and therefore accelerate, the surrounding fluid, thus increasing the inertia of the parcel. The oscillating parcel we have just described is not a solution of the full equations of motion of the fluid. A further caveat is that any real oscillating parcel would exchange matter with its environment (*entrainment*) and thus effectively reduce its entropy contrast. Both these effects indicate that the Brunt–Väisälä frequency in Eq. 4.44 is an upper bound on observed parcel frequencies.

The Brunt–Väisälä frequency is an important parameter for wave phenomena in a stratified fluid. For example, it can be shown that a fluid with a constant Brunt–Väisälä frequency *N* supports linear internal gravity waves with a frequency ω equal to

$$\omega^2 = \frac{k^2 N^2}{k^2 + l^2}. \tag{4.47}$$

Here, k and l are the horizontal and vertical wavenumbers of the wave. This dispersion relation shows that for such actual solutions of the equations of motion the Brunt–Väisälä frequency is indeed an upper bound on parcel frequencies: $\omega < N$.

4.5 DRY STATIC ENERGY AND BERNOULLI FUNCTION

For a parcel moving hydrostatically in a geopotential field, the specific enthalpy varies according to

$$dh = T\,\mathrm{d}s + v\,\mathrm{d}p = T\,\mathrm{d}s - \mathrm{d}\phi. \tag{4.48}$$

We can move the differential $\mathrm{d}\phi$ to the left-hand side to form the differential of the combination

$$\blacktriangleright \qquad \eta = h + \phi. \tag{4.49}$$

This expression is called the *generalized enthalpy* or the *static Bernoulli function.* The first order variation in generalized enthalpy is

$$\mathrm{d}\eta = \mathrm{d}h + \mathrm{d}\phi = T\,\mathrm{d}s. \tag{4.50}$$

For an ideal gas with isobaric heat capacity c_p we have $\mathrm{d}h = c_p\,\mathrm{d}T$, so we can integrate the expression for $\mathrm{d}h + \mathrm{d}\phi$ to

$$\blacktriangleright \qquad h^\star = c_p T + g_0 Z. \tag{4.51}$$

Here, $h^\star$ is called the *dry static energy.* The dry static energy differs from the generalized enthalpy for an ideal gas by an integration constant. This integration constant becomes important in the discussion of mixtures. Note that $\mathrm{d}h^\star/\mathrm{d}z = 0$ defines the dry adiabatic lapse rate for an ideal gas.

Equation 4.50 is the energy budget for a parcel in an external potential field: for adiabatic changes, any reduction in potential energy is credited to the enthalpy of the parcel. Why does a reduction in potential energy not simply translate into a gain of internal energy? The difference is due to the macroscopic nature of rearrangements. The displacement of an air parcel requires *flow work,* see Section 3.5. So for adiabatic rearrangements, a reduction in potential energy is used to boost the sum of internal energy and flow work,

$$-\mathrm{d}\phi = \mathrm{d}u + \mathrm{d}(pv) = \mathrm{d}h. \tag{4.52}$$

Flow work has to be part of the budget because if the potential energy of a parcel is to change, it has no choice but to change its vertical position.

Equation 4.50 refers to slow, reversible changes. If the change is fast then the parcel will also gain kinetic energy. The kinetic energy per unit mass is $\mathscr{V}^2/2$, with $\mathscr{V}$ the macroscopic speed of the air parcel. This then has to be included in the budget for adiabatic motion. Written as the energy budget of a fluid parcel, we find *Bernoulli's equation:*

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(h + \phi + \frac{1}{2}\mathscr{V}^2\right) = 0, \tag{4.53}$$

that is, for adiabatic motions the Bernoulli function (the term in brackets) is constant following a fluid parcel. This budget equation is only true if there are no other sources of energy for the parcel. We need to exclude diffusion processes (which may diffuse heat or kinetic energy to or from the parcel and thus change its specific entropy or kinetic energy) and non-steady conditions (in which case the budget would not need to be closed). A more familiar form of Bernoulli's equation follows for incompressible fluids where the density of each fluid parcel is constant, ρ_0, so that $\mathrm{d}h = \mathrm{d}p/\rho_0$. In this case we find

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{p}{\rho_0} + \phi + \frac{1}{2}\mathscr{V}^2\right) = 0. \tag{4.54}$$

Bernoulli's equation has many applications and the best known is probably the 'explanation' of why a plane can fly: the wings are shaped such that the flow speed over the wing is higher than the flow speed below the wing. Bernoulli's equation then states that there has to be a pressure difference between top and the bottom of the wing, which provides the necessary lift. However, this is only a consistency argument and not an explanation giving cause and effect. All that Bernoulli's equation implies is that in a steady adiabatic flow changes in pressure and changes in flow speed go hand in hand; we cannot say that one causes the other. Furthermore, air is compressible, so any argument needs to be properly phrased in terms of enthalpy rather than pressure, see also Problem 4.7.

In the dynamic meteorology literature, the generalized enthalpy is usually called the Montgomery function.[27] Its relevance to dynamics becomes clear when considering the total specific force on a fluid in the equations of motion

---

[27] Because in dynamic meteorology applications only gradients of the Montgomery function play a role, we normally use the dry static energy, Eq. 4.51, instead of the generalized enthalpy to define the Montgomery function for the atmosphere.

(Navier–Stokes equations or Euler equations); this specific force $\boldsymbol{f}$ is given by

$$\boldsymbol{f} = -v\,\nabla p - \nabla\phi. \tag{4.55}$$

With the definition of the Montgomery function this becomes

$$\boldsymbol{f} = -\nabla\eta + T\,\nabla s. \tag{4.56}$$

This expression becomes particularly simple when *isentropic coordinates* are used; that is, the specific entropy is used as a vertical coordinate. This is only possible when there is a stable stratification so that $\mathrm{d}s/\mathrm{d}z > 0$ everywhere and z can be mapped uniquely onto s. The 'horizontal' part of the specific force $\boldsymbol{f}_h$ in isentropic coordinates becomes

$$\boldsymbol{f}_h = -\nabla_h\eta \tag{4.57}$$

because horizontal gradients are evaluated at constant s.

The vertical part of the specific force $\boldsymbol{f}_v$ is

$$\boldsymbol{f}_v = \left(-\frac{\partial\eta}{\partial z} + T\frac{\partial s}{\partial z}\right)\hat{\boldsymbol{k}}, \tag{4.58}$$

with $\hat{\boldsymbol{k}}$ the unit vector in the vertical. Because the entropy s is the vertical coordinate it is an increasing function of height z only. Therefore, the vertical specific force becomes

$$\boldsymbol{f}_v = \frac{\mathrm{d}s}{\mathrm{d}z}\left(-\frac{\partial\eta}{\partial s} + T\right)\hat{\boldsymbol{k}}. \tag{4.59}$$

In hydrostatic balance, vertical accelerations are ignored so that the vertical specific force has to vanish,

$$\frac{\partial\eta}{\partial s} = T. \tag{4.60}$$

This is the isentropic formulation of the hydrostatic equation Eq. 4.13. Note that this balance is here only used as an approximation for the vertical force balance; for a strictly motionless fluid we have $\mathrm{d}\eta = T\,\mathrm{d}s$ everywhere, as in Eq. 4.50.

Isentropic coordinates have a couple of advantages. They give an expression for the horizontal specific force as a simple gradient, Eq. 4.57. Furthermore, the vertical velocity in isentropic coordinates is $\mathrm{D}s/\mathrm{D}t$. In the absence of diabatic effects we have $\mathrm{D}s/\mathrm{D}t = 0$, so there is no vertical motion in isentropic coordinates. Because the large-scale flow in the atmosphere is to a good approximation adiabatic, the flow in isentropic coordinates is quasi-horizontal.

A disadvantage of isentropic coordinates is that the atmosphere is assumed to be stable ($\mathrm{d}s/\mathrm{d}z > 0$) everywhere so that the geometric height can be mapped uniquely onto the entropy. Where the atmosphere becomes near neutral, an isentropic coordinate representation becomes inaccurate because here very few isentropic levels correspond to a large height difference. For example, isentropic coordinates are not optimal in the planetary boundary layer. In practice isentropic coordinates are mainly used to simulate strongly stratified flows, such as the stratosphere. Alternatively, models can use hybrid coordinates: the vertical coordinate transforms smoothly from entropy s aloft to a more appropriate variable near the Earth's surface (typically a σ-coordinate with $\sigma = p/p_{\mathrm{sfc}}$).

4.6 STATISTICAL MECHANICS

The isothermal atmosphere has a pressure profile as in Eq. 4.16. Because temperature is constant this also gives the density profile, the so-called *barometric distribution,*

$$\rho = \rho_0 \exp\left(-\frac{Z}{H}\right) = \rho_0 \exp\left(-\frac{\phi}{RT_0}\right), \tag{4.61}$$

with ϕ the geopotential or, equivalently, the potential energy per unit mass. By definition

$$R = \frac{R^\star}{\mu} = \frac{N_A k_B}{\mu} \tag{4.62}$$

with $R^\star$ the universal gas constant, N_A the Avogadro number, k_B the Boltzmann constant, and μ the molar mass. Note that μ/N_A is the mass per molecule so the energy per molecule E equals

$$E = \frac{\mu}{N_A}\phi. \tag{4.63}$$

The mass density ρ is proportional to the probability of finding a molecule at a particular location. Thus the probability P of finding a molecule at some potential energy level E equals

$$P = P_0 \exp\left(-\frac{E}{k_B T}\right). \tag{4.64}$$

The probability is proportional to the well-known *Boltzmann factor* $\exp(-E/k_B T)$, a result that is traditionally arrived at by methods of statistical mechanics.

The Boltzmann factor for general thermodynamic systems follows from considering a system A to be kept at constant volume V and constant temperature T by interacting with a reservoir R (a system with infinitely large

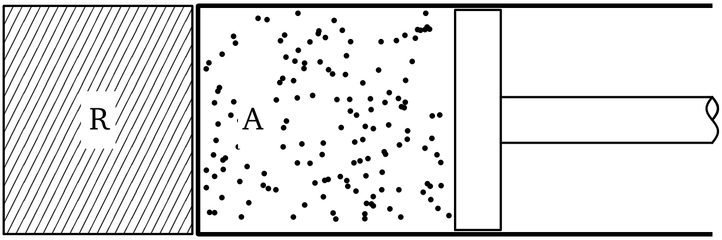


FIGURE 4.5 System A is in contact with a reservoir R which has constant temperature and volume.

heat capacity) at temperature T, see Figure 4.5. System A is kept at constant temperature through heat flows between the reservoir and the system.

What is the probability of finding system A in a particular microstate i? According to the Boltzmann entropy this can be expressed in terms of the entropy of the whole system, see Eq. 2.38. Assume that when system A is in microstate i, the entropy of the reservoir is $S_{R,i}$. The entropy of system A in microstate i is by definition 0, because it corresponds to a single state. That means that the total entropy when system A is in microstate i equals $S_{R,i}$. So the probability P_i of finding system A in microstate i equals

$$P_i = P_0 \exp\left(\frac{S_{R,i}}{k_B}\right). \tag{4.65}$$

The reservoir entropy varies according to the first law at constant volume

$$\mathrm{d}S_R = \frac{\mathrm{d}U_R}{T} = -\frac{\mathrm{d}U_A}{T}, \tag{4.66}$$

where, in the second equality, it is assumed that energy is conserved so $U_A + U_R = \text{constant}$. Because T is constant, the entropy of the reservoir S_R is a linear function of the energy of the system U_A. We thus find that the probability of finding system A in microstate i, with given temperature T, equals

$$\blacktriangleright \qquad P_i = \frac{\exp\left(-U_{A,i}/k_BT\right)}{Z}, \tag{4.67}$$

where we have rewritten the normalization P_0 as $1/Z$, a standard notation in statistical mechanics, see below. The probability is proportional to the Boltzmann factor.

The Boltzmann factor is a central result in classical statistical mechanics (quantum systems at low temperatures have different statistics, depending on whether they are made of fermions or bosons). We will come across instances of the Boltzmann factor in interpreting equilibrium states of droplets. We also referred to the Boltzmann factor in Section 3.3, in the discussion of the heat capacity of ideal gases.

As an application, consider an ideal gas with molecules moving freely inside a fixed volume that is kept at temperature T. The ideal gas is made up of non-interacting molecules. This means that the probability distribution for finding a molecule in a particular energy state is proportional to the Boltzmann factor, with the energy representing the energy of the single molecule. The molecule has internal energy (in any accessible rotational vibrational modes) and kinetic energy. The molecule will always be in some internal energy state (it has a probability of 1 of being in any of the internal energy states), so the probability for the molecule to have a velocity $\boldsymbol{U} = (\mathscr{U}, \mathscr{V}, \mathscr{W})$ will be proportional to

$$\exp\left(-\frac{M_1\mathscr{U}^2 + M_1\mathscr{V}^2 + M_1\mathscr{W}^2}{2k_BT}\right) = \exp\left(-\frac{M_1|\boldsymbol{U}|^2}{2k_BT}\right), \tag{4.68}$$

with M_1 the mass of the molecule. We have assumed here that each velocity vector $\boldsymbol{U}$ represents a possible microstate of the system, which is true for macroscopic systems. The number of microstates corresponding to a speed $|\boldsymbol{U}|$ is proportional to $4\pi|\boldsymbol{U}|^2$, the area of a spherical shell in three-dimensional $\boldsymbol{U}$-space that corresponds to the given speed. Therefore, the probability distribution $P(|\boldsymbol{U}|)$ of finding a molecule at a given speed $|\boldsymbol{U}|$ is

$$\blacktriangleright \qquad P(|\boldsymbol{U}|) = \left(\frac{M_1}{2\pi k_BT}\right)^{3/2} 4\pi|\boldsymbol{U}|^2 \exp\left(-\frac{M_1|\boldsymbol{U}|^2}{2k_BT}\right), \tag{4.69}$$

where the prefactor was calculated by the normalization condition

$$\int_0^\infty P(|\boldsymbol{U}|)\ \mathrm{d}|\boldsymbol{U}| = 1, \tag{4.70}$$

using the identity

$$\int_0^\infty x^2 \exp\left(-a^2x^2\right) = \frac{\sqrt{\pi}}{4a^3}. \tag{4.71}$$

The probability distribution for molecular speeds in an ideal gas, Eq. 4.69, is called the *Maxwell distribution,* see Figure 4.6. The Maxwell distribution has been verified experimentally. The probability distribution for a single component of the velocity, say $\mathscr{U}$ (which can be positive as well as negative), can be found by normalizing the single component factor of Eq. 4.68. We find the distribution $P_1(\mathscr{U})$ for a single velocity component,

$$P_1(\mathscr{U}) = \sqrt{\frac{M_1}{2\pi k_BT}} \exp\left(-\frac{M_1\mathscr{U}^2}{2k_BT}\right). \tag{4.72}$$

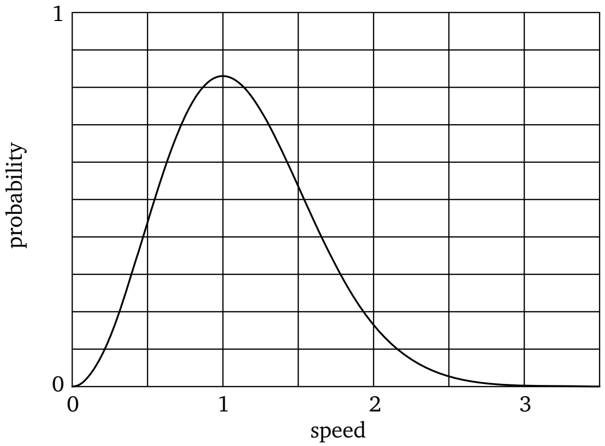


FIGURE 4.6 Maxwell velocity distribution in units of $\sqrt{2k_BT/M_1}$.

With this distribution we can calculate the expectation value of the root-mean-square of $\mathscr{U}$, again using the identity Eq 4.71,

$$\langle \mathscr{U}^2 \rangle = \int_{-\infty}^{\infty} P_1(\mathscr{U})\, \mathscr{U}^2 \, \mathrm{d}\mathscr{U} = \frac{k_B T}{M_1}, \tag{4.73}$$

which is most usefully written as the average kinetic energy for motion in the x-direction,

$$\frac{1}{2} M_1 \langle \mathscr{U}^2 \rangle = \frac{1}{2} k_B T, \tag{4.74}$$

and equivalently for the other two velocity components. We thus find that each translational degree of freedom has on average $k_BT/2$ of kinetic energy, and we are back at the microscopic definition of temperature introduced in Chapter 1, Eq. 1.6.

The Boltzmann factor links the microscopic world of microstates to the macroscopic world of thermodynamics. The probabilities of all microstates have to add up to unity. This then defines the normalization constant Z in Eq. 4.67, which is called the *partition function,*

► $$Z = \sum_i \exp\left(-\frac{U_{A,i}}{k_B T}\right). \tag{4.75}$$

It can now be verified that the mean energy $\langle U_A \rangle$ of system A can be written as

$$\langle U_A \rangle = \sum_i P_i U_{A,i} = k_B T^2 \frac{\partial \ln Z}{\partial T}. \tag{4.76}$$

An assumption of statistical mechanics is that this mean value of the system energy $\langle U_A \rangle$ is the actual observed value U_A. If we define the free energy F_A of the system as

$$F_A = -k_B T \ \ln Z, \tag{4.77}$$

it can now also be verified that this free energy satisfies

$$S_A = -\left(\frac{\partial F_A}{\partial T}\right)_V = k_B \ln Z + \frac{\langle U_A \rangle}{T}, \tag{4.78}$$

which is the expected thermodynamic relationship between F_A and U_A,

$$F_A = \langle U_A \rangle - TS_A. \tag{4.79}$$

This demonstrates that the above definition of F_A in terms of the partition function is consistent with thermodynamics. All the usual thermodynamic quantities can be related in a similar way to the partition function and, for differently constrained systems, to its modifications.

We are now also able to prove the equipartition theorem, which is central to our understanding of heat capacity. The energy of the system U is a function of all the degrees of freedom, usually called *generalized coordinates*. If we integrate all the degrees of freedom out except a single one, say q, then we can write the probability distribution over q as

$$P(q) = \frac{1}{Z} \exp\left(-\frac{U(q)}{k_B T}\right) \tag{4.80}$$

Here we assume that q is a continuous variable. For a discrete variable, the derivation is not strictly valid. Indeed, this is the reason why, for example, the vibrational degrees of freedom do not contribute in the usual way to the heat capacity of a diatomic gas: their energy levels are too far apart to be well approximated by a continuous variable. Using partial integration, we can now demonstrate that the expectation value of $q \, \mathrm{d}U/\mathrm{d}q$ is

$$\left\langle q \frac{\mathrm{d}U}{\mathrm{d}q} \right\rangle = k_B T. \tag{4.81}$$

We assume that for $q \to \infty$ we have $U \to \infty$, so that the boundary term in the partial integration vanishes. Now typically the energy is quadratic in its generalized coordinates. For example, the kinetic energy in the x-direction for a particle can be written as $p^2/2M_1$ with p the particle momentum in the

x-direction. Generally we can write

$$U(q) = U_0 + \alpha q^2, \tag{4.82}$$

where the constant α depends on the nature of the generalized coordinate q. For such a quadratic dependence on the generalized coordinate q, Eq. 4.81 reduces to the equipartition theorem,

$$\langle \alpha q^2 \rangle = k_B T/2. \tag{4.83}$$

For a more general representation of the dependence of the energy on the generalized coordinate of the form

$$U(q) = U_0 + \alpha q^n \tag{4.84}$$

we find the generalized equipartition theorem,[28]

$$\blacktriangleright \qquad \langle \alpha q^n \rangle = k_B T/n. \tag{4.85}$$

The altitude z of a particle in a gravitational field is an example of a degree of freedom with $n = 1$ because its potential energy is $M_1 g z$. According to the generalized equipartition theorem, the mean potential energy of a particle in an isothermal atmosphere is therefore $k_B T$. This result is consistent with the barometric distribution, Eq. 4.61. The generalized equipartition theorem is also used in the derivation of the Rayleigh–Jeans radiation law, see Section 9.9.

PROBLEMS

4.1. Newton's law of gravity states that the gravitational acceleration outside a spherical body of mass M is

$$\boldsymbol{g} = -\hat{\boldsymbol{r}}\, G\frac{M}{r^2},$$

with G the universal gravitational constant, $\hat{\boldsymbol{r}}$ the unit radial vector pointing outward, and r the radial distance to the centre of the sphere. Show that the geopotential around such a body can be written as in Eq. 4.5 with $g_0 = GM/a^2$ where a is the radius of the sphere.

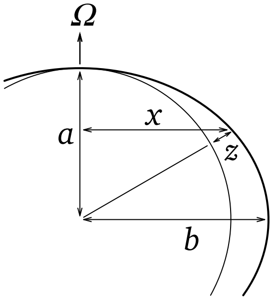
4.2. The centrifugal acceleration points away from the rotation axis $\boldsymbol{\Omega}$ and has magnitude $\Omega^2 x$, with x the distance to the rotation axis and Ω the rotation rate. This corresponds to an additional apparent geopotential

---

[28] Tolman, R. C. (1918) *Phys. Rev.* **11**, 261–275.

of $\Omega^2x^2/2$. With the result from Problem 4.1, the total effective geopotential around a rotating spherical planet can be written as

$$\phi = g_0 \frac{az}{a+z} - \frac{1}{2}\Omega^2 x^2,$$



with z, as before, the vertical height above the sphere of radius a. Show that if the planet's surface is a geopotential, its semi-major axis b and semi-minor axis a are related by

$$b - a = qa, \text{ with } q = \Omega^2 a/2g_0.$$

The observed value for Earth is $q \approx 1/300$. The present model is not accurate because it assumes the gravitational potential of a spherical planet instead of an oblate planet.

4.3. The adiabatic lapse Γ_a is the temperature change with height in a hydrostatic, isentropic fluid. From this definition one would expect that

$$\blacktriangleright \quad \Gamma_a = -\left(\frac{\partial T}{\partial p}\right)_s \frac{\mathrm{d}p}{\mathrm{d}z}. \tag{4.86}$$

Use the reciprocity relation for partial derivatives and a Maxwell relation to show that this is equivalent to Eq. 4.29.

4.4. Assuming water has a constant heat capacity c and a constant thermal expansivity α, show that the temperature profile for an adiabatic lapse rate in water is

$$T = T_0 \exp\left(-g_0 \alpha Z/c\right).$$

The scale height for this profile is very large so a linear profile is appropriate for large height ranges. For water $\alpha \approx 2 \times 10^{-4}\,\mathrm{K}^{-1}$ and $c \approx 4.2 \times 10^3\,\mathrm{J\,kg^{-1}\,K^{-1}}$ so that at typical temperatures the adiabatic lapse rate is less than about $0.15\,\mathrm{K\,km^{-1}}$.

4.5. Estimate typical values for the geopotential thickness of the 1000–500 hPa layer when the surface temperature is 0°C and when the surface temperature is 20°C. By how much would the geopotential thickness of the 1000–500 hPa layer increase on a uniform increase in temperature of 1°C?

4.6. Show that, for an isothermal atmosphere, the Brunt–Väisälä frequency is given by:

$$N^2 = \frac{g^2}{c_p T}.$$

The isothermal atmosphere is one of the most stable profiles observed in the atmosphere (for example, in the lower stratosphere). This provides a typical upper bound for the buoyancy frequency. Hence show that typical buoyancy periods are not shorter than about 5 minutes.

4.7. Total air temperature. The air at the skin of an aircraft has the same speed as the aircraft itself because viscosity makes the air 'stick' to the aircraft. In the frame of reference moving with the aircraft the air then has to decelerate from the aircraft speed to a standstill at the skin of the aircraft. Use the Bernoulli equation to show that the temperature T_t of this decelerated air is related to the actual air temperature T by

$$\frac{T_t}{T} = 1 + \frac{\gamma - 1}{2}\left(\frac{\mathscr{V}}{\mathscr{C}}\right)^2,$$

with $\gamma = c_p/c_v$, $\mathscr{V}$ the flight speed of the aircraft, and $\mathscr{C} = \sqrt{\gamma RT}$ the speed of sound. The ratio $\mathscr{V}/\mathscr{C}$ is called the *Mach number.* The temperature T_t of the decelerated air is called the *stagnation temperature* or the *total air temperature.* In this context, the actual air temperature T is also called the *static air temperature.* A probe on an aircraft measures the total air temperature and uses the flight speed to calculate the static air temperature.

Calculate typical values of the total air temperature for an aircraft at cruising altitude. Would an aircraft expand appreciably due to this heating effect? The difference between T_t and T can be quite large: the same effect is responsible for heating up spacecraft or meteorites when they (re-)enter the Earth's atmosphere. In popular literature it is often said that this heating is due to friction. This is a confusing way of describing the phenomenon. Friction only ensures that the air sticks to the space ship; the heating itself is due to the Bernoulli effect, or its supersonic equivalent (at very low densities ballistic effects of individual molecules need to be taken into account).

4.8. Entropy in statistical mechanics. Using the definition of the probability P_i of a microstate i, from Eq. 4.67, show that the entropy of a system, Eq. 4.78, can be expressed as

$$\blacktriangleright \qquad S_A = -k_B \sum\nolimits_i P_i \ln P_i. \qquad (4.87)$$

This expression for entropy is called the *Gibbs entropy*. Besides its central role in statistical mechanics, it is also the relevant expression of Shannon's information entropy, see footnote 14. Show that if the microstates i have an equal probability, the Gibbs entropy reduces to the Boltzmann entropy, Eq. 2.35.

5

Water in the atmosphere

Water vapour is the most important variable constituent of the atmosphere. For this reason, water vapour is normally treated as a separate constituent alongside the well-mixed gases (nitrogen, oxygen, and argon) that make up the rest of the bulk of the atmosphere. When water vapour cools down enough it will condense to a liquid or even freeze to a solid; it will experience a phase transition. In this chapter we examine the thermodynamics of phase transitions and derive its key result: the Clausius–Clapeyron equation, which describes the equilibrium pressure of the vapour at coexistence with the liquid. Although we use these results to understand the properties of water in the Earth's atmosphere, we can use them equally to understand freezing of CO_2 in the Martian atmosphere or the formation of methane clouds on Titan.

One result of phase transitions in the atmosphere is the formation of clouds and precipitation. However, the formation of drops from vapour depends critically on additional factors such as surface tension and solute concentration. This is the subject of Chapter 7.

Phase transitions are also responsible for a large amount of heat transfer in the atmosphere. To evaporate water requires energy, the latent heat of evaporation. This latent heat is given back to the atmosphere on condensation. So vapour is a storage and transport medium of energy. Latent heat is one of the most important contributions to the global energy cycle. For example, the atmosphere is dominantly heated in the tropics and the main source of heat for the tropical atmosphere is latent heat in convective clouds.

Water vapour is sometimes confused with moist air or steam. Water vapour is a gas made up of H_2O molecules. Moist air is a mixture of dry air and water vapour. Steam is technically a vapour at high temperature (above boiling point) and pressure. In informal language, steam is a mixture of air and vapour where the water vapour is condensed into small haze droplets.

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

5.1 THE CLAUSIUS–CLAPEYRON EQUATION

Consider a closed cylinder with a liquid and its vapour, but nothing else, co-existing.[29] All the time some of the liquid molecules will evaporate to vapour, while some of the vapour molecules will condense to liquid. This situation will equilibrate: when excess molecules are leaving the fluid to go into the vapour, the pressure of the vapour will increase and consequently the number of molecules moving back into the liquid will increase, reducing the excess pressure until equilibrium is restored.

For systems in equilibrium, where the liquid and the vapour phases coexist, the specific Gibbs functions g_l and g_v for the liquid and vapour phases have to be the same. An informal argument for this was given in Section 3.6. The equality of the specific Gibbs functions in turn determines how the vapour pressure changes with temperature, as described by the Clausius–Clapeyron equation. Now we give a more complete argument why the specific Gibbs functions for the liquid and the vapour have to be the same when they coexist.

Consider a cylinder with a piston, which is filled with a liquid and its vapour and held under a constant temperature T and pressure e (the pressure of a vapour is denoted by the symbol e to distinguish it from the total pressure p of air). Because the system is held at constant pressure and temperature, the Gibbs function for the whole system

$$G = U - TS + eV \tag{5.1}$$

has a constant value as it is a function of temperature and pressure only, see Section 3.1. But the total Gibbs function has contributions from both the liquid and the vapour,

$$G = M_l g_l + M_v g_v, \tag{5.2}$$

with M_l and M_v the masses of the liquid and vapour and g_l and g_v the specific Gibbs functions for the liquid and the vapour phases. If now, by some fluctuation, some amount δM of liquid evaporates to the vapour, the total Gibbs function changes by

$$\begin{aligned} \delta G &= M_l\,\delta g_l + M_v\,\delta g_v + (g_v - g_l)\,\delta M \\ &= M_l\,(-\,s_l\,\delta T + v_l\,\delta e) + M_v\,(-\,s_v\,\delta T + v_v\,\delta e) + (g_v - g_l)\,\delta M. \end{aligned} \tag{5.3}$$

Because the total Gibbs function is kept constant, its first order variation has to vanish, $\delta G = 0$. The pressure and the temperature are prescribed, so we

---

[29] This requires the cylinder to be below the critical temperature, the temperature above which the gas and liquid phases become indistinguishable and no phase separation can occur no matter how high the pressure is. The critical temperature for water is 647 K.

have $\delta e = 0$ and $\delta T = 0$. It then follows that

$$\blacktriangleright \quad g_v = g_l. \tag{5.4}$$

The specific Gibbs functions for the liquid and vapour phases are the same if they coexist in the cylinder.

At first sight it appears that the equality of g_v and g_l depends on the fact that the system was in a cylinder at constant temperature and pressure. This is not the case. Suppose we had fixed the piston and the system was at constant temperature T and constant volume V. In that case the total free energy

$$F = U - TS \tag{5.5}$$

would have to stay constant. Using the same arguments as above (with $F = M_l f_l + M_v f_v$) it then follows that

$$\begin{aligned}\delta F &= M_l\,\delta f_l + M_v\,\delta f_v + (f_v - f_l)\,\delta M\\ &= M_l\,(-\,s_l\,\delta T - e\,\delta v_l) + M_v\,(-\,s_v\,\delta T - e\,\delta v_v) + (f_v - f_l)\,\delta M.\end{aligned} \tag{5.6}$$

Now we have to take into account that, although the total volume V is kept constant, the specific volumes v_l and v_v can change when there is a mass transfer δM from the liquid to the vapour phase. We have

$$M_l\,\delta v_l = \delta V_l - v_l\,\delta M_l, \quad M_v\,\delta v_v = \delta V_v - v_v\,\delta M_v. \tag{5.7}$$

Substituting this in Eq. 5.6 we find

$$\begin{aligned}\delta F &= -(M_l s_l + M_v s_v)\,\delta T - e\,(\delta V_l + \delta V_v) + (f_v + e v_v - f_l - e v_l)\,\delta M\\ &= -(M_l s_l + M_v s_v)\,\delta T - e\,\delta V + (g_v - g_l)\,\delta M,\end{aligned} \tag{5.8}$$

where we have used the fact that $g = f + ev$. Because the free energy is kept constant, its first order variation has to vanish, $\delta F = 0$. the temperature and volume are prescribed so we have $\delta V = 0$ and $\delta T = 0$. It again follows that $g_v = g_l$, that is, the specific Gibbs functions for the two phases have to be the same if they coexist.

It turns out we can choose any set of constraints and still find that the specific Gibbs functions for both phases have to be the same, see Problem 5.1. The Gibbs function is the only thermodynamic potential that depends on two intensive variables: pressure and temperature. This is the ultimate reason why for coexisting phases their specific Gibbs functions, rather than any of the other thermodynamic potentials, are the same.

Having established that when the phases coexist their specific Gibbs functions have to be the same, we will now change the temperature T of the

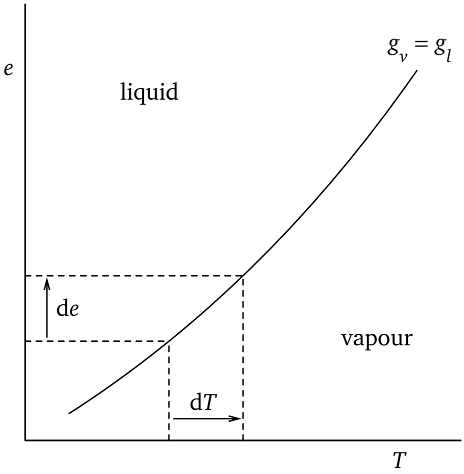


FIGURE 5.1 Derivation of the Clausius–Clapeyron equation: the coexistence curve $g_v = g_l$ determines how much the vapour pressure needs to change (de) when the temperature is changed (dT).

cylinder by dT. This temperature change will lead to a change in vapour pressure e by de, see Figure 5.1. On changing e and T we have for the specific Gibbs functions

$$\mathrm{d}g_l = v_l \,\mathrm{d}e - s_l \,\mathrm{d}T, \quad \mathrm{d}g_v = v_v \,\mathrm{d}e - s_v \,\mathrm{d}T. \tag{5.9}$$

Because for the coexisting phases the specific Gibbs functions g_l and g_v are equal, we must have $\mathrm{d}g_l = \mathrm{d}g_v$. From this follows

$$\frac{\mathrm{d}e}{\mathrm{d}T} = \frac{s_v - s_l}{v_v - v_l}. \tag{5.10}$$

The entropy difference between the liquid and the vapour is related to the heat required to evaporate a certain amount δM of liquid into the vapour phase. This is most easily formalized by noting that at equilibrium

$$\begin{aligned} g_v &= g_l && \text{by Eq. 5.4,} \\ h_v - Ts_v &= h_l - Ts_l && \text{by Eq. 3.12,} \\ L &= T\,(s_v - s_l) && \text{by Eq. 3.58.} \end{aligned} \tag{5.11}$$

Substituting this in Eq. 5.10, we find the *Clausius–Clapeyron equation*,

► $$\frac{\mathrm{d}e_s}{\mathrm{d}T} = \frac{L}{T(v_v - v_l)}. \tag{5.12}$$

This equation describes how the pressure of the vapour changes when the temperature changes; the subscript s has been added to emphasize that the vapour is saturated, that is, the vapour is in equilibrium with the coexisting liquid. (If no liquid is available, the vapour pressure is not determined by the Clausius–Clapeyron equation but by the equation of state for the vapour.) Because $v_v > v_l$ and $L > 0$, we find that the saturated vapour pressure is an increasing function of temperature; at higher temperatures more molecules can escape the attractive potential in the liquid.

5.2 CALCULATION OF SATURATED VAPOUR PRESSURE

The Clausius–Clapeyron equation needs to be integrated to find the vapour pressure at a certain temperature. To do this, assume that the vapour is an ideal gas, so we have $e_s v_v = R_v T$, with R_v the specific gas constant for the vapour (in the case of water vapour $R_v = 461.5\,\mathrm{J\,kg^{-1}\,K^{-1}}$). Further assume that $v_v \gg v_l$, which is true away from the critical temperature. The Clausius–Clapeyron equation then becomes

▶ $$\frac{\mathrm{d}e_s}{\mathrm{d}T} = \frac{L e_s}{R_v T^2}. \tag{5.13}$$

Further assuming that L is constant (this is not a very good assumption, see below) this integrates to

$$e_s = e_{s0} \exp\left[\frac{L}{R_v}\left(\frac{1}{T_0} - \frac{1}{T}\right)\right], \tag{5.14}$$

with e_{s0} the saturated vapour pressure at temperature T_0. Because the specific latent heat, L, reduces with temperature this integration is only valid for small temperature differences.

It can be seen that the saturated vapour pressure is a strongly increasing function of temperature. This aspect is independent of the presence of any other gases alongside the vapour. By Dalton's law, the total pressure of a mixture is the sum of the partial pressures of the constituents; the partial pressure of vapour above a liquid surface is determined by the Clausius–Clapeyron equation, independent of any other gases present. This clarifies the common misconception that 'warm air holds more water'; the air does not 'hold' the water, and its presence does not change the vapour pressure.

If the vapour pressure is the same as the atmospheric pressure, the liquid has reached its boiling point. Above the boiling point, the vapour cannot coexist with the liquid as its equilibrium vapour pressure is higher than the prescribed pressure in the system. From the Clausius–Clapeyron equation it then follows that the boiling point of a liquid reduces if the environmental pressure reduces, see Problem 5.5. The boiling point of water is set at 100°C because at that temperature the vapour pressure is equal to the standard atmospheric pressure of 1013.25 hPa.

In Section 3.6 we derived that the specific latent heat of evaporation can be approximated as

$$L = L_0 - (c_{pl} - c_{pv})\,(T - T_0), \tag{5.15}$$

with L_0 the latent heat of evaporation at some reference temperature T_0. With this approximation for the specific latent heat of evaporation as a function of temperature we can now rewrite the Clausius–Clapeyron equation, Eq. 5.13, as

$$\frac{\mathrm{d}e_s}{e_s} = \frac{L_0 + (c_{pl} - c_{pv})\,T_0}{R_v}\,\frac{\mathrm{d}T}{T^2} - \frac{c_{pl} - c_{pv}}{R_v}\,\frac{\mathrm{d}T}{T}. \tag{5.16}$$

This integrates to

$$e_s = e_{s0}\,\exp\left[\frac{L_0}{R_v}\left(\frac{1}{T_0} - \frac{1}{T}\right)\right]\left(\frac{T_0}{T}\right)^{\alpha}\exp\left[\alpha\left(1 - \frac{T_0}{T}\right)\right], \tag{5.17}$$

with $\alpha = (c_{pl} - c_{pv})/R_v$. At $T_0 = 25^\circ\mathrm{C}$ we have $L_0 = 2.444 \times 10^6\,\mathrm{J\,kg^{-1}}$, $c_{pv} = 1865.1\,\mathrm{J\,kg^{-1}\,K^{-1}}$, and $c_{pl} = 4179.9\,\mathrm{J\,kg^{-1}\,K^{-1}}$. Using these values, the accuracy of Eq. 5.17 is better than 1 part in two thousand when $0^\circ\mathrm{C} < T < 40^\circ\mathrm{C}$ and better than 1.5 parts in a hundred when $0^\circ\mathrm{C} < T < 100^\circ\mathrm{C}$; an impressive result as the saturated vapour pressure varies by more than two orders of magnitude over this temperature range.[30]

For realistically varying L and without ignoring v_l there are several empirical formulae which give values of the saturated vapour pressure as a function of temperature. A particularly simple empirical formula is *Tetens' formula*,

$$e_s(\mathrm{hPa}) = 6.112\,\exp\left(\frac{17.67\,T(^\circ\mathrm{C})}{T(^\circ\mathrm{C}) + 243.5}\right), \tag{5.18}$$

with $T(^\circ\mathrm{C})$ the temperature in degrees celsius.[31] This formula has an accuracy level similar to the analytical expression, Eq. 5.17.

Such formulae can also be produced for the saturated vapour pressure over ice. The argument leading to the Clausius–Clapeyron equation can also be applied to the balance between sublimation and deposition of vapour over ice. Equation 5.14 with $L_{\mathrm{ice}} = 2.826 \times 10^6\,\mathrm{J\,kg^{-1}}$ gives the vapour pressure over ice accurate to two parts in a hundred down to temperatures of –50°C. Again, accurate empirical formulae exist to calculate the saturated vapour pressure over ice.

---

[30] In these equations, the enthalpy of vaporization L always occurs in the fraction L/R_v with R_v the specific gas constant for the vapour. It is often more convenient to rewrite this fraction as $L/R_v = \Lambda/R^\star$ with Λ the *molar* enthalpy of vaporization and $R^\star$ the universal gas constant.

[31] This version of Tetens' formula was published in Bolton, D. (1980) *Mon. Wea. Rev.* **108**, 1046–1053.

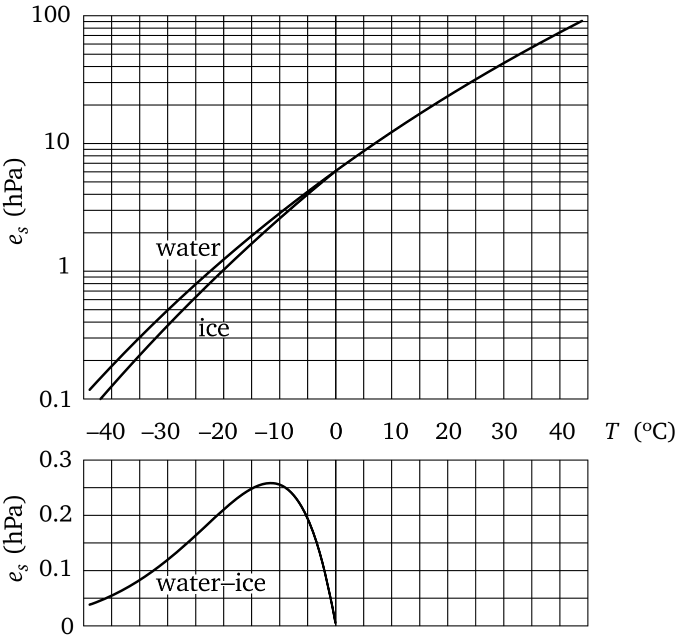


FIGURE 5.2 Saturation vapour pressure e_s over water and over ice as a function of temperature (top), and the difference between the vapour pressures over water and over ice (bottom).

Figure 5.2 shows a graph of the saturated vapour pressures over water and ice, and their difference. The vapour pressure over ice is not defined above the melting point. However, the vapour pressure over water *is* defined below the freezing point because of the existence of supercooled water at those temperatures.

The vapour pressure increases approximately linearly in this logarithmic graph: a temperature difference of 10°C corresponds roughly to a doubling of the saturated vapour pressure. So a temperature increase from 30°C to 31°C has much greater consequences for the saturated vapour pressure than an increase from 5°C to 6°C, see also Table 5.1.

The saturated vapour pressure over ice is smaller than that over water, because it is easier for a molecule to escape water than to escape ice: the

TABLE 5.1 Values of the saturated vapour pressure over water (in hPa) as a function of temperature (sum of the row index and the column index, in °C). Data from Haar, L. et al. (1984) *NBS/NRC Steam Tables*. Hemisphere Publishing, New York.

	00	*10*	*20*	*30*
0	6.113	12.28	23.39	42.46
2	7.061	14.03	26.45	47.58
4	8.136	15.99	29.85	53.23
6	9.354	18.19	33.63	59.45
8	10.73	20.64	37.82	66.30

latent heat of evaporation over supercooled water is lower than the latent heat of sublimation over ice. This has important consequences in cloud physics: if supercooled water drops and ice crystals are present at the same time in a cloud then the ice crystals will grow at the expense of the supercooled drops. This is called the *Bergeron–Findeisen process.*

5.3 HUMIDITY VARIABLES

In atmospheric physics several different variables are used to describe the amount of water vapour in the air. These tend to be used in different contexts and here we put the most commonly encountered variables together.

The first variable is the concentration by mass of water vapour, usually called the *specific humidity* q,

$$\blacktriangleright \qquad q = \frac{\rho_v}{\rho_d + \rho_v}, \tag{5.19}$$

with ρ_v and ρ_d the local densities of the vapour and the dry air respectively. It was introduced in section 1.3 to define the virtual temperature. The specific humidity is dimensionless but it is often given 'units' of $\mathrm{kg\,kg^{-1}}$ or, because there is usually much less water vapour than dry air, units of $\mathrm{g\,kg^{-1}}$. A related quantity is the mass *mixing ratio* r_v,

$$\blacktriangleright \qquad r_v = \frac{\rho_v}{\rho_d}. \tag{5.20}$$

The mass mixing ratio is also dimensionless but is often given units of $\mathrm{kg\,kg^{-1}}$ or $\mathrm{g\,kg^{-1}}$. From their definitions it follows that the two variables can be transformed by

$$r_v = \frac{q}{1-q} \quad \text{and} \quad q = \frac{r_v}{1+r_v}. \tag{5.21}$$

Because generally $q \ll 1$ we find that $r_v \approx q$ (or more precisely: $r = q + q^2 + q^3 + q^4 + \ldots$). In the Earth's atmosphere the two are the same to within one part in a hundred.

From the ideal gas law, $p = \rho RT$, we can relate the partial densities ρ_v and ρ_d to the partial pressures of the water vapour e and the dry air $p - e$, with p the total pressure. It then follows that

$$q = \frac{e}{(\mu_d/\mu_v)\,(p-e) + e}, \quad r_v = \frac{e}{(\mu_d/\mu_v)\,(p-e)}, \tag{5.22}$$

with μ_d and μ_v the effective molar masses of dry air and water respectively, so

$$\mu_d/\mu_v = R_v/R_d = 1.61. \tag{5.23}$$

The *relative humidity* (RH) is defined as the ratio of the actual vapour pressure e to the saturated vapour pressure vapour e_s at the given temperature,

$$\blacktriangleright \quad \mathrm{RH} = \frac{e}{e_s(T)}. \tag{5.24}$$

The relative humidity is dimensionless but it is most commonly expressed as a percentage. Because e_s increases with temperature, the relative humidity at constant specific humidity will decrease at increasing temperature. Relative humidity indicates how far we are away from saturation. For example, just above the sea surface the relative humidity is usually very close to 100% while the specific humidity varies strongly with temperature.

Another commonly used humidity variable is the *dewpoint temperature*. It is defined as follows:

> *Dewpoint temperature T_d is that temperature to which moist air has to be cooled isobarically to achieve saturation.*

Following Eq. 5.22, the vapour pressure also remains constant in such a process, because r_v remains the same. In equations, the dewpoint temperature T_d is defined implicitly by

$$\blacktriangleright \quad e_s(T_d) = e, \tag{5.25}$$

that is, the vapour pressure equals the saturated vapour pressure at the dewpoint temperature. The relative humidity can now be expressed as

$$\mathrm{RH} = e_s(T_d)/e_s(T). \tag{5.26}$$

The difference between the actual temperature and the dewpoint temperature, $T - T_d$, is called the *dewpoint depression*. Air at low relative humidity has a large dewpoint depression; air at 100% relative humidity has a dewpoint depression of 0°C.

The definition of dewpoint temperature suggests a way to measure the humidity in air: cool it isobarically until saturation – the saturated vapour pressure at that temperature is the actual vapour pressure. This is the principle behind accurate measurements of humidity. However, it does require a fairly complex apparatus such as, for example, a chilled mirror hygrometer which works by cooling a mirror until condensation is optically detected.

A simpler way is to put water at the same temperature in contact with the air parcel and to cool the air parcel isobarically by evaporating the water into the parcel. Doing so will change the actual water vapour content of the parcel but this can be taken into account. It forms the basis of a classic humidity measurement using the so-called *wet-bulb temperature*. It is defined as follows:

> *Wet-bulb temperature T_w is that temperature to which air can be cooled isobarically by evaporating water into it.*

At T_w the air is saturated. If it were not we could evaporate more water into it and cool the parcel further. Because in the process the specific humidity of the parcel has increased, the wet-bulb temperature is higher than the dewpoint temperature.

Here we present a quick calculation to indicate how the wet-bulb temperature relates to the vapour mixing ratio; a more accurate calculation is presented in the next section. To evaporate a unit mass of water we need energy L, the latent heat of evaporation. Let us assume that this energy is provided by the internal energy of the dry air, which forms the bulk of an air parcel. Further assume the dry air is an ideal gas. Because the process is isobaric we have

$$L\,\mathrm{d}M_v = -c_p M_d\,\mathrm{d}T, \tag{5.27}$$

with M_v the mass of the vapour in the parcel and M_d the mass of the dry air in the parcel. Divide this equation by M_d to get an equation in terms of vapour mixing ratio r_v,

$$L\,\mathrm{d}r_v = -c_p\,\mathrm{d}T. \tag{5.28}$$

Further assuming L and c_p to be constant, this equation can be integrated from the initial, dry-bulb temperature T to the final, wet-bulb temperature T_w, when the air is saturated. This gives the so-called *psychrometric equation,*

$$\blacktriangleright \quad r_v - r_{vs}(T_w) = -\frac{c_p}{L}(T - T_w), \tag{5.29}$$

with the additional subscript s indicating saturated values of the vapour mixing ratio. The difference $T - T_w$ is called the *wet-bulb depression*. The factor c_p/L is called the *psychrometric constant*.

Figure 5.3 illustrates the construction of the wet-bulb temperature on a graph of vapour mixing ratio versus temperature. It also illustrates the construction of the dewpoint temperature.

Dewpoint temperature is usually measured with a whirling psychrometer (or sling psychrometer). A whirling psychrometer has two thermometers, one of which has a wet wick (piece of fabric) around its bulb. The thermometers are whirled round so as to ventilate air through the wet wick. The air in the wick cools down and saturates, and the temperature of the air in the wick now equals the wet-bulb temperature. The difference between the temperatures of the two thermometers is the wet-bulb depression. The psychrometric equation is then solved to find the mixing ratio from the wet-bulb depression and the temperature. There are tables, charts, slide rules, or computer programs to solve the psychrometric equation; Table 5.2 and Figure 5.4 are examples. If $T_w < 0^\circ\mathrm{C}$ there is a chance the wet bulb might freeze, so these tables and charts have limited validity in that regime although they typically continue

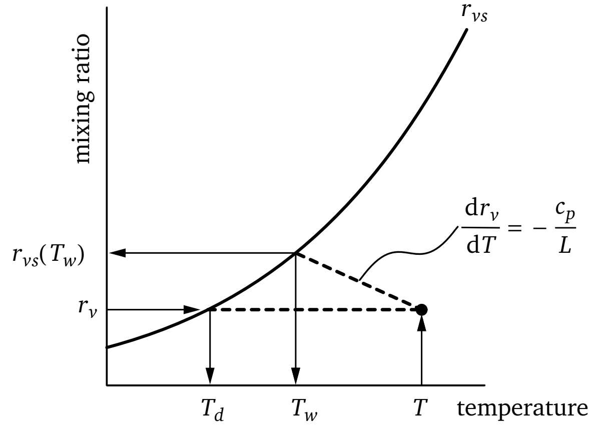


FIGURE 5.3 Construction of dewpoint temperature T_d and wet-bulb temperature T_w. The dewpoint temperature is found by cooling a parcel at fixed vapour mixing ratio r_v until saturation is achieved, $r_v = r_{vs}(T_d)$. The wet-bulb temperature is found by cooling the parcel by evaporating water into it until saturation is achieved. In this process the vapour mixing ratio changes according to $dr_v/dT = -c_p/L$; for constant c_p/L we can infer from the geometry in the figure below that $(T - T_w)\,(c_p/L) = r_{vs}(T_w) - r_v$, which is the psychrometric equation.

working down to $T_w \approx -2$°C. For below freezing T_w we need to use an *ice-bulb* thermometer.

Modern radiosondes use a capacitor with a porous dielectric material inside. The capacitance changes according to the amount of vapour the dielectric absorbs. An electric circuit is then used to determine the capacitance and from this any humidity variable can be determined. Other instruments rely

TABLE 5.2 Psychrometric table giving the relative humidity (in %) at 1000 hPa as a function of dry-bulb temperature and wet-bulb depression, based on the psychrometric equation, Eq. 5.41.

	$T - T_w$ (°C)													
T (°C)	*0.5*	*1*	*1.5*	*2*	*2.5*	*3*	*3.5*	*4*	*5*	*6*	*8*	*10*	*12*	*14*
0	91	82	74	65	57	49	40	32	16	1				
2	92	84	76	68	61	53	46	38	24	10				
4	93	85	78	71	64	57	50	43	30	17			RH (%)	
6	93	86	80	73	67	60	54	48	36	24	1			
8	94	87	81	75	69	63	57	52	40	30	9			
10	94	88	83	77	71	66	60	55	45	35	15			
12	94	89	84	78	73	68	63	58	48	39	21	4		
15	95	90	85	80	76	71	66	62	53	44	28	13		
18	95	91	86	82	77	73	69	65	57	49	34	20	7	
22	96	92	87	84	80	76	72	68	61	54	41	28	17	6
26	96	92	89	85	81	78	74	71	64	58	46	35	24	14
32	97	93	90	87	83	80	77	74	68	63	52	42	33	24

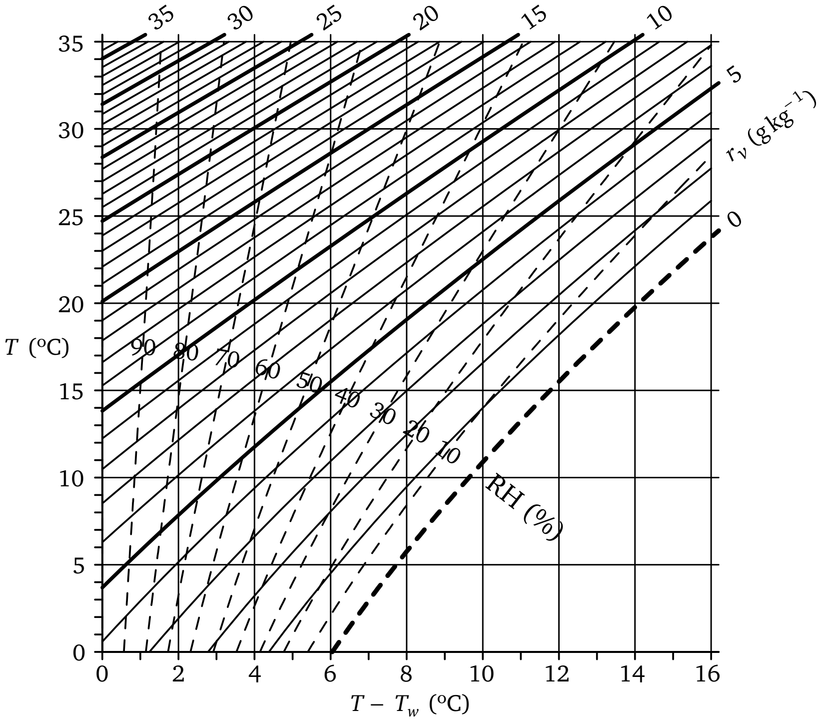


FIGURE 5.4 Psychrometric chart giving the water vapour mixing ratio r (in g kg^{-1}, solid lines) at 1000 hPa and relative humidity (in %, dashed lines) as a function of dry-bulb temperature and wet-bulb depression, based on the psychrometric equation, Eq. 5.41.

on the change of resistance of a material with humidity and use measured resistance to determine the humidity variables.

5.4 MOIST STATIC ENERGY

In Section 4.5 we saw how the generalized enthalpy can be used to define the energy budget of a parcel moving under hydrostatic balance. Here we will examine how to define the energy budget for a moist air parcel.

Equation 4.50 is generally valid for an air parcel that only exchanges heat and work with its environment. In particular we can think of the air parcel as consisting of dry air, water vapour, and liquid water (and possibly ice; we shall ignore this here for simplicity). The total mass M of a parcel can therefore be written as

$$M = M_d + M_v + M_l, \tag{5.30}$$

with the subscripts referring to the three constituents of the parcel. The total enthalpy H of the parcel now is

$$H = h_d M_d + h_v M_v + h_l M_l. \tag{5.31}$$

The specific enthalpy for the parcel therefore is

$$h = (1 - q_w)\, h_d + q_v h_v + q_l h_l, \tag{5.32}$$

where specific humidities (concentrations by mass) for the vapour, liquid, and total water content are

$$q_v = M_v/M, \quad q_l = M_l/M, \quad q_w = q_v + q_l. \tag{5.33}$$

Using the definition of the latent heat of evaporation, $L = h_v - h_l$, the specific enthalpy can be written

▶ $$h = (1 - q_w)\, h_d + q_w h_l + L q_v, \tag{5.34}$$

The generalized enthalpy for a moist parcel now becomes

$$\eta = (1 - q_w)\, h_d + q_w h_l + L q_v + \phi. \tag{5.35}$$

All we have done here is to write out how the generalized enthalpy of the parcel is made up of different components; we have not introduced any new physics. The above expressions are also valid for unsaturated air where $q_l = 0$ and $q_w = q_v$.

For an ideal gas $\mathrm{d}h_d = c_{pd}\,\mathrm{d}T$ and for a liquid, to a good approximation, $\mathrm{d}h_l = c_{pl}\,\mathrm{d}T$ with c_{pl} the specific heat capacity of the liquid. With these approximations, the generalized enthalpy budget, $\mathrm{d}\eta = T\,\mathrm{d}s$ (see Eq. 4.50), can be written

$$(1 - q_w)\, c_{pd}\,\mathrm{d}T + q_w c_{pl}\,\mathrm{d}T + \mathrm{d}(L q_v) + \mathrm{d}\phi = T\,\mathrm{d}s. \tag{5.36}$$

The term on the left-hand side can be integrated to define the *moist static energy, $h_v^\star$*:

▶ $$h_v^\star = \big((1 - q_w)\, c_{pd} + q_w c_{pl}\big)\; T + L q_v + g_0 Z. \tag{5.37}$$

The moist static energy equals the generalized enthalpy for a moist ideal gas parcel apart from an integration constant. The distinction between moist static energy and dry static energy is somewhat artificial. They both represent the fluctuating part of the total specific enthalpy of a parcel.

For adiabatic motions, the moist static energy of a parcel is constant. The term 'adiabatic' is used here to signify that the parcel does not exchange energy or water with its environment; the water in the parcel may still undergo phase transitions with the associated internal heat transfers.

Let us apply this formalism to a derivation of an improved psychrometric equation. Here a parcel is cooled isobarically (although strictly speaking a constant geopotential, rather than constant pressure, is assumed) by evaporating water into it until saturation. So with a constant geopotential ϕ we therefore have that

$$(c_{pd} + c_{pl}r_w)\,T + Lr_v = \text{constant}, \tag{5.38}$$

where we have divided the moist static energy by $1 - q_w$ to convert concentrations by mass to mixing ratios, see Eq. 5.21. Comparing the initial state, with dry-bulb temperature T and vapour mixing ratio r_v, with the final state, with wet-bulb temperature T_w and mixing ratio $r_{vs}(T_w)$ (by definition of T_w this is the saturated mixing ratio), we see that conservation of moist static energy requires that

$$(c_{pd} + c_{pl}r_w)T + L(T)r_v = (c_{pd} + c_{pl}r_w)T_w + L(T_w)r_{vs}(T_w). \tag{5.39}$$

Using the fact that $L(T_w) = L(T) - (T_w - T)(c_{pl} - c_{pv})$, this equation can be rearranged to give

$$L(T)(r_v - r_{vs}(T_w)) = -\left(c_{pd} + c_{pl}(r_w - r_{vs}(T_w)) + c_{pv}r_{vs}(T_w)\right)(T - T_w). \tag{5.40}$$

There is an arbitrary total water mixing ratio r_w in this equation which influences the result. To set r_w we can assume that we provide just enough water to cool down the parcel to saturation. This means that $r_w = r_{vs}(T_w)$ and we find the following improved psychrometric equation:

► $$r_v - r_{vs}(T_w) = -\frac{c_{pd} + c_{pv}r_{vs}(T_w)}{L(T)}\,(T - T_w). \tag{5.41}$$

This equation forms the basis for the psychrometric chart, Figure 5.4 and the psychrometric table, Table 5.2. This equation introduces only small corrections to the approximate psychrometric equation, Eq 5.29, and for small vapour mixing ratios or low temperatures, this equation reduces to Eq. 5.29. The differences are usually smaller than the accuracy at which a wet-bulb thermometer can be operated to measure humidity variables.

PROBLEMS

5.1. When two phases coexist in a cylinder which has constant volume V and which is isolated from its environment ($dQ = 0$ so S is constant) then the total internal energy U of the system cannot change. Show that under these constraints the specific Gibbs functions for the two phases are the same.

5.2. Linearize Eq. 5.14 around T_0 to demonstrate that, for water, the saturated vapour pressure doubles approximately every 10°C.

5.3. Use the Clausius–Clapeyron equation in the form of Eq. 5.13 to demonstrate that the expected fractional change in the atmospheric water content is about 6.5% per degree global mean surface warming. What assumptions are made in arriving at this result?[32]

5.4. For mercury (Hg) the vapour pressure at 500 K is 5.239 kPa and at 550 K is 19.4348 kPa. Estimate the vapour pressure for mercury at 530 K and at room temperature, using Eq. 5.14. Will the room temperature estimate be an overestimate or an underestimate?

5.5. Use Eq. 5.18 to estimate the boiling point of water at environmental pressure of 700 hPa. Assuming blood is largely made up of water, estimate at what height blood begins to boil spontaneously. Use Table 4.1 for a standard atmospheric profile. In reality, blood in the body would be at a somewhat higher pressure than the environmental pressure. Nevertheless, at these heights humans would start to suffer ebullism, gas bubbles in the body fluids, and would need a space suit to survive.

5.6. Mixing clouds. Calculate the temperature and the vapour mixing ratio of a mixture of two air parcels of mass M_i, temperature T_i, and vapour mixing ratio r_i with $i = 1, 2$. Ignore any variations in heat capacity with moisture content and assume that the mixing occurs isobarically. Demonstrate that even if the two parcels are initially subsaturated, their mixture might be supersaturated. This process is responsible for aircraft contrails and steaming breath in cold weather.

---

[32] The effects of global warming on water vapour and the hydrological cycle are discussed in Boer, G. J. (1993) *Clim. Dynam.* **8**, 225–239; Allen, M. R. & Ingram, W. J. (2002) *Nature* **419**, 224–232; Held, I. M. & Soden, B. J (2006) *J. Clim.* **19**, 5686–5699.

6

Vertical structure of the moist atmosphere

In this chapter we consider how the presence of water modifies the vertical structure in the atmosphere. The key features are a change in the adiabatic lapse rate compared to the dry adiabatic lapse rate and a more complex stability behaviour for finite vertical displacements. Both of these are due to the latent heat release on lifting a saturated parcel.

6.1 ADIABATIC LAPSE RATE FOR MOIST AIR

The adiabatic lapse rate for a saturated parcel is expected to be different from that of an unsaturated parcel. If we lift a saturated parcel, it will expand and cool down, leading to vapour condensation. This condensation will release latent heat, which will partially offset the cooling. The adiabatic lapse rate for a saturated parcel is therefore lower than that for an unsaturated parcel.

First consider moist but unsaturated air. There is no liquid water present in the parcel and the entropy budget is that of an ideal gas made up of dry air and vapour. All the tools developed for ideal gases then remain valid. For example, the isobaric expansion coefficient α_p of unsaturated air equals $1/T$, as is the case for all ideal gases. The isobaric specific heat capacity of moist air is the weighted mean of the heat capacities for dry air and water vapour,

$$c_p = (1 - q)\, c_{pd} + q c_{pv}. \tag{6.1}$$

We therefore find that the adiabatic lapse rate, Eq. 4.29, for an unsaturated parcel is the dry adiabatic lapse rate g/c_p with the heat capacity c_p as above. Because $c_{pd} = 1004\,\mathrm{J\,kg^{-1}\,K^{-1}}$ and $c_{pv} = 1865\,\mathrm{J\,kg^{-1}\,K^{-1}}$, we find that the specific heat capacity of moist air is slightly higher than that of dry air, typically by less than 1%. Consequently, the dry adiabatic lapse rate of moist air is slightly lower than that of dry air.

Now consider moist saturated air. The adiabatic lapse rate can be found from the general definition of Eq. 4.29. We need to know the isobaric thermal

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

expansivity and the specific isobaric heat capacity for moist saturated air. An approximate equation of state for a parcel of moist saturated air is

$$(p - e_s)\, V = M_d R_d T. \tag{6.2}$$

Here, M_d is the mass of dry air in the parcel of volume V, and R_d is the specific gas constant for dry air. This equation of state expresses that the partial pressure for the dry air component of the parcel satisfies the equation of state for an ideal gas. We have ignored any contributions of the liquid water to the volume of the parcel. From this equation of state we can show that the isobaric thermal expansivity α_p for moist saturated air is

$$\alpha_p = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p = \frac{1}{T}\left(1 + \frac{L r_{vs}}{R_d T}\right). \tag{6.3}$$

The isobaric heat capacity for moist saturated air can be found from the specific enthalpy h for a saturated parcel, Eq. 5.34. For small total water content this can be approximated as

$$h = h_d + L r_{vs}. \tag{6.4}$$

We will ignore variations in L and we will set

$$r_{vs} = \frac{\mu_v}{\mu_d}\frac{e_s(T)}{p}, \tag{6.5}$$

which is consistent with the assumed low total water content. We can now determine the isobaric heat capacity from $c_p = (\partial h/\partial T)_p$. Using the Clausius–Clapeyron equation for an ideal gas we find

$$c_p = c_{pd} + \frac{L^2 r_{vs}}{R_v T^2}. \tag{6.6}$$

Now we can use the general definition of the adiabatic lapse rate Eq. 4.29 to find the *moist-adiabatic lapse rate,*

► $$-\frac{\mathrm{d}T}{\mathrm{d}z} = \Gamma_s = \frac{g\alpha_p T}{c_p} = \frac{g}{c_{pd}}\frac{1 + L r_{vs}/R_d T}{1 + L^2 r_{vs}/c_{pd} R_v T^2}. \tag{6.7}$$

The moist-adiabatic lapse rate is lower than the dry adiabatic lapse rate g/c_{pd} for dry air. The difference is, to first order, proportional to e_s/T. This means that at low temperatures, the moist-adiabatic lapse rate is close to the dry adiabatic lapse rate. For high temperatures, the moist adiabatic lapse rate is considerably lower than the dry adiabatic lapse rate. For example, at standard pressure and temperature we have $\Gamma_s = 0.49\, g/c_{pd} = 4.8\,\mathrm{K\,km^{-1}}$.

We can use the full expression for the moist static energy for a saturated parcel and the temperature dependence of L to derive a more accurate equation for the isobaric heat capacity,

$$c_p = (1 - q_{vs})\, c_{pd} + q_{vs} c_{pv} + q_l\,(c_{pl} - c_{pd}) \\ + (1 + (\mu_d/\mu_v - 1)\, q_{vs}) \frac{L^2 q_{vs}}{R_v T^2}. \qquad (6.8)$$

This improved expression for the heat capacity can be used to find a more accurate equation for the adiabatic lapse rate. For modest water content, the heat capacity is accurately represented by Eq. 6.6.

The moist-adiabatic lapse rate depends on the total water content of the parcel, except in its approximate form, Eq. 6.7. Therefore it cannot be expressed as a function of pressure and temperature alone. To overcome this problem we often consider *pseudo-adiabatic* processes where any liquid water leaves the parcel, say, through some idealized instantaneous precipitation.[33] In this case the properties of the saturated parcel are a function of pressure and temperature alone. For example, the heat capacity is given by Eq. 6.8, where we set $q_l = 0$. The associated lapse rate is called the *pseudo-adiabatic lapse rate.* Figure 6.1 shows values of the pseudo-adiabatic lapse rate as a

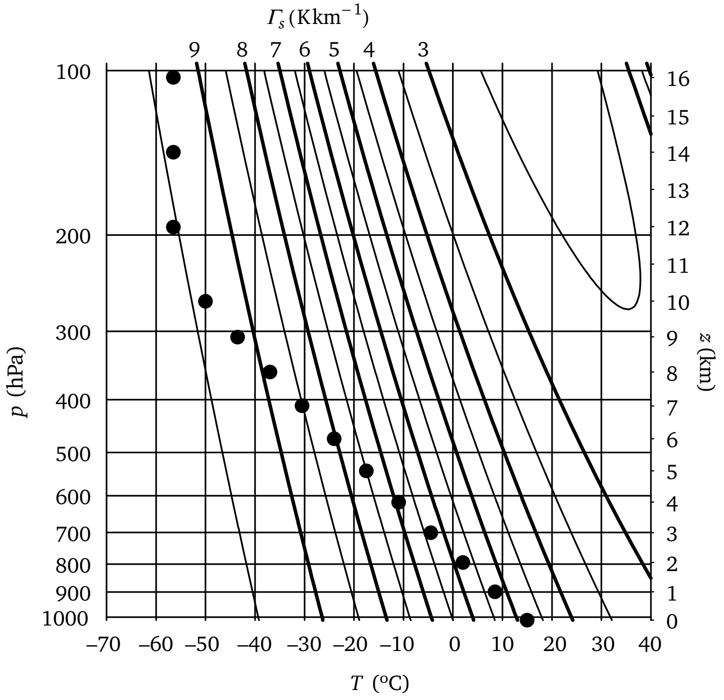


FIGURE 6.1 Contour lines: pseudo-adiabatic lapse rate Γ_s as a function of pressure and temperature. Points: pressure and temperature of the US Standard Atmosphere.

---

[33] Pseudo-adiabatic processes are somewhat artificial for descending parcels: this would require the external provision of just enough liquid water at the right temperature to keep the parcel saturated.

function of temperature and pressure. It should be emphasized that local values of pseudo-adiabatic lapse rates and moist adiabatic lapse rates are very close indeed for normal atmospheric conditions. In turn, Eq. 6.7 represents a very good approximation to both of these.

With the pseudo-adiabatic lapse rate given as a function of temperature and pressure (or the moist-adiabatic lapse rate, if $q_w = q_v + q_l$ is given) we can also write the pressure lapse rate $\mathrm{d}T/\mathrm{d}p = \Gamma_s/\rho g$ as a function of temperature and pressure. We can then integrate this expression to find the temperature as a function of pressure by following a parcel pseudo-adiabatically or moist adiabatically from a starting pressure and temperature to a new pressure as

$$T(p_1) = T(p_0) + \int_{p_0}^{p_1} \frac{\Gamma_s}{\rho g}\,\mathrm{d}p. \tag{6.9}$$

This procedure defines a *pseudo-adiabat* or a *moist adiabat*. It is the temperature a saturated parcel would have if moved pseudo-adiabatically or moist adiabatically from pressure p_0 to pressure p_1. Although the above equation defines pseudo-adiabats or moist adiabats, for graphical representation it is easier to plot isolines of the entropy of a saturated parcel, as discussed in the next section.

6.2 ENTROPY BUDGET FOR SATURATED AIR

Following similar arguments leading to Eq. 5.32, the specific entropy s for a parcel consisting of dry air, vapour (concentration by mass q_v), and liquid (concentration by mass q_l) equals

$$s = (1 - q_w)\, s_d + q_v s_v + q_l s_l, \tag{6.10}$$

with, as before, $q_w = q_v + q_l$ the total water concentration. At equilibrium we have

$$T\,(s_v - s_l) = L, \tag{6.11}$$

with L the latent heat of evaporation. Equilibrium means that the vapour is in a Clausius–Clapeyron equilibrium with the liquid in the parcel. In other words, the vapour is always saturated, $q_v = q_{vs}$. So we have

$$s = (1 - q_w)\, s_d + q_w s_l + L q_{vs}/T. \tag{6.12}$$

The adiabatic lapse rate is defined by $\mathrm{d}s/\mathrm{d}z = 0$. When lifting a parcel, its temperature will decrease and therefore also its (saturated) vapour mixing ratio. The third term in the entropy expression then decreases so the first two terms will have to increase. This corresponds to the latent heating these constituents experience due to condensation of water vapour.

There is a potential confusion in the terminology used to describe the latent heating of a saturated parcel. A lifted parcel will experience latent heating but does not increase its entropy. The latent heating is a transfer of heat released by condensation of vapour to the internal energy of the parcel. There is no external heat input, so no production of reversible entropy. The parcel is assumed to be in equilibrium, so there is no irreversible production of entropy. The confusion possibly arises from the common practice in computer models of separating the dry air component of the model from the water components. In such a model, condensation provides heating for the dry air component.

Assume dry air is an ideal gas. Its specific entropy is then given by Eq. 3.40, with the pressure the partial pressure of the dry air $p - e_s(T)$. The specific entropy of the liquid water is given as in Problem 3.6. We therefore have

$$s_d = c_{pd} \ln (T/T_0) - R_d \ln (p/p_0) - R_d \ln (1 - e_s/p) \tag{6.13}$$

$$s_l = c_{pl} \ln (T/T_0) \tag{6.14}$$

with T_0 and p_0 constant reference values. The third term in the specific entropy of the dry air component is called the *mixing entropy*; Chapter 8 covers properties of mixtures in more detail. Figure 6.2 shows how the entropy for a saturated parcel compares with that of a dry parcel. We actually plot the

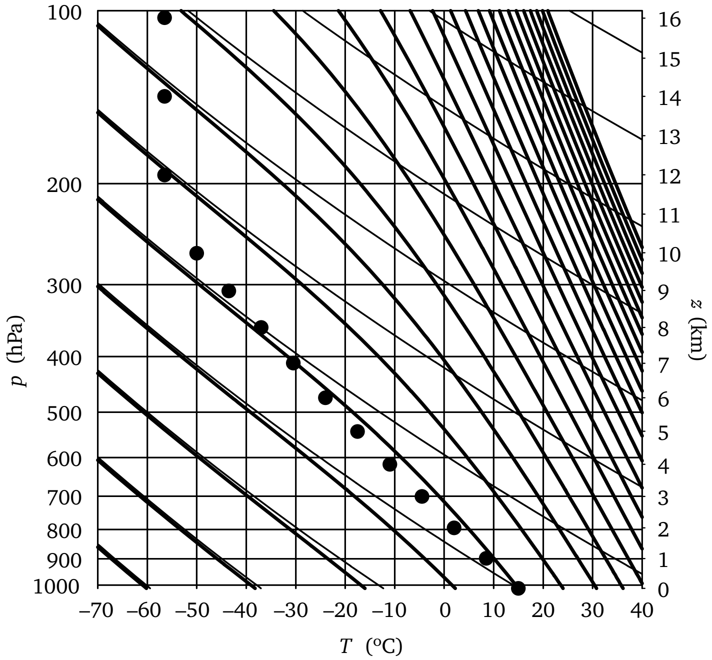


Figure 6.2 Thick contour lines: isolines of specific entropy for a saturated parcel with no liquid water (pseudo-adiabats). Thin contour lines: isolines of specific entropy for a dry parcel (dry adiabats). In both cases, the entropy constant is such that at standard pressure (1013.25 hPa) and temperature (15°C) the entropy vanishes. The contour interval is 100 J kg^{-1} K^{-1}. Points: temperature and pressure of the US Standard Atmosphere.

pseudo-entropy, that is, the entropy of a saturated parcel without liquid water. The pseudo-entropy can be found from Eq. 6.12 if we choose $q_w = q_{vs}$. We can see in Figure 6.2 that the pseudo-adiabats, isolines of pseudo-entropy, become parallel to the dry adiabats at low temperatures. We can also see that at all pressures and temperatures the pseudo-adiabats have a lower lapse rate than the dry adiabats.

It is common usage in meteorology to translate the entropy for a saturated parcel into temperature-like quantities, inspired by the potential temperature. To this end, we rewrite the specific entropy of the saturated parcel as

$$s = (1 - q_w)\, c_{pd} \ln(\theta_e / T_0), \tag{6.15}$$

where we have introduced the *equivalent potential temperature*, θ_e

$$\theta_e = \theta_d \exp\left(\frac{L r_{vs}}{c_{pd} T}\right) \left(\frac{T}{T_0}\right)^{r_w c_{pl}/c_{pd}} \left(1 - \frac{e_s}{p}\right)^{-R_d/c_{pd}}, \tag{6.16}$$

with

$$\theta_d = T \left(\frac{p_0}{p}\right)^{R_d/c_{pd}} \tag{6.17}$$

the potential temperature of dry air. For a saturated parcel, the equivalent potential temperature θ_e is *not* the actual potential temperature of the parcel, as defined by Eq. 3.78. It is just a convenient measure of the entropy of a parcel. As can be seen from its definition, at low temperatures and small water mixing ratios, the equivalent potential temperature is the same as the dry potential temperature.

If we assume $r_w \ll 1$, and therefore $e_s \ll p$, we can approximate the above expression. The equivalent potential temperature now becomes

$$\blacktriangleright \quad \theta_e = \theta_d \exp\left(\frac{L r_{vs}}{c_{pd} T}\right). \tag{6.18}$$

This expression is commonly used to calculate the equivalent potential temperature if high accuracy is not essential.

A moist adiabat is defined by a constant value of the specific entropy and thus a constant value of the equivalent potential temperature. Again, the definition depends on the total amount of water in a parcel. As before, we can overcome this by considering the entropy in a pseudo-adiabatic process. The value of the entropy in a pseudo-adiabatic process can now be expressed as the logarithm of a *pseudo-equivalent potential temperature*. Emanuel's book provides further details.[34] Because a parcel that is lifted to great heights will lose all its water in a pseudo-adiabatic process, its entropy will then be given

---

[34] Emanuel, K. (1994) *Atmospheric convection*. Oxford University Press, Oxford.

in terms of its usual dry potential temperature. We can therefore define the pseudo-equivalent potential temperature as follows:

> *The pseudo-equivalent potential temperature is the potential temperature a parcel would have if lifted pseudo-adiabatically to great heights.*

The pseudo-equivalent potential temperature of a lifted parcel is lower than the equivalent potential temperature because the liquid water lost in the pseudo-adiabatic ascent carries internal energy away from the parcel.

It is important to remember that the pseudo-equivalent potential temperature is again just a convenient label of a pseudo-adiabat; it is not a true potential temperature of a saturated parcel. A more commonly used label for a pseudo-adiabat is the temperature where the pseudo-adiabat reaches the reference pressure. This temperature is called the *wet-bulb potential temperature,* θ_w.[35] Of all the potential temperature-like variables used in meteorology, the wet-bulb potential temperature is perhaps the one that is closest to the original general definition of potential temperature, in Eq. 3.78. It is not quite a true potential temperature because the process used to define it is pseudo-adiabatic.

This definition of the pseudo-equivalent potential temperature and wet-bulb potential temperature is also used for unsaturated parcels: if the parcel is initially unsaturated, the lifting parcel will conserve its potential temperature. When the parcel has cooled down enough it will become saturated. The level at which this happens is called the *lifting condensation level (LCL)*. Beyond the LCL the pseudo-equivalent potential temperature will be conserved.

6.3 FINITE AMPLITUDE INSTABILITIES

In Section 4.4 the stability of a dry atmosphere was discussed as the result of the vertical gradient of specific entropy. For positive $\mathrm{d}s/\mathrm{d}z$ the atmosphere is stable for infinitesimal parcel displacements, for negative $\mathrm{d}s/\mathrm{d}z$ the atmosphere is unstable, and for $\mathrm{d}s/\mathrm{d}z = 0$ the atmosphere is neutrally stable. Insofar as the air can remain saturated, this is still the case for saturated air, as long as the entropy is calculated from Eq. 6.12.

When an air parcel is displaced over a finite distance (rather than an infinitesimal distance) new effects can occur. The linearizations of Section 4.4 are not valid anymore and we need to reconsider the buoyancy of a parcel, now assuming finite displacements.

Suppose we have a stable profile of subsaturated air, with the specific entropy increasing with height. If we lift the parcel beyond the lifting condensation level, latent heat will be released. This latent heat release will increase the temperature above that expected from the dry adiabatic lapse rate; although the parcel total entropy remains the same, the entropy released by

---

[35] For accurate expressions for the wet-bulb potential temperature see Bolton, D. (1980) *Mon. Wea. Rev.* **108**, 1046–1053.

condensation is used to increase the temperature of the parcel. It is now possible to lift the parcel so far that its temperature will become higher than that of the environment and the parcel becomes positively buoyant. Strictly speaking, the specific volume of the parcel has to become larger than that of the environment, see Eqs. 4.35 and 4.36. The level at which the parcel becomes positively buoyant is called the *level of free convection* (LFC).

The situation sketched above is called *conditional instability.* The instability is conditional on saturation after a finite displacement of the parcel. A conditionally unstable atmosphere is stable for infinitesimal displacements; it has a real buoyancy frequency. The mechanism of conditional instability is sketched in Figure 6.3.

From inspection of Figure 6.3 we find that conditional instability cannot occur if the environmental temperature lapse rate is everywhere lower than the moist adiabatic lapse rate. In this case the parcel can nowhere cross the environmental profile. A formal way of stating this is that stability is ensured if for the profile $\mathrm{d}\theta_e^\star/\mathrm{d}z > 0$, where $\theta_e^\star$ is the *saturated equivalent potential temperature*, the equivalent potential temperature if we assume the parcel is saturated. If there are layers in the profile where the temperature lapse rate is larger than the moist adiabatic lapse rate, the potential (in-)stability depends on the whole profile and which parcel is being lifted. A careful analysis of the full profile is then needed.

The final type of instability is called *potential instability* or, confusingly, *convective instability*. This type of instability is due to the lifting of a whole layer of air rather than a single parcel. This typically happens when air flows

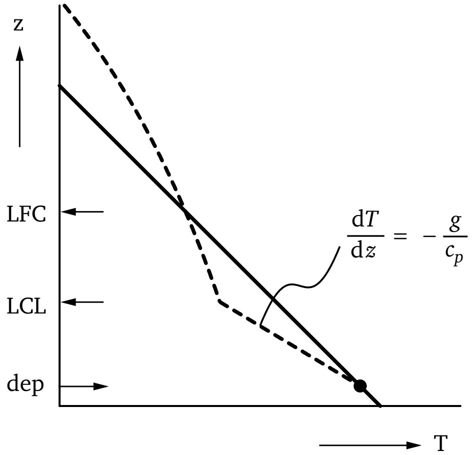


FIGURE 6.3 Illustration of conditional instability. On lifting the parcel from its departure point (dep) it will follow the dry adiabatic lapse rate (dashed straight line) and become negatively buoyant, that is, colder than the environment (solid line). Beyond the lifting condensation level the parcel will follow a moist adiabat (dashed curved line). It is now possible that the parcel will reach the same temperature as the environment. Beyond this level, the level of free convection (LFC), the parcel is positively buoyant.

over a hill. Suppose the layer was initially stable; that is, the potential temperature (or specific entropy) increases with height in the layer. If there is no humidity in the air, then the layer will always be stable because the potential temperature is conserved on lifting and a positive gradient cannot change into a negative gradient. But now suppose that the lowest part of the layer is more humid than the upper part. When lifting the layer, the lower part can condense sooner and start following a moist adiabat, that is, cool at a lower rate than the dry adiabatic lapse rate. We may then encounter a situation where the lower part of the layer gets a higher dry potential temperature than the upper part and the layer becomes unstable. This is potential instability.

The easiest way to diagnose potential instability is to consider the equivalent potential temperature. As stated before, this is the potential temperature of a parcel when it is lifted to great heights such that all the water vapour has condensed out (strictly speaking, this defines the pseudo-equivalent potential temperature). Suppose the layer is initially subsatured and stable, $d\theta/dz > 0$, while the vapour content in the lower part of the layer is much higher than in the upper part such that $d\theta_e/dz < 0$. This means that if we lift the layer to great heights, the potential temperature will decrease with height in the layer, $d\theta/dz < 0$. So at some point in the lifting process the layer must have become unstable.

6.4 VERTICAL STRUCTURE IN THERMODYNAMIC DIAGRAMS

Here we will take a graphical approach to the analysis of vertical structure, to complement the formal approach in the previous sections. We will use a *tephigram,* to illustrate this. A tephigram is a Ts diagram which has been adjusted for meteorological use. Appendix C discusses the precise structure of a tephigram as well as the relationship between tephigrams and skew$T$–log$p$ diagrams, the other commonly used thermodynamic diagram. The analysis below works essentially the same for either of these diagrams.

First we will consider the meaning of a single sounding data point as plotted on a tephigram. A radiosonde produces measurements of temperature T, dewpoint temperature T_d, and pressure p. We can plot those measurements on the tephigram. First we plot a single measurement of T, T_d, and p. We can then use the tephigram to deduce several other variables.

On the tephigram in Figure 6.4 we have marked such a single measurement: the sonde measured at 850 hPa a temperature $T = 8.6°\mathrm{C}$ and dewpoint temperature $T_d = -4°\mathrm{C}$.

The potential temperature θ is found by following the dry adiabat (θ constant) to 1000 hPa and we can read off that $\theta = 22°\mathrm{C}$. The value of the potential temperature can be read off anywhere along the dry adiabat, as the dry adiabat is labelled by its potential temperature.

The dewpoint temperature gives the water vapour mixing ratio. Remember that $e_s(T_d) = e$ so if the air parcel at $p = 850$ hPa and $T = 8.6°\mathrm{C}$ were cooled

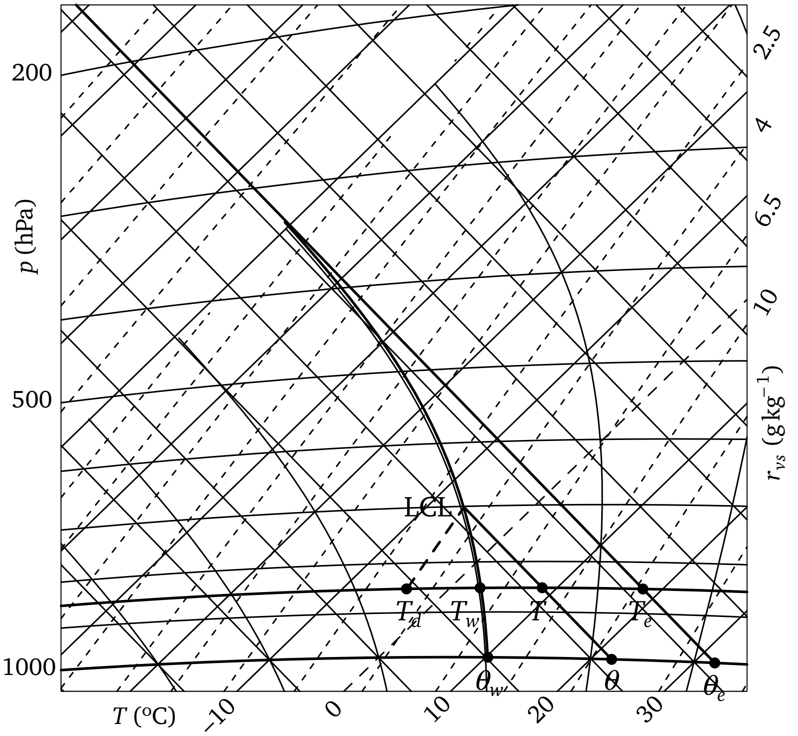


FIGURE 6.4 Normand's construction. Given the pressure p, temperature T, and dewpoint temperature T_d, we can derive the wet-bulb temperature T_w, wet-bulb potential temperature θ_w, potential temperature θ, the pseudo-equivalent potential temperature θ_e, and the equivalent temperature T_e. The lifting condensation level is also found.

isobarically, it would reach saturation at $T_d = -4°\text{C}$. On the tephigram we can read off that the saturated mixing ratio at $T = -4°\text{C}$ and $p = 850\,\text{hPa}$ is about $3.5\,\text{g}\,\text{kg}^{-1}$. So this is the vapour mixing ratio of the air parcel: $r_v = 3.5\,\text{g}\,\text{kg}^{-1}$.

The saturated vapour mixing ratio at $T = 8.6°\text{C}$ and $p = 850\,\text{hPa}$ is read off as about $8.5\,\text{g}\,\text{kg}^{-1}$. This means that the relative humidity of the air parcel is $3.5/8.5 \times 100\% \approx 40\%$.

If we lift the air parcel, then it follows the dry adiabat upward. It will cool down until its mixing ratio of $3.5\,\text{g}\,\text{kg}^{-1}$ becomes the saturated mixing ratio. This happens at the vertex labelled lifting condensation level in Figure 6.4, the lifting condensation level. We can read off that the lifting condensation level for this air parcel is about 700 hPa. So the lifting condensation level is found where the dry adiabat crosses the isoline of the air parcel's mixing ratio.

If we lift the parcel further it will now follow the moist adiabat, or the pseudo-adiabat, depending on what happens to the liquid water after condensation. Note that the tephigram has pseudo-adiabats labelled, and we will always assume pseudo-adiabatic processes. Lifting it further until all moisture is condensed out, the parcel will reach the potential temperature, defined as its pseudo-equivalent potential temperature θ_e. It can be read off to be about 32°C.

The temperature the parcel would have if all its water vapour were condensed out isobarically is called the equivalent, or pseudo-equivalent temperature T_e. It is therefore also the temperature a parcel would have at the same pressure level if its potential temperature were equal to θ_e.

It is of interest to locate the wet-bulb temperature on the tephigram. Remember that the wet-bulb temperature is defined as that temperature to which an air parcel can be cooled isobarically by evaporating water into it. We can get the wet-bulb temperature in an equivalent but more roundabout way: first lift the parcel adiabatically up to the lifting condensation level. The parcel is now saturated. Then lower the parcel back to its original pressure level while at the same time evaporating water into it to keep it saturated. This process occurs pseudo-adiabatically because the energy used for evaporation is extracted from the internal energy of the parcel, like in a wet-bulb thermometer. At the end of this process the parcel will have its original pressure, it will be saturated, and it will have cooled by evaporating water into the parcel. We must conclude that at the end of this process the parcel will be at its wet-bulb temperature T_w.

The process can be easily followed on the tephigram: follow the dry adiabat up to its lifting condensation level. Then lower the parcel along the pseudo-adiabat (the parcel is saturated in this process) until its original pressure is achieved. The point on the tephigram is labelled T_w. In our present case we can read it off as about $T_w = 3$°C.

Following the pseudo-adiabat further down to 1000 hPa, we can read off the wet-bulb potential temperature, θ_w. Wet-bulb potential temperature is simply a label for a pseudo-adiabat, just as potential temperature is a label for a dry adiabat. On the tephigram we can read off that $\theta_w = 10$°C. Pseudo-equivalent potential temperature is also a label for a moist adiabat. Note that these two labels have different values for any chosen moist adiabat; in our case $\theta_w = 10$°C and $\theta_e = 32$°C. These two labels refer to the same moist adiabat.

The construction above is called *Normand's construction:*

> *The dry adiabat through point* (p, T)*, the isoline of* r_{vs} *through point* (p, T_d)*, and the pseudo-adiabat through point* (p, T_w) *meet in one point at the lifting condenstation level.*

Let us now continue and plot a full profile on the tephigram. The solid line in Figure 6.5 represents the temperature T of the sounding, the dashed line represents the dewpoint temperature T_d. The Normand construction for the parcel at 870 hPa is also drawn in. We can perform Normand's construction at every point on this sounding to get any of the moisture variables at any level, but we can also view the more general structure of the profile. For example:

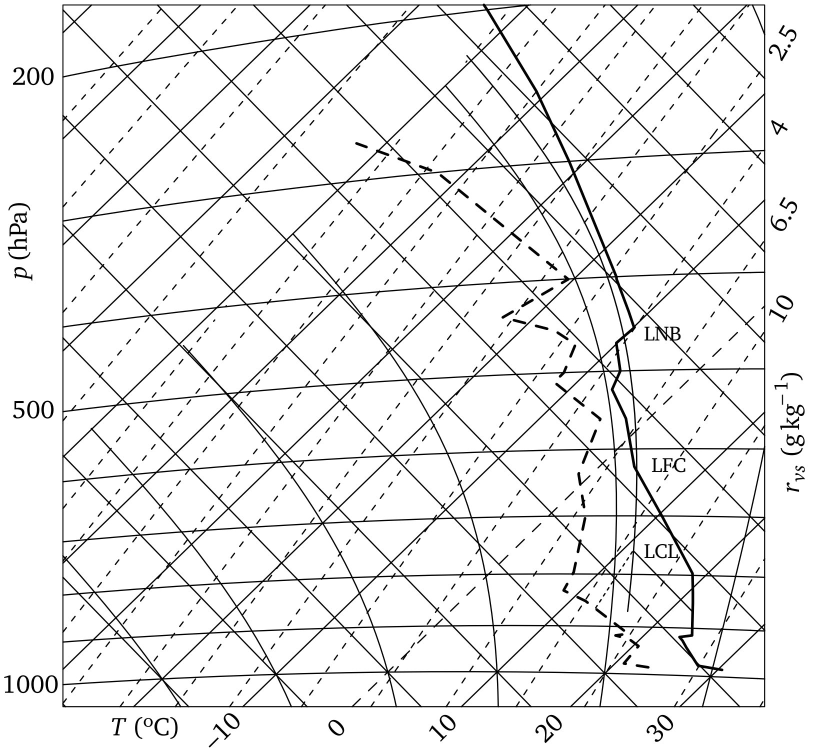


FIGURE 6.5 Sounding data plotted on a tephigram, including Normand's construction for the parcel at 870 hPa.

- Near the surface, below 980 hPa, the atmosphere is unstable ($d\theta/dz < 0$) in a thin layer, probably because of cold air flowing over a warm surface.

- Below 920 hPa the atmosphere is well mixed and near-neutral, close to a dry adiabatic lapse.

- On top of the boundary layer is a stable and shallow inversion with a strong increase in temperature with height.

- The layer between about 900 and 800 hPa is potentially unstable as its equivalent potential temperature decreases with height ($\theta_e(900\,\text{hPa}) \approx 70°\text{C}$ and $\theta_e(800\,\text{hPa}) \approx 60°\text{C}$) although the layer would have to be lifted very high for it to become unstable. There are further potentially unstable layers, such as above 400 hPa.

- In the mid-troposphere (500–600 hPa) the air becomes more humid (smaller dewpoint depressions) with nearly moist adiabatic lapse rates, probably indicative of clouds.

- At 460 hPa there is another strong inversion most likely due to radiative cooling at the top of the cloud.
- The air becomes dry in the upper troposphere.
- The tropopause is higher than 200 hPa.

It is a bit of an art to tease out all the information out of any sounding. Usually more additional information is required to uniquely pin down all the features visible in a profile on a tephigram.

Let us now consider in more detail the air parcel at 870 hPa. From Normand's construction we can see that its lifting condensation level is at about 780 hPa. Above this level the parcel will be saturated and will follow the pseudo-adiabat ($\theta_e = 68°\mathrm{C}$). We can also see that initally the parcel will be colder than its environment, as deduced from the profile, and the parcel is therefore negatively buoyant.

If we lift the parcel further up, to about 630 hPa, we see that it reaches the same temperature as the environment; it will be neutrally buoyant. If we were to lift the parcel further it would be warmer than its environment and would therefore be positively buoyant and would start to rise by itself. Cearly, this parcel is conditionally unstable. The level where the parcel moves from negatively buoyant to positively buoyant is called the *level of free convection*, labelled LFC on the tephigram. The parcel under consideration has a level of free convection of 630 hPa.

The parcel would continue to rise by itself until it became cooler than the environment again. For our particular parcel this happens at about 460 hPa, at the inversion above the cloud. We call this level the *level of neutral buoyancy*, labelled LNB in the tephigram. So our parcel has a level of neutral buoyancy of 460 hPa. If the parcel were made to convect by lifting it, it would stop convecting above 460 hPa.

We have to keep in mind that such analyses mainly give potential idealized scenarios. Whether a parcel would really be lifted in the way described above depends on many factors; it is in fact a highly unlikely scenario.

6.5 CONVECTIVE AVAILABLE POTENTIAL ENERGY

When a parcel is lifted above its level of free convection it experiences a specific (i.e. per unit mass) upward buoyancy force of

$$f = g\frac{T_p - T_e}{T_e}, \tag{6.19}$$

as in Eq. 4.36; we make the ideal gas assumption here.[36] T_p is the virtual temperature of the parcel (the pseudo-adiabat drawn in on Figure 6.5)

---

[36] In these arguments we use virtual temperature, but note that the tephigram only shows temperatures.

and T_e the environmental virtual temperature (the sounding profile in Figure 6.5). Between the LFC and LNB we see that $T_p > T_e$ so that $f > 0$, positive buoyancy. Moving the parcel upward by an amount dz would then release a specific energy of f dz. The height differential dz is by the hydrostatic relationship related to a pressure differential dp by $\mathrm{d}p = -\rho_e g dz$, with ρ_e the density in the environment. We can now calculate the total specific energy released when the parcel moves between the LFC and the LNB. We call this released energy the *convective available potential energy* (CAPE). So we find:

$$\mathrm{CAPE} = \int_{\mathrm{LFC}}^{\mathrm{LNB}} f\,\mathrm{d}z = -\int_{\mathrm{LFC}}^{\mathrm{LNB}} g\frac{T_p - T_e}{T_e}\frac{\mathrm{d}p}{\rho_e g}, \tag{6.20}$$

or

▶ $$\mathrm{CAPE} = -R\int_{\mathrm{LFC}}^{\mathrm{LNB}} (T_p - T_e)\frac{\mathrm{d}p}{p}. \tag{6.21}$$

The convective available potential energy has units of $\mathrm{J\,kg^{-1}}$; it is a *specific* energy.

CAPE corresponds to the area between the environmental profile and the parcel profile: we should not be surprised at this because a tephigram is a proper thermodynamic diagram where area corresponds to energy. The differential of enthalpy,

$$\mathrm{d}h = T\,\mathrm{d}s + v\,\mathrm{d}p \tag{6.22}$$

for an ideal gas ($\mathrm{d}h = c_p\,\mathrm{d}T$ and $pv = RT$) can be rewritten as:

$$T\,\mathrm{d}s = c_p\,\mathrm{d}T - RT\,\mathrm{d}p/p. \tag{6.23}$$

From this equation we see that the area on a tephigram, which is the integral over a closed loop of $T\,\mathrm{d}s$, can be written as the sum of two integrals. The integral of the term $c_p\,\mathrm{d}T$ vanishes as it is a simple differential. The integral of the term $RT\,\mathrm{d}p/p$ can be written as Eq. 6.21.

A rough estimate of the CAPE for our parcel can be found by using the following estimates: $T_p - T_e \approx 1.5$°C, $p \approx 550\,\mathrm{hPa}$, $\Delta p \approx 170\,\mathrm{hPa}$. So we find that CAPE $\approx 130\,\mathrm{J\,kg^{-1}}$.

Before the parcel reaches its LFC, it has to be lifted against its negative buoyancy. This will cost energy. The total energy required to lift the parcel from its starting level to the level of free convection is called the *convective inhibition* (CIN). Following the analogous calculations as above we readily find that:

▶ $$\mathrm{CIN} = -R\int_{p_{\mathrm{start}}}^{\mathrm{LFC}} (T_e - T_p)\frac{\mathrm{d}p}{p}. \tag{6.24}$$

CIN has the same units as CAPE. It is again the area between the environmental and parcel profiles but now below the LFC.

A rough estimate of the CIN for our parcel can be found by using the following estimates: $T_e - T_p \approx 2.5°\mathrm{C}$, $p \approx 750\,\mathrm{hPa}$, $\Delta p \approx 250\,\mathrm{hPa}$. So we find that $\mathrm{CIN} \approx 240\,\mathrm{J\,kg^{-1}}$.

Often CIN is included in the calculation of CAPE. So in this definition CAPE becomes the positive area between the LFC and LNB, minus the CIN. With this definition it is possible to have a negative CAPE. Indeed, our parcel starting from 870 hPa has a greater negative area (about $240\,\mathrm{J\,kg^{-1}}$) than positive area (about $130\,\mathrm{J\,kg^{-1}}$). So it would cost energy to get the parcel up from 870 hPa to its level of neutral buoyancy.

The specific energy released between the LFC and the LNB, as given by Eq. 6.21, can be translated into specific kinetic energy, $\mathcal{W}^2/2$ with $\mathcal{W}$ the parcel speed; this speed is assumed to be upward as the buoyancy force points upward. For our parcel, starting at 870 hPa, this would lead to a speed of about $16\,\mathrm{m\,s^{-1}}$. This is quite a severe overestimate of the parcel speed at the level of neutral buoyancy. The two main causes of the overestimate are (i) the parcel always mixes partially with its environment, which reduces the temperature contrast and therefore its buoyancy, and (ii) the rising air parcel has to move the environmental air out of the way and thus use part of its buoyant energy to impart kinetic energy onto the environment; it acquires an added mass. These caveats were also pointed out in the context of the Brunt–Väisälä frequency, Section 4.4. We should also ask ourselves whether the CAPE can be fully transformed into kinetic energy. In more complete theories, CAPE is linked to the thermodynamic efficiency of convective motion.[37]

So CAPE is not obviously related to parcel kinetic energy at the level of neutral buoyancy. High values of CAPE, or CAPE − CIN, also turn out to be fairly poor predictors of convection. Nonetheless, CAPE is used as a measure of how stable or unstable a particular profile is. Indeed, the onset of thunderstorms is usually associated with a strong increase of CAPE ahead of the storm with values of CAPE easily over $1000\,\mathrm{J\,kg^{-1}}$. Furthermore, high values of CAPE are indicative of vigorous convection *if* the convection occurs. High CAPE is a necessary, but not a sufficient, condition for the onset of convection.

---

[37] See Renno, N. O. & Ingersoll, A. P.(1996) *J. Atmos. Sci.* **53**, 572–585.

7

Cloud drops

In this chapter we introduce the thermodynamics of drops. Droplet nucleation is described by modifications of the Clausius–Clapeyron equation due to surface tension. These modifications also provide the framework for understanding droplet nucleation due to electric charge. We also introduce some of the processes involved in droplet growth.

7.1 HOMOGENEOUS NUCLEATION: THE KELVIN EFFECT

Naively, we would expect vapour to start to condense and thus form cloud droplets if the relative humidity is above 100%. However, we have to be careful: the saturation vapour pressure as calculated from the Clausius–Clapeyron equation is only valid for *flat* water surfaces. Cloud droplets have strongly curved surfaces and as a consequence the surface tension needs to be incorporated in the free energy budgets leading to the Clausius–Clapeyron equation. The surface tension effect is named after William Thomson, Lord Kelvin, who made many seminal contributions to thermodynamics and physics in general.

Any liquid–gas interface experiences a surface tension: the surface tries to contract like an elastic membrane. Therefore, increasing the surface area of such an interface requires work energy. The amount of work required dW is proportional to the increment in surface area dA with the proportionality constant called *surface tension,* γ,

$$\blacktriangleright \qquad dW = \gamma\,\mathrm{d}A. \tag{7.1}$$

The unit of surface tension is $\mathrm{J\,m^{-2}}$ or, equivalently, $\mathrm{N\,m^{-1}}$. It is the surface tension that makes droplets (and bubbles) spherical. The surface tension tries to minimize the surface area of a drop for the volume of the drop given. The shape that has minimum surface area for a given volume is a sphere.

A typical value for the surface tension of a water–air interface is

$$\gamma = 75 \times 10^{-3}\,\mathrm{N\,m^{-1}}. \tag{7.2}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum
ISBN: 978-0-470-74515-1

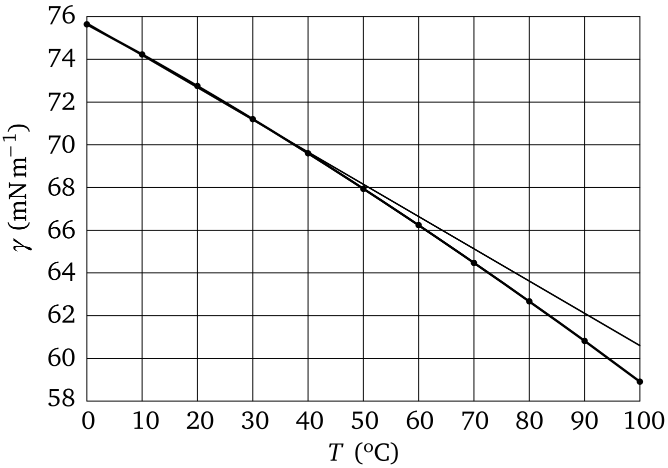


Figure 7.1 Points: measured surface tension γ for water as a function of temperature. Thin line: surface tension γ from linear fit of Eq. 7.3. Data from Vargaftik, N.B. et al. (1983) *J. Phys. Chem. Ref. Data* **12** 817–820.

Surface tension reduces with increasing temperature but the reduction is small. The linear approximation

$$\gamma = (75.7 - 0.151\,T(^\circ\mathrm{C})) \times 10^{-3}\,\mathrm{N\,m^{-1}} \qquad (7.3)$$

is accurate to within one part per thousand in the range of $0^\circ\mathrm{C} < T < 40^\circ\mathrm{C}$, see Figure 7.1. For many practical purposes we take γ to be constant.

The origin of surface tension is the relatively strong attractive force between molecules in a liquid; it takes energy to separate the molecules of a liquid. In Figure 7.2 it can be seen that molecules at the surface of the liquid have fewer neighbours than molecules in the bulk of the liquid. In order to reside at the surface the molecule must separate from some of its surrounding

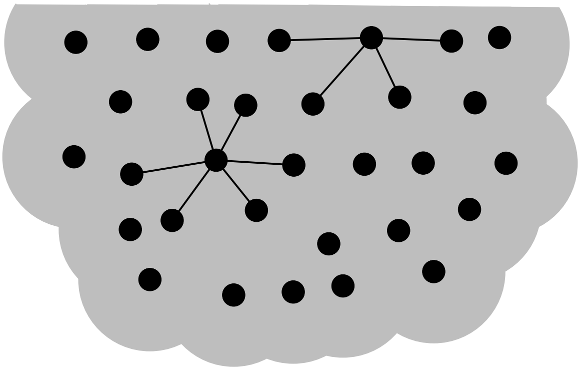


Figure 7.2 The molecules in the bulk of the liquid are attracted by more neighbours than molecules at the surface and therefore have a lower potential energy. The total potential energy of the liquid is minimized when the surface area is minimum for a given bulk volume. The increase of potential energy with surface area manifests itself as surface tension.

molecules. This separation costs energy and the total energy involved will be proportional to the number of molecules that reside at the surface, which is again proportional to the surface area of the liquid. This explains that any increase of surface area will correspond to a proportional increase in energy, as in Eq. 7.1.

On increasing the temperature of the liquid it becomes easier to sever the link between the liquid molecules because the kinetic energy of the molecules helps to overcome the energy barrier between the bound and separated states. This means that at higher temperatures the excess energy to enable a molecule to reside at the surface becomes lower. As a result, the surface tension decreases with increasing temperature, as is indeed seen in Eq. 7.3.

We will now repeat the derivation of the Clausius–Clapeyron equation for a liquid that has a spherical surface. Any evaporation from such a surface would decrease the surface area and thus release energy. The internal energy budget for a liquid drop is therefore modified to

$$\mathrm{d}U_l = T\,\mathrm{d}S_l - p\,\mathrm{d}V_l + \gamma\,\mathrm{d}A. \tag{7.4}$$

Transforming this to a Gibbs function budget we find

$$\mathrm{d}G_l = -S_l\,\mathrm{d}T + V_l\,\mathrm{d}p + \gamma\,\mathrm{d}A. \tag{7.5}$$

This modified Gibbs function budget forms the basis of a modified Clausius–Clapeyron equation. A flat surface is of course also under surface tension, but any evaporation from a flat surface does not change its surface area, so does not contribute to the Gibbs function budget. This is why in the derivation of the Clausius–Clapeyron equation in Chapter 5 the surface tension did not play a role.

For a spherical liquid drop, the area change δA is related to a mass transfer δM from the liquid to the vapour phase (so a positive δM corresponds to a reduction in droplet volume V) as

$$\delta A = \frac{\mathrm{d}A}{\mathrm{d}V}\,\delta V = -\frac{\mathrm{d}A}{\mathrm{d}V} v_l\,\delta M = -\frac{2v_l}{r}\,\delta M, \tag{7.6}$$

with v_l the specific volume of the liquid phase and r the radius of the droplet. Here we take it that for a spherical droplet $V = (4/3)\pi r^3$ and $A = 4\pi r^2$ and write $\mathrm{d}A/\mathrm{d}V = (\mathrm{d}A/\mathrm{d}r)/(\mathrm{d}V/\mathrm{d}r)$. Following on from Eq. 5.3, the change of the Gibbs function δG at fixed temperature and pressure now becomes

$$\begin{aligned} \delta G = {} & M_l\,(-s_l\,\delta T + v_l\,\delta e) + M_v\,(-s_v\,\delta T + v_v\,\delta e) \\ & + \; (g_v - g_l - 2\gamma v_l/r)\,\delta M. \end{aligned} \tag{7.7}$$

Because by construction $\delta G = 0$, $\delta T = 0$, and $\delta e = 0$, we find that

$$g_v = g_l + 2\gamma v_l/r. \tag{7.8}$$

This equation can be used to calculate the change in the vapour pressure for a given temperature due to surface tension. At a fixed temperature we find

$$\left(\frac{\partial g_v}{\partial p}\right)_T \delta e = \left(\frac{\partial g_l}{\partial p}\right)_T \delta e + 2\gamma v_l \, \delta(1/r). \tag{7.9}$$

Using $(\partial g/\partial p)_T = v$ and assuming $v_l \ll v_v$ we find

$$v_v \, \delta e = 2\gamma v_l \, \delta(1/r). \tag{7.10}$$

We can next use the ideal gas law to write $v_v = R_v T/e$ and arrive at

$$R_v T \, \delta e/e = 2\gamma v_l \, \delta(1/r) \tag{7.11}$$

At fixed temperature T and assuming v_l, γ and R_v are not functions of the radius, this can be integrated between the case of a flat surface with radius $r \to \infty$ and a curved surface of radius r. The vapour pressure e then varies between its flat surface value $e_s(\infty)$, as determined by the Clausius–Clapeyron equation, and its curved surface value $e_s(r)$, which we are trying to calculate. The integral becomes

$$R_v T \ln \frac{e_s(r)}{e_s(\infty)} = \frac{2\gamma v_l}{r}. \tag{7.12}$$

This is written in terms of a *saturation ratio* S_K as

▶ $$S_K = \frac{e_s(r)}{e_s(\infty)} = \exp\left(\frac{2\gamma v_l}{R_v T r}\right). \tag{7.13}$$

The curvature effect increases the saturated vapour pressure compared to its flat surface value, that is, $S_K > 1$. This effect is called the *Thomson effect* or the *Kelvin effect*. A physical picture for the Kelvin effect is that for a drop under surface tension it is easier to evaporate water because the associated reduction of droplet radius releases surface energy. The effective energy barrier for evaporation has therefore reduced. Figure 7.3 shows a plot of the saturation ratio as a function of droplet radius.

The relative humidity and the saturation ratio can be compared. For example, if

$$\mathrm{RH} > S_K \tag{7.14}$$

the drop is in a supersaturated environment and vapour will condense on it; the drop will grow. A flat surface has a saturation ratio of 1 and therefore the vapour will condense on it when the relative humidity is above 100%. Curved drops have higher saturation ratios and therefore need a higher relative humidity (above 100%) to grow.

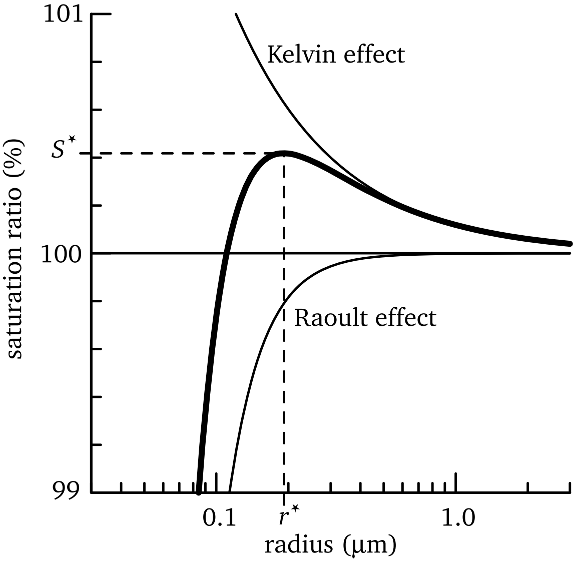


FIGURE 7.3 Saturation ratio as a function of droplet radius ('Köhler curve', thick line) for a droplet at 0°C and with 10^{-16} g of NaCl solute. The Köhler curve is the product of the Kelvin effect and the Raoult effect (thin lines).

The Kelvin effect is small for large droplet radii. It becomes very large for radii r smaller than the *Kelvin radius* r_K with

$$r_K = \frac{2\gamma v_l}{R_v T}. \tag{7.15}$$

For water drops at typical atmospheric temperatures, $r_K \approx 1.2\,\text{nm}$, about 10 times the size of a water molecule. A drop of water of size r_K would therefore contain several thousand water molecules; this is getting close to the regime where equilibrium thermodynamics becomes less applicable and we need to use molecular dynamics calculations. However, in the atmosphere, typical saturation ratios are only slightly larger than 1. For such saturation ratios we can write

$$S_K \approx 1 + \frac{2\gamma v_l}{R_v T r} = 1 + \frac{r_K}{r}. \tag{7.16}$$

This can be rearranged to

$$\frac{r}{r_K} = \frac{1}{S_K - 1}. \tag{7.17}$$

So for a droplet with a radius 100 times larger than the Kelvin radius, the Kelvin effect corresponds to a supersaturation of 1% ($S_K = 1.01$). At such a droplet radius ($\approx 0.12\,\mu\text{m}$) the drop contains enough molecules to be accurately described by macroscopic physics.

We can rewrite the right-hand side of Eq. 7.9 as

$$\left(\frac{\partial g_l}{\partial p}\right)_T \delta e + 2\gamma v_l\, \delta(1/r) = \left(\frac{\partial g_l}{\partial p}\right)_T \delta(e + 2\gamma/r). \tag{7.18}$$

So the surface tension term in Eq. 7.8 can be incorporated in the Gibbs function for the drop as

$$g_l(e, T) + 2\gamma v_l/r = g_l(e', T), \tag{7.19}$$

with

$$e' = e + 2\gamma/r. \tag{7.20}$$

In other words, the surface tension increases the pressure inside the drop by $2\gamma/r$, something that should come as no surprise. So according to Eq. 7.8, the specific Gibbs functions for the two phases have to be the same as long as the pressure inside the drop is augmented by the *capillary pressure*

$$\blacktriangleright \qquad p_\gamma = 2\gamma/r \tag{7.21}$$

compared to the external vapour pressure.

With the interpretation of pressure as a volumetric energy density, see Section 2.2, the Kelvin effect can be interpreted as a Boltzmann factor corresponding to the excess energy per droplet molecule due to the surface tension,

$$\exp\left(\frac{2\gamma v_l}{R_v T r}\right) = \exp\left(-\frac{\Delta E}{k_B T}\right), \tag{7.22}$$

with the excess energy per molecule written in terms of capillary pressure,

$$\Delta E = -p_\gamma V/N, \tag{7.23}$$

and with N the number of molecules in the drop and V its volume. So the capillary pressure provides a reduced energy barrier between the liquid and the vapour. This interpretation in terms of the Boltzmann factor emphasizes the statistical nature of the vapour–liquid equilibrium at a fixed temperature.

So can water molecules clump together to form water droplets when the relative humidity is large enough? Such a process is called *homogeneous nucleation.* Initial clusters of water molecules may form by chance but their radius is quite a bit smaller than 1 nm. This means that the Kelvin effect is important and in practice would correspond to saturation ratios $S_K = 4$ or larger. For growth to occur we would need a relative humity in excess of 400% or so, see Eq. 7.14. This never happens in the atmosphere. Homogeneous nucleation cannot be the source of cloud droplets in the atmosphere. Somehow, we need to counteract the Kelvin effect.

7.2 HETEROGENEOUS NUCLEATION: THE RAOULT EFFECT

The atmosphere contains particles that can serve as condensation nuclei for drops. Many of these particles can partially dissolve in the water they attract. It turns out that the dissolved nucleus in the water reduces the saturation vapour pressure enough to counteract the Kelvin effect.

Suppose the water drop contains number concentrations c_i of solute i; that is, c_i of the molecules in the droplet are made up of the solute i. *Raoult's law* now states that the saturated vapour pressure is the sum of the individual saturated vapour pressures of the constituents in the liquid weighted with their number concentrations. A derivation of Raoult's law requires introduction of so-called *chemical potentials*, the generalized forces that correspond to changes in composition of a substance. This is set out in detail in Chapter 8.

An informal justification is as follows. Suppose c is the number concentration of all solute molecules. Then $1 - c$ is the number concentration of water molecules. That means that according to Raoult's law the saturated vapour pressure $e_s(c)$ for a water with solute concentration c is

$$e_s(c) = (1 - c)\, e_s(0), \tag{7.24}$$

with $e_s(0)$ the saturated vapour pressure for pure water (with the Kelvin effect included for spherical droplets). From a microscopic point of view this equation makes sense. According to the equipartition theorem, the kinetic energy will be equally distributed amongst all the molecules, because they are at the same temperature. This means that at some temperature the water molecules have the same chance of escaping the solution. It is assumed that the solute does not change the energy barrier for escaping the solution; this is, implicitly, one of the assumptions in deriving Raoult's law: the assumption of an *ideal solution*. However, the number of water molecules per unit number of molecules has reduced by a factor $1 - c$, thus reducing the evaporation rate, and therefore the saturation vapour pressure of the water by the same factor.

We can rewrite Eq. 7.24 in a more explicit form. Write the total number of molecules N_t as $N_t = N_l + N_s$ with N_l the number of molecules of liquid water, the solvent, and N_s the number of solute molecules. We then find that

$$1 - c = \frac{N_t - N_s}{N_t} = \frac{N_l}{N_l + N_s} = \frac{1}{1 + N_s/N_l}. \tag{7.25}$$

The fraction N_s/N_l can be expressed as

$$\frac{N_s}{N_l} = i \frac{M_s}{\rho_l V} \frac{\mu_l}{\mu_s}, \tag{7.26}$$

with M_s the total mass of the solute in the drop, V the volume of the drop, being nearly equal to the volume of all the water molecules, and μ_l and μ_s the molar weights of the water and solute molecules, respectively. The factor i is the *Van't Hoff factor,* which takes into account the dissociation of the solute molecules. For example, in solution common salt NaCl will dissociate into two ions, Na^+ and Cl^-. The Van 't Hoff factor i is typically equal to 2 for many of the relevant solutes.

This expression can be substituted in Eq. 7.24 to find for the saturation ratio due to the Raoult effect

$$S_R = \frac{e_s(M_s)}{e_s(0)} = \frac{1}{1 + i\frac{3M_s}{\rho_l 4\pi r^3}\frac{\mu_l}{\mu_s}}, \tag{7.27}$$

where we have now written $V = (4/3)\pi r^3$ for spherical drops of radius r. Figure 7.3 shows a plot of the saturation ratio as a function of droplet radius.

7.3 KÖHLER THEORY

Combining the Kelvin and Raoult effects, we get an expression for the ratio of the saturated vapour pressure over a flat surface of pure water, as derived in Section 5.1, to the saturated vapour pressure of a solution in a spherical droplet of radius r at temperature T. This saturation ratio is:

▶

$$S = S_K\, S_R = \frac{\exp\left(a/rT\right)}{1 + b/r^3} \tag{7.28}$$

with, for water drops in air,

$$a = \frac{2\gamma v_l}{R_v} \approx 3.3 \times 10^{-7}\,\mathrm{K\,m}, \tag{7.29a}$$

$$b = \frac{3v_l\mu_l}{4\pi}\frac{iM_s}{\mu_s} \approx 4.3 \times 10^{-6}\,(iM_s/\mu_s)\,\mathrm{m}^3. \tag{7.29b}$$

A plot of S as a function of radius is called a *Köhler curve.* Figure 7.3 shows an example of a Köhler curve highlighting the Kelvin and Raoult effects. The Köhler curve defines at what relative humidity the droplet would be in equilibrium with the environment.

The maximum of the Köhler curve plays an important role for droplet growth. The radius at the maximum is called the *activation radius,* denoted $r^\star$, and the value of the supersaturation at the activation radius is called the *activation saturation ratio* or *critical saturation ratio,* denoted $S^\star$.

The values of $r^\star$ and $S^\star$ follow from setting $\mathrm{d}S/\mathrm{d}r = 0$ in the Köhler equation. An exact calculation requires finding the root of a cubic polynomial. However, from the above graph it can be seen that both the Kelvin and Raoult effects are typically quite small around the activation radius (of order one percent). This suggests that we can use a Taylor expansion for both effects and find an approximate result. To do this, define a non-dimensional radius x as

$$r = x\sqrt{\frac{bT}{a}}, \tag{7.30}$$

and a non-dimensional parameter ϵ as

$$\epsilon = \sqrt{\frac{(a/T)^3}{b}}. \tag{7.31}$$

We can write the Köhler curve equation, Eq. 7.28, as

$$S(x) = \frac{\exp(\epsilon/x)}{1+\epsilon/x^3}. \tag{7.32}$$

For typical atmospheric values of the parameters a, b, and T it can be verified that ϵ is quite small. Now assuming that x is not much bigger or much smaller than 1 – to be verified in hindsight – we can use a Taylor expansion for S in small ϵ,

$$S(x) = 1 + \frac{\epsilon}{x} - \frac{\epsilon}{x^3} + \mathcal{O}(\epsilon^2). \tag{7.33}$$

We thus have

$$\frac{\mathrm{d}S}{\mathrm{d}x} = -\frac{\epsilon}{x^2} + 3\frac{\epsilon}{x^4} + \mathcal{O}(\epsilon^2). \tag{7.34}$$

Setting this to zero gives the non-dimensional activation radius $x^\star$,

$$x^\star = \sqrt{3} + \mathcal{O}(\epsilon), \tag{7.35}$$

which is indeed not much larger than 1, validating the Taylor expansion.[38] The corresponding critical saturation ratio is

$$S^\star = 1 + \frac{2\epsilon}{3\sqrt{3}}. \tag{7.36}$$

[38] A full solution of $\mathrm{d}S/\mathrm{d}x = 0$ without approximation leads to

$$x^\star = 2\cos\left[\frac{1}{3}\cos^{-1}\left(-\frac{\epsilon}{2}\right)\right].$$

The difference between this exact expression for $x^\star$ and the approximated expression is small; it can be shown that $x^\star = \sqrt{3} - \epsilon/6 + \mathcal{O}(\epsilon^2)$.

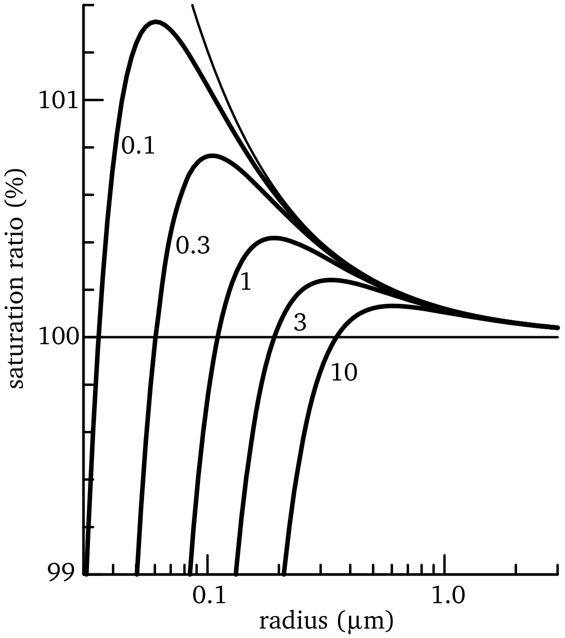


FIGURE 7.4 Köhler curves for different amounts of NaCl solute (in 10^{-16} g). In the limit of vanishing solute amount, the Köhler curve approaches that of the Kelvin effect (thin line).

So the non-dimensional parameter ϵ is a measure of the critical supersaturation.

Substituting the definitions of x and ϵ, we find the activation radius $r^\star$

$$\blacktriangleright \qquad r^\star = \sqrt{\frac{3bT}{a}}, \tag{7.37}$$

and the critical saturation ratio $S^\star$

$$\blacktriangleright \qquad S^\star = 1 + \sqrt{\frac{4(a/T)^3}{27b}}. \tag{7.38}$$

The *critical supersaturation*, defined as $S^\star - 1$, decreases when b increases, which is to say when the amount of solute increases in the droplet. At the same time the activation radius $r^\star$ increases. Figure 7.4 shows various Köhler curves for different amounts of solute, illustrating the dependencies of $r^\star$ and $S^\star$ on the solute amount. Note that the activation radius is always very small. For any drops larger than about 1 μm both the Kelvin and Raoult effects can be ignored.

The Köhler curve contains a lot of information about initial cloud droplet growth. Consider Figure 7.5, where the Köhler curve and the level of the critical saturation ratio divides the figure into three different areas, labelled A, H, and N.

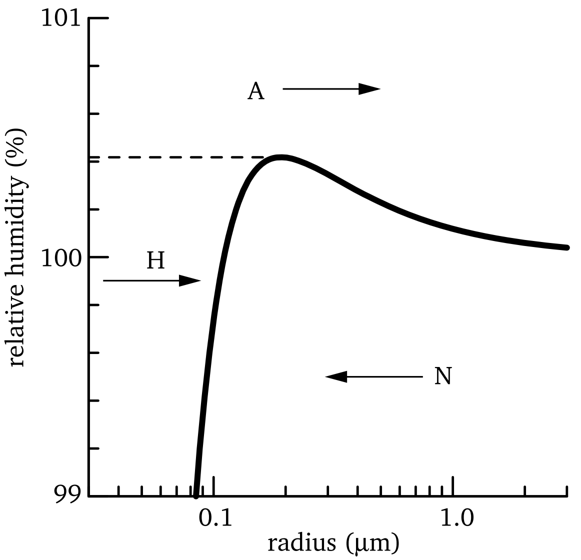


FIGURE 7.5 Three distinct regions, A, H, and N, in the (r, RH) space, as defined by the Köhler curve.

Now suppose the nucleus has attracted some water and forms a small droplet of a particular radius r. Further suppose this droplet is moved into air of a particular relative humidity RH. This initial droplet then finds itself at the point (r, RH) in the figure and will be in one of the labelled areas.

Suppose the initial droplet was somewhere in area H. In this area the droplet is in an environment with a higher saturation ratio than the equilibrium saturation ratio for the droplet, as given by the Köhler curve. This means that the droplet is in a supersaturated environment and water vapour will start to condense onto the drop. The drop will grow; the point (r, RH) in the figure will move to the right, as indicated by the arrow. This growth will continue until the point hits the Köhler curve, where RH $= S$; here the droplet is in equilibrium with its environment and will remain steady. Such droplets are called *haze*.

Suppose the initial droplet was somewhere in area N. In this area the droplet is in an environment with lower relative humidity than the equilibrium saturation ratio for the droplet. The droplet will therefore start to evaporate and reduce in radius. The point (r, RH) for the droplet will move to the left in the figure until it hits the Köhler curve, where the drop is again in equilibrium, $RH = S$. The droplet has again become haze. To find initial droplets in area N requires the existence of relatively large drops in subsaturated air, which is possible when turbulence is present or when raindrops fall in subsaturated air.

Haze droplets on the boundary between areas H and N, given by the Köhler curve below the activation radius, are stable: if for some reason a haze droplet were to move away from the boundary the tendency would be for the droplet to be pushed back to the boundary. Haze is a stable state for droplets. For larger condensation nuclei the steady radius of haze droplets can be up to several hundreds of nanometres, and therefore become visible.

Suppose the initial droplet was somewhere in area A. When considering cloud droplet growth from initial condensation nuclei we start at very small radii. So typically, droplets in area A have to be in an environment with a relative humidity higher than the critical saturation ratio $S^\star$. The droplet is in a supersaturated environment and will start to grow. But now the growing droplet cannot hit the Köhler curve at any point; it will continue to grow indefinitely and will eventually become a cloud drop. Droplets in region A are said to be *activated*.

For droplets to become activated and grow into cloud drops the environmental supersaturation has to be larger than the critical supersaturation. Typical critical supersaturations are less than a couple of percent. This is the reason why in the atmosphere the relative humidity will never exceed 100% by very much. With any temporary increase in relative humidity, haze droplets would activate and form clouds. These growing cloud droplets would reduce the vapour content of the air until a relative humidity of 100% was achieved and droplets could no longer use up water vapour.

So at relative humidities below 100% any condensation nuclei will form stable haze. Suppose that this air is forced to lift by some external process; this will expand and cool down the air, thus increasing the relative humidity. Because of the stability of the haze part of the Köhler curve, the haze droplets move up on the Köhler curve and get a larger radius.[39] This process

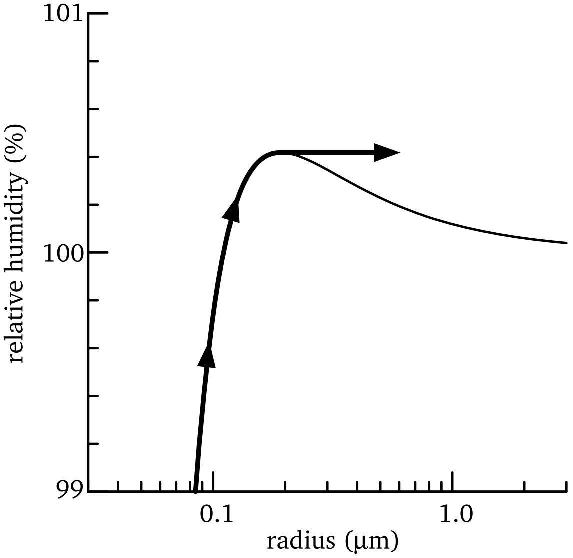


FIGURE 7.6 Growth of a haze drop in an updraft.

---

[39] The Köhler curve itself is a weak function of temperature. However, this does not change the main argument here.

continues until the air reaches critical supersaturation; the haze droplets will then have grown to the activation radius. Any further adiabatic cooling will activate the droplet and make it grow into cloud drops, and thus remove water vapour so as to keep the relative humidity down. Figure 7.6 shows this process on the Köhler curve.

7.4 CHARGE-ENHANCED NUCLEATION

Droplets are often charged. There can be many sources of this charge, typically charge separation by the differential motion of drops of different sizes, or the always-present electric current between ionosphere and the Earth's surface, which can charge up individual droplets at cloud boundaries. A net charge corresponds to an electrostatic energy on the drop which modifies the Gibbs free energy budget and therefore the saturated vapour pressure around the drop.

A spherically symmetric charge distribution has an electrostatic potential V_e at radius r of

$$V_e = \frac{Q}{4\pi\epsilon r}, \tag{7.39}$$

with Q the charge and ϵ the electric permittivity; for all practical purposes we can use ϵ_0, the permittivity of the vacuum.[40] Increasing the charge by an amount $\mathrm{d}Q$ requires work $\mathrm{d}W$ against the electrostatic potential of $\mathrm{d}W = V_e\,\mathrm{d}Q$ (see also Section 2.1). So to charge up a sphere from no charge to a total charge Q requires a total electrostatic energy W of

$$W = \frac{Q^2}{8\pi\epsilon r}. \tag{7.40}$$

For a given charge, a change in the radius of the drop would change the electrostatic energy. That means that on evaporating a mass δM from a charged drop, its electrostatic energy will change. Analogous to Eq. 7.7, the Gibbs function budget is modified to

$$\delta G = (g_v - g_l)\,\delta M - \frac{Q^2}{8\pi\epsilon r^2}\,\delta r. \tag{7.41}$$

We have ignored any effects of solute or surface tension here in order to isolate the specific effects of the charge and we have omitted the terms proportional to δe and δT as these are zero by construction. As before, δr is related to δM

---

[40] The relevant inverse permittivity is $1/\epsilon = 1/\epsilon_{\text{air}} - 1/\epsilon_{\text{water}}$ but the permittivity of water is about 80 times larger than that of air, which in turn is very similar to that of the vacuum, ϵ_0.

by

$$\delta r = \frac{\delta V}{A} = -\frac{v_l}{4\pi r^2}\,\delta M, \tag{7.42}$$

with V the volume of the sphere, A its area and v_l the specific volume of the liquid phase (remember that positive δM corresponds to a reduction in droplet volume). For the variations in the total Gibbs function to vanish, we therefore have

$$g_v(e,T) = g_l(e,T) - \frac{Q^2 v_l}{32\pi^2 \epsilon r^4}. \tag{7.43}$$

At infinite radius the charge effect vanishes and the vapour pressure is the same as the flat surface value from the Clausius–Clapeyron equation. We can now perform the same integration over radius as was performed following Eq. 7.8. The result is a new charge-induced saturation ratio S_Q,

$$\blacktriangleright \qquad S_R = \frac{e_s(Q)}{e_s(0)} = \exp\left(-\frac{Q^2 v_l}{32\pi^2 \epsilon R_v T r^4}\right), \tag{7.44}$$

where have returned to the usual notation of e_s for saturated vapour pressure and with $e_s(0)$ the saturated vapour pressure for an uncharged drop. We see that the charge of the drop decreases the vapour pressure around the drop by an exponential factor, $S_R < 1$. This effect is called the *Rayleigh effect*. The physical picture is that on a charged drop it is more difficult to evaporate water because the associated reduction of droplet radius requires extra electrostatic energy.

Analogous to Eq. 7.19, the electrostatic term can be absorbed in the pressure dependency of the Gibbs function,

$$g_l(e,T) - \frac{Q^2 v_l}{32\pi^2 \epsilon r^4} = g_l(e',T), \tag{7.45}$$

with

$$e' = e - \frac{Q^2}{32\pi^2 \epsilon r^4}. \tag{7.46}$$

In other words, the effective pressure inside the drop is decreased by the charge. This decrease of pressure inside a drop should, again, come as no surprise: the repulsive electrostatic charges try to expand the drop. The pressure drop due to the charge is p_Q, with

$$p_Q = \frac{Q^2}{32\pi^2 \epsilon r^4}. \tag{7.47}$$

The charge effect can be interpreted as a Boltzmann factor analogous to Eq. 7.22,

$$\exp\left(-\frac{Q^2 v_l}{32\pi^2 \epsilon R_v T r^4}\right) = \exp\left(-\frac{\Delta E}{k_B T}\right), \tag{7.48}$$

with the excess energy per molecule written in terms of the pressure drop p_Q due to the charge,

$$\Delta E = \frac{p_Q V}{N}, \tag{7.49}$$

with N the number of molecules in the drop and V its volume.

The charge helps to form droplets by counteracting the Kelvin effect. It is of interest to see when the two effects exactly compensate. Using Eqs. 7.13 and 7.44 we find compensation when $S_K S_R = 1$ or equivalently when

$$p_\gamma = p_Q. \tag{7.50}$$

For a given charge Q this compensation will occur at a critical radius r_R, the so-called *Rayleigh radius*, with

► $$r_R^3 = \frac{Q^2}{64\pi^2 \epsilon \gamma}. \tag{7.51}$$

Rayleigh showed that for droplets with a radius smaller than r_R the repulsive electrostatic force becomes so strong that the droplets disintegrate explosively;[41] thermodynamically this can be interpreted as the destabilizing charge pressure overcoming the stabilizing capillary pressure.

With the definition of the Rayleigh radius r_R and the Kelvin radius r_K we can rewrite the combined Rayleigh and Kelvin effect to the saturation ratio S as

$$S = S_K S_R = \exp\left(\frac{r_K}{r}\left(1 - \frac{r_R^3}{r^3}\right)\right). \tag{7.52}$$

From this equation it becomes clear that the charge effect is only important when the droplet radius is close to the Rayleigh radius. Only for droplets smaller than the Rayleigh radius, the charge effect dominates the Kelvin effect and saturation can occur at relative humidities below 100%.

So what is a typical value for the Rayleigh radius? For a unit charge e (an electron has charge $-e$), we can put in standard values for a water droplet

---

[41] For a derivation of the onset of Rayleigh explosions, see Peters, J. M. H. (1980) *Eur. J. Phys.* **1**, 143–146.

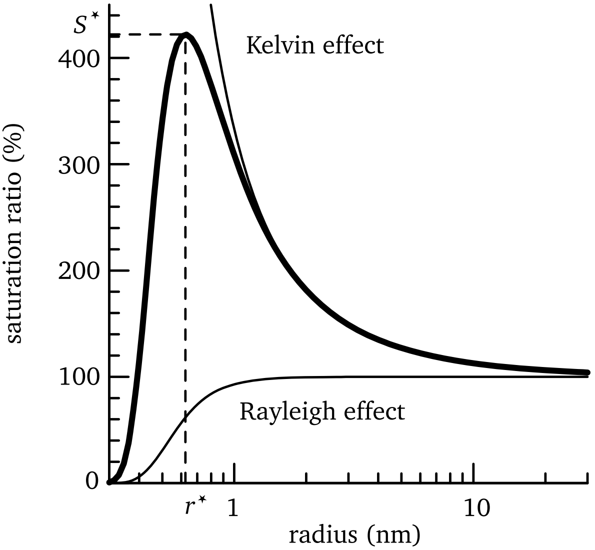


FIGURE 7.7 Saturation ratio as a function of charged droplet radius for a clean droplet at 0°C with one unit charge.

to find $r_R \approx 0.4\,\text{nm}$. Very tiny indeed; in fact this is so small that quantum effects become important.

Like the Köhler curve, the saturation ratio for a charged drop has a maximum. Figure 7.7 shows the saturation ratio for a droplet of pure water with a unit charge as a function of radius. It is straightforward to show that this maximum occurs at a critical radius of $r^\star = 4^{1/3} r_R$, about 0.6 nm for a unit charge droplet. Substituting this radius back in the equation for the saturation ratio, we find a critical saturation ratio of about $S^\star = 4$.

For droplets with such high saturation ratios to grow we need relative humidities in excess of RH = 400%;[42] this does not occur in nature. We are forced to conclude that in the atmosphere single charge ions cannot serve as condensation nuclei. This is of importance in the present discussion on influence of cosmic rays on climate. Cosmic rays generally produce single charge ions in the atmosphere and it has been argued that these may form cloud condensation nuclei. It is clear from the argument above that this is impossible. In order for cosmic rays to influence cloud formation other processes such as ion clustering or charge accumulation have to occur.

Supersaturations of 400% are fairly easy to achieve in the laboratory. By expanding a vessel which has water and its vapour in it we can achieve very high supersaturations: on adiabatic expansion the pressure as well as the

---

[42] There is an asymmetry between positively charged ions and negatively charged ions: positively charged ions require a relative humidity of about 600%; this is presumably due to the geometry of water molecules. See Wilson, C. T. R. (1899) *Proc. Roy. Soc. London* **65**, 289–290.

temperature will drop. By the Clausius–Clapeyron equation the saturation vapour pressure will also drop and it is found to drop more than the pressure itself. This means that vapour that was originally saturated will become supersaturated. In Problem 7.2 this situation is examined and it is shown that an expansion by a factor of about 1.3 is enough to reach a supersaturation of 400%.

This supersaturation in the laboratory is the basis of the Wilson cloud chamber. Wilson used a closed chamber with saturated vapour and by expansion made the vapour supersaturated. Any condensation nuclei would rapidly saturate and grow into drops. Wilson allowed these drops to settle out. This was repeated until no condensation nuclei were present anymore. Wilson then found that there were still cloud traces in his chamber. Further experimentation with X-ray sources made him realize that these traces were due to charged particles in the cloud chamber, the residual traces being due to natural radioactivity and cosmic rays. The cloud chamber is used as a detector in particle physics. Wilson received the Nobel prize for physics in 1927 ‘for his method of making the paths of electrically charged particles visible by condensation of vapour.’

7.5 DROPLET GROWTH

If a cloud droplet is activated it is out of equilibrium: the environment is supersaturated with respect to the droplet and the droplet will grow by condensation of vapour onto the droplet. However, this process is limited by the speed with which the condensed vapour is replenished by new vapour from the environment. If there is no such replenishment, the immediate vicinity of the drop will run out of vapour and will become subsaturated. Consequently, the drop would stop growing.

The process that replenishes the vapour is a diffusive flux of water vapour. To a good approximation, this diffusive flux is proportional to the gradient in water vapour density (*Fick’s law*). So the vapour flux $\boldsymbol{F}_v$ is

$$\boldsymbol{F}_v = -D\,\nabla\rho_v, \tag{7.53}$$

where ρ_v is the mass density of the vapour and the constant of proportionality D is called the *diffusion coefficient*. The direction of the flux is opposite to the gradient of the vapour density: mass flows from high densities to low densities. If a drop is activated, vapour molecules will condense on the drop thus reducing the vapour density in the immediate vicinity of the drop. This sets up a gradient in vapour density between the immediate vicinity of the drop and the far field. This will lead to a diffusive flux of water vapour towards the drop.

The value of the diffusion coefficient varies with the species being diffused (approximately with the square root of the mass) and it increases with temperature. For diffusion of water vapour at 5°C we have

$$D = 22 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-1}, \tag{7.54}$$

with an increase of about $0.15 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-1}$ per 1°C temperature increase.

The origin of the vapour flux is the random motion of the vapour molecules. If there is a density gradient there will be fewer molecules moving from the low density to the high density region than the other way around simply because the low density region has fewer molecules to move to the high density region than the other way around. This discrepancy results in a net motion of molecules from the high density area to the low density area. It is also clear that the discrepancy is proportional to the difference in molecule densities between the high and the low density regions.

Assuming the situation is spherically symmetric, all variables will be functions of the radial coordinate r and of time t. We now make the assumption that away from the drop the density field is constant in time (but not necessarily constant in space). The vapour flux $\boldsymbol{F}_v$ is according to Fick's law

$$\boldsymbol{F}_v = -D \frac{\mathrm{d}\rho_v}{\mathrm{d}r} \hat{\boldsymbol{r}}, \tag{7.55}$$

with $\hat{\boldsymbol{r}}$ the unit vector in the radial direction. The total mass flux F_m into a shell of radius r then is

$$F_m = -4\pi r^2 \boldsymbol{F}_v \cdot \hat{\boldsymbol{r}} = 4\pi D r^2 \frac{\mathrm{d}\rho_v}{\mathrm{d}r}. \tag{7.56}$$

Note the sign here: if the vapour density increases away from the drop there will be a positive mass flux towards the drop.

F_m is the total vapour mass entering any spherical shell, so it is also the total vapour mass entering the droplet. So the mass growth $\mathrm{d}M_d/\mathrm{d}t$ of the droplet is

$$\frac{\mathrm{d}M_d}{\mathrm{d}t} = F_m. \tag{7.57}$$

The drop grows by condensing the vapour onto its surface. This condensation releases latent heat, which will heat up the drop. In a steady state situation this heat will be conducted away by a total heat flux F_q equal to

$$F_q = -L \frac{\mathrm{d}M_d}{\mathrm{d}t} = -L F_m, \tag{7.58}$$

with L the latent heat of condensation and F_q the total inward heat flux through a spherical shell. Note that as the drop grows (F_m positive) the heat flux is directed outward (F_q negative). According to Fourier's law, the heat

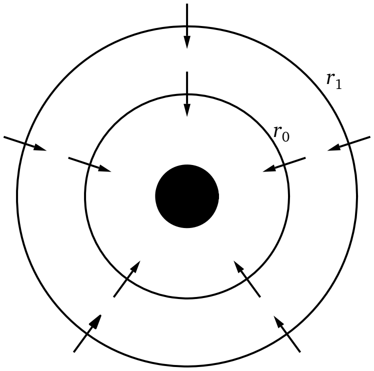


FIGURE 7.8 Flux of vapour through spherical shells towards a drop. In a steady state the flux through all spherical shells has to be equal.

flux is proportional to the gradient of the temperature. Therefore total inward heat flux is, analogous to Eq. 7.56,

$$F_q = 4\pi K r^2 \frac{\mathrm{d}T}{\mathrm{d}r}, \tag{7.59}$$

with K the *heat conduction coefficient.* The thermal conductivity of air increases with increasing temperature. A typical value of K at 10°C is

$$K = 25 \times 10^{-3}\,\mathrm{W\,m^{-1}\,K^{-1}}. \tag{7.60}$$

Between -5°C and 25°C, K varies only by about 4%.

In a steady state, the total mass flux into a spherical shell is independent of the radius of the shell. Consider the volume between an outer shell of radius r_1 and an inner one of radius r_0, see Figure 7.8. For a steady state the total vapour mass flux into this volume needs to be the same as the total flux out of this volume. Because this is true for any pair of radii, we have

$$F_m = \text{constant}. \tag{7.61}$$

We can combine this constraint with Fick's law to find a simple expression for the mass flux. First, rewrite Eq. 7.56 as

$$\frac{F_m}{r^2} = 4\pi D \frac{\mathrm{d}\rho_v}{\mathrm{d}r}. \tag{7.62}$$

Integrating this expression between the droplet radius $r = r_d$ and the far field $r = \infty$ we find

$$F_m / r_d = 4\pi D\,(\rho_v - \rho_{v,d}), \tag{7.63}$$

where we have written ρ_v for the vapour density in the far field (that is, the vapour density of the air not too close to the drop) and $\rho_{v,d}$ the vapour density at the drop surface.

The analogous argument holds for a steady heat flux: the total energy between radii r_0 and r_1 needs to remain constant, so the heat entering the volume must be the same as the heat exiting the volume. This can only be true if F_q is not a function of the radius. This then leads to

$$F_q/r_d = 4\pi K\,(T - T_d), \tag{7.64}$$

with T the air temperature in the far field and T_d the temperature at the drop surface. Substituting these expressions in the steady state condition of Eq. 7.58 we find

$$T - T_d = -\frac{LD}{K}(\rho_v - \rho_{v,d}). \tag{7.65}$$

The temperature difference and the vapour density difference are proportional. For large heat conductivity K, the temperature difference will be small; the latent heat can be efficiently transported away from the drop. For small heat conductivity, the latent heat will accumulate on the drop and larger temperature gradients are required to transport the heat away. The precise meaning of ‘large’ or ‘small’ will become clear later on.

To calculate the droplet growth we need to know the difference between the vapour density at the drop surface and the vapour density of the air. We assume that vapour near the drop is saturated. Using the ideal gas law we thus have

$$\rho_{v,d} = \rho_{vs}(T_d) = \frac{e_s(T_d)}{R_v T_d}. \tag{7.66}$$

We have suppressed the radius dependence of the saturated vapour pressure so this argument is only accurate for drops substantially larger than the activation radius. It is fairly straightforward to include this radius dependence but this would just make the equations more convoluted without providing any further insight.

Let us for the moment assume the drop temperature T_d is the same as the air temperature T. We then have

$$\rho_{v,d} = \rho_{vs}(T) = \frac{e_s(T)}{R_v T}. \tag{7.67}$$

Using the relative humidity, the vapour density in the far field can be written in terms of the saturated vapour density,

$$\rho_v = \mathrm{RH}\,\rho_{vs}(T). \tag{7.68}$$

With these expressions the total mass flux F_m in Eq. 7.63 becomes

$$F_m = 4\pi D r_d \, \rho_{vs} \, (\mathrm{RH} - 1), \tag{7.69}$$

where ρ_{vs} is evaluated at temperature T. In other words, the droplet growth is due to supersaturation: if the relative humidity is larger than 100% (positive supersaturation) the drop will grow; if it is smaller than 100% the drop will shrink. So for relative humidities below 100% the above equation can be used to calculate how quickly a droplet evaporates. Note also that the growing droplets extract water vapour from the air and thus reduce its relative humidity; droplet growth by condensation is a self-limiting process.

In the above derivation we took $T_d = T$. As the droplet is heated up by condensation this assumption is not valid. However, assuming that T_d and T are not too far apart we can linearize the saturated vapour pressure in Eq. 7.66 around temperature T,

$$\rho_{v,d} = \rho_{vs}(T_d) = \rho_{vs}(T) + \frac{\mathrm{d}\rho_{vs}}{\mathrm{d}T}(T_d - T). \tag{7.70}$$

Using the Clausius–Clapeyron equation in the form of Eq. 5.13 we have

$$\frac{\mathrm{d}\rho_{vs}}{\mathrm{d}T} = \left(\frac{L}{R_v T} - 1\right)\frac{\rho_{vs}}{T}. \tag{7.71}$$

We therefore find

$$\rho_{v,d} = \rho_{vs}\left(1 + \left(\frac{L}{R_v T} - 1\right)\frac{T_d - T}{T}\right). \tag{7.72}$$

The temperature difference between the drop and the far field can be expressed as a vapour density difference by Eq. 7.65. We now get for the vapour density difference

$$\rho_v - \rho_{v,d} = \rho_{vs}\left(\mathrm{RH} - 1 - \frac{\rho_v - \rho_{v,d}}{\rho_K}\right), \tag{7.73}$$

where we introduced a conduction density scale

$$\rho_K = \frac{KT}{LD}\left(\frac{L}{R_v T} - 1\right)^{-1}. \tag{7.74}$$

This equation can be rearranged to find the vapour density difference $\rho_v - \rho_{v,d}$ in terms of the saturated vapour density ρ_{vs}. A particularly compact way to write the ensuing expression is

$$\rho_v - \rho_{v,d} = \rho_r \, (\mathrm{RH} - 1), \tag{7.75}$$

where we have introduced a reduced vapour density ρ_r as

$$\frac{1}{\rho_r} = \frac{1}{\rho_{vs}} + \frac{1}{\rho_K}. \tag{7.76}$$

This density difference then leads to the total mass flux,

$$\blacktriangleright \quad F_m = 4\pi D r_d \rho_r\,(\mathrm{RH} - 1). \tag{7.77}$$

As can be seen, Eq. 7.77 is the same as Eq. 7.69 with the far-field saturated vapour density replaced by the reduced vapour density ρ_r.

It now becomes clear that latent heating slows the growth of the of the drop because for all ρ_K we have

$$\rho_r < \rho_{vs}. \tag{7.78}$$

Physically this makes sense: a warmer drop reduces the supersaturation with respect to the far field and thus reduces the vapour diffusion towards the drop. There are two limiting cases of interest. If the heat conduction coefficient K is very large, then any excess heat is immediately diffused away from the drop. We thus expect the heating effect to be small. Indeed, for $K \to \infty$ we have $\rho_K \to \infty$ and we find

$$\rho_r \to \rho_{vs}. \tag{7.79}$$

So the mass flux in Eq. 7.77 reduces to Eq. 7.69. The other limiting case is for a small heat conduction coefficient. In this case the latent heat cannot be transported away from the drop and we expect a strong reduction of the supersaturation. Indeed, for $K \to 0$ we have $\rho_K \to 0$ and we find

$$\rho_r \to \rho_K \to 0, \tag{7.80}$$

so the mass flux to the drop vanishes. Realistic values of the diffusion and conduction coefficients give a value of ρ_K of about $7 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$ and it is typically similar or somewhat larger than ρ_{vs}. The heating effect therefore reduces the growth by a factor of two to four compared to the case without heating.

The reduction becomes larger for higher temperatures. In fact, the reduced vapour density is a much weaker function of temperature than the saturated vapour density, see Figure 7.9. As can be seen in this figure, the reduced density is very nearly a linear function of the temperature. The approximation

$$\rho_r = (2.80 + 0.11\,T(^\circ\mathrm{C})) \times 10^{-3}\,\mathrm{kg\,m^{-3}} \tag{7.81}$$

is accurate to within 4% between temperatures of –5°C and 50°C.

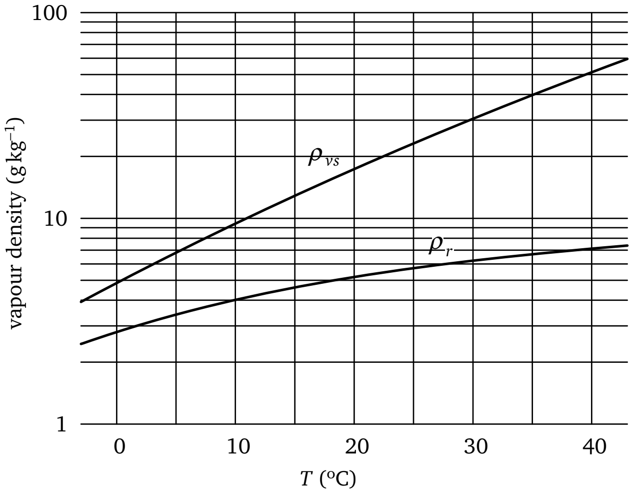


FIGURE 7.9 Saturated vapour density ρ_{vs} and reduced vapour density ρ_r as a function of temperature T.

Now we have an expression for the mass-flux, we can calculate the change in droplet radius r_d. We have

$$F_m = \frac{\mathrm{d}M}{\mathrm{d}t} = \rho_l \frac{\mathrm{d}V}{\mathrm{d}t} = 4\pi\rho_l r_d^2 \frac{\mathrm{d}r_d}{\mathrm{d}t} \tag{7.82}$$

with ρ_l the density of the liquid and V the droplet volume, $V = (4/3)\pi r_d^3$. Substituting the expression for the mass flux, we find

$$r_d \frac{\mathrm{d}r_d}{\mathrm{d}t} = D\frac{\rho_r}{\rho_l}\,(\mathrm{RH}-1). \tag{7.83}$$

The right-hand side of Eq. 7.83 is a constant at given temperature and relative humidity, so we can integrate the equation to find:

$$\blacktriangleright \quad r_d(t) = \sqrt{r_d^2(0) + 2\mathscr{D}t} \ \text{ with } \ \mathscr{D} = D\frac{\rho_r}{\rho_l}\,(\mathrm{RH}-1). \tag{7.84}$$

The droplet radius grows as the square root of time. When the relative humidity is smaller than 100%, this equation describes how the droplet radius reduces in time through evaporation.

Square root change of scale with time is typical of phenomena that are driven by diffusive processes. For example, the typical size of a cloud of tracer released by a local source grows with the square root of time and heat penetrates a conductor to a depth that grows with the square root of time.

An interesting property of Eq. 7.84 is that the radii of two droplets that start growing from different initial radii become more and more similar. From

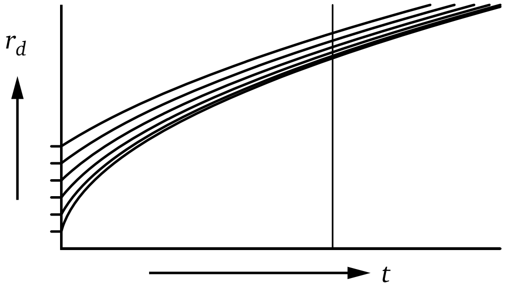


FIGURE 7.10 Growth of droplet radius through vapour diffusion. The distribution of droplet radii becomes ever narrower.

Eq. 7.84 we have

$$r_a^2(t) - r_b^2(t) = r_a^2(0) - r_b^2(0), \tag{7.85}$$

with r_a and r_b the radii of two different droplets growing in the same background. This is equivalent to

$$r_a(t) - r_b(t) = \frac{r_a^2(0) - r_b^2(0)}{r_a(t) + r_b(t)}. \tag{7.86}$$

Because both r_a and r_b will grow with time, their radii will get closer with time. This effect is illustrated in Figure 7.10.

Note that with changing droplet radius, the mass flux F_m is not constant in time. We then need to ask ourselves whether the steady state assumption leading to Eq. 7.61 is accurate. Any density fluctuations in the vapour field around the drop will adjust to a steady state through a diffusion process with diffusion coefficient D. The boundary condition change through the changing droplet radius occurs with an effective diffusion coefficient $\mathscr{D}$. Because

$$D \gg \mathscr{D}, \tag{7.87}$$

the vapour field adjusts to a steady state on a much faster time scale than that of the changing boundary (the relevant time scales are r_d^2/D and $r_d^2/\mathscr{D}$). We can therefore safely assume that the vapour flux is constant for any given droplet radius during the growth process.

Let us now put some numbers in Eq. 7.84. A typical supersaturation at activation is about 2% or less. At 10°C, the reduced vapour density ρ_r is about $4 \times 10^{-3}\,\mathrm{kg\,m^{-3}}$. The time taken to grow from a small droplet to a typical rain drop of radius about 1 mm is several days. Even for growth to modest droplet size of several tens of micrometres we find times of several hours. The typical time scale for rain to form in a convective cloud is very much shorter, perhaps half an hour or less. Diffusive growth cannot be the complete story to go from activation to raindrop.

The next stage in droplet growth is described by the process of *collision and coalescence*. In its simplest form, drops at different radii will fall at different terminal velocities. This then allows the drops to collide and possibly coalesce. This is a more efficient mechanism for droplet growth than diffusional growth.

The terminal velocity of a drop increases with increasing size. Now consider the following idealized case: A relatively large drop of radius r_d is falling through an environment with a number of smaller drops with a relative speed of $\mathscr{V}$. The large drop would sweep out a volume of $\mathscr{V}\pi r_d^2$ per unit time. The smaller drops represent a total liquid water density, ρ_b. We assume that a fraction E of the small drops in the volume swept out by the large drop would collide and coalesce with the large drop. The mass increase of the large drop would then be

$$\frac{\mathrm{d}M}{\mathrm{d}t} = E\,\mathscr{V}\pi r_d^2 \rho_b. \tag{7.88}$$

The effective volume swept out by the drop is somewhat larger than this because its effective radius has to be increased with the radius of the small droplets. The fraction E is called the *collection efficiency* and it is a strong function of the radius of the large drop and the radius of the smaller drops. The collection efficiency is typically a few per cent if the small droplets are much smaller than the larger droplets. This is because the air flow around the large drop drags most small droplets around the large drop and thus prevents them from colliding with the large drop. For small droplets larger than about 20 μm the collection efficiency is typically 90% or higher.

As in Eq. 7.82, this mass increase is related to a radius increase and we can rewrite the above equation as

$$\frac{\mathrm{d}r_d}{\mathrm{d}t} = E\,\frac{\mathscr{V}\rho_b}{4\rho_l}. \tag{7.89}$$

This is not a closed equation because the terminal velocity $\mathscr{V}$ is a function of the radius of the drop. For small drops (typical cloud drops with radii smaller than about 30 μm), the terminal velocity is low and the Reynolds number of the flow around the drop is small. This is the viscous Stokes regime, where the terminal velocity is proportional to the square of the radius. For large drops (larger than about a millimetre), the terminal velocity is large and the Reynolds number of the flow is large. This is the turbulent regime where the terminal velocity is proportional to the square root of the droplet radius.[43] In the intermediate regime, an empirical match is found for the terminal

---

[43] In the Stokes regime, the terminal velocity is a function of the effective droplet buoyancy g', the viscosity ν, and the droplet radius r_d. The only combination of these variables with the dimension of speed is $(g'\nu)\,r_d^2$, which is, up to a constant factor, the Stokes formula. For high Reynolds numbers, the terminal velocity cannot be a function of the viscosity. The only possible combination of buoyancy and droplet radius with the dimension of speed is $\sqrt{g' r_d}$. The effective buoyancy is $g' = g\,(\rho_l - \rho)/\rho$. Note that the effective droplet inertia in the denominator is given by the density of the displaced air, ρ.

velocity. In its simplest form, a terminal velocity proportional to the radius is chosen. Approximate formulae for droplet terminal velocity are

$$\mathscr{V} = 1.2 \times 10^8 r_d^2\,\mathrm{m\,s^{-1}} \qquad \text{if } r_d \lesssim 30\,\mu\mathrm{m}, \tag{7.90a}$$

$$\mathscr{V} = 250\sqrt{r_d}\,\mathrm{m\,s^{-1}} \qquad \text{if } r_d \gtrsim 1\,\mathrm{mm}, \tag{7.90b}$$

$$\mathscr{V} = 8 \times 10^3\, r_d\,\mathrm{m\,s^{-1}} \qquad \text{otherwise.} \tag{7.90c}$$

These fall speed equations can be combined with Eq. 7.89 to find the growth of drops by collisions and coalescence.

This theory has many approximations. The collection efficiency is not a constant but a complex function of the sizes of the two colliding drops as well as the relative speed of the colliding drops and of the turbulence in the surrounding air. Also, to achieve realistic droplet growths from this model, we need to assume a whole spectrum of initial droplet sizes. The largest of the initial droplets will then grow at the expense of the smaller drops because this process corresponds to the highest collection efficiencies.

We have only discussed the formation and growth of water droplets. So we have been mainly looking at processes that occur in *warm clouds*, that is, clouds below the freezing level. However, many clouds are above the freezing level and are therefore likely to be at least partly made up of ice. Cold cloud physics has a more empirical character as the ice processes are so complex.

Ice crystal activation is similar to droplet activation. Homogeneous nucleation does not usually occur; we need a substrate for the ice crystal to nucleate on. The most obvious choice would be for water droplets to freeze. It turns out that for small drops this happens when the water droplet is severely supercooled. All drops would be frozen typically at around $T < -40^\circ\mathrm{C}$. Droplets can also freeze at higher temperatures, especially if they contain a solid nucleus. However, many clouds contain substantial amounts of supercooled water. Analogous to heterogeneous nucleation of water drops, ice crystals can form on suitable substrates. Wind-blown clay-dust is a typical cause of nucleation. Ice nuclei become more efficient at lower temperatures. Indeed, the effective ice nucleus concentration increases exponentially with decreasing temperatures.

When ice crystals are activated they can grow through diffusional growth, just like water drops. In fact we find an analogue of Eq. 7.63 for ice crystals,

$$F_m/C = 4\pi D\,(\rho(\infty) - \rho(r_d)). \tag{7.91}$$

Here the radius of the droplet in Eq. 7.63 is replaced by a so-called *capacitance* C, which takes into account the shape of the ice crystal.[44] The capacitance of a sphere is r. The word 'capacitance' has its origin in electrostatics: the calculation for diffusional flux around a crystal turns out to be analogous to

---

[44] See Westbrook, C. D. et al. (2008) *J. Atmos. Sci.* **65**, 206–219.

the calculation of the electric field around a capacitor of a certain shape with capacitance C.

It is important to note that the growth rate of the crystal is now proportional to the supersaturation over ice rather than over water. The saturated vapour pressure over ice is lower than that over water, which means that the supersaturation with respect to ice is higher than the supersaturation with respect to water. It may then happen that in clouds where ice crystals and water droplets are present, the ice crystals grow faster than the water droplets, perhaps up to the point where the water droplets become subsaturated because the ice crystals will have scavenged the water vapour. At this point the ice crystals will grow at the expense of the water drops. This process is called the *Bergeron–Findeisen* process. Hole clouds are thought to be spectacular manifestations of the Bergeron–Findeisen process: an aeroplane flying through a supercooled water cloud can initiate ice crystal growth at the expense of the water drops. The cloud is left with a hole filled with ice crystals.

The droplet can grow further by the collision of ice crystals with supercooled drops, a process usually called *accrection*, or the clumping together of different ice crystals, a process usually called *aggregation*. The collection efficiencies for these processes are hard to determine as they depend strongly on crystal shape and temperature.[45]

PROBLEMS

7.1. For a drop which grows by condensation, calculate a typical temperature difference between the drop and the far field.

7.2. Wilson cloud chamber. Take a cylinder with some water and its saturated vapour. Now expand the cylinder rapidly. The adiabatic expansion is so fast the vapour does not immediately adjust to its new equilibrium. Using the Clausius–Clapeyron equation in its approximate form, show that the relative humidity after rapid expansion from volume V_0 to volume V_1 satisfies

$$RH = \left(\frac{V_0}{V_1}\right)^{c_{pv}/c_{vv}} \exp\left(\frac{L}{R_v T_0}\left(\left(\frac{V_1}{V_0}\right)^{R_v/c_{vv}} - 1\right)\right),$$

For water vapour, the specific gas constant is $R_v = 461\,\mathrm{J\,kg^{-1}\,K^{-1}}$ and the specific heat capacity at constant volume is $c_{vv} = 1410\,\mathrm{J\,kg^{-1}\,K^{-1}}$. Hence show that an expansion by a factor of 1.3 is sufficient to achieve a relative humidity of 400%. How does this result change when the cylinder is filled with air at standard pressure and temperature, some water, and its saturated vapour?

---

[45] For further reading on cloud physics, see Mason, B. J. (1971) *The physics of clouds*, 2nd edn. Oxford University Press, Oxford; Rogers R. R. & Yau, M. K. (1989) *A short course in cloud physics*, 3rd edn. Butterworth–Heinemann; Pruppacher, H. R. & Klett, J. D. (1997) *Microphysics of clouds and precipitation*, 2nd edn. Kluwer, Dordrecht.

7.3. Supersaturation in updrafts. How far do we need to lift an initially saturated air parcel to achieve a supersaturation of 1%? What is a typical updraft velocity needed to prevent a cloud-drop from falling from the air?

7.4. Evaporation of raindrops. Assume that raindrops have a terminal velocity of $\mathscr{V} = Cr_d$ with r_d the drop radius and $C = 8 \times 10^3\,\mathrm{s}^{-1}$. Show that, if we ignore the temperature dependence in the reduced density ρ_r, the radius of a raindrop falling through an environment with relative humidity RH changes with height z as

$$r_d^3(z) = r_d^3(z_0) + \frac{3D\rho_r}{C\rho_l}\,(\mathrm{RH} - 1)\,(z_0 - z),$$

with z_0 the initial height. For a cloud base of 1 km, an environmental temperature of 280 K, and relative humidity RH = 70%, how large does the initial raindrop need to be to reach the ground? Estimate the error made by assuming a constant ρ_r. The effective diffusion coefficient for a falling drop will be increased because of ventilation. How will this change the answer?

8

Mixtures and solutions

Variations in the composition of substances have thermodynamic implications. This contrasts with our previous emphasis on so-called simple substances, substances where the composition does not vary during the processes considered. Here we will mainly concentrate on how the presence of solutes in water affects the properties of the water. Although of considerable interest to atmospheric science, we will not consider the thermodynamic aspects of chemical reactions.

8.1 CHEMICAL POTENTIALS

Consider a substance which is made up of N different composites with the i-th composite having mass M_i. So the total mass M is

$$M = \sum_i M_i. \tag{8.1}$$

The Gibbs function G of such a composite substance will be a function of the pressure p, temperature T and the composition,

$$G = G(p, T, M_1, M_2, M_3, \dots). \tag{8.2}$$

The differential of G can be written as

$$\begin{aligned} dG &= \left(\frac{\partial G}{\partial p}\right)_{T,M_i} \mathrm{d}p + \left(\frac{\partial G}{\partial T}\right)_{p,M_i} \mathrm{d}T + \sum_i \left(\frac{\partial G}{\partial M_i}\right)_{p,T} \mathrm{d}M_i \\ &= V\,\mathrm{d}p - S\,\mathrm{d}T + \sum_i g_i\,\mathrm{d}M_i, \end{aligned} \tag{8.3}$$

where the g_i are defined by the partial differentiation of the Gibbs function with respect to M_i,

► $$g_i = \left(\frac{\partial G}{\partial M_i}\right)_{T,p}. \tag{8.4}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

The g_i are called *chemical potentials;* the chemical potentials are intensive quantities that are functions of the pressure, temperature, and the composition of the substance.

Note that the chemical potentials are not functions of the absolute values of the M_i but only of the ratios M_i/M_j: any uniform scaling of the masses M_i by some factor does not change the definition of the chemical potentials, as the Gibbs function G itself also scales with this factor. Such a uniform scaling of the masses can be written as

$$\mathrm{d}M_i = M_i\,\mathrm{d}\alpha, \tag{8.5}$$

where $\mathrm{d}\alpha = \mathrm{d}M_i/M_i$ is the same for all composites. So for constant composition we find that at fixed temperature and pressure

$$\mathrm{d}G = \sum\nolimits_i g_i M_i\,\mathrm{d}\alpha. \tag{8.6}$$

Because the g_i do not change in this process, this equation can be integrated to find

$$G = \sum\nolimits_i M_i\,g_i. \tag{8.7}$$

It now becomes clear that the chemical potentials g_i are the multi-component analogues of the specific Gibbs function for simple substances. The chemical potentials g_i are not the same as the specific Gibbs functions for the pure components, as the chemical potentials in general will depend on the composition of the substance.

The analogy between chemical potentials and specific Gibbs functions also applies to the equilibrium between different phases. As shown in Chapter 5 this equilibrium follows from $\mathrm{d}G = 0$ at constant pressure and temperature. For simple substances this leads to the result that the specific Gibbs functions of the substance in the two phases have to be the same, see Eq. 5.4. This can be generalized to multi-component substances.

The phase equilibrium at constant pressure and temperature is still defined by $\mathrm{d}G = 0$, where G is the Gibbs function for the total system containing the two phases. We can now consider the variation in the Gibbs function due to the transfer of component i from one phase to another. For the sake of argument, consider a system made up of a liquid and its vapour, and the evaporation of a mass δM_i of component i from the liquid to the vapour. At constant pressure and temperature the change in Gibbs function δG is then given by

$$\delta G = \sum\nolimits_i (g_{i,v} - g_{i,l})\,\delta M_i. \tag{8.8}$$

As δG has to vanish for any variation in the system, we conclude that for all components the chemical potentials have to be the same between the two phases in equilibrium,

$$\blacktriangleright \qquad g_{i,v} = g_{i,l}. \tag{8.9}$$

This is the multi-component generalization of Eq. 5.4. Note that on adding a component we also add another equilibrium relation for the chemical potential so the equilibrium state remains well defined.

The reliance on the use of the Gibbs function G to define chemical potentials and phase equilibria is only apparent; the choice is one of convenience, not necessity. The situation is analogous to that in the derivation of the Clausius–Clapeyron equation in Section 5.1. For example, consider the enthalpy H of a substance made up of several components. Because $H = G + TS$ we can use Eq. 8.3 to write for the differential of the enthalpy

$$\mathrm{d}H = \mathrm{d}G + T\,\mathrm{d}S + S\,\mathrm{d}T = V\,\mathrm{d}p + T\,\mathrm{d}S + \textstyle\sum_i g_i\,\mathrm{d}M_i. \tag{8.10}$$

So we can also define the chemical potentials g_i as partial derivatives of the enthalpy $g_i = (\partial H/\partial M_i)_{S,p}$. Analogous arguments work for the other thermodynamic potentials and they lead to four equivalent definitions of the chemical potentials:

$$g_i = \left(\frac{\partial U}{\partial M_i}\right)_{S,V} = \left(\frac{\partial H}{\partial M_i}\right)_{S,p} = \left(\frac{\partial F}{\partial M_i}\right)_{T,V} = \left(\frac{\partial G}{\partial M_i}\right)_{T,p}. \tag{8.11}$$

In contrast, the arguments leading to Eq. 8.7 cannot be straightforwardly translated to the other thermodynamic potentials: only the Gibbs function has two intensive variables as natural variables (T and p); the other potentials have at least one extensive natural variable, which would change on a uniform scaling of the masses of all constituents. So although we can use any of the four standard thermodynamic potentials to define the chemical potential, the chemical potential is the multi-component generalization of the specific Gibbs function.

8.2 IDEAL GAS MIXTURES AND IDEAL SOLUTIONS

For ideal gas mixtures we can write down explicit expressions for the chemical potentials. In order to calculate the chemical potentials in an ideal gas mixture we will first calculate its total Gibbs function. We will take it that for an ideal gas mixture the individual components contribute independently to the extensive variables as each component, by definition of the ideal gas, is not influenced by the presence of the other components.

For an ideal gas mixture the total internal energy U is given by

$$U = \textstyle\sum_i M_i\,u_i. \tag{8.12}$$

Remember that for ideal gases the specific internal energies u_i are functions of the temperature only. The above expression is valid for a mixture because an ideal gas is made up of non-interacting molecules so there is no contribution of interaction energy between the different components.

The entropy S of an ideal gas is also the sum of the individual contributions of the components in the mixture:

$$S = \sum_i M_i s_i. \tag{8.13}$$

The specific entropies s_i can be written as functions of the partial pressure of the component p_i and the temperature T, see Eq. 3.40,

$$s_i = c_{pi} \ln (T/T_0) - R_i \ln (p_i/p_0), \tag{8.14}$$

with c_{pi} the specific heat capacity at constant pressure for component i and $R_i = R^\star/\mu_i$ its specific gas constant. The T_0 and p_0 are the reference temperature and pressure. This expression for specific entropy can be rewritten as

$$\begin{aligned} s_i &= c_{pi} \ln (T/T_0) - R_i \ln (p/p_0) - R_i \ln (p_i/p) \\ &= s_i(p, T) - R_i \ln (n_i/n), \end{aligned} \tag{8.15}$$

where in the last step the ideal gas law for each component $p_i V = n_i R^\star T$ is used with n_i the number of moles of component i and $n = \sum_i n_i$ the total number of moles.

The total Gibbs function G of the ideal gas mixture now is

$$\begin{aligned} G &= U + pV - ST \\ &= \sum_i M_i u_i + p_i V - M_i s_i T \\ &= \sum_i M_i (u_i(T) + R_i T - s_i(p, T)\, T + R_i T \ln (n_i/n)) \\ &= \sum_i M_i (g_{0i}(p, T) + R_i T \ln (n_i/n)). \end{aligned} \tag{8.16}$$

In the second step the ideal gas law was used in the form $p_i V = M_i R_i T$. In the last step we introduced the notation g_{0i} for the specific Gibbs function of the pure component i. The chemical potentials can now be derived by either comparing the above equation with Eq. 8.7 or by calulating the partial derivatives $g_i = \partial G/\partial M_i$. We find that for an ideal gas mixture the chemical potentials are given by

▶ $$g_i = g_{0i}(p, T) + R_i T \ln (n_i/n). \tag{8.17}$$

So the chemical potential can be expressed as a contribution from the specific Gibbs function of the pure component at the given temperature and pressure and a mixing contribution. Note that this mixing contribution only appears because the pure component Gibbs function was expressed as a function of total pressure rather than partial pressure. For a single component ideal gas, the mixing contribution will vanish as in this case $n_i = n$.

An *ideal solution* is the analogue of an ideal gas for liquid mixtures. In an ideal gas the interaction potential between different molecules vanishes; in an ideal solution the interaction potential between different molecules is the same, irrespective of their nature. In other words, each molecule interacts with the rest of a solution as if it was a pure substance made up of a single component. This means that, for example, the specific internal energy of a component in a solution cannot be a function of the composition of that solution. We therefore find for the total internal energy of an ideal solution

$$U = \sum\nolimits_i M_i \, u_{0i}, \tag{8.18}$$

with u_{0i} the specific internal energy of the pure substance made of component i. For the total volume V we can write

$$V = \sum\nolimits_i M_i \, v_{0i}, \tag{8.19}$$

with $v_{0i} = V/M$ the specific volume of the pure substance made of component i (for the pure substance $M_i = M$). We can then use the first law to write the differential of the entropy S of an ideal solution,

$$\mathrm{d}S = \sum\nolimits_i M_i \, \frac{\mathrm{d}u_{0i} + p \, \mathrm{d}v_{0i}}{T}. \tag{8.20}$$

This differential can be integrated taking into account that at fixed composition the u_{0i} and the v_{0i} are only functions of p and T. We find

$$S = \sum\nolimits_i M_i \, s_{0i}(p, T) + C(M_1, M_2, M_3, \ldots), \tag{8.21}$$

where s_{0i} are the specific entropies for the pure components at the given pressure and temperature and C is an integration constant that can only depend on the composition of the solution. Only for ideal solutions are the s_{0i} the specific entropies for the pure components; for non-ideal solutions they are also functions of the composition.

So what is the value of the integration constant C? The above arguments are valid for any temperature or pressure as long as the solution is ideal, that is, the interaction between molecules does not depend on the nature of the molecules. Therefore, if we choose high enough temperature and low enough pressure, the solution would evaporate and become an ideal gas and the entropy would be given by Eq. 8.15. The integration constant C does not depend on temperature or pressure, which means that the integration constant for the ideal solution is the same as that for the ideal gas. We therefore find

$$C = -\sum_i M_i R_i \ln(n_i/n). \tag{8.22}$$

We have thus shown that the entropy of an ideal solution is

$$S = \sum_i M_i (s_{0i}(p,T) - R_i \ln(n_i/n)). \tag{8.23}$$

The Gibbs function for an ideal solution follows from $G = U + pV - ST$ and can be written as

$$G = \sum_i M_i (g_{0i}(p,T) + R_i T \ln(n_i/n)). \tag{8.24}$$

This expression is identical in structure to the ideal gas result, Eq. 8.16. From this we find the chemical potentials for an ideal solution as:

► $$g_i = g_{0i}(p,T) + R_i T \ln(n_i/n). \tag{8.25}$$

As in the ideal gas case, Eq. 8.17, the chemical potential for an ideal solution is the sum of the specific Gibbs function for the pure component and a mixing contribution.

8.3 RAOULT'S LAW REVISITED

Now consider an ideal solution with its saturated vapour. The equilibrium condition for each component in the solution and the vapour is given by Eq. 8.9. For the solution, the chemical potential for each component is

$$g_{i,l} = g_{0i,l}(p,T) + R_i T \ln(n_i/n), \tag{8.26}$$

where we added the subscript l to make explicit that we refer to the solution. The vapour is also considered ideal so we can use Eq. 8.17. However, in the present case it is convenient to write the chemical potentials for the ideal vapour, not as functions of the total pressure p, but as functions of the saturated vapour pressure for the pure components, denoted e_{s0i}. We can then readily show that for the ideal vapour mixture we have

$$g_{i,v} = g_{0i,v}(e_{s0i},T) + R_i T \ln(p_i/e_{s0i}). \tag{8.27}$$

It is of interest to write the equilibrium condition for a pure substance in this notation. We find for a substance made up of pure component i,

$$g_{0i,v}(e_{s0i},T) = g_{0i,l}(e_{s0i},T) \tag{8.28}$$

as for a pure substance $p, p_i \rightarrow e_{s0i}$ and $n_i \rightarrow n$. The general equilibrium condition, $g_{i,v} = g_{i,l}$, can now be written as

$$g_{0i,l}(e_{s0i}, T) - g_{0i,l}(p, T) = R_i T \ln\left(\frac{n_i e_{s0i}}{n p_i}\right). \tag{8.29}$$

The left-hand side of this equation has a magnitude of about $(\partial g/\partial p)\,\Delta p = v_l\,\Delta p$ where Δp is the pressure difference between e_{s0i} and p. The $R_i T$ on the right-hand side is certainly much larger as it scales with v_v rather than v_l. We therefore have to conclude that the logarithm has to be small, of the order v_l/v_v. Therefore the argument of the logarithm has to be close to unity, an approximation which is valid up to order v_l/v_v. This can be written as

$$\blacktriangleright \qquad p_i = \frac{n_i}{n}\, e_{s0i}. \tag{8.30}$$

This equation is known as Raoult's law. It assumes that the solution and the vapour are ideal and that the density of the solution is much greater than the density of the vapour. Raoult's law says that the vapour pressure of each component over a solution is the vapour pressure over the pure substance weighted by its number concentration in the solution.

As an example we can take a water-based ideal solution with a number concentration c of solute. The water then has a number concentration of $1 - c$. According to Raoult's law the saturated water vapour pressure $e_s(c)$ over this solution is then related to the saturated vapour pressure over pure water $e_s(0)$ by

$$e_s(c) = (1 - c)\, e_s(0), \tag{8.31}$$

where we have used the same notation as in Eq. 7.24. An alternative way to read this version of Raoult's law is to observe that the relative humidity of a vapour over a a solution can never exceed

$$\mathrm{RH}_{\mathrm{max}} = (1 - c) \times 100\%, \tag{8.32}$$

where the maximum is achieved when the vapour is saturated.

8.4 BOILING AND FREEZING OF SOLUTIONS

Water changes its boiling point and freezing point on dissolving some solute in it. The use of common salt to free roads from ice and snow is a well-known application. We have now developed all the tools to understand and calculate the effects of solute on boiling points and freezing points.

For the boiling point of a solution, we make the simplifying assumption that the solution is made of a solvent which will evaporate while the solute will not evaporate; the solute is *non-volatile.* This is well known from boiling salty water: the water evaporates, while the salt stays behind. In the context of ideal solutions, the situation is equivalent to the solutes having a very low equilibrium saturation vapour pressure, so that for all practical purposes they do not contribute to the total vapour pressure over the solution.

We could equally have assumed that the amount of solvent is small. In this case we can drop the ideal solution approximation and still get essentially the same results as in the previous two sections: the chemical potentials are, to first order in the solute concentrations, only functions of the pressure p and temperature T. Any dependency of the chemical potentials on solute concentrations would correspond to second order (in solute concentration) corrections to expressions such as Eq. 8.24. With the assumption of small solute concentrations, the expression of Raoult's law in the form of Eq. 8.31 remains valid, even for non-ideal solutions.

The change of the saturated vapour pressure δe_s due to a solute with number concentration c equals

$$\delta e_s = -c\, e_s(0), \tag{8.33}$$

with $e_s(0)$ the vapour pressure for the pure solvent. For the solution to boil, the vapour pressure needs to equal the atmospheric pressure. So because the solute reduces the vapour pressure, we need to increase the boiling temperature by an amount δT such that the increase in vapour pressure of the solvent, $e_s(0)$, compensates for the decrease in vapour pressure due to the solute, δe_s. The increase of $e_s(0)$ due to a temperature increase δT is obtained from the Clausius–Clapeyron equation, Eq. 5.13,

$$\delta T = \frac{R_v T_0^2}{L} \frac{\delta e_s(0)}{e_s(0)}, \tag{8.34}$$

with T_0 the boiling point of the pure solvent. The $\delta e_s(0)$ is to compensate for the reduction in vapour pressure due to the solute,

$$\delta e_s(0) = c\, e_s(0). \tag{8.35}$$

Substituting this in the linearized Clausius–Clapeyron equation we find

$$\blacktriangleright \qquad \delta T = c\, \frac{R_v T_0^2}{L}. \tag{8.36}$$

So this is the linear change in the boiling point of a solution with a change in the number concentration of a non-volatile solute. The change in temperature is always positive and it is called the *boiling point elevation*.

A more direct route to calculate the boiling point of a solution can be found by equating the chemical potentials of the solvent in the solution and in the vapour. We assume again an ideal solution (or small solute concentration) and we assume a non-volatile solute. Equating the chemical potentials for the solvent, with number concentration $1 - c$, and the solvent vapour we find

$$g_{0,l}(p, T_1) + R_v T_1 \ln(1 - c) = g_{0,v}(p, T_1). \qquad (8.37)$$

Here, T_1 is the boiling point of the solution and p is the atmospheric pressure, which is equal to the vapour pressure at the boiling point. For the pure solvent we would have at the boiling point

$$g_{0,l}(p, T_0) = g_{0,v}(p, T_0), \qquad (8.38)$$

where T_0 is the boiling point of the pure solvent and p is again the atmospheric pressure. Subtracting these two equations and assuming small variations in the boiling point, we find

$$(s_{0,v} - s_{0,l})(T_1 - T_0) = -R_v T_1 \ln(1 - c) \qquad (8.39)$$

where we have taken it that $(\partial g/\partial T)_p = -s$. At equilibrium, the difference in specific entropy for the vapour and liquid is related to the enthalpy of vaporization L,

$$T_0\,(s_{0,v} - s_{0,l}) = L. \qquad (8.40)$$

This can be substituted in Eq. 8.39 to find the familiar result for the boiling point elevation,

$$\delta T = T_1 - T_0 = c\,\frac{R_v T_0^2}{L}, \qquad (8.41)$$

where we have taken it that c is small so $\ln(1 - c) \approx -c$ and $T_1 \approx T_0$.

The advantage of this more direct route is that it does not employ the ideal gas assumption for the vapour. It can therefore also be applied to the equilibrium between a solid and an ideal solution, as long as the solid only contains molecules of the solvent. Salty water is again a good example: on freezing salty water, the ice only contains water and the salt remains in the solution.

It costs energy for the solvent to move from the solid to the solution. The enthalpy of freezing is therefore negative $-L_f$ where L_f is commonly called the *enthalpy of fusion*. So, analogous to the boiling point elevation of Eq. 8.41, we now find the *freezing point depression* δT,

$$\blacktriangleright \qquad \delta T = -c\,\frac{R_v T_0^2}{L_f}, \qquad (8.42)$$

with T_0 now the freezing point of the pure solvent and L_f the enthalpy of fusion. For water, the enthalpy of fusion at 0°C is $0.334 \times 10^6\,\mathrm{J\,kg^{-1}}$.

Both the freezing point depression and the boiling point elevation depend on the solution being ideal or dilute. So for large solute concentrations these equations become less accurate. However, there is a limit to how much solute will dissolve in a solvent. For example, about 350 g of common salt can be dissolved in a litre of water. Such a concentration would correspond to a freezing point depression of about 18 K, using the ideal solution equations. This is in fact very close to the observed maximum freezing point depression of 21.1 K.

The boiling point elevation and the freezing point depression for dilute solutions are not dependent on the type of solute; they only depend on the type of solvent. Such properties are called *colligative properties*. Note that both effects depend on the *number* concentration c of the solute in the solvent. To find the number concentration from the mass of dissolved solute we need to take into account the dissociation of solute molecules in the solution. For example, common salt NaCl in water will nearly completely dissociate into Na^+ and Cl^- ions, so each mole of NaCl will contribute nearly 2 moles to the solute concentration (a more precise value is 1.8). The level of dissociation varies with solute and solvent type. The dissociation effect was included as the Van 't Hoff factor i in the discussion of the Raoult effect in Section 7.2; common salt in water has a Van 't Hoff factor of about 1.8. An accurate measurement of the freezing point depression of a solution can in fact be used to measure the level of dissociation of the solute.

PROBLEMS

8.1. Mixing entropy. Consider a cylinder divided in two equal compartments by a wall and kept at temperature T. Each compartment contains n moles of a gas. When the separating wall is removed, the two gases will mix. By comparing this process to expansion into vacuum (remember that the gases are ideal so do not interact with each other) show that the mixing entropy is given by the second term in Eq. 8.15. What happens if the two gases are the same?

8.2. Number concentrations. Dissolve a mass M_2 of solute into a mass M_1 of solvent. Show that the number concentration c of solute molecules is

$$c = \frac{iM_2}{iM_2 + M_1\mu_2/\mu_1},$$

with i the Van 't Hoff factor, and μ_1 and μ_2 the molar masses of the solvent and the solute respectively.

8.3. How many grams of common salt (NaCl) need to be dissolved in a litre of pure water in order to increase its boiling point by one degree Celsius? The molar masses of water and common salt are $18.02\,\mathrm{g\,mol^{-1}}$ and $58.44\,\mathrm{g\,mol^{-1}}$, respectively.

How many grams of common salt (NaCl) need to be dissolved in a litre of pure water in order to reduce its freezing point by one degree Celsius?

Ocean water typically contains 35 grams of salts (mainly common salt) per kilogram of water. Calculate the typical freezing point of ocean water.

Refering to the Köhler theory in Chapter 7, what are typical freezing point depressions for drops near activation?

8.4. The boiling point elevation δT_b and freezing point depression δT_f are often expressed as functions of *molality*, c_m, defined as the number of moles of solute per kilogram of solvent. For dilute solutions we have

$$\delta T_b = iK_b\,c_m \quad \text{and} \quad \delta T_f = -iK_f\,c_m,$$

with K_b the boiling point elevation constant, K_f the freezing point depression constant, and i the relevant van 't Hoff factor. Find expressions for K_b and K_f and show that for water $K_b = 0.52\,\mathrm{K\,kg\,mol^{-1}}$ and $K_f = 1.86\,\mathrm{K\,kg\,mol^{-1}}$.

9
Thermal radiation

The Sun's radiation is the ultimate source of nearly all energy on the Earth. The differential heating of the Earth makes the atmosphere unstable and sets it in motion. Here we introduce some of the key radiative processes involved. We mainly concentrate on the thermodynamic aspects of radiation rather than the molecular structure of of absorption and emission spectra.[46]

9.1 THERMAL RADIATION AND KIRCHHOFF'S LAW

All bodies emit electromagnetic radiation, *thermal radiation*, by virtue of their temperature. Before deriving its detailed characteristics, we will first explore some of the basic features of this radiation. A useful picture to keep in mind is that in quantum mechanics, electromagnetic radiation is mediated by massless particles called photons. Many properties of the radiation field can be understood by interpreting it as a photon gas.

Let us consider a vacuum vessel, the walls of which are kept at a certain temperature T. What is the nature of the equilibrium thermal radiation in the vessel? How much energy is associated with the thermal radiation? It is convenient to think of the volumetric radiative energy density; radiative energy per unit mass cannot be defined because photons have no mass. The only variables that can set the energy density then are the vessel wall temperature, the wavelength under consideration, and, at first sight, the vessel size and geometry.

However, the vessel size and geometry cannot be important for the equilibrium energy density. We can connect two different vessels at the same temperature with a connection that can let through radiation of a certain wavelength. If those vessels contain a different radiative energy density (in other words, a different photon density at the chosen wavelength) then

---

[46] There are quite a few specialized texts on atmospheric radiation. Notable examples are Goody R. M. & Yung, Y. L. (1989) *Atmospheric radiation. Theoretical basis*, 2nd edn. Oxford University Press, Oxford; Petty, G. W. (2004) *A first course in atmospheric radiation* Sundog Publishing, Madison.

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

energy (photons) will flow from the high density to the low density vessel. However, this contradicts thermodynamic equilibrium: bodies at the same temperature are in thermal equilibrium and do not spontaneously exchange energy. Similarly, we can argue that the energy density cannot be a function of the wave direction; the radiation is *isotropic.* We have to conclude that the volumetric radiative energy density per unit wavelength $\tilde{u}_\lambda$ is a scalar function of wavelength λ and temperature T,

$$\tilde{u}_\lambda = f(\lambda, T). \tag{9.1}$$

The thermal radiation in the vessel impinges on the walls with a certain energy flux. Because the radiative energy density is an isotropic function of wavelength and temperature, this energy flux can also be only a function of wavelength and temperature. The radiative energy incident on the vessel walls per unit area and per unit wavelength is therefore some radiative flux $B_\lambda(T)$ (units $\mathrm{W\,m^{-2}\,m^{-1}}$). In Appendix D we demonstrate that

$$B_\lambda(T) = \tilde{u}_\lambda\, c/4, \tag{9.2}$$

with c the speed of light.

Now assume that the vessel walls absorb a fraction α_λ, the *absorptivity*, of the incoming radiation. In equilibrium, the vessel walls emit energy $E_\lambda(T)$ equal to the energy absorbed,

$$E_\lambda(T) = \alpha_\lambda B_\lambda(T). \tag{9.3}$$

A *black body* is now defined as a body that has absorptivity equal to one for every wavelength, $\alpha_\lambda = 1$. So black bodies must emit $B_\lambda(T)$, the *black body emission.* For general bodies we can define an emissivity ϵ_λ which expresses the emission as a fraction of the black body emission,

$$E_\lambda(T) = \epsilon_\lambda B_\lambda(T). \tag{9.4}$$

In equilibrium, as expressed by Eq. 9.3, this gives *Kirchhoff's law,*

► $$\epsilon_\lambda = \alpha_\lambda. \tag{9.5}$$

So Kirchhoff's law states that at equilibrium the emissivity and the absorptivity of a surface have to be the same at each wavelength.

Materials which have an constant value of α_λ lower than one are called *grey bodies.* Materials with varying absorptivity are called *coloured.* Most real materials are coloured. The atmosphere has a complicated absorption structure associated with quantum mechanical absorption bands of the different atmospheric constituents. For approximate calculations, we often use the grey body approximation in wide bands of the spectrum. That is, we carve up the spectrum into bands with a constant average absorptivity. Such models are called wide-band models.

In Section 9.9 we will show that the black body emission per unit wavelength and per unit area $B_\lambda(T)$, equals

$$\blacktriangleright \qquad B_\lambda(T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \tag{9.6}$$

with *h Planck's constant*,

$$h = 6.626 \times 10^{-34} \mathrm{J\,s}, \tag{9.7}$$

c the speed of light, and k_B the Boltzmann constant. The above expression for the black body emission is called *Planck's law* and it played a pivotal role in the early development of quantum mechanics.

Black body radiation is a strong function of temperature. The graph in Figure 9.1 shows, on a logarithmic scale, the black body emission as a function of wavelength for two different temperatures: 5780 K, the temperature of the Sun's surface, and 280 K, a typical temperature at the Earth's surface. Not only is the total emission (the integral under the Planck curve) much greater for the higher temperature, its peak is also at shorter wavelengths.

The peak for the solar temperature is in the visible spectrum. This is to be expected on evolutionary grounds: any life form would develop 'eyes' that could see the radiation that the local star predominantly produced, which then by definition is 'visible'. Additionally, the atmosphere is nearly transparent at these wavelengths so all this radiation reaches the Earth's surface.

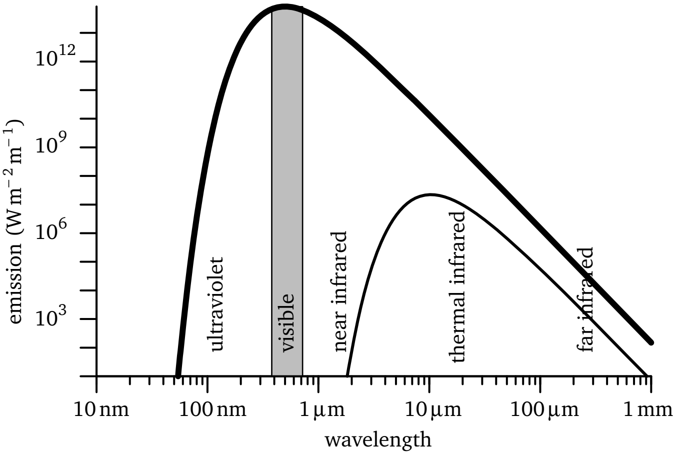


FIGURE 9.1 Black body emission as a function of wavelength for two different temperatures, 5780 K, the temperature of the Sun's surface (thick curve), and 280 K, a typical temperature at the Earth's surface (thin curve).

The fact that the Sun emits nearly equally at visible wavelengths would make it white in colour. Because the atmosphere preferentially scatters blue light, the Sun looks yellow when viewed from the Earth's surface. For temperatures typical for the Earth's surface and the atmosphere, the emission is mainly in the thermal infrared, a property that is used in infrared thermometers.

9.2 THE STEFAN–BOLTZMANN AND WIEN DISPLACEMENT LAWS

The Stefan–Boltzmann law expresses how much total radiative energy flux a black body emits; the Wien displacement law expresses at what wavelength the black body has its maximum emission. These laws can be derived from thermodynamic arguments as well as from the Planck law. The former method is covered in section 9.8. Here we discuss how both laws follow from the Planck law.

The Planck law can be integrated over all wavelengths to get the *total emission B* of a black body,

$$B = \int B_\lambda \, \mathrm{d}\lambda = \int_0^\infty \frac{2\pi hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \, \mathrm{d}\lambda. \tag{9.8}$$

It is convenient to introduce the *thermal wavelength* Λ,

$$\blacktriangleright \quad \Lambda = hc/k_B T. \tag{9.9}$$

We can now non-dimensionalize the wavelength with the thermal wavelength by defining the variable

$$\xi = \lambda/\Lambda. \tag{9.10}$$

In terms of this non-dimensional wavelength, the total black body emission can be written

$$B = \frac{2\pi k_B^4 T^4}{h^3 c^2} \int_0^\infty \frac{1/\xi^5}{\exp(1/\xi) - 1} \, \mathrm{d}\xi. \tag{9.11}$$

The integral is a fixed number and it can be shown that it equals $\pi^4/15$. We now find the *Stefan–Boltzmann law*,

$$\blacktriangleright \quad B = \sigma T^4, \tag{9.12}$$

with

$$\sigma = \frac{2\pi^5 k_B^4}{15 h^3 c^2} = 5.670 \times 10^{-8} \mathrm{W\,m^{-2}\,K^{-4}} \tag{9.13}$$

the *Stefan–Boltzmann constant.* The Stefan–Boltzmann law states that the total radiative energy per unit area emitted by a black body increases with the fourth power of its temperature. The units of B are $\mathrm{W\,m^{-2}}$.

The Planck curve shows a maximum at a certain wavelength that can be determined from $dB_\lambda/d\lambda = 0$. It is again easiest to non-dimensionalize the problem and find the non-dimensional wavelength ξ for which $dB_\xi/d\xi = 0$. After some algebra, we find that at the maximum ξ has to satisfy

$$(5 - 1/\xi)\, e^{1/\xi} = 5. \tag{9.14}$$

This is a trancedental equation which can be solved numerically and its solution is $1/\xi_m = 4.96511423\ldots$ So the Planck function $B_\lambda(T)$ has a maximum when $\xi = \xi_m$;[47] that is, for a wavelength λ_m of

$$\lambda_m = \xi_m \frac{hc}{k_B T}, \tag{9.15}$$

which can be written as

▶ $$\lambda_m = b/T, \tag{9.16}$$

with

$$b = 2.898 \times 10^{-3}\mathrm{m\,K}. \tag{9.17}$$

Equation 9.16 is called *Wien's displacement law* and the constant b is called the *displacement constant*. So the higher the temperature, the shorter the wavelength at which the maximum emission occurs.

For the surface of the Sun at $T = 5780\,\mathrm{K}$ we find $\lambda_m = 500\,\mathrm{nm}$, well inside the visible spectrum (from about 400 nm to about 700 nm, see Figure 9.1). In fact, about 40% of the solar radiative output is in the visible spectrum – see also Problem 9.2 and Figure 9.2. For $T = 280\,\mathrm{K}$ we find $\lambda_m = 10\,\mu\mathrm{m}$, well in the thermal infrared. Thermography is based on this principle: an object at different temperatures emits predominantly at different wavelengths and has therefore a different 'colour' in the infrared.

In an atmospheric context, radiation around the thermal infrared is often called *long-wave radiation* and radiation around the visible spectrum and shorter wavelengths is often called *short-wave radiation*.

---

[47] The Planck law can also be expressed in frequency instead of wavelength. The maximum emission in the frequency spectrum corresponds to $1/\xi'_m = 2.821439\ldots$ which is the solution to $(3 - 1/\xi)\, e^{1/\xi} = 3$.

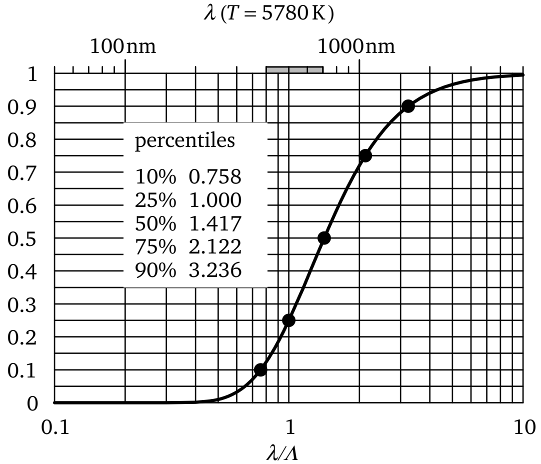


FIGURE 9.2 Cumulative fraction of black body emission as a function of wavelength λ, expressed in thermal wavelengths Λ (bottom axis) and in nanometres for a radiation temperature of 5780 K (top axis; the grey bar corresponds to the visible part of the spectrum).

9.3 GLOBAL ENERGY BUDGET AND THE GREENHOUSE EFFECT

The Sun's radius is about $R_S = 6.96 \times 10^8$m and its surface temperature is about $T_S = 5780$K. The Sun can be approximately treated as a black body. From Stefan's law we can then find the total radiative energy output of the Sun. When this radiation has reached the Earth's orbit (average orbital radius $r_E = 149.5 \times 10^9$m) this energy is used to irradiate a much larger area, $4\pi r_E^2$ instead of $4\pi R_S^2$. So the radiative energy flux per unit area S_0 at the distance of the Earth can be found from

$$4\pi r_E^2\, S_0 = 4\pi R_S^2\, \sigma T_S^4. \tag{9.18}$$

The flux S_0 is called the *solar constant*, or the *total solar irradiance*, and it has a value of about

$$\blacktriangleright \qquad S_0 = 1366\,\text{W}\,\text{m}^{-2}. \tag{9.19}$$

The solar constant is not really constant. There is a small variation, typically less than $1\text{W}\,\text{m}^{-2}$, due to the change of the solar output over an 11-year solar cycle, as well as variations on cosmological time scales.

The solar constant is defined as the amount of solar radiation received outside the Earth's atmosphere at the Earth's mean distance from the Sun. The Earth's orbit is slightly eccentric, which leads to a variation of solar irradiance from $1412\,\text{W}\,\text{m}^{-2}$ in early January to $1321\,\text{W}\,\text{m}^{-2}$ in early July, see Problem 9.1.

The solar radiation reaching the Earth will irradiate the Earth disk of size πR_E^2 with $R_E = 6.371 \times 10^6$m the mean radius of the Earth. However, the total surface of the Earth is $4\pi R_E^2$. That means that the average radiation per unit area on Earth, the so-called average *insolation*, is $S_0/4 = 342\,\mathrm{W\,m^{-2}}$.

Because the total energy content of the Earth system is constant on long time scales, the Sun's heat input is compensated by the Earth's heat output; therefore Earth will have to export, on average, $S_0/4 = 342\,\mathrm{W\,m^{-2}}$. About 30% of this export is accounted for by reflection of the incoming short-wave radiation from the clouds and from the Earth's surface. The fraction α of insolation that is reflected is called the *albedo*; the Earth system has an average albedo of $\alpha = 0.3$. The radiation that is not reflected back to space, about $239\,\mathrm{W\,m^{-2}}$ on average, is exported as long-wave radiation. Using Stefan's law, this radiative export can be expressed as the effective temperature of a black body with the same energy output. This effective temperature is called the *radiation temperature* or *bolometric temperature*, T_b. For the Earth we find a radiation temperature of $T_b = 255\,\mathrm{K}$.

The radiation temperature is lower than the observed average surface temperature on Earth, T_E, of about 288 K. This high surface temperature is due to the *greenhouse effect* produced by the atmosphere. The total greenhouse effect on Earth is about $288\,\mathrm{K} - 255\,\mathrm{K} = 33\,\mathrm{K}$. Physically, the atmosphere can be viewed as a long-wave blanket over the Earth. The top of the atmosphere will have to radiate at a temperature of 255 K. To emit at this temperature, the Earth's surface beneath the blanket has to be warmer. This is the greenhouse effect. If the blanket insulates better, the Earth's surface has to become hotter to enable the same radiation temperature at the top of the atmosphere.

In its simplest possible setting, the greenhouse effect can be understood by assuming that the atmosphere is relatively transparent in the short-wave spectrum ($\epsilon_{\mathrm{SW}} \approx 0$) and opaque in the long-wave spectrum ($\epsilon_{\mathrm{LW}} \approx 1$). The schematic in Figure 9.3 shows how the radiation is transmitted in such a system, where we assume the atmosphere to consist of a single, cloud-free slab of uniform temperature T_A. In equilibrium, the energy fluxes into and

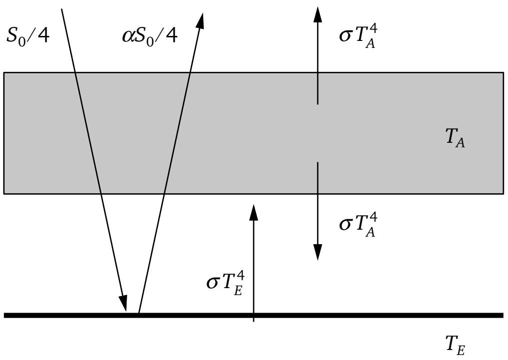


FIGURE 9.3 Paths of radiative fluxes in a system made of a single, cloud-free slab atmosphere at temperature T_A, with unit emissivity in the long-wave and zero emissivity in the short-wave part of the spectrum. The insolation is $S_0/4$. The planetary short-wave albedo α is here provided by the Earth's surface alone.

out of the components of the system must balance. The energy budget can be evaluated for the space surrounding the Earth system at the top of the atmosphere (TOA), in the atmosphere itself (ATM), and at the Earth's surface (SFC). This leads to the following three budget equations:

$$\begin{aligned} & \text{energy output} = \text{energy input} & \\ \text{TOA:}\quad & S_0/4 = \alpha\, S_0/4 + \sigma T_A^4 & (9.20\text{a}) \\ \text{ATM:}\quad & 2\sigma T_A^4 = \sigma T_E^4 & (9.20\text{b}) \\ \text{SFC:}\quad & \sigma T_E^4 = \sigma T_A^4 + (1-\alpha)\, S_0/4. & (9.20\text{c}) \end{aligned}$$

The factor of two in the atmospheric budget corresponds to the fact that the atmosphere radiates to the Earth as well as to space.

In this energy budget, we have three equations with two unknowns but the three budgets are dependent; we can substitute Eq. 9.20b into Eq. 9.20c to find Eq. 9.20a. So effectively we have two equations with two unknowns, T_E and T_A, which have as solution

$$T_A = \left(\frac{(1-\alpha)\, S_0}{4\sigma}\right)^{1/4} \quad \text{and} \quad T_E = 2^{1/4}\, T_A. \tag{9.21}$$

Putting in numbers, we find $T_A = 255\,\text{K}$; indeed, it has to be the same as the previously calculated radiation temperature, T_b, because the atmosphere in this model radiates as a black body to space. For the Earth's surface temperature we find $T_E = 303\,\text{K}$, rather higher than the 255 K calculated without an atmosphere. This value is also higher than observed (about 288 K) because the atmosphere is not completely opaque to long-wave radiation and because there are other heat exchange processes between the Earth's surface and the atmosphere.

It is enlightening to work out several budget scenarios based on simple slab models. One obvious extension of the above model would be to allow for a finite long-wave emissivity, ϵ_{LW}, and for a latent and sensible heat transfer, H, between the Earth's surface and the atmosphere. The budget equations for this modified system are

$$\begin{aligned} \text{TOA:}\quad & S_0/4 = \alpha\, S_0/4 + \epsilon_{\text{LW}}\, \sigma T_A^4 + (1-\epsilon_{\text{LW}})\, \sigma T_E^4 & (9.22\text{a}) \\ \text{ATM:}\quad & 2\epsilon_{\text{LW}}\, \sigma T_A^4 = \epsilon_{\text{LW}}\, \sigma T_E^4 + H & (9.22\text{b}) \\ \text{SFC:}\quad & \sigma T_E^4 + H = \epsilon_{\text{LW}}\, \sigma T_A^4 + (1-\alpha)\, S_0/4. & (9.22\text{c}) \end{aligned}$$

Here it is taken that an atmospheric long-wave emissivity of ϵ_{LW} corresponds to a long-wave transmissivity of $1-\epsilon_{\text{LW}}$. As before, only two of these equations are independent. The solution for T_E is

$$(1-\epsilon_{\text{LW}}/2)\sigma T_E^4 = (1-\alpha)\, S_0/4 - H/2. \tag{9.23}$$

From this equation we can see that the Earth's surface temperature increases with increasing $\epsilon_{\rm LW}$, the greenhouse effect, and with decreasing H.

The greenhouse effect is due to a mix of several greenhouse gases in the atmosphere, gases that absorb strongly in the long-wave part of the spectrum. The most dominant greenhouse gases are H_2O and CO_2. It is important to note that we cannot simply add contributions of each individual gas to find the total long-wave absorption. Most greenhouse gases absorb strongly in certain wavelength bands and when bands of two different gases overlap, the total absorption is less than the sum of the individual absorptions. For this reason it is impossible to uniquely define the contribution of each greenhouse gas to the total. Nonetheless, the water vapour contribution to the total long-wave absorption is about four times larger than the CO_2 contribution.

Increases in these greenhouse gases make the atmosphere more opaque to long-wave radiation (they will increase $\epsilon_{\rm LW}$) and will increase the surface temperature, as in Eq. 9.23. This is called the *enhanced greenhouse effect.* It requires detailed radiative calculations to quantitatively determine the effect of adding a certain amount of CO_2 to the atmosphere. One of the key difficulties comes in estimating the size of the so-called *water vapour feedback*. If CO_2 in the atmosphere increases, the temperature will go up. This will enhance saturation vapour pressure in the atmosphere and its water vapour content. This will then further increase the opacity of the atmosphere.

We can devise a simple model for the water vapour feedback mechanism based on Eq. 9.23. Assuming the latent and sensible heat fluxes remain the same, we can estimate the change in temperature in terms of the change in long-wave emissivity. We find for small variations in $\epsilon_{\rm LW}$

$$4\frac{{\rm d}T_E}{T_E} = \frac{{\rm d}\epsilon_{\rm LW}}{2-\epsilon_{\rm LW}}. \tag{9.24}$$

We now linearize our radiation code so as to express variations in the long-wave emissivity in terms of changes in CO_2 concentration and changes in water vapour concentration. The latter is written as proportional to the saturated vapour pressure at temperature T_E – this assumes that the mean relative humidity does not change, something which is generally confirmed by general circulation models. So we have

$${\rm d}\epsilon_{\rm LW} = P\,{\rm dCO_2} + Q\,{\rm d}e_s(T_E), \tag{9.25}$$

with P and Q some constants that follow from linearizing the total radiation effects. With the Clausius–Clapeyron equation the change in saturated water vapour pressure can be written in terms of the change in surface temperature. Substitution in Eq. 9.24 gives

$$(8-4\epsilon_{\rm LW})\frac{{\rm d}T_E}{T_E} = P\,{\rm dCO_2} + Q\frac{e_s L}{R_v T_E}\frac{{\rm d}T_E}{T_E}. \tag{9.26}$$

Rearranging gives

$$\frac{dT_E}{dCO_2} = \frac{PT_E}{8 - 4\epsilon_{LW}} \left(1 + \frac{x}{1 - x}\right), \tag{9.27}$$

with

$$x = \frac{Qe_s L}{R_v T_E \left(8 - 4\epsilon_{LW}\right)}. \tag{9.28}$$

If the radiative effect of increased water vapour is ignored we have $Q = 0, x = 0$. So the first term on the right-hand side of Eq. 9.27 represents the direct effect of the CO_2 forcing. The second term on the right-hand side represents the water vapour feedback. It is positive and can easily be larger than the direct effect. Note that the linearization constants P and Q depend on the presence of CO_2, water vapour and all the other greenhouse gases in the atmosphere, so the direct effect of CO_2 depends on the presence of water vapour.

The water vapour feedback becomes infinite when $x \rightarrow 1$. This is what is called the *runaway greenhouse effect*. This is a possibility in our model: remember that the saturated vapour pressure increases strongly with temperature. In a runaway state, the surface temperature shoots up until all the water vapour on the planet has evaporated. The present one-layer model is much too simple to accurately estimate the limits of the runaway greenhouse effect. From more comprehensive models it is believed that the Earth cannot get to a runaway state by pumping CO_2 into the atmosphere.

Although didactically enlightening, these simple slab models are ultimately insufficient to accurately describe the energy balance of the atmosphere and to predict useful values for surface or atmospheric temperatures. To achieve any level of accuracy for the radiative part of the description we need to move to comprehensive radiative transfer models. We need to relax the assumption of horizontal homogeneity. We also need to take into account the true emissivity structure of the atmosphere. This structure is complex and depends on molecular absorption spectra and density profiles of absorbers, such as water vapour or ozone, in the atmosphere. This is crucial to understanding the true effect of radiation on the atmosphere and key to understanding the climate change response to changed greenhouse gases. These subjects are covered in texts that are specifically devoted to atmospheric radiation.

Furthermore, these slab models do not generally provide adequate descriptions of the many additional complexities in the atmospheric energy budget. For example, the sensible and latent heat fluxes are not set by any of the variables in the models. We can devise parameterizations. We could for example take the flux from the Earth's surface to the atmosphere to be proportional to their temperature difference. These types of parameterizations, so-called *bulk aerodynamic parameterizations*, are widely used in general circulation models, although the constants of proportionality are usually complex functions of atmospheric parameters, such as vertical wind shear, that are not included in simple slab models. More complex parameterizations may be found in *radiative–convective equilibrium* models. Here the latent and sensible heat fluxes are estimated based on approximations of convective activity in the atmosphere.

Additionally, the Earth system contains many feedback mechanisms. The water vapour feedback, discussed above, is only one of them. Particularly, the albedo is a complex function of the state of the atmosphere. For example, in a warmer world, we generally expect less ice cover of the poles and therefore a reduced albedo in the polar regions. This will lead to enhanced absorption of solar radiation leading to further warming. This is called the *ice–albedo feedback*; it is a positive feedback. This is again a very simplified picture of the ice–albedo feedback: it is not obvious to what extent a warmer world would actually have less ice, because the polar ice cover is a complex function of the ocean currents and atmospheric circulation patterns, both of which will change in a warmer world. Changing cloud cover provides another feedback mechanism for which we do not even know for sure whether it is on average positive or negative. Clouds account for about three-quarters of the albedo of planet Earth[48] and at the same time provide additional long-wave heating of the Earth's surface. It is one of the great challenges of climate science to determine how these competing effects will change in a changing climate.

9.4 HORIZONTAL VARIATIONS

In the previous section we examined simple slab models for the global energy budget by taking global mean values of all relevant parameters, such as insolation and albedo. The actual value of insolation is a function of latitude, season, and time of day. At any time, the insolation S equals

$$S = S_0(\langle r_E\rangle / r_E(t))^2 \cos\theta, \tag{9.29}$$

with $r_E(t)$ the distance from the Earth to the Sun as a function of time, $\langle r_E\rangle$ the average distance from the Earth to the Sun, and θ the *zenith angle*, that is, the angle between the rays of the Sun and the local vertical. The $\cos\theta$ factor accounts for the projection of the incident rays onto the surface.

With this equation, the insolation averaged over a day can be calculated as a function of latitude and season; this requires standard astronomical equations for the declination of the Sun (the angle between the rays of the Sun and the plane of the Earth's equator) and the distance between the Earth and the Sun as a function of day number. Figure 9.4 shows the result of this calculation. It can be seen that the insolation vanishes in the polar nights. The maximum insolation is achieved at the poles in the summer hemispheres. This is due to the fact that, although the solar zenith angle increases when going poleward, the length of the day also increases in the summer hemisphere. In the winter hemisphere there is a strong insolation contrast between the equator and the

---

[48] See Trenberth, K. E., Fasullo, J. T. and Kiehl, J. (2009) *Bull. Am. Meteor. Soc.* **90**, 311–324.

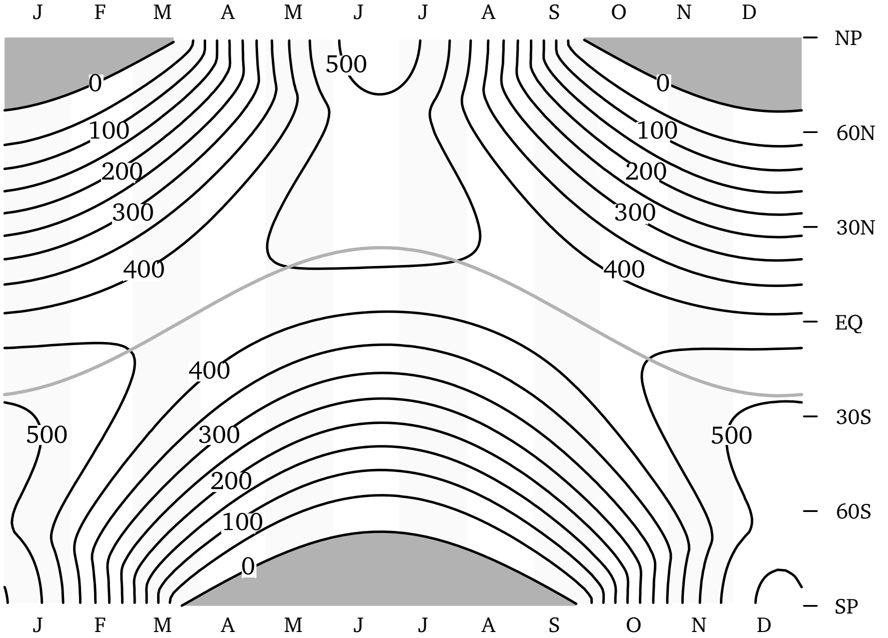


FIGURE 9.4 Insolation for current astronomical conditions averaged over a day as a function of season and latitude. The contour interval is $50\,\mathrm{W\,m^{-2}}$. The shaded areas have vanishing insolation. The grey line indicates the declination of the Sun as a function of season.

pole; in the summer hemisphere there is a much weaker, indeed reversed, insolation contrast.

The short-wave albedo is the fraction of incoming short-wave radiation that is reflected. It is also a strong function of local conditions such as cloud cover, cloud type, surface type, solar zenith angle. The average short-wave albedo for Earth is $\alpha \approx 0.3$. Figure 9.5 shows the average albedo of the Earth as a function of geographical location for the year 1986. It is determined from satellite measurements of the outgoing radiation. As can be seen in the figure, the albedo increases strongly from the tropics ($\alpha \approx 0.2$) to the poles ($\alpha \approx 0.6$). This is partly due to the increased average zenith angle going towards the poles, and partly due to the presence of snow and ice at high latitudes. The regions of lowest albedo correspond to relatively clear regions over the tropics.

The combined effect of the varying solar insolation and the increased albedo towards the higher latitudes implies that the absorbed solar radiation at the Earth's surface is larger at the equator (about $300\,\mathrm{W\,m^{-2}}$ on average) than at the poles (about $100\,\mathrm{W\,m^{-2}}$ on average), see Figure 9.5; the largest contrast is achieved in the winter hemisphere. This contrast in absorbed radiation sets up a temperature gradient from pole to equator and a subsequent build-up of potential energy. The build-up of potential energy makes the atmosphere unstable and this results in poleward transport of heat. The oceans are driven by varying heat inputs and the atmospheric winds. They

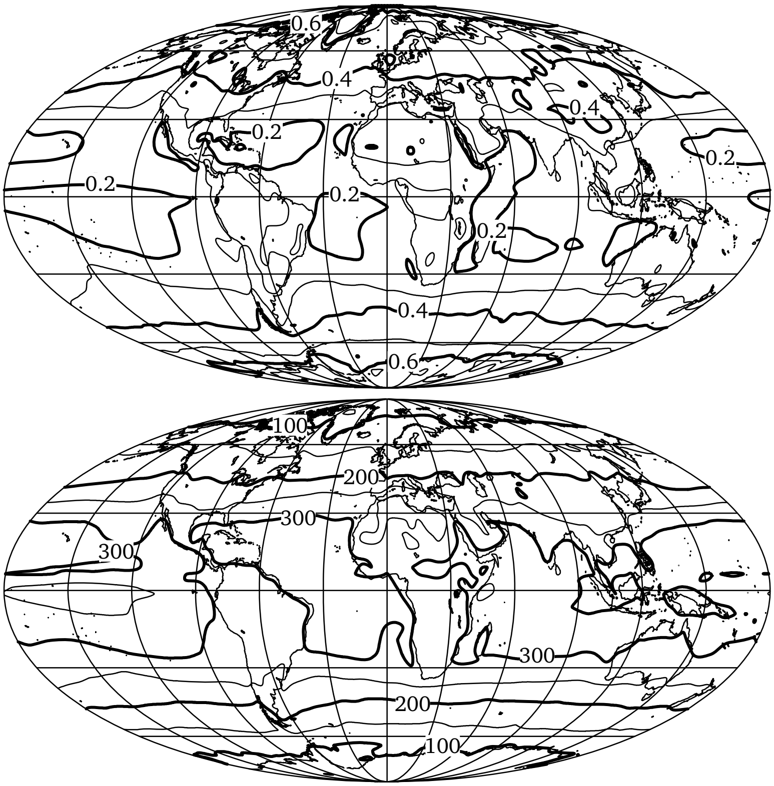


FIGURE 9.5 Annual (1986) mean albedo (top panel) and absorbed short-wave radiation in W m^{-2} (bottom panel), as determined from ERBE satellite measurements. ERBE data provided by Marc Michelsen, University of Washington.

also transport heat poleward. Figure 9.6 shows the observed northward transport of heat by the atmosphere and oceans as a function of latitude, averaged over a year. This transport of heat tends to reduce the temperature contrast between the poles and the equator and releases the generated potential energy. The overall structure of this transport is called the *general circulation.* All motions in the atmosphere ultimately have their origin in the contrast in absorbed solar radiation between equator and poles.[49]

9.5 RADIATIVE INTENSITY

Up to now we have considered systems where radiative properties only vary in the vertical and we have only been concerned with upward or downward transport of radiative energy, as given by the radiative flux. However, the flux through some surface does not contain information on the direction of the radiation. In general we want to consider the direction of the radiation as well.

---

[49] For further reading, see James, I. N. (1995) *Introduction to circulating atmospheres*. Cambridge University Press, Cambridge.

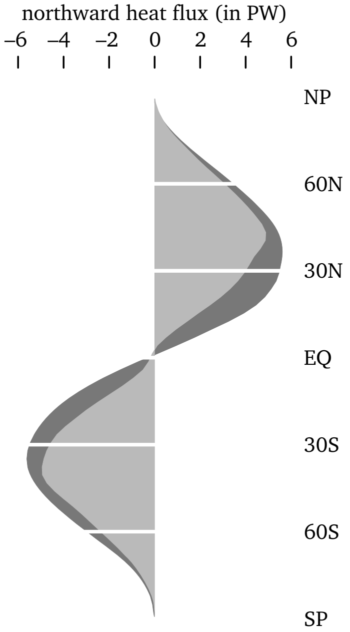


FIGURE 9.6 Annual average northward heat transport, in petawatts, by the atmosphere (light grey) and oceans plus atmosphere (dark grey) as a function of latitude. After Trenberth, K. E. & Caron, J. M. (2001) *J. Clim.* **14**, 3433–3443.

Consider a unit sphere drawn around a fixed point. Any radiation coming from that point will pass through the unit sphere in a location which can be identified with the usual zenith angle (θ) and azimuth angle (ϕ). If we identify a small solid angle $\delta\Omega = \sin\theta\,\delta\phi\,\delta\theta$ around this direction we can identify a small part δF of the radiative flux that passes through this solid angle. The *radiative intensity I* is now defined as the flux per unit solid angle in the given direction. In the limit of $\delta\Omega \to 0$ this is written as

$$I = \frac{\mathrm{d}F}{\mathrm{d}\Omega}. \tag{9.30}$$

The units of flux, F, are $\mathrm{W\,m^{-2}}$, so the units of intensity, I, are $\mathrm{W\,m^{-2}\,sr^{-1}}$ with solid angles measured in steradians (sr). The intensity is normally a function of the direction. *Isotropic radiation* is constant in all directions. Black body emission is an example of isotropic radiation.

The flux and the intensity are related. Let us calculate the ‘upward’ flux due to radiation of intensity $I(\phi,\theta)$; that is, the flux through a surface that is perpendicular to the upward direction $\theta = 0$. The upward flux F is the integral of the upward component of the intensity I over all directions. The upward component of $I(\phi,\theta)$ is $I(\phi,\theta)\cos\theta$. The upward flux F therefore is

$$F = \int_0^{2\pi}\int_0^{\pi/2} I(\phi,\theta)\,\cos\theta\,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi. \tag{9.31}$$

We only integrate over zenith angles θ up to $\pi/2$, as larger zenith angles correspond to downward radiation. We can generalize the above expression for flux in any direction; that is, flux through surfaces with any orientation. For isotropic radiation, I is not a function of direction and we can readily evaluate the double integral. We find

$$F = \pi I. \tag{9.32}$$

As black body radiation is isotropic, the Planck law, Eq. 9.6, can also be expressed in terms of radiative intensity. The difference between the expressions for black body intensity and flux is a factor π, as above, but remember that the black body intensity has units of $\mathrm{W\,m^{-2}\,sr^{-1}}$.

The solar constant S_0 can be related to the radiation intensity coming from the Sun. If the radiative flux at the surface of the Sun is σT_S^4, then its intensity will be $I = \sigma T_S^4/\pi$, assuming isotropic radiation. The Sun subtends a solid angle of about $\pi R_S^2/r_E^2$ in the sky. The flux hitting the surface normal to the solar radiation will therefore be $\sigma T_S^4\, R_S^2/r_E^2$, which is the definition of the solar constant S_0, Eq. 9.18. This calculation is only valid for small values of R_S/r_E. A more accurate calculation demonstrates the equivalent result for finite values of R_S/r_E if the radiating sphere is directly overhead. From the geometry in Fig 9.7 we can see that the flux hitting the surface from isotropic radiation from a sphere equals

$$F = \int_0^{2\pi} \int_0^{\sin^{-1} R_S/r_E} I \cos\theta\, \sin\theta \,\mathrm{d}\theta\,\mathrm{d}\phi, \tag{9.33}$$

where $I\cos\theta$ is the intensity component perpendicular to the surface and the integral is over the solid angle subtended by the sphere. This integral is

$$F = \pi I\, R_S^2/r_E^2. \tag{9.34}$$

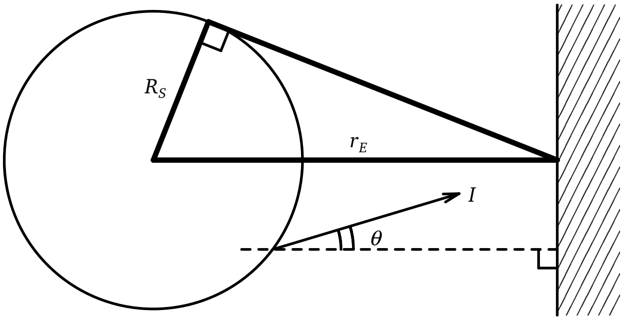


FIGURE 9.7 Geometry used to calculate the flux from a radiating sphere of radius R_S at a distance r_E.

For the Sun we have $\pi I = \sigma T_S^4$ so we arrive again at the definition of the solar constant.

9.6 RADIATIVE TRANSFER

A layer of air will absorb a certain fraction of the incoming radiation. The rate of absorption will depend on the wavelength. For an infinitesimally thin layer of air, the change in the intensity of a beam passing through the layer can be written as

$$\mathrm{d}I_\lambda = -I_\lambda \, \mathrm{d}\delta_\lambda, \tag{9.35}$$

where we have introduced the infinitesimal *optical depth* $\mathrm{d}\delta$. The optical depth depends on the wavelength. The right-hand side is negative because the intensity decreases through absorption. This equation can be integrated through a layer of finite depth to find a relation between the incoming intensity I_{in} and the outgoing intensity I_{out},

► $$I_{\mathrm{out}} = I_{\mathrm{in}} \, e^{-\delta}, \tag{9.36}$$

where δ is the total optical depth of the layer and where subscripts λ have been omitted for brevity. Note that optical depth is not a physical 'depth' (it has no units); it defines the fraction of the incoming radiation being absorbed along a path.

A layer of air will have a density of absorbers $\tilde{n}$ and each absorber will have an absorptivity, expressed by its molecular absorption cross-section σ_λ. The absorption cross-section measures the area of an incoming beam of wavelength λ that would be taken out by absorption by a single molecule in its path. The optical depth δ is related to the number density of absorbers and their absorption cross-section by the *Beer–Lambert law* which states that

► $$\delta_\lambda = \tilde{n}\sigma_\lambda l, \tag{9.37}$$

with l the geometric depth of the layer. It is trivial to extend this law for a mixture of various absorbing species. This law is strictly only valid for low densities of absorbers: for higher densities, the absorbing molecules will sit in each-other's 'shade' and will thus contribute less to the total absorption as expressed by the optical depth.

In the atmosphere, the number density of absorbers will often be a function of location. In such situations, the Beer–Lambert law is generalized to

► $$\delta_\lambda = \int \tilde{n}\sigma_\lambda \, \mathrm{d}l, \tag{9.38}$$

where the integral is now along the path of the beam. In differential form, this becomes

$$\mathrm{d}\delta_\lambda = \tilde{n}\sigma_\lambda \, \mathrm{d}l. \tag{9.39}$$

This expression of the Beer–Lambert law can be used to rewrite Eq. 9.35 in terms of geometric depth; the Beer–Lambert law provides the transformation between geometric depth and optical depth. Note that this transformation is different for different wavelengths.

It is of interest to consider a vertically directed beam. First, the number density $\tilde{n}$ is rewritten as a concentration by mass c using

$$c = \tilde{n}M_1/\rho, \tag{9.40}$$

with M_1 the molecular mass of the absorber and ρ the local density of the air. Substituting this expression in the Beer–Lambert law for a vertically directed beam, for which $\mathrm{d}l$ is replaced by $\mathrm{d}z$, we find

$$\delta_\lambda = \int \frac{c\sigma_\lambda}{M} \rho \, \mathrm{d}z. \tag{9.41}$$

Now using hydrostatic balance, $\rho \, \mathrm{d}z = -\mathrm{d}p/g$ we find the Beer–Lambert law for vertical beams

► $$\delta_\lambda = \int_{p_0}^{p_1} \frac{c\sigma_\lambda}{M_1 g} \, \mathrm{d}p. \tag{9.42}$$

For beams that come in at a zenith angle θ from the vertical, we need to divide this expression by $\cos\theta$ – see below. The relevance of the last expression is that for a well-mixed absorber, c is constant and we find that the optical thickness along a path is proportional to the pressure difference along the path.

We typically study the optical depth over a band of wavelengths. Indeed, we will only discuss rather crude wide-band approximations in this section. In these situations the width of the individual absorption lines is important to the total absorption in a layer. The width of an absorption line increases with pressure due to the collisions between molecules, an effect called *pressure broadening*. At increased width, the absorption cross-section decreases in the centre of the line and increases in the tails of the line. However, for strong absorption lines the extinction, as expressed by Eq. 9.36, will remain strong at the centre of the line, while it will appreciably increase in the tails of the line. The effective optical depth integrated over the wavelength band will therefore increase with increased line width. A reasonable way to account for the pressure broadening is to use an effective absorption cross-section [50]

$$\sigma = \sigma_r \, (p/p_r)^{\kappa}, \tag{9.43}$$

with κ sometimes taken equal to 1 but, more accurately, smaller than 1, and σ_r the effective absorption cross-section over the wavelength band at reference pressure p_r. Using a value of $\kappa = 1$ we can show that the optical thickness between two pressure levels for a well-mixed absorber can be written as

$$\delta = \delta_r \frac{(p_1 - p_0)(p_1 + p_0)/2}{p_r^2}, \tag{9.44}$$

with

$$\delta_r = \frac{c\sigma_r p_r}{M_1 g}. \tag{9.45}$$

In other words, for small pressure differences the optical thickness is proportional to the pressure difference but for larger pressure differences the optical thickness increases at a slower rate.

The factor exp $(-\delta)$ in Eq. 9.36 is called the *transmittance* of the layer: it is the fraction of radiation that is transmitted through the layer and is sometimes given the symbol τ,[51]

$$\tau = \exp(-\delta). \tag{9.46}$$

Like the optical depth, transmittance generally depends on the wavelength. If the radiation traverses two layers of air with transmittances τ_1 and τ_2, the combined transmittance τ_{1+2} equals

$$\tau_{1+2} = \tau_1 \tau_2. \tag{9.47}$$

Figure 9.8 illustrates the set up. According to the Beer–Lambert law, the combined optical depth δ_{1+2} of two adjacent layers is the sum of the individual optical depths of the layers,

$$\delta_{1+2} = \delta_1 + \delta_2. \tag{9.48}$$

This is consistent with the formal relation between transmittance and optical depth, Eq. 9.46. This property of the transmittance can be used to work out refined versions of the atmospheric slab models in the previous section, with several layers of finite transmittance and where the emittance/absorptance ϵ of a layer is the complement of its transmittance $\tau = 1 - \epsilon$, as illustrated in Figure 9.8.

---

[50] See Manabe, S. & Strickler, R. F. (1964) *J. Atmos. Sci.* **21**, 361–385; Manabe, S. & Wetherald, R. T. (1967) *J. Atmos. Sci.* **24**, 241–259

[51] Optical depth is also often given the symbol τ.

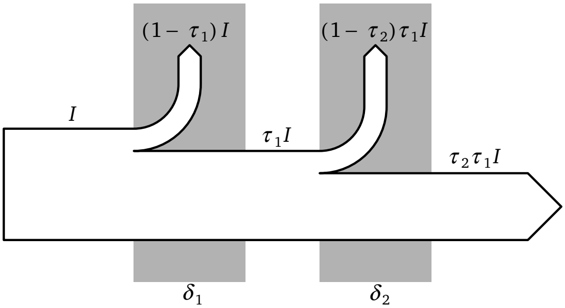


FIGURE 9.8 Transmittance of radiation through two adjacent layers of optical depths δ_1 and δ_2. The transmittance of each layer is $\tau_i = \exp(-\delta_i)$.

A parcel of air will also emit long-wave radiation. By Kirchhoff's law, the emissivity of air is the same as its absorptivity. This means we can use the Beer–Lambert law to find that the emissivity along an infinitesimal path dl equals $d\delta_\lambda = \tilde{n}\sigma_\lambda \, dl$. So the radiation budget for an infinitesimal path is

$$dI_\lambda = (I_{B,\lambda}(T) - I_\lambda)\, d\delta_\lambda, \tag{9.49}$$

with $I_{B,\lambda}$ the black body intensity for temperature T (remember, $I_B = B/\pi$). The above expression is usually written as the *Schwarzschild equation,*

▶ $$\frac{dI}{d\delta} = I_B - I, \tag{9.50}$$

where we have again omitted the subscript λ for brevity. The formal solution for the Schwarzschild equation is

$$I(\delta) = I(0)\, e^{-\delta} + \int_0^\delta I_B(\delta - \delta')\, e^{-\delta'}\, d\delta'. \tag{9.51}$$

The first term on the right-hand side is the transmitted radiation through a layer of optical depth δ. The integrand in the second term is the contribution of the black body radiation emitted by the atmosphere at an optical depth δ' away from the present location. This equation can also be written in terms of physical path length, by using the Beer–Lambert law.

An application of the above formalism is *Langley's method,* a method to infer insolation at the top of the atmosphere by surface measurements only. Consider a horizontally uniform atmosphere with optical depth δ_0 at zero zenith angle. A beam of intensity I_{in} hitting the atmosphere at a zenith angle θ will then experience an optical depth of δ with

$$\delta = \frac{\delta_0}{\cos\theta}. \tag{9.52}$$

The geometry of this situation is illustrated in Figure 9.9.

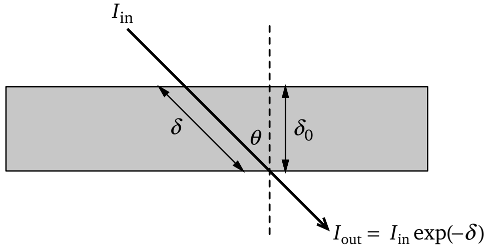


FIGURE 9.9 Relationship between path optical depth, δ, layer optical depth, δ_0, and zenith angle, θ.

At the surface we can measure the *beam*, S_b, which is the radiative energy flux incident on a plane normal to the direction of the incident solar radiation. This flux is to be measured in a narrow waveband where we expect the atmospheric absorption to be more-or-less constant as a function of measured wavelength. From the geometry in Figure 9.9 (with I_{in} replaced by S_0 and I_{out} replaced by S_b) it becomes clear that the measured S_b is related to the solar irradiance at the top of the atmosphere S_0 and the optical depth δ by

$$S_b = S_0 \exp(-\delta) = S_0 \exp(-\delta_0/\cos\theta), \tag{9.53}$$

with, as before, δ_0 the optical thickness of the atmosphere at a zero zenith angle.

When making surface measurements, we have two unknowns δ_0 and S_0. From standard astronomical equations, or even direct measurement, we can infer the solar zenith angle with great accuracy. If we measure the beam S_b at different zenith angles, but at constant normal optical thickness δ_0, we can statistically fit the above equation to the measurements and thus estimate both δ_0 and S_0. Typically, the measurements take place after sunrise or before sunset on clear days where the optical atmospheric conditions do not change during the measurement period. For the statistical fit we can divide the above equation by some fixed reference irradiance S_r and take the logarithm. The resulting equation can be rearranged to

$$\cos\theta \ln(S_b/S_r) = \cos\theta \ln(S_0/S_r) - \delta_0 \tag{9.54}$$

If we now plot $\cos\theta \ln(S_b/S_r)$ versus $\cos\theta$ we find a straight line with slope $\ln(S_0/S_r)$ and intercept $-\delta_0$.[52]

This technique only works accurately in narrow bands where the atmosphere is relatively transparent. For wide-band measurements the absorption varies substantially with measured wavelength and the dependence of the

---

[52] The value of the reference irradiance S_r does not modify the outcome of the fit but does change the graphical representation of the data. The default choice of 1 W m^{-2} misrepresents the scatter in the data; it is better to choose S_r to be similar to S_0 or typical values of S_b.

measured flux on the zenith angle is different for these different absorptions. It is therefore impossible to fit a single optical depth to these measurements. Furthermore, wide-band measurements will usually also measure in bands where the atmosphere is practically opaque. That means that for these wavelengths we do not measure any flux and we cannot therefore infer the incoming flux at the top of the atmosphere.

In the discussion of absorption of radiation we have ignored the effect of scattering. Scattering is the change of direction of a beam of radiation. Attenuation by scattering for a single beam can be modeled largely, like absorption, with the Beer–Lambert law, and including density of scatterers and the scattering coefficient for a single molecule. Scattering not only attenuates a beam: the scattered radiation from other beams may contribute to the intensity of the beam under consideration. Scattering in the atmosphere is the dominant process of extinction in clouds or aerosol regions.

The physical origins of scattering are manifold. We typically discriminate between *Rayleigh scattering*, where the wavelength of the radiation is much larger than the scatterers and *Mie scattering* where the wavelength of the radiation is of comparable size to or smaller than the scatterers.

Rayleigh scattering is due to the electric dipole moment of scatterers interacting with the electromagnetic field. Rayleigh scattering is strongly wavelength dependent, with the scattering cross-section (the effectivity of the scattering) typically proportional to λ^{-4}. Molecules can Rayleigh-scatter visible light. Due to the wavelength dependence of the scattering cross-section, blue light is scattered more strongly than red light. This is the origin of the blue colour of the sky, the yellow colour of the Sun, and the orange colour of the sunset. Cloud drops can Rayleigh-scatter microwaves, which is the physical basis of cloud radar.

Unlike Rayleigh scattering, Mie scattering is highly directional, with most radiation being scattered in the forward direction. Scattering of sunlight by cloud drops is an example of Mie scattering. So Mie scattering is the reason we can actually see clouds. The fact that most scattering is in the forward direction makes the clouds white, unless they are deep. Photons in clouds typically experience multiple scattering events and in most clouds photons are backscattered enough to make clouds white when viewed from the top.

9.7 RADIATIVE–CONVECTIVE EQUILIBRIUM

It is of interest to consider the radiative transfer problem in an atmosphere which is uniform in the horizontal. We are then only interested in the vertical flux of radiation in either the upward or downward direction. We will consider a wide-band approximation, with the long-wave and short-wave fluxes in separate bands. As a vertical coordinate we will use the effective long-wave optical depth δ between the surface and some level in the atmosphere. It is related to the geometrical height z by

$$d\delta = e_{\mathrm{LW}}\, dz. \tag{9.55}$$

Here, e_{LW} is an effective fractional emissivity in the long-wave part of the spectrum. So the atmosphere is assumed to be a grey body in the long-wave part. At the Earth's surface we have $\delta = 0$ and at the top of the atmosphere we set $\delta = \delta_{\mathrm{TOA}}$.

The Schwarzschild equations, Eq. 9.50, for the upward long-wave flux $L_\uparrow$ and the downward long-wave flux $L_\downarrow$ are

$$dL_\uparrow/d\delta = B - L_\uparrow, \tag{9.56a}$$
$$-dL_\downarrow/d\delta = B - L_\downarrow. \tag{9.56b}$$

Here B is the usual black body flux, which, for atmospheric temperatures, is mostly in the long-wave part of the spectrum. The Schwarzschild equations for the shortwave fluxes $S_\uparrow$ and $S_\downarrow$ are

$$dS_\uparrow/d\delta = -(e_{\mathrm{SW}}/e_{\mathrm{LW}})\, S_\uparrow, \tag{9.57a}$$
$$-dS_\downarrow/d\delta = -(e_{\mathrm{SW}}/e_{\mathrm{LW}})\, S_\downarrow. \tag{9.57b}$$

The black body emission by the atmosphere does not contribute to the short-wave part of the fluxes. The pre-factor $e_{\mathrm{SW}}/e_{\mathrm{LW}}$ is ratio of the short-wave optical depth and the long-wave optical depth, δ.

Radiative equilibrium follows from combining a steady state condition with the radiative transfer equations. The steady state condition follows by setting the local heating rate to zero. The local heating is given by the convergence of the radiative fluxes and any enthalpy fluxes, which we will denote by H. These enthalpy fluxes are normally associated with convective motion. The steady state condition then is

$$\frac{d}{dz}(L_n + S_n + H) = 0, \tag{9.58}$$

where the net upward long-wave L_n and short-wave S_n fluxes are

$$L_n = L_\uparrow - L_\downarrow, \quad S_n = S_\uparrow - S_\downarrow. \tag{9.59}$$

In a steady state the total upward heat flux $L_n + S_n + H$ therefore must be constant. We know that at the top of the atmosphere the upward heat flux must vanish, so this constant is zero. We therefore find

$$\blacktriangleright \qquad L_n + S_n + H = 0. \tag{9.60}$$

Taking the derivative of this equation with respect to the long-wave optical depth and substituting the Schwarzschild equations for the long-wave components we find

$$2B - L_{\downarrow} - L_{\uparrow} - \rho\dot{q}_{\mathrm{SW}}/\epsilon_{\mathrm{LW}} + \mathrm{d}H/\mathrm{d}\delta = 0. \tag{9.61}$$

Here we have rewritten the derivative of the net upward short-wave component in terms of the short-wave heating $\dot{q}_{\mathrm{SW}}$; that is,

$$\rho\dot{q}_{\mathrm{SW}} = -\mathrm{d}S_n/\mathrm{d}z = -\epsilon_{\mathrm{LW}}\mathrm{d}S_n/\mathrm{d}\delta. \tag{9.62}$$

We can take a further derivative with respect to the optical depth and again use the Schwarzschild equations for the long-wave components to find

▶ $$2\frac{\mathrm{d}B}{\mathrm{d}\delta} + L_n - \frac{\mathrm{d}}{\mathrm{d}\delta}\left(\frac{\rho\dot{q}_{\mathrm{SW}}}{\epsilon_{\mathrm{LW}}}\right) + \frac{\mathrm{d}^2H}{\mathrm{d}\delta^2} = 0. \tag{9.63}$$

This equation expresses the temperature lapse rate with optical depth in terms of the radiative and enthalpy fluxes. It is straightforward to verify that the temperature lapse rate with geometric height $\Gamma = -\mathrm{d}T/\mathrm{d}z$ is related to $\mathrm{d}B/\mathrm{d}\delta$ by

$$\mathrm{d}B/\mathrm{d}\delta = -4\sigma T^3\Gamma/\epsilon_{\mathrm{LW}}. \tag{9.64}$$

We can now consider the implications of these budget equations for the atmospheric profile.

First consider the case where the atmosphere is transparent to short-wave radiation and there are no enthalpy fluxes, $H = 0$. There is no convergence of the short-wave flux, $\mathrm{d}S_n/\mathrm{d}z = 0$, as there is no black body contribution to the short-wave flux. The equilibrium condition, Eq. 9.60, then becomes

$$L_n = -S_n = \text{constant} = \sigma T_b^4, \tag{9.65}$$

where we have expressed the constant net upward short-wave flux in terms of the radiation temperature of the planet ($T_b = 255\,\mathrm{K}$ in the case of the Earth). We conclude that at radiative equilibrium

$$2\frac{\mathrm{d}B}{\mathrm{d}\delta} = -\sigma T_b^4. \tag{9.66}$$

This equation can be integrated with the integration constant set by applying Eq. 9.61 at the top of the atmosphere: $2B(\mathrm{TOA}) = \sigma T_b^4$. The solution is

▶ $$T = T_b\left(\frac{1 + \delta_{\mathrm{TOA}} - \delta}{2}\right)^{1/4}. \tag{9.67}$$

The surface temperature T_E in this radiative equilibrium model equals

$$T_E = T_b \left(\frac{1 + \delta_{\mathrm{TOA}}}{2} \right)^{1/4}. \tag{9.68}$$

The above set-up gives a more complete description of the greenhouse effect: in radiative equilibrium the temperature increases going downward into the atmosphere. The more optically thick the atmosphere is, the larger this temperature increase is. This explains the high surface temperatures of Venus: with its high albedo, Venus has a relatively low radiation temperature ($T_b \approx 227\,\mathrm{K}$) but the Venusian atmosphere is optically so thick that its surface temperature can go above 700 K.

If we take the long-wave absorbers to be well mixed, as is the case for CO_2, we can use Eq. 9.44 to express the optical depth coordinate in terms of pressure. Taking p_0 the surface pressure, then the optical depth from the surface to some pressure p is

$$\delta = \delta_{\mathrm{TOA}} \frac{(p_0 + p)(p_0 - p)}{p_0^2}. \tag{9.69}$$

Indeed, at the surface $p = p_0$, the optical depth vanishes and at the top of the atmosphere $p = 0$ the optical depth becomes δ_{TOA}. So in terms of pressure, the radiative equilibrium profile, Eq. 9.67, becomes

$$T = T_b \left(\frac{1 + \delta_{\mathrm{TOA}} p^2 / p_0^2}{2} \right)^{1/4}. \tag{9.70}$$

Figure 9.10 shows a typical profile for radiative equilibrium with Earth-like parameters. Its lapse rate near the surface is so large it is convectively unstable. We need convective enthalpy fluxes to get more realistic profiles.[53]

Next we will introduce an enthalpy flux, H, while the atmosphere remains transparent to short-wave radiation. The equilibrium condition, Eq. 9.60, then becomes

$$L_n = -S_n - H = \sigma T_b^4 - H. \tag{9.71}$$

Substituting in Eq. 9.63 we find

$$2 \frac{\mathrm{d}B}{\mathrm{d}\delta} = -\sigma T_b^4 + H - \frac{\mathrm{d}^2 H}{\mathrm{d}\delta^2}. \tag{9.72}$$

---

[53] For more comprehensive model calculations of radiative-convective equilibrium, see Manabe, S. & Wetherald, R. T. (1967) *J. Atmos. Sci.* **24**, 241–259; Ramanathan V. & Coakley, J. A. (1978) *Rev. Geoph. Space Phys.* **16**, 465–489

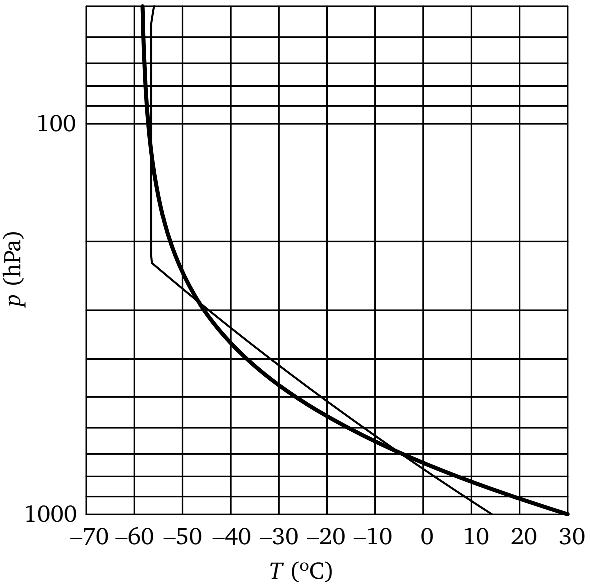


FIGURE 9.10 Radiative equilibrium profile (thick line) as predicted from Eq. 9.70 using an long-wave optical depth of $\delta_{\mathrm{TOA}} = 3.0$. The thin line shows the US Standard Atmosphere, 1976 for comparison.

Integrating this equation between δ and δ_{TOA} and using the boundary condition $2B(\mathrm{TOA}) = \sigma T_b^4$ (from Eq. 9.61) we find the solution

$$2\sigma T^4 = \sigma T_b^4\,(1 + \delta_{\mathrm{TOA}} - \delta) + \frac{\mathrm{d}H}{\mathrm{d}\delta} - \int_{\delta}^{\delta_{\mathrm{TOA}}} H\,\mathrm{d}\delta. \tag{9.73}$$

For $H = 0$ this equation reduces to Eq. 9.67. The surface temperature T_E in this model is given by

$$2\sigma T_E^4 = \sigma T_b^4\,(1 + \delta_{\mathrm{TOA}}) + \frac{\mathrm{d}H_{\mathrm{sfc}}}{\mathrm{d}\delta} - \int_{0}^{\delta_{\mathrm{TOA}}} H\,\mathrm{d}\delta. \tag{9.74}$$

Upward enthalpy fluxes act to cool the surface: the second term on the right-hand side is opposite to the local heating rate at the surface due to the enthalpy fluxes. Because enthalpy fluxes tend to be maximal at the surface, this term tends to be negative. The third term is also always negative for upward fluxes.

It is instructive to write Eq. 9.72 in terms of a temperature lapse rate by using Eq. 9.64. The lapse rate is given by

$$\frac{8\sigma T^3}{\epsilon_{\mathrm{LW}}}\Gamma = \sigma T_b^4 - H + \frac{\mathrm{d}^2 H}{\mathrm{d}\delta^2}. \tag{9.75}$$

From this it becomes clear how the presence of enthalpy fluxes reduces the lapse rate. Indeed, the radiative equilibrium profile of Eq. 9.67 is convectively unstable and the ensuing enthalpy fluxes act to reduce the lapse rate. Strictly speaking, the enthalpy fluxes act to reduce the combination $T^3\varGamma$, but this is generally equivalent to a reduction in $\varGamma$.

To account for the negative lapse rates of the stratosphere we need to take into account short-wave absorption. The ozone layer is the main contributor to short-wave absorption in the stratosphere. The heating due to the short-wave absorption, $\dot{q}_{\mathrm{SW}}$, follows from the convergence of the net downward shortwave flux as in Eq. 9.62. For simplicity we set the enthalpy flux H to zero (this is a good approximation in the stratosphere). In this case the equilibrium equation, Eq. 9.58, and the lapse rate equation, Eq. 9.63, become

$$\frac{\mathrm{d}L_n}{\mathrm{d}\delta} = \frac{\rho \dot{q}_{\mathrm{SW}}}{\epsilon_{\mathrm{LW}}}, \tag{9.76a}$$

$$2\frac{\mathrm{d}B}{\mathrm{d}\delta} = -L_n + \frac{\mathrm{d}}{\mathrm{d}\delta}\left(\frac{\rho \dot{q}_{\mathrm{SW}}}{\epsilon_{\mathrm{LW}}}\right). \tag{9.76b}$$

The first equation expresses how the heating by the shortwave absorption is compensated by a long-wave cooling. The second equation expresses how the lapse rate (expressed in terms of $\mathrm{d}B/\mathrm{d}\delta$) is modified: although the net downward long-wave flux remains negative, the lapse rate is reduced below regions of short-wave absorption where the second term on the right-hand side of Eq. 9.76b is positive.

The lapse rate below a region of short-wave heating is lower than the lapse rate above a region of short-wave heating. To demonstrate this, take a further derivative of Eq. 9.76b with respect to the optical depth to find

$$2\frac{\mathrm{d}^2B}{\mathrm{d}\delta^2} = -\frac{\rho \dot{q}_{\mathrm{SW}}}{\epsilon_{\mathrm{LW}}} + \frac{\mathrm{d}^2}{\mathrm{d}\delta^2}\left(\frac{\rho \dot{q}_{\mathrm{SW}}}{\epsilon_{\mathrm{LW}}}\right). \tag{9.77}$$

Now for simplicity assume the short-wave heating is confined between levels at optical depths δ_0 and δ_1. We choose these bounding levels to be just outside the region of short-wave heating itself. Integrating the above equation between these two levels we find

$$2\frac{\mathrm{d}B_1}{\mathrm{d}\delta} - 2\frac{\mathrm{d}B_0}{\mathrm{d}\delta} = -\int_{z_0}^{z_1} \rho \dot{q}_{\mathrm{SW}}\,\mathrm{d}z. \tag{9.78}$$

Here we changed variables in the integral by noting that $\epsilon_{\mathrm{LW}}\mathrm{d}z = \mathrm{d}\delta$. The right-hand side of this equation is negative. Now using Eq. 9.64 to express $\mathrm{d}B/\mathrm{d}\delta$ in term of lapse rate $\varGamma$, we find that the lapse rate at level 1 is larger than the lapse rate at level 0. Outside this region of short-wave absorption, the lapse rate is set by Eq. 9.66. Above the region the net upward long-wave flux is still σT_b^4. This means that the introduction of a single region of short-wave absorption in a radiative equilibrium model reduces the lapse rate *below* that region. This reduced lapse rate effectively defines the stratosphere.

9.8 THERMODYNAMICS OF A PHOTON GAS

Thermodynamics can be applied to a gas of photons, particles that mediate the electromagnetic field. In this section we will derive the Stefan–Boltzmann law and the Wien displacement law from thermodynamic arguments only.[54]

In Chapter 1 we used a kinematic argument to derive how the pressure in a gas can be written as an average of the momentum transfer in the direction of a wall over all particles that collide with the wall, see Eq. 1.11. For a photon gas this equation needs to be reinterpreted: despite being massless, a photon does have momentum. We can derive an equivalent version of Eq. 1.11 where we replace $M_1\mathscr{U}$ by P_x, the momentum of the particle in the direction of the wall. We then find for the pressure p

$$p = \tilde{n}\,\langle \mathscr{U} P_x \rangle, \tag{9.79}$$

with, as before, $\tilde{n}$ the volumetric particle number density and $\mathscr{U}$ the particle velocity in the direction of the wall. Because the particles on average have no directional preference we could have used any velocity and momentum component in the above equation. We can therefore rewrite the pressure as the average of the contributions from all directions,

$$p = \frac{1}{3}\tilde{n}\,\langle \boldsymbol{U} \cdot \boldsymbol{P} \rangle, \tag{9.80}$$

with $\boldsymbol{U}$ the vector velocity of the particle and $\boldsymbol{P}$ its vector momentum.

A photon moves with the speed of light, c, and has momentum of magnitude P, which is related to the energy e of the photon by

$$e = cP. \tag{9.81}$$

We can substitute this in the above kinematic expression to find the pressure p of the photon gas,

$$p = \frac{1}{3}\tilde{n}\langle cP \rangle = \frac{1}{3}\tilde{n}\langle e \rangle. \tag{9.82}$$

---

[54] The present derivation is partly based on Adkins, C. J. (1983) *Equilibrium thermodynamics*, 3rd edn. Cambridge University Press, Cambridge.

The volumetric energy density $\tilde{u}$ of the photon gas can be written as $\tilde{u} = \tilde{n}\langle e\rangle$. So we arrive at

$$\blacktriangleright \quad p = \tilde{u}/3, \tag{9.83}$$

which is the equation of state for a photon gas. The Stefan–Boltzmann law and the Wien displacement law can be derived from this equation of state.

We use the following expressions for the total entropy S, total energy U, and pressure p of the equilibrium radiation field:

$$S = \tilde{s}V, \quad U = \tilde{u}V, \quad p = \tilde{u}/3, \tag{9.84}$$

with $\tilde{s}$ and $\tilde{u}$ the volumetric energy densities of entropy and energy. With these expressions, the first law in differential form

$$\mathrm{d}S = \frac{\mathrm{d}U + p\,\mathrm{d}V}{T} \tag{9.85}$$

can be rewritten as

$$V\,\mathrm{d}\tilde{s} + \tilde{s}\,\mathrm{d}V = V\,\frac{\mathrm{d}\tilde{u}}{T} + \frac{4}{3}\frac{\tilde{u}}{T}\,\mathrm{d}V. \tag{9.86}$$

Because $\tilde{u}$ and, for the same reason, $\tilde{s}$ are only functions of temperature,[55] independent variations in temperature and volume lead to the following equalities:

$$\tilde{s} = \frac{4}{3}\frac{\tilde{u}}{T}, \quad \mathrm{d}\tilde{s} = \frac{\mathrm{d}\tilde{u}}{T}. \tag{9.87}$$

In order for both equations to be true, the differential version of the first equation must be equivalent to the second equation. The differential of the first equation is

$$\mathrm{d}\tilde{s} = \frac{4}{3}\left(\frac{\mathrm{d}\tilde{u}}{\mathrm{d}T} - \frac{\tilde{u}}{T}\right)\frac{\mathrm{d}T}{T}. \tag{9.88}$$

This equals $\mathrm{d}\tilde{u}/T$, as in the second Eq. 9.87, if and only if

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}T} = 4\frac{\tilde{u}}{T}. \tag{9.89}$$

This equation can be integrated to find

$$\blacktriangleright \quad \tilde{u} = k\,T^4, \tag{9.90}$$

---

[55] We can add a constant S_0 to the total entropy S by adding S_0/V to the entropy density $\tilde{s}$ and still satisfy Eq. 9.87. By the third law, this constant is zero.

with k an integration constant. As a corollary we find a relationship between entropy density and energy density of the equilibrium radiation field,

$$\blacktriangleright \quad \tilde{s} = \frac{4}{3}\frac{\tilde{u}}{T} = \frac{4}{3}kT^3. \tag{9.91}$$

The Stefan–Boltzmann equation now follows from the result derived in Appendix D, which states that the total radiative energy flux J_u is related to the volumetric energy density $\tilde{u}$ by

$$J_u = \tilde{u}c/4, \tag{9.92}$$

with c the speed of light. For a photon gas in equilibrium we therefore have

$$\blacktriangleright \quad J_u = \frac{kc}{4}T^4 \tag{9.93}$$

with k as in Eq. 9.90. This is the Stefan–Boltzmann law with J_u the total black body emission and $kc/4$ the Stefan–Boltzmann constant.

The Wien displacement law follows from a more general thermodynamic prediction on the possible shape of the spectrum of thermal radiation. We proceed by performing a thought experiment where we adiabatically compress the radiation in a cubic vessel with perfectly reflecting walls and with edge size L. Any single wavemode with wavenumber λ will adjust its wavelength to fit the size of the vessel, see Figure 9.11. We therefore have

$$\lambda \propto L. \tag{9.94}$$

The total energy of the vessel changes according to the first law under adiabatic conditions,

$$\mathrm{d}U = -p\,\mathrm{d}V. \tag{9.95}$$

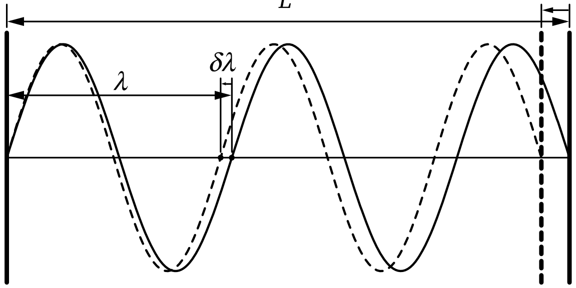


FIGURE 9.11 Compressing the vessel of size L leads to a proportional compression of a wave with wavelength λ.

If we substitute $U = \tilde{u}V$ and $p = \tilde{u}/3$ we find

$$4\tilde{u}\,\mathrm{d}V + 3V\,\mathrm{d}\tilde{u} = 0, \tag{9.96}$$

which integrates to

$$\tilde{u}^3V^4 = \text{constant} \tag{9.97}$$

or, because $V = L^3$ and the wavelength λ of any mode is proportional to the edge size L,

$$\tilde{u}\lambda^4 = \text{constant.} \tag{9.98}$$

From the derivation of the Stefan–Boltzmann law, Eq. 9.90, we know that $\tilde{u}$ is proportional to T^4. We conclude that for a single wavemode

$$\lambda T = \text{constant} \tag{9.99}$$

on adiabatic compression.

The first law can also be applied to parts of the spectrum. So if we take the total energy $\tilde{u}_\lambda\,\Delta\lambda\,V$ between wavelengths λ and $\lambda + \Delta\lambda$ then the first law for adiabatic compression reads

$$\mathrm{d}(\tilde{u}_\lambda\,\Delta\lambda\,V) = -p_\lambda\,\Delta\lambda\,\mathrm{d}V, \tag{9.100}$$

with $p_\lambda\Delta\lambda = \tilde{u}_\lambda\,\Delta\lambda/3$ the partial pressure due to the wavemodes in that wavelength range. We arrive at the analogous expression of Eq. 9.98,

$$\tilde{u}_\lambda\lambda^4\,\Delta\lambda = \text{constant.} \tag{9.101}$$

But because the wavelength interval $\Delta\lambda$ will compress proportional to L, just as λ, we find that on adiabatic compression

$$\tilde{u}_\lambda\lambda^5 = \text{constant.} \tag{9.102}$$

Finally, in Section 9.1 we argued that $\tilde{u}$ can only be a function of λ and T.[56] We have now gathered enough properties of $\tilde{u}$ to predict some important properties of the spectrum. We have during the adiabatic compression of the radiation that for any wavemode λ

$$\lambda T = \text{constant}, \tag{9.103a}$$

$$\tilde{u}_\lambda\lambda^5 = \text{constant}, \tag{9.103b}$$

$$\tilde{u}_\lambda = f(\lambda, T). \tag{9.103c}$$

---

[56] We have assumed that the radiation remains black during the compression.

To satisfy these properties for all λ, we see that $\tilde{u}_\lambda$ must be of the form

$$\blacktriangleright \quad \tilde{u}_\lambda = \lambda^{-5} g(\lambda T), \tag{9.104}$$

for some function g. The Planck law, Eq. 9.6, is of this form (remember that the energy density $\tilde{u}_\lambda$ is proportional to the energy flux B_λ). An equivalent expression of the form of the spectrum is

$$\blacktriangleright \quad \tilde{u}_\lambda = T^5 g'(\lambda T), \tag{9.105}$$

for some other function g' where $g'(x) = g(x)x^{-5}$. From this form of the spectral distribution we find that the only dependency on the wavelength is contained in the proportionality to $g'(\lambda T)$. The maximum of $g'(x)$ corresponds to some x_m. This then corresponds to some wavelength λ_m with

$$\blacktriangleright \quad \lambda_m = x_m / T. \tag{9.106}$$

This is Wien's displacement law with x_m the displacement constant.

9.9 DERIVATION OF THE PLANCK LAW

One cannot help but be impressed by the thermodynamic arguments leading to Eq. 9.104, or the equivalent Eq. 9.105. These equations contain the Stefan–Boltzmann law and the Wien displacement law. However, thermodynamics alone cannot take us any further. The derivation of the Planck law itself requires quantum mechanical concepts. We present its derivation here as it is one of the classic results in physics. This by now canonical derivation is based on statistical mechanics, which employs the Boltzmann factor introduced in Section 4.6.[57]

We consider a standing electromagnetic wave of wavelength λ in a cubic vacuum vessel with perfectly reflecting walls and with edge size L. The electromagnetic wave corresponds to a photon with energy, e, which is related to the wavelength, λ, as

$$e = hc/\lambda. \tag{9.107}$$

Introducing the total wavenumber K ,with $K\lambda = 2\pi$, this is written

$$e = \frac{h}{2\pi} Kc. \tag{9.108}$$

The wavenumber K is the length of the three-dimensional wave vector $\boldsymbol{K} = (K_x, K_y, K_z)$.

---

[57] For further reading see Kittel C. & Kroemer, H. (1980) *Thermal physics* 2nd edn. W. H. Freeman, New York; Baierlein, R. (1999) *Thermal physics* Cambridge University Press, Cambridge.

We first need to know how many wavevectors of total wavenumber K there are. In a cubic vessel with reflecting walls, the components K_i of the wavevector can only have discrete values such that precisely N_i half wavelengths fit inside the vessel size L; for example, see Figure 9.11. We therefore have[58]

$$K_i = \pi N_i / L. \tag{9.109}$$

So for each one-dimensional wavenumber π/L we can fit one wavenumber K_i. An elementary volume of the size $(\pi/L)^3$ in three-dimensional wavenumber space therefore contains one three-dimensional wavenumber mode. Photons also have an internal degree of freedom called *spin*. For a photon the spin can be either up or down. So each elementary volume in wavenumber space corresponds to two possible photon modes – each wavenumber mode has a *degeneracy* of 2.

The wave vectors with total wavenumber between K and $K+\mathrm{d}K$ represent a three-dimensional spherical shell in wave-vector space with volume $4\pi K^2\,\mathrm{d}K$. Because all wavenumber components K_i are assumed positive, only 1/8 of this shell contributes to the total wavenumber volume. An elementary volume of $(\pi/L)^3$ contains two degenerate photon modes so the number of modes $g_K\,\mathrm{d}K$ corresponding to the shell between total wavenumber K and $K+\mathrm{d}K$ is

$$g_K\,\mathrm{d}K = \frac{K^2 L^3}{\pi^2}\,\mathrm{d}K. \tag{9.110}$$

We now express this as the number of modes $g_e\,\mathrm{d}e$ with photon energies between e and $e+\mathrm{d}e$. We have

$$g_e\,\mathrm{d}e = g_K\,\mathrm{d}K. \tag{9.111}$$

Using Eq. 9.108 we now find that

$$g_e = \frac{8e^2\pi L^3}{h^3 c^3}. \tag{9.112}$$

This function is called the *density of states* of the energy level.

So what is the probability of finding N photons at some mode with energy e? The Boltzmann factor $\exp(-e/k_B T)$ gives the probability of finding a photon in a single mode of energy e, see Section 4.6. The probability $P_e(n)\,\mathrm{d}e$ of finding n photons between energies e and $e+\mathrm{d}e$ therefore is

$$P_e(n) = g_e\,\frac{\exp(-\beta n e)}{Z}, \tag{9.113}$$

---

[58] The argument presented here is not complete in quantum mechanics as we ignore the ground level of each wave mode. However, this does not change the phase space density g_K, which is the purpose of the quantization presented here.

where the partition function Z provides the normalization and we have used the standard notation

$$\beta = 1/k_B T. \tag{9.114}$$

The density of states g_e multiplies the Boltzmann factor to account for the fact that there are $g_e\,\mathrm{d}e$ photon modes corresponding to the given photon energy.

The partition function Z is

$$Z = g_e \sum\nolimits_{n=0}^{\infty} \exp\left(-\beta n e\right). \tag{9.115}$$

This summation has the standard form for summation of powers of quantities $a < 1$,

$$\sum\nolimits_{n=0}^{\infty} a^n = \frac{1}{1-a}. \tag{9.116}$$

Using this formula we find that the partition function for photons with energy e is

$$Z = \frac{g_e}{1 - \exp\left(-\beta e\right)}. \tag{9.117}$$

The expectation value for the total energy E_e, representing photons at energy e, is

$$E_e = \sum\nolimits_{n=0}^{\infty} P_e(n)\, n e = \sum\nolimits_{n=0}^{\infty} g_e \frac{n e \exp\left(-\beta n e\right)}{Z}. \tag{9.118}$$

This can be written as

$$E_e = -\frac{\partial}{\partial \beta} \ln Z. \tag{9.119}$$

Using the explicit form of the partition function, Eq. 9.117 we then find the expectation value for the total energy E_e of photons of energy e,

► $$E_e = \frac{g_e\, e}{\exp\left(\beta e\right) - 1}. \tag{9.120}$$

To derive the Planck law in the form of Eq. 9.6 we express the total energy expectation value in terms of wavelength instead of photon energy. We introduce a transformed density of states g_λ in wavelength space which is defined in terms of the usual density of states g_e by

$$g_e\, \mathrm{d}e = g_\lambda\, \mathrm{d}\lambda. \tag{9.121}$$

Using Eq. 9.107, we find

$$g_\lambda = 8\pi L^3/\lambda^4 \tag{9.122}$$

The total energy expressed in terms of wavelength therefore is

$$E_\lambda = \frac{g_\lambda\, e(\lambda)}{\exp\left(\beta e(\lambda)\right) - 1} \tag{9.123}$$

The volumetric energy density $\tilde{u}_\lambda$ is

$$\tilde{u}_\lambda = E_\lambda/L^3. \tag{9.124}$$

Substituting the explicit expressions for g_λ and $e(\lambda)$ we find

$$\tilde{u}_\lambda = \frac{8\pi hc}{\lambda^5}\frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \tag{9.125}$$

According to Eq. 9.92 this corresponds to a surface energy flux of $B_\lambda = \tilde{u}_\lambda c/4$, which is the Planck law in the form of Eq. 9.6.

The fact that we have assumed a cubic vessel instead of a vessel with arbitrary shape is not important. For arbitrary shapes the analogous derivation becomes technically more challenging but gives the same result. However, it is easier to use the thermodynamic result of Section 9.1, which states that at equilibrium the radiative energy density $\tilde{u}_\lambda$ cannot be a function of vessel shape or size. The radiation law derived for the cubic vessel is therefore valid for a vessel of any shape.

It is of interest to note that E_e can be written as the product of the photon energy e and the expectation value n_e of the number of photons at this particular energy. We thus find that

$$n_e = \frac{g_e}{\exp\left(\beta e\right) - 1}. \tag{9.126}$$

This is the *Bose–Einstein distribution* for a system with zero chemical potential, such as a photon gas. Note that the argument βe equals Λ/λ, with Λ the thermal wavelength introduced in Eq. 9.9. For values of λ shorter than about $\Lambda/3$ the Bose–Einstein distribution becomes indistinguishable from the Boltzmann distribution, $g_e \exp\left(-\Lambda/\lambda\right)$. In the derivation of the Wien displacement law we found that the Planck function peaks at around $\lambda \approx \Lambda/5$. This means that the Bose–Einstein correction is only important in the long wavelength part of the black body spectrum; say wavelengths longer than the peak wavelength. For all other wavelengths the Bose–Einstein distribution reduces to the Boltzmann approximation

$$n_e \approx g_e \exp\left(-\beta e\right). \tag{9.127}$$

It is not surprising that we find the Boltzmann distribution for short wavelengths. The Bose–Einstein distribution follows from adding all the probabilities of finding any number of particles in a given mode. However at short wavelengths the photon energy e is much larger than the thermal energy $k_B T$, so that each mode of energy e has only a low chance of being occupied. The chance of multiple occupancy can therefore be ignored and the Bose–Einstein distribution reduces to the probability of occupying the mode with energy e, which is the Boltzmann distribution.

At short wavelengths, the Planck law expressed as a function of wavelength therefore becomes

$$B_\lambda(T) \approx \frac{2\pi h c^2}{\lambda^5} \exp\left(-\frac{hc}{\lambda k_B T}\right). \tag{9.128}$$

This equation is called *Wien's distribution law,* and it is accurate except at long wavelengths, see Figure 9.12. The spectral peak in Wien's distribution law corresponds to a wavelength of $\lambda_m = \Lambda/5$, which is very close to that predicted by the Planck law, $\lambda_m = \Lambda/4.965\ldots$

The approximate result for long wavelengths follows from a Taylor expansion of the Planck law,

$$B_\lambda(T) \approx \frac{2\pi c k_B T}{\lambda^4}. \tag{9.129}$$

This approximation is called the *Rayleigh–Jeans law,* see Figure 9.12. Like Wien's distribution law, the Rayleigh–Jeans law was derived independently from the Planck law. Both approximations satisfy the thermodynamically predicted shape of Eq. 9.104.

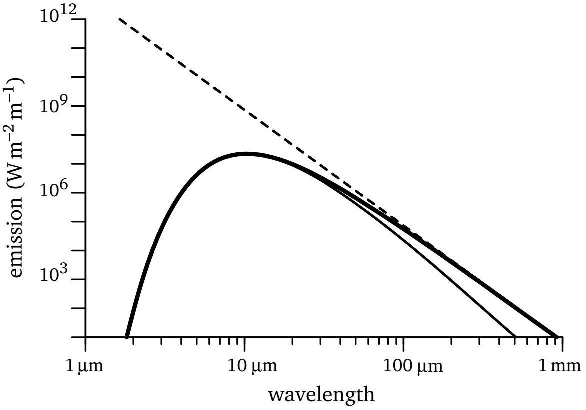


FIGURE 9.12 Black body emission as a function of wavelength for $T = 280\,\mathrm{K}$ according to the Planck law (thick solid line), the Wien distribution law (thin solid line), and the Rayleigh–Jeans law (thin dashed line).

The Rayleigh–Jeans law is interesting because it is based on purely classical physics. The derivation starts from the density of modes in wavelength, g_λ. Using equipartition of energy we would associate an energy of k_BT with each mode: the energy of a mode is proportional to its wavenumber, so we expect from the generalized equipartition theorem, Eq. 4.85, an amount of k_BT in each mode. This leads to a volumetric energy density of $g_\lambda k_BT/L^3$. Multiplying this by $c/4$ to transform the energy density to an energy flux, we find the Rayleigh–Jeans law.

The Rayleigh–Jeans law does not permit calculation of the total emitted power: the integral of the Rayleigh–Jeans $B_\lambda(T)$ over λ is divergent due to the $1/\lambda^4$ singularity at low wavelengths. This divergence is called the *ultraviolet catastrophe*. This was an early indication that classical physics alone could not account for the behaviour of thermal radiation. (The failure of equipartition to predict the heat capacity for diatomic gases was another indication of the problems with classical physics, see Section 3.3.)

PROBLEMS

9.1. The eccentricity e of the Earth's orbit is 0.0167, where eccentricity is defined by

$$\frac{r_{\min}}{r_{\max}} = \frac{1-e}{1+e},$$

and $r_{\min}$ and $r_{\max}$ are the minimum and maximum distances between the Earth and the Sun. Show that for a solar constant S_0 of 1366 W m^{-2}, the eccentricity leads to a variation of the solar irradiance between about 1321 W m^{-2} and 1412 W m^{-2}.

9.2. Show that the fraction, f, of the total black body emission which is emitted around the peak wavelength λ_m can be approximated as

$$f = 0.66\Delta\lambda/\lambda_m$$

with $\Delta\lambda$ the chosen width of the wavenumber band. Hence show that about 40% of the radiative output of the Sun occurs in the visible part of the spectrum.

9.3. Radiative balance in the stratosphere. Recall the simple radiative one-layer black body slab model of the atmosphere discussed in Section 9.3. Now put on top of this model another slab with a short wave emissivity ϵ_s and a longwave emissivity ϵ_l. Show that the equilibrium temperature T_S in the stratosphere is given by

$$\epsilon_l(2-\epsilon_l)\,\sigma T_S^4 = \frac{S_0}{4}\left(1-\alpha(1-\epsilon_s)^2-(1-\alpha)(1-\epsilon_l)(1-\epsilon_s)\right).$$

Hence show that, contrary to the troposphere, for small values of ϵ_l the temperature of the stratosphere *decreases* with increasing ϵ_l. What is the physical picture for this decrease? Show that for small ϵ_l the stratospheric temperature is generally higher than the tropospheric temperature.

9.4. Multilayer atmosphere. Consider an atmosphere made up of N atmospheric layers that are black body ($\epsilon_l = 1$) in the long-wave regime and completely transparent in the short-wave regime ($\epsilon_s = 0$). The layers are numbered 1 (nearest to the planet) to N (furthest from the planet). Show that for the temperature T_i in layer i we have

$$\sigma T_i^4 = \sigma T_{i+1}^4 + \delta F,$$

with δF some fixed energy flux. Show that

$$\sigma T_N^4 = \delta F \quad \text{and} \quad \sigma T_N^4 = (1-\alpha)\, S_0/4.$$

Hence show that

$$\sigma T_E^4 = (N+1)\,(1-\alpha)\, S_0/4,$$

with T_E the surface temperature of the planet. Compare this result to the continuous atmosphere result of Eq. 9.68.

9.5. Show that for the radiative equilibrium model of Eq. 9.70 the temperature lapse rate $\Gamma = -\mathrm{d}T/\mathrm{d}z$ is given by

$$\Gamma = \frac{g}{2R}\frac{x}{1+x}, \quad \text{with } x = \delta_{\mathrm{TOA}} p^2/p_0^2.$$

Hence show that for $\delta_{\mathrm{TOA}} > 2R/(c_p - 2R)$ the dry atmosphere becomes convectively unstable.

10

Non-equilibrium processes

For individual, homogeneous volumes of a substance, gas or liquid, in local thermodynamic equilibrium, we can apply, by definition, the machinery of equilibrium thermodynamics. The majority of processes described in this book fall under this heading. When the substance is non-homogeneous, it is not in thermodynamic equilibrium and we need to consider non-equilibrium processes. Clearly, the atmosphere is profoundly non-homogeneous.

The signature of non-equilibrium thermodynamics is the irreversible increase of entropy in a system. The task at hand is to ascribe the irreversible increase of entropy to specific physical processes.

Remember that for an individual homogeneous parcel we can still describe entropy changing (diabatic) processes. However, we would consider such processes reversible as they result from heat exchange with the environment. This distinction was first described in Chapter 2. Irreversible entropy change only occurs when the full non-homogeneous system is analyzed.

10.1 ENERGETICS OF MOTION

In continuum mechanics, specific variables (quantities expressed per unit mass) satisfy equations of the shape

$$\frac{\mathrm{D}c}{\mathrm{D}t} = S, \tag{10.1}$$

where c is some specific variable and S is some source term. Here $\mathrm{D}/\mathrm{D}t$ represents the Lagrangian derivative

$$\mathrm{D}/\mathrm{D}t = \partial/\partial t + \boldsymbol{U} \cdot \nabla, \tag{10.2}$$

with $\boldsymbol{U}$ the flow velocity. The Lagrangian derivative is the time derivative following a fluid parcel. Equation 10.1 expresses how a property c of a fluid parcel changes through some source term S.

Thermal Physics of the Atmosphere Maarten H. P. Ambaum
ISBN: 978-0-470-74515-1

An example of such an equation is the *continuity equation* for the specific volume v,

$$\frac{\mathrm{D}v}{\mathrm{D}t} = v\nabla \cdot \boldsymbol{U}. \tag{10.3}$$

This is essentially an expression for conservation of mass: the volume over which the mass of a parcel is distributed changes due to the expansion of the parcel, as expressed by the divergence of the velocity field $\nabla \cdot \boldsymbol{U}$. The more familiar form of the continuity equation is expressed in terms of density,

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho\nabla \cdot \boldsymbol{U}. \tag{10.4}$$

We can use the latter form of the continuity equation to transform the equation for the specific variable c into an equation for the variable per unit volume ρc. For example, if c is a concentration by mass, ρc is a concentration by volume. We find

$$\frac{\partial \rho c}{\partial t} + \nabla \cdot (\rho \boldsymbol{U} c) = \rho S. \tag{10.5}$$

Here, the local change in the volumetric density is a result of a volumetric source term ρS and a flux divergence $\nabla \cdot (\rho \boldsymbol{U} c)$. This form of the equations will now be used to analyze the energy budget of fluids in motion.[59]

First consider the equation for the kinetic energy per unit volume $\rho|\boldsymbol{U}|^2/2$. This follows from the momentum equations. For notational simplicity we will first consider the equation for one component, say $\mathscr{U}$, the velocity component in the x-direction. We then complete the other equations by analogy. The relevant momentum equation for $\mathscr{U}$ is

$$\rho\frac{\mathrm{D}\mathscr{U}}{\mathrm{D}t} = -\frac{\partial p}{\partial x} - \rho\frac{\partial \phi}{\partial x} - \nabla \cdot \boldsymbol{F}_{\mathscr{U}}, \tag{10.6}$$

with the first term on the right-hand side the pressure gradient force, the second term the force due to gravity, or any other body force defined by a potential ϕ, and the third term the viscosity force which is expressed in a general way as the divergence of a viscous momentum flux.

We have not included any rotation effects in the momentum equation: the Coriolis force does not feature in the kinetic energy budget as it is perpendicular to the velocity, and the centrifugal force is taken to be part of the effective geopotential ϕ, see Section 4.1.

---

[59] Here we restrict ourselves to atmospheric applications. More general treatments can be found in De Groot, S. R. & Mazur, P. (1984) *Non-equilibrium thermodynamics.* Dover, New York; Woods, L. C. (1975) *The thermodynamics of fluid systems*. Oxford University Press, Oxford.

Multiply the momentum equation by $\mathscr{U}$ to get an equation for the specific kinetic energy contribution of the speed in the x-direction, $\mathscr{U}^2/2$,

$$\rho\frac{\mathrm{D}\mathscr{U}^2/2}{\mathrm{D}t} = -\mathscr{U}\frac{\partial p}{\partial x} - \rho\mathscr{U}\frac{\partial \phi}{\partial x} - \mathscr{U}\nabla\cdot\boldsymbol{F}_{\mathscr{U}}. \tag{10.7}$$

By analogy we find the corresponding equations for the $\mathscr{V}$ and $\mathscr{W}$ components. Now define the specific kinetic energy k as

$$k = \left(\mathscr{U}^2 + \mathscr{V}^2 + \mathscr{W}^2\right)/2. \tag{10.8}$$

The equation for the specific kinetic energy then is

$$\rho\frac{\mathrm{D}k}{\mathrm{D}t} = -\boldsymbol{U}\cdot\nabla p - \rho\boldsymbol{U}\cdot\nabla\phi - \sum\nolimits_i \mathscr{U}_i\nabla\cdot\boldsymbol{F}_{\mathscr{U}_i} \tag{10.9}$$

where the $\mathscr{U}_i$ with $i = 1, 2, 3$ represent the three velocity components. Using the continuity equation we can rewrite this equation in the flux form,

$$\blacktriangleright \quad \frac{\partial\rho k}{\partial t} + \nabla\cdot(\rho\boldsymbol{U}k) = -\boldsymbol{U}\cdot\nabla p - \rho\boldsymbol{U}\cdot\nabla\phi - \sum\nolimits_i \mathscr{U}_i\nabla\cdot\boldsymbol{F}_{\mathscr{U}_i}. \tag{10.10}$$

The first term on the right-hand side represents the loss of kinetic energy due to motion against the pressure gradient force and the second term represents the loss of kinetic energy due to motion against gravity. By convention the latter direction is usually *defined* to be the z-direction, in which case this term reduces to $-\rho\mathscr{W}g$. The third term represents the effects of viscosity. This will be further analyzed in the next section.

Conservation of energy dictates that the energy source terms in the kinetic energy equation must come from somewhere. They are compensated by opposite terms in the budget equations for the internal energy and the geopotential.

Next consider the budget for internal energy. From the first law of thermodynamics in differential form,

$$\mathrm{d}u = T\,\mathrm{d}s - p\,\mathrm{d}v, \tag{10.11}$$

the equation for the specific internal energy u is

$$\frac{\mathrm{D}u}{\mathrm{D}t} = T\frac{\mathrm{D}s}{\mathrm{D}t} - p\frac{\mathrm{D}v}{\mathrm{D}t} = T\frac{\mathrm{D}s}{\mathrm{D}t} - pv\nabla\cdot\boldsymbol{U}, \tag{10.12}$$

where we have used the continuity equation in the form Eq. 10.3. For the internal energy per unit volume ρu we now find

$$\frac{\partial\rho u}{\partial t} + \nabla\cdot(\rho\boldsymbol{U}u) = \rho T\frac{\mathrm{D}s}{\mathrm{D}t} - p\nabla\cdot\boldsymbol{U}. \tag{10.13}$$

The first term on the right-hand side is the diabatic heating per unit volume, as follows from the second law of thermodynamics. The last term in this equation can be written as a flux and a residual,

$$p\nabla \cdot \boldsymbol{U} = \nabla \cdot (p\boldsymbol{U}) - \boldsymbol{U} \cdot \nabla p. \tag{10.14}$$

Substituting this expression we find

▶
$$\frac{\partial \rho u}{\partial t} + \nabla \cdot (\rho \boldsymbol{U} h) = \rho T \frac{\mathrm{D}s}{\mathrm{D}t} + \boldsymbol{U} \cdot \nabla p, \tag{10.15}$$

with the specific enthalpy $h = u + pv$. In this form the second term on the right-hand side compensates the opposite term in the budget equation for the kinetic energy per unit volume, Eq. 10.10.

The internal energy flux is given by the advective transport of enthalpy, not the transport of internal energy. This is consistent with the arguments in Section 3.5, where it was explained how this difference arises from the additional flow work required to move matter from one volume to another.

Finally, the equation for the specific potential energy ϕ is an expanded version of the Lagrangian derivative of a field that does not vary in time,

$$\frac{\mathrm{D}\phi}{\mathrm{D}t} = \boldsymbol{U} \cdot \nabla \phi. \tag{10.16}$$

The equation for the potential energy per unit volume $\rho\phi$ therefore is

▶
$$\frac{\partial \rho \phi}{\partial t} + \nabla \cdot (\rho \boldsymbol{U} \phi) = \rho \boldsymbol{U} \cdot \nabla \phi. \tag{10.17}$$

The right-hand side of this equation compensates for the second source term in the kinetic energy budget, Eq 10.10.

The compensating terms in the above budget equations are usually called *conversion terms*. The energy budget contains two conversion terms. Firstly, we have the conversion between internal and kinetic energy, $-\boldsymbol{U} \cdot \nabla p$, due to flow along the pressure gradient. Secondly, we have conversion between potential and kinetic energy, $-\rho \boldsymbol{U} \cdot \nabla \phi$, due to downward (that is, in the negative z direction) motion in a gravity field. The conversion terms are reversible: they can be of either sign.

We can add the three energy equations to find an equation for the total specific energy $e = k + u + \phi$,

▶
$$\frac{\partial \rho e}{\partial t} + \nabla \cdot \boldsymbol{F}_e = \rho T \frac{\mathrm{D}s}{\mathrm{D}t} - \sum\nolimits_i \mathscr{U}_i \nabla \cdot \boldsymbol{F}_{\mathscr{U}_i} \tag{10.18}$$

with the total energy flux

▶
$$\boldsymbol{F}_e = \rho \boldsymbol{U}(k + h + \phi). \tag{10.19}$$

That is, the total energy in some volume can only change by diabatic heating or viscosity.

In a steady state and in the absence of diabatic effects (heating or viscosity), the total energy budget reduces to

$$\nabla \cdot (\rho \boldsymbol{U}(k+h+\phi)) = \rho \boldsymbol{U} \cdot \nabla (k+h+\phi) = 0, \tag{10.20}$$

where we have used the steady state form of the continuity equation, $\nabla \cdot (\rho \boldsymbol{U}) = 0$. This equation states that, following the flow, the term between brackets remains constant. This is Bernoulli's equation, discussed in Section 4.5. When kinetic energy is small, this reduces to conservation of the generalized enthalpy or dry static energy following a fluid parcel.

For a steady flow with closed streamlines (for example, idealized overturning circulations), the energy budget, Eq. 10.18, can be integrated along a closed streamline. The momentum diffusion term on the right-hand side removes energy along this closed streamline, see next section. This means that the entropy contribution on the right-hand side has to put in energy along this closed streamline. In equations,

$$\oint T \,\mathrm{d}s > 0, \tag{10.21}$$

where the contour integration is along the closed streamline. In other words, to keep a steady state going against dissipation, we need to heat the flow ($\mathrm{d}s > 0$) at higher temperatures than where we cool the flow ($\mathrm{d}s < 0$). This expression becomes particularly interesting when we use the potential temperature definition for an ideal gas, Eqs. 3.80 and 3.84. In this case, the above contour integral becomes

$$\oint c_p \,(p/p_0)^{\kappa} \,\mathrm{d}\theta > 0. \tag{10.22}$$

The ideal gas flow needs to be heated at a higher pressure than where it is cooled. The application of Eq. 10.21 to incompressible fluids leads to the statement that heating (expansion) needs to occur at higher pressures than the cooling (contraction). Oceanographers traditionally refer to this as the Sandström theorem.[60]

The atmosphere receives most of its heat near the surface and it loses most of its heat through long-wave radiation at higher altitudes: heat is received at pressures higher than those at which it is lost and the atmosphere can therefore stay in motion against friction. For thermally driven circulations in the ocean, this is less obvious. The external heat input and output tend to occur at the same pressure level, and we need diffusion and stirring to deposit

---

[60] See Vallis, G. K. (2006) *Atmospheric and oceanic fluid dynamics*. Cambridge University Press, Cambridge; Kuhlbrodt, T. (2008) *Tellus* **60A**, 819–836.

the heat at deeper levels in order to keep the thermally driven circulation going against dissipation.

10.2 DIABATIC EFFECTS AND THE SECOND LAW

The energy budgets, as described in the previous section, represent the generalization of the first law of thermodynamics for continuous systems. Next we will consider the second law.

We will start by analyzing the viscous momentum flux $\boldsymbol{F}_{\mathcal{U}}$, see Eq. 10.6, as it plays a special role in the energy budget. The viscous momentum flux is always opposite to the momentum gradient, and for incompressible fluids is usually taken to be linear in the momentum gradient,

$$\boldsymbol{F}_{\mathcal{U}} = -\mu \nabla \mathcal{U}, \tag{10.23}$$

with analogous versions for the other momentum components. The coefficient μ is called the *shear viscosity.*

The contribution of the viscous momentum flux to the energy budget, Eq. 10.18, can be written as a flux and a residual,

$$\sum_i \mathcal{U}_i \nabla \cdot \boldsymbol{F}_{\mathcal{U}_i} = \nabla \cdot \left(\sum_i \mathcal{U}_i \boldsymbol{F}_{\mathcal{U}_i} \right) - \sum_i \boldsymbol{F}_{\mathcal{U}_i} \cdot \nabla \mathcal{U}_i \tag{10.24}$$

The first term is the divergence of the viscous kinetic energy flux, $\boldsymbol{F}_\nu$,

$$\boldsymbol{F}_\nu = \sum_i \mathcal{U}_i \boldsymbol{F}_{\mathcal{U}_i} \tag{10.25}$$

and the second term is the *viscous dissipation rate,* $\dot{\epsilon}$, defined by

$$\rho \dot{\epsilon} = -\sum_i \boldsymbol{F}_{\mathcal{U}_i} \cdot (\nabla \mathcal{U}_i). \tag{10.26}$$

Because the viscous momentum flux $\boldsymbol{F}_{\mathcal{U}}$ is always opposite to the momentum gradient, the viscous dissipation rate is locally always positive (that is, this term always dissipates kinetic energy). For example, if the viscous momentum flux is linear in the momentum gradient, as in Eq. 10.23, we find

$$\rho \dot{\epsilon} = \mu \left| \nabla \mathcal{U}_i \right|^2, \tag{10.27}$$

which is evidently positive definite.

With these definitions the total energy budget, Eq. 10.18, becomes

$$\frac{\partial \rho e}{\partial t} + \nabla \cdot (\boldsymbol{F}_e + \boldsymbol{F}_\nu) = \rho \left(\dot{q} - \dot{\epsilon} \right) \tag{10.28}$$

where we have written the local entropy change as a diabatic heating rate $\dot{q}$ defined by

$$\dot{q} = T\,\mathrm{D}s/\mathrm{D}t. \tag{10.29}$$

The heating rate $\dot{q}$ has several contributions. Firstly, the local dissipation of kinetic energy $\dot{\epsilon}$ is a local source of internal energy. Secondly, the second law of thermodynamics states that any temperature gradient will locally set up a heat flux $\boldsymbol{F}_q$. Thirdly, the absorption and emission of radiation provides a non-local source of energy that can be written as the divergence of a radiative flux $\boldsymbol{F}_r$. Finally, there may be energy sources related to phase changes in the volume, the latent heating $\dot{q}_L$. In equations,

$$\rho\dot{q} = \rho\dot{\epsilon} - \nabla\cdot\boldsymbol{F}_q - \nabla\cdot\boldsymbol{F}_r + \rho\dot{q}_L. \tag{10.30}$$

Many numerical models ignore the mechanical dissipation rate $\dot{\epsilon}$ in this equation. This is entirely unnecessary and leads to small but systematic errors.

Next we will consider how these four contributions to the heating rate satisfy the second law of thermodynamics. In order to do this we need to evaluate their contributions to the production of total entropy S in the whole system under consideration, which is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \int_V \frac{\rho\dot{q}}{T}\,\mathrm{d}V, \tag{10.31}$$

with V the volume of the system.

The viscous dissipation of kinetic energy $\dot{\epsilon}$ is a local source of internal energy. Because it is a conversion from kinetic energy to internal energy, it cancels in the total energy budget, see Eq. 10.28. Contrary to the kinematic conversion term, $-\boldsymbol{U}\cdot\nabla p$, the viscous dissipation $\dot{\epsilon}$ can only remove kinetic energy and produce internal energy. The local change in entropy due to viscous dissipation equals

$$\rho\,Ds/\mathrm{D}t = \rho\dot{\epsilon}/T. \tag{10.32}$$

Because $\dot{\epsilon}$ is positive definite, the total entropy production due to this process, Eq. 10.31, is positive. Therefore, the viscous dissipation source satisfies the second law; it produces an irreversible entropy increase.

The next contribution to consider is the heating due to heat conduction. The heat flux $\boldsymbol{F}_q$ is always opposite to the temperature gradient and it is usually taken to be linear in the gradient,

$$\boldsymbol{F}_q = -K\nabla T, \tag{10.33}$$

with K the *diffusion coefficient* or *thermal conductivity*. For small temperature gradients this linear approximation is very accurate. Note that the exchange of sensible heat with the Earth's surface also has this form and thus no new effects are introduced at the boundaries. The heat conduction flux leads to a total entropy source

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\int_V \frac{\nabla \cdot \boldsymbol{F}_q}{T}\,\mathrm{d}V = -\int_V \left(\nabla \cdot \left(\frac{\boldsymbol{F}_q}{T}\right) + \frac{\boldsymbol{F}_q \cdot \nabla T}{T^2}\right)\mathrm{d}V. \tag{10.34}$$

Next we use Gauss' theorem to transform the volume integral of the first integrand into a surface integral of the area A around the system,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\int_A \frac{\boldsymbol{F}_q \cdot \hat{\boldsymbol{n}}}{T}\,\mathrm{d}A - \int_V \frac{\boldsymbol{F}_q \cdot \nabla T}{T^2}\,\mathrm{d}V, \tag{10.35}$$

with $\hat{\boldsymbol{n}}$ the unit normal vector perpendicular to A pointing outward of V. The first term on the right-hand side is the conduction flux of entropy $\boldsymbol{F}_q/T$ through the boundary of the system; it therefore represents the reversible part of the entropy change due to external input of heat,

$$\frac{\mathrm{d}_e S}{\mathrm{d}t} = -\int_A \frac{\boldsymbol{F}_q \cdot \hat{\boldsymbol{n}}}{T}\,\mathrm{d}A. \tag{10.36}$$

The second term on the right-hand side is the irreversible internal generation of entropy due to heat conduction. It is positive definite because the heat flux is opposite to the temperature gradient,

$$\frac{\mathrm{d}_i S}{\mathrm{d}t} = -\int_V \frac{\boldsymbol{F}_q \cdot \nabla T}{T^2}\,\mathrm{d}V \geq 0. \tag{10.37}$$

We thus find that the heat conduction also satisfies the second law.

The above discussion of heat conduction is the continuous system generalization of the system discussed in Section 2.3.

An analogous equation for the radiative flux $\boldsymbol{F}_r$, the third contribution in Eq. 10.30, leads to a radiative flux of entropy through the system boundary and a radiative generation of entropy, consistent with the second law. For a closed system, the net black body radiation flux will be opposite to temperature gradients so this again provides a positive definite source of entropy. The Earth's atmosphere also exchanges radiation with the Sun and with space, which are taken as being outside the system under consideration, and therefore qualify as reversible entropy changes. The exchange of radiation with the Sun and with space provides an important constraint on the total entropy production on the climate system, which will be explored further in Section 10.4.

The latent heating processes $\dot{q}_L$, the fourth contribution in Eq. 10.30, requires more careful thought. As explained in Section 6.1, any phase change in a parcel in local thermodynamic equilibrium is reversible. Such a phase change does not produce heat, it just transfers heat between the vapour and the liquid (or ice) phase in the same parcel.

We therefore need to introduce irreversible processes to produce diabatic heating. The most common processes are the removal of water from the parcel via precipitation and any disequilibrium between the vapour and the liquid. Ice processes behave in an analogous fashion.

Precipitation irreversibly heats the parcel from which the liquid water is removed. Consider a saturated parcel, with no water, moving pseudo-adiabatically up over height δz. It will reduce its temperature by $\Gamma_s\, \delta z$, with Γ_s the pseudo-adiabatic lapse rate. Some of the water vapour will have condensed into liquid water and precipitated out. Now move the parcel down to the original level; because the parcel is now subsaturated it will follow the dry adiabat and increase its temperature by $\Gamma_d\, \delta z$. In this precipitation process the parcel will therefore have increased its temperature by $(\Gamma_d - \Gamma_s)\, \delta z$ at the same pressure. The specific entropy increase is, approximately,

$$\delta s = c_p\, (\Gamma_d - \Gamma_s)\, \delta z / T. \tag{10.38}$$

Because of the precipitation, the heat that is released by condensation to a liquid is not used to evaporate the same liquid again. This process is irreversible and increases the entropy of the parcel. The ultimate source of the irreversibility is the gravitational separation of liquid water and air parcel. The entropy production can be best described in terms of the precipitation rate. The entropy production due to precipitation is approximately

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\frac{L}{T}\frac{\mathrm{d}q_l}{\mathrm{d}t}, \tag{10.39}$$

with $-\mathrm{d}q_l/\mathrm{d}t$ the precipitation rate expressed as a reduction in liquid water concentration by mass.

The other mechanism of irreversible entropy production is a disequilibrium between the vapour and the liquid. This can have several forms. Firstly, the liquid and the vapour can have different temperatures. In this case heat conduction will occur in the system with the usual consequences for the entropy budget. Secondly, the vapour can be at a partial pressure different from the saturation pressure. In this case the specific Gibbs function (or chemical potential) for the vapour and the liquid are different; the difference between the Gibbs functions for the vapour and the liquid is sometimes called the *affinity of vaporization*.[61] On evolution to equilibrium, the chemical potentials become equal and the affinity of vaporization vanishes. This evolution to equilibrium is an entropy increasing process, see Appendix B.

The above discussion of disequilibrium between vapour and liquid is rather incomplete. A more comprehensive description would require a precise definition of the processes that lead to the establishment of equilibrium. These

---

[61] See Dutton, J. A (1995) *Dynamics of atmospheric motion*. Dover, New York.

would include diffusion or advective transport of vapour and heat. In any case, we find that all diabatic changes in entropy can be written as a diabatic entropy flux and a diabatic entropy generation that is positive definite. It may be worth noting that not all entropy generating processes involve diabatic heating. An example is the increase in mixing entropy when a tracer diffuses into a liquid or a gas, see Chapter 8.

As discussed in Chapter 2, a system in thermodynamic equilibrium is defined by maximizing the entropy. Maximization of entropy under different constraints is discussed in Appendix B. One of the great open challenges of thermodynamics is to find an analogous formalism for systems out of equilibrium. Here we will discuss one such formalism.

Consider a system where only the conductive heat flux $\boldsymbol{F}_q$ plays a role. The local temperature change in such a system is given by

$$\rho c\, \partial T/\partial t = -\nabla \cdot \boldsymbol{F}_q, \tag{10.40}$$

with c the local specific heat capacity. Let us assume that the conductive heat flux is given by

$$\boldsymbol{F}_q = \kappa \nabla (1/T), \tag{10.41}$$

with $\kappa > 0$ and possibly a function of space. For small relative variations in T this expression becomes equivalent to the more familiar version of Eq. 10.33 (note that κ has a different dimension to K in Eq. 10.33). Fixing the temperature on the boundary and using the Gauss theorem, the rate of change of the irreversible entropy production can now be proven to be

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}_i S}{\mathrm{d}t}\right) = \int_V \frac{2}{T^2}\frac{\partial T}{\partial t} \nabla \cdot \boldsymbol{F}_q \,\mathrm{d}V. \tag{10.42}$$

From Eq. 10.40 it now becomes clear that the irreversible entropy production rate $\mathrm{d}_i S/\mathrm{d}t$ reduces over time until the local heating rate vanishes. It also means that the steady state is a state of minimum entropy production. This is a special instance of the *Prigogine theorem,* which states that a thermodynamic system evolves to a state of minimum entropy production if the system has linear relationships between fluxes and their driving gradients and if it has fixed boundary conditions.

The Prigogine theorem is sometimes invoked as a possible extremum principle that can be used to make non-trivial statements about systems that are not in equilibrium. However, even for the heat conduction case the general applicability of the theorem hinges on the precise way the heat conduction depends on the temperature gradients. It is the lack of generality, even in the simple heat conduction case, which makes the Prigogine theorem of limited use in climate science. There are generalizations that are more widely applicable but that provide only weak constraints on the system. We will not discuss those generalizations here.[62]

10.3 THERMODYNAMICS OF FORCED DISSIPATIVE SYSTEMS

The first law for a volume of fluid is

$$\mathrm{d}U/\mathrm{d}t = \dot{Q} + \dot{W} \tag{10.43}$$

with U the total internal energy, $\dot{Q}$ the heating rate, and $\dot{W}$ the work rate (performed on the volume). More specifically, for a volume that only exchanges heat and pressure work with the environment,

$$\dot{Q} = -\int_A \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}}\,\mathrm{d}A \quad \text{and} \quad \dot{W} = -\int_A p\boldsymbol{U} \cdot \hat{\boldsymbol{n}}\,\mathrm{d}A. \tag{10.44}$$

Here $\boldsymbol{F}_q$ is the heat flux, $\boldsymbol{U}$ is the local flow velocity, and $\hat{\boldsymbol{n}}$ the unit normal vector pointing outward from the area A surrounding the volume. Using Gauss' theorem, these surface integrals can be transformed into local energy budgets as discussed in Section 10.1.

We now write the normal heat flux $\boldsymbol{F}_q \cdot \hat{\boldsymbol{n}}$ as the sum of an inward directed and an outward directed flux, defined as follows:

$$F_{\text{in}} = \begin{cases} 0 & \text{if } \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} > 0 \\ -\boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} & \text{if } \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} < 0 \end{cases} \qquad F_{\text{out}} = \begin{cases} \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} & \text{if } \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} > 0 \\ 0 & \text{if } \boldsymbol{F}_q \cdot \hat{\boldsymbol{n}} < 0 \end{cases} \tag{10.45}$$

Both F_{in} and F_{out} are positive. We can write the heating rate now as

$$\dot{Q} = \int_A F_{\text{in}}\,\mathrm{d}A - \int_A F_{\text{out}}\,\mathrm{d}A. \tag{10.46}$$

The entropy budget can be written in an analogous fashion. Recall from the previous section that the change in total entropy S can be written as

$$\mathrm{d}S/\mathrm{d}t = \mathrm{d}_e S/\mathrm{d}t + \mathrm{d}_i S/\mathrm{d}t \tag{10.47}$$

with

$$\frac{\mathrm{d}_e S}{\mathrm{d}t} = -\int_A \frac{\boldsymbol{F}_q \cdot \hat{\boldsymbol{n}}}{T}\,\mathrm{d}A \quad \text{and} \quad \frac{\mathrm{d}_i S}{\mathrm{d}t} \geq 0. \tag{10.48}$$

We need not specify the origins of the irreversible entropy production here; we only use the second law result that it must be non-negative. Using the

---

[62] See Kondepudi, D. & Prigogine, I. (1998) *Modern Thermodynamics* J. Wiley & Sons, Chichester.

above definitions for F_{in} and F_{out} we can write for the reversible entropy change

$$\frac{\mathrm{d}_e S}{\mathrm{d}t} = \frac{1}{T_{\text{in}}} \int_A F_{\text{in}}\, \mathrm{d}A - \frac{1}{T_{\text{out}}} \int_A F_{\text{out}}\, \mathrm{d}A, \tag{10.49}$$

where we have defined the temperatures T_{in} and T_{out} as

$$\frac{1}{T_{\text{in}}} = \left(\int_A F_{\text{in}}\, \mathrm{d}A \right)^{-1} \int_A \frac{F_{\text{in}}}{T}\, \mathrm{d}A, \tag{10.50}$$

with the analogous equation for T_{out}, replacing F_{in} by F_{out}. So T_{in} is the weighted harmonic mean[63] temperature where the system receives heat from the environment, and T_{out} is the weighted harmonic mean temperature where the system loses heat to the environment.

These formulations become particularly useful when considering a system that, on average, is steady. Its energy will be constant and, because entropy is a state variable, its entropy will be constant. In this case we have, on average,

$$\dot{L} = \dot{Q} \quad \text{and} \quad \mathrm{d}_e S/\mathrm{d}t = -\mathrm{d}_i S/\mathrm{d}t, \tag{10.51}$$

where we have introduced the *work output*, $\dot{L} = -\dot{W}$. Now substituting Eq. 10.46 for $\dot{Q}$, we can write F_{out} in terms of $\dot{L}$ and F_{in}. Substituting the ensuing expression for F_{out} and Eq. 10.49 in $\mathrm{d}_e S/\mathrm{d}t$, above, we find

▶ $$\dot{L} = \left(1 - \frac{T_{\text{out}}}{T_{\text{in}}} \right) \int_A F_{\text{in}}\, \mathrm{d}A - T_{\text{out}} \frac{\mathrm{d}_i S}{\mathrm{d}t}. \tag{10.52}$$

Because $\mathrm{d}S_i/\mathrm{d}t \geq 0$, the theoretical maximum output of work $\dot{L}_{\text{rev}}$ is achieved when the system is reversible, and it equals

▶ $$\dot{L}_{\text{rev}} = \left(1 - \frac{T_{\text{out}}}{T_{\text{in}}} \right) \int_A F_{\text{in}}\, \mathrm{d}A. \tag{10.53}$$

This is the continuous version of a classical result by Carnot that states that the work output of a heat engine as a fraction of its heat input has an upper bound and that this upper bound is achieved for reversible systems.

The factor relating the maximum work output to the heat input is called the *Carnot efficiency,* η,

▶ $$\eta = 1 - T_{\text{out}}/T_{\text{in}}. \tag{10.54}$$

We encountered it in Chapter 2 in the problem about the Stirling engine. For simple heat engines, which operate between two fixed temperatures, the generalized version here reduces to the classical Carnot expression.

---

[63] The harmonic mean c of a set of N numbers a_i is defined as $c^{-1} = N^{-1} \sum_i a_i^{-1}$.

The difference between the actual work output and the maximum theoretical work output is sometimes called the *lost useful work output,* $\dot{L}_{\text{lost}}$. From the above expressions we have

$$\dot{L}_{\text{lost}} = \dot{L}_{\text{rev}} - \dot{L} = T_{\text{out}}\, \mathrm{d}_i S/\mathrm{d}t \geq 0. \tag{10.55}$$

In other words, the system destroys useful work at a rate which is proportional to its irreversible entropy production. This is the continuous version of the *Guoy–Stodola theorem,* see Appendix B.

Because $\dot{L}_{\text{rev}}$ represents an upper bound to the work output, we can see that the system cannot perform work when $T_{\text{out}} > T_{\text{in}}$. When this is the case, the system cannot maintain motion against dissipative, irreversible processes. We conclude that for motion to be maintained in a fluid, such as the atmosphere or the ocean, it must receive its energy at temperatures higher than those at which it loses its energy. We have already encountered this property in Section 10.1.

10.4 CLIMATE THERMODYNAMICS

Ignoring secular climate trends, the Earth system is on average in a steady state. Because it can only interact with space through radiative heat fluxes, the steady state condition, Eq. 10.51, reduces to

$$\dot{R}_{\text{in}} = \dot{R}_{\text{out}} \quad \text{and} \quad \mathrm{d}_e S/\mathrm{d}t = -\mathrm{d}_i S/\mathrm{d}t, \tag{10.56}$$

where we have defined the total radiative heat input and output rates

$$\dot{R}_{\text{in}} = \int_A F_{\text{in}}\, \mathrm{d}A \quad \text{and} \quad \dot{R}_{\text{out}} = \int_A F_{\text{out}}\, \mathrm{d}A. \tag{10.57}$$

The total short-wave radiative heat input rate $\dot{R}_{\text{in}}$ is set by the solar constant S_0, the mean albedo α, and the area of the Earth disk πR_E^2 (R_E is the Earth's radius),

$$\dot{R}_{\text{in}} = (1-\alpha)\, S_0 \pi R_E^2 = 122\,\text{PW}, \tag{10.58}$$

($1\,\text{PW} = 10^{15}\,\text{W}$). Per unit area on Earth, this corresponds to $239\,\text{W}\,\text{m}^{-2}$. The long-wave radiative heat output rate $\dot{R}_{\text{out}}$ has the same value, on average.

The irreversible entropy production in the Earth system can be estimated from the expression for $\mathrm{d}_e S/\mathrm{d}t$, Eq. 10.49. In this case this expression reduces to

$$\frac{\mathrm{d}_i S}{\mathrm{d}t} = -\frac{\mathrm{d}_e S}{\mathrm{d}t} = \dot{R}_{\text{in}} \left(\frac{1}{T_{\text{out}}} - \frac{1}{T_{\text{in}}} \right). \tag{10.59}$$

The precise determination of T_{in} and T_{out} is difficult but we can give estimates. As a first approximation we can argue that the radiation is absorbed or emitted at either the surface temperature of 288 K or the atmospheric radiation temperature of 255 K. It turns out that about 1/3 of the incoming short-wave radiation is absorbed by the atmosphere and 2/3 is absorbed by the surface. About 1/6 of the outgoing long-wave radiation is emitted by the surface and the rest is emitted by the atmosphere.[64] These fractions can be used in the definitions of T_{in} and T_{out}, Eq. 10.50, to find

$$T_{\mathrm{in}} \approx 276\,\mathrm{K} \quad \text{and} \quad T_{\mathrm{out}} \approx 260\,\mathrm{K}. \tag{10.60}$$

The total irreversible entropy production in the Earth system (commonly expressed per unit area on Earth) is now

$$\mathrm{d}_i S/\mathrm{d}t \approx 53\,\mathrm{mW\,m^{-2}\,K^{-1}}. \tag{10.61}$$

This estimate is in line with much more complex analyses of general circulation models.[65]

The irreversible entropy production, above, is also referred to as the *material* entropy production because it refers to the actual energy and entropy throughput of the material components of the Earth system. However, one could argue that the short-wave radiation enters the Earth system at the Solar radiation temperature of 5780 K and leaves it at the Earth's radiation temperature 255 K. If we use these values for T_{in} and T_{out} we find an irreversible entropy production of 896 $\mathrm{mW\,m^{-2}\,K^{-1}}$. The vast majority of this entropy production occurs at the point of short-wave absorption in the Earth system. It has no influence on the material workings of the climate system; the entropy production resides in the photon field. In fact, this number underestimates the photon contribution to the entropy production by a factor 4/3, the factor that relates the energy content and the entropy content of the radiation field, see Eq. 9.91. Unfortunately, the literature contains much confusing discussion regarding this matter and we will not pursue it further here. Suffice it to say that the entropy production occurring in the photon field is immaterial to the workings of the climate engine.

Next we will consider just one process contributing to the irreversible entropy production in the Earth system: the meridional heat flux. As discussed in Section 9.4, the absorbed radiation in the equatorial regions is substantially larger than that in the polar regions. The region equatorward of 30 degrees north or south spans half the Earth's surface yet it receives about 2/3 of the incoming absorbed radiation, that is, about 80 PW; the other half of the Earth's surface, poleward of 30 degrees latitude, receives about 40 PW. The total heat input is therefore about 120 PW, as before. Most of this energy

---

[64] See Trenberth, K. E., Fasullo, J. T. and Kiehl, J. (2009) *Bull. Amer. Meteor. Soc.* **90**, 311–324.

[65] See Pascale et al., (2010) *Clim. Dyn.*, doi:10.1007/s00382-009-0781-1.

is radiated away to space in the same region. However, about 10 PW is first transported from the equatorial region to the polar regions, see Figure 9.6 (note that the poleward transport occurs in both hemispheres).

If the meridional heat transport were to increase to 20 PW then the equatorial region and the polar regions would receive the same amount of energy, about 60 PW. Their radiation temperature would have to be the same, about 255 K, to radiate this energy input back to space. For larger meridional heat transports, the radiation temperature in the polar regions would become higher than in the equatorial region. Assuming that the radiation temperature is some increasing function of the mean temperature in each region, the climate engine would then transport heat from lower to higher temperatures. This is prohibited by the second law. We conclude that the theoretical maximum meridional heat transport on Earth is about 20 PW.

Here we will describe a simple two-box model of the meridional heat transport in the Earth system where the two boxes correspond to the equatorial region and the polar regions. Figure 10.1 gives a schematic representation of the two-box model. In order to estimate the entropy production due to the meridional heat transport, we need to know at what temperature each region operates. As a crude estimate we will use the mean radiation temperature for each region. The radiation temperature is set by the balance between radiative heat input and output. The total heat input $\dot{Q}_{\mathrm{in},i}$ is set by the radiative input $\dot{R}_{\mathrm{in},i}$ for each region i and the poleward meridional heat flux J,

$$\dot{Q}_{\mathrm{in},1} = \dot{R}_{\mathrm{in},1} - J \quad \text{and} \quad \dot{Q}_{\mathrm{in},2} = \dot{R}_{\mathrm{in},2} + J, \tag{10.62}$$

where region 1 is the equatorial region and region 2 is the polar region. For the Earth we have

$$\dot{R}_{\mathrm{in},1} = 80\,\mathrm{PW} \quad \text{and} \quad \dot{R}_{\mathrm{in},2} = 40\,\mathrm{PW}. \tag{10.63}$$

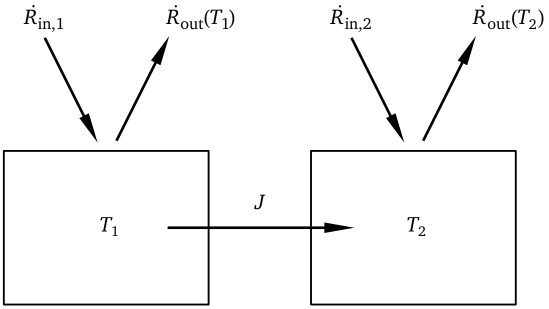


FIGURE 10.1 Schematic of the two-box model of the meridional heat transport in the Earth system. The equatorial region is represented by a box at radiation temperature T_1, the polar region by a box at radiation temperature T_2. The incoming radiation $\dot{R}_{\mathrm{in},i}$ is fixed, the outgoing radiation $\dot{R}_{\mathrm{out}}$ is a function of the radiation temperature. The meridional heat transport is represented by a single energy flux J.

For computational simplicity, we approximate the radiative output $\dot{R}_{\text{out},i}$ in each region i by a linear function of the radiation temperature. The appropriate linearization of Stefan's law is

$$\dot{R}_{\text{out},i} = \frac{\dot{R}_{\text{in}}}{2}\left(\frac{T_b + 4\,(T_i - T_b)}{T_b}\right), \tag{10.64}$$

with $\dot{R}_{\text{in}} = \dot{R}_{\text{in},1} + \dot{R}_{\text{in},2}$ the global radiative heat input, T_b the global mean radiation temperature, and T_i the radiation temperature in region i. For Earth we have

$$\dot{R}_{\text{in}} = 120\,\text{PW} \quad \text{and} \quad T_b = 255\,\text{K}. \tag{10.65}$$

The above linear approximation produces a radiative output of $\dot{R}_{\text{in}}/2 = 60\,\text{PW}$ per region when the radiation temperature is $T_b = 255\,\text{K}$, and it has the same variation with temperature as Stefan's law at this point. It is possible to use linear fits to real data to get more accurate versions for the Earth system.

The system is in equilibrium when the heat input is the same as the heat output for each region,

$$\dot{Q}_{\text{in},i} = \dot{R}_{\text{out},i}. \tag{10.66}$$

Substituting Eqs. 10.62 and 10.64 we can solve for the equilibrium temperature in each region as a function of meriodional heat flux J,

$$T_1 = T_b\left(1 + \frac{1}{4}\left(\frac{\dot{R}_{\text{in},1} - J - \dot{R}_{\text{in}}/2}{\dot{R}_{\text{in}}/2}\right)\right), \tag{10.67a}$$

$$T_2 = T_b\left(1 + \frac{1}{4}\left(\frac{\dot{R}_{\text{in},2} + J - \dot{R}_{\text{in}}/2}{\dot{R}_{\text{in}}/2}\right)\right). \tag{10.67b}$$

Figure 10.2 shows the two temperatures as a function of J for Earth parameters. As expected, for the maximum poleward heatflux J_{max}, given by

$$J_{\text{max}} = (\dot{R}_{\text{in},1} - \dot{R}_{\text{in},2})/2, \tag{10.68}$$

we find that the radiation temperatures in the two regions become the same. For Earth parameters, we have $J_{\text{max}} = 20\,\text{PW}$.

The entropy production $\text{d}_iS/\text{d}t$ due to the heat flux can be estimated as

$$\frac{\text{d}_iS}{\text{d}t} = J\left(\frac{1}{T_2} - \frac{1}{T_1}\right), \tag{10.69}$$

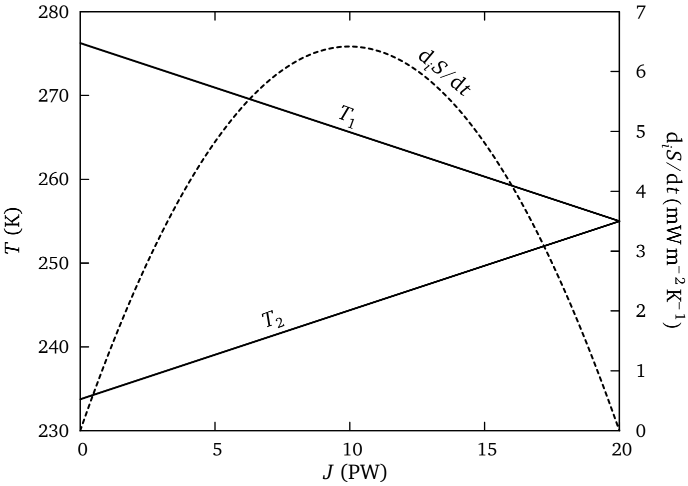


FIGURE 10.2 Results for the two-box model of the Earth system for varying poleward heat flux J. Solid lines: radiation temperature of equatorial region, T_1, and polar region, T_2. Dashed line: irreversible entropy production due to the poleward meridional heat flux. The entropy production is expressed per unit area on Earth.

which is equivalent to the expression for the heat conduction example of Section 2.3, Eq. 2.32. Using the above expressions for T_1 and T_2 we find

$$\frac{\mathrm{d}_iS}{\mathrm{d}t} = \frac{J\,(J_{\max} - J)}{\dot{R}_{\mathrm{in}}} \frac{T_b}{T_1 T_2} \approx \frac{J\,(J_{\max} - J)}{\dot{R}_{\mathrm{in}} T_b}, \tag{10.70}$$

where it should be noted that both T_1 and T_2 are also functions of J. The approximation on the right-hand side is better than one part in a hundred for the relevant range of values for J. It can be seen that the irreversible entropy production vanishes when $J = 0$ or when $J = J_{\max}$. In the intermediate range, the entropy production is positive, with a maximum production rate when $J \approx J_{\max}/2$, see Fig 10.2.

For Earth parameters, the maximum of entropy production of about 6.5 mW m^{-2} K^{-1} occurs for $J \approx 10$ PW. This value of the poleward heat flux is close to the sum of the observed values at 30° N and 30° S. This coincidence may well be fortuitous. However, problem 10.3 shows how maximum entropy production can be used to predict the sensible and latent heat flux between the surface and the atmosphere. Also, much more complicated models have been produced that appear to indicate that observed fluxes are close to those that maximize the production of entropy.[66] Maximum entropy production is perhaps the principle of non-equilibrium thermodynamics that parallels

[66] Paltridge, G. W. (1975) *Quart. J. Roy. Met. Soc.* **101**, 475–484; Ozawa, H. et al. (2003) *Rev. Geoph.* **41(4)**, 1–24; Kleidon, A. & Lorenz, R., eds. (2005) *Non-equilibrium thermodynamics and the production of entropy.* Springer, Berlin.

the maximum entropy principle of equilibrium thermodynamics. However, as yet, no convincing theoretical grounding exists and competing principles have been formulated.

The thermodynamics of the Earth system presents one of the great challenges of physics.

PROBLEMS

10.1. Rainfall rates. Equations 10.38 and 10.39 can be used to get an interesting estimate of typical rainfall rates. We get a low estimate of the entropy production rate if we assume that the up-and-down motion leading to Eq. 10.38 is produced by a buoyancy oscillation. Then the entropy production can be estimated as

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{c_p}{T}\left(\Gamma_d - \Gamma_s\right) N\delta z,$$

with δz the amplitude of the buoyancy oscillation. Combine this equation with Eq. 10.39 to show that a typical rainfall rate for such motion, $\dot{r}$, expressed in $\mathrm{mm\,h^{-1}}$, is

$$\dot{r} = \left(3.6 \times 10^6\right) \frac{\rho g N \delta z H}{\rho_l L} \left(1 - \frac{\Gamma_s}{\Gamma_d}\right)$$

with ρ the density of air in the cloud region, ρ_l the density of liquid water, and H the depth of the cloud. Does this equation produce realistic rainfall rates? How can this equation be improved?

10.2. Use Gauss' theorem to show that the global budgets of Eqs. 10.43 and 10.44 lead to the local budgets of Section 10.1.

10.3. Vertical heat fluxes and maximum entropy production. Consider the cloud-free single slab model of the atmosphere discussed in Section 9.3. Now include in the budget a latent and sensible heat flux H from the surface to the slab. Show that, if the long-wave emissivity of the slab is $\epsilon_{\mathrm{LW}} = 1$, the entropy production of the latent and sensible heat flux is maximal when $H \approx 133\,\mathrm{W\,m^{-2}}$. At this value we find a surface temperature $T_E = 280\,\mathrm{K}$ and an entropy production $\mathrm{d}_i S/\mathrm{d}t = 46\,\mathrm{mW\,m^{-2}\,K^{-1}}$. The observed value of H is about $100\,\mathrm{W\,m^{-2}}$ and of the surface temperature T_E is about 288 K. The entropy production is in line with what complex general circulation models predict. What happens if we change the long-wave emissivity of the slab to a more realistic $\epsilon_{\mathrm{LW}} = 0.8$?

Appendix A

Functions of several variables

Here we review some aspects of partial differentiation that are of relevance to applications in thermodynamics. Consider a functional relationship between variables x, y, and z (the arguments here can be straightforwardly extended to more variables)

$$F(x, y, z) = 0. \tag{A.1}$$

An equation of state is an example of such a relationship; for an ideal gas we have $F(p, v, T) = pv - RT$. Assuming F to be suitably smooth, we can, at least locally, define z to be a function of the two variables x and y, $z = z(x, y)$ (this is formalized in the *implicit function theorem*). We can use a Taylor expansion to calculate what happens to z if we perturb the variables by a small amount δx and δy:

$$z(x + \delta x, y + \delta y) = z(x, y) + \left(\frac{\partial z}{\partial x}\right)_y \delta x + \left(\frac{\partial z}{\partial y}\right)_x \delta y + \ldots \tag{A.2}$$

The dots here indicate terms of higher order in δx or δy that are small when variations are small. The subscripts indicate explicitly which parameters are to be kept constant in the differentiation; this is useful for thermodynamic applications, where it is common to use different combinations of variables. For infinitesimal variations $\mathrm{d}x$ and $\mathrm{d}y$ in variables x and y the infinitesimal difference $\mathrm{d}z$ (also called the *differential* of z) becomes

$$\mathrm{d}z = z(x + \mathrm{d}x, y + \mathrm{d}y) - z(x, y) = \left(\frac{\partial z}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial z}{\partial y}\right)_x \mathrm{d}y \tag{A.3}$$

This equation is really a shorthand notation for expressing the local dependence of z on x and y. A typical use of this shorthand notation is the following: suppose variables x and y are functions of time t. The derivative of z with

---

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

respect to t then is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \left(\frac{\partial z}{\partial x}\right)_y \frac{\mathrm{d}x}{\mathrm{d}t} + \left(\frac{\partial z}{\partial y}\right)_x \frac{\mathrm{d}y}{\mathrm{d}t}. \tag{A.4}$$

Equation A.3 expresses in an efficient way that z is a function of x and y only and how z varies locally with variations in x or y.

From Eq. A.1, we can also express x locally as a function of y and z, $x = x(y, z)$. The differential of x then is

$$\mathrm{d}x = \left(\frac{\partial x}{\partial y}\right)_z \mathrm{d}y + \left(\frac{\partial x}{\partial z}\right)_y \mathrm{d}z. \tag{A.5}$$

Substituting Eq. A.3 in Eq. A.5 we find

$$\mathrm{d}x = \left(\frac{\partial x}{\partial y}\right)_z \mathrm{d}y + \left(\frac{\partial x}{\partial z}\right)_y \left[\left(\frac{\partial z}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial z}{\partial y}\right)_x \mathrm{d}y\right]. \tag{A.6}$$

This equation can be rearranged to get

$$\left[\left(\frac{\partial x}{\partial z}\right)_y \left(\frac{\partial z}{\partial x}\right)_y - 1\right] \mathrm{d}x + \left[\left(\frac{\partial x}{\partial y}\right)_z + \left(\frac{\partial x}{\partial z}\right)_y \left(\frac{\partial z}{\partial y}\right)_x\right] \mathrm{d}y = 0. \tag{A.7}$$

As this has to be true for any dx and dy, the terms between the square brackets each have to vanish. This then leads to the *reciprocal theorem,*

► $$\left(\frac{\partial x}{\partial z}\right)_y \left(\frac{\partial z}{\partial x}\right)_y = 1, \tag{A.8}$$

and the *reciprocity theorem,*

► $$\left(\frac{\partial x}{\partial y}\right)_z = -\left(\frac{\partial x}{\partial z}\right)_y \left(\frac{\partial z}{\partial y}\right)_x. \tag{A.9}$$

Finally, it can be verified by Taylor expansion that for multiple derivatives the order of differentiation is unimportant,

► $$\frac{\partial}{\partial y}\left(\left(\frac{\partial z}{\partial x}\right)_y\right)_x = \frac{\partial}{\partial x}\left(\left(\frac{\partial z}{\partial y}\right)_x\right)_y. \tag{A.10}$$

This property is used in the derivation of the Maxwell relations in Chapter 3.

Appendix B
Exergy and thermodynamic stability

Given the total internal energy U and the volume V of a system, the second law states that a system will move to a state of maximum total entropy S. This defines the equilibrium state of the system: any spontaneous internal rearrangement of the system would correspond to an increase in entropy, by the second law. However, if the entropy is maximum, no such rearrangements can exist. The system must therefore be stable. In this section we will examine how to generalize this argument to understand thermodynamic stability if there are different constraints.

Consider a closed system interacting with its environment through heat fluxes $\mathrm{d}Q$ and work exchange $\mathrm{d}W$. The environment is assumed to be 'large': its temperature and pressure will not change through interactions with the system. The environment is a heat and work reservoir at fixed temperature T_0 and pressure p_0. Engineers sometimes call this environment 'the atmosphere'. For most engineering applications, systems are in contact with an atmosphere that does not change its temperature or pressure while interacting with the system. This is also the case for an air parcel interacting with its environment.

We first assume that the system under consideration (engine, air parcel, drop) is not necessarily in thermal equilibrium with its environment. It will be characterized by its own intensive properties, temperature T and pressure p, and its extensive properties, volume V and entropy S.

On any interaction with the environment the system will change its internal energy U by

$$\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W, \tag{B.1}$$

with $\mathrm{d}Q$ the heat exchange and $\mathrm{d}W$ the work exchange with the environment. The second law in the form of Eq. 2.28 defines the heat exchange $\mathrm{d}Q$ in terms of the entropy change of the system,

$$T_0\,\mathrm{d}S = \mathrm{d}Q + T_0\,\mathrm{d}_iS, \tag{B.2}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

with $\mathrm{d}_i S$ the irreversible entropy change of the system, which, according to the second law, is always non-negative, $\mathrm{d}_i S \geq 0$. The relevant temperature is T_0 because the exchange is with the environment at fixed temperature.

For environmental applications we only consider reversible exchange of work with the environment. This means that the work exchange $\mathrm{d}W$ can be written

$$\mathrm{d}W = -p_0 \, \mathrm{d}V. \tag{B.3}$$

Again, the relevant pressure for the work exchange is p_0, the fixed pressure of the environment.

We can now combine the above equations to

$$\mathrm{d}U - T_0 \, \mathrm{d}S + p_0 \, \mathrm{d}V = -T_0 \, \mathrm{d}_i S \leq 0. \tag{B.4}$$

This can be compactly written as the differential of the *exergy*, A,[67]

$$\blacktriangleright \qquad A = U - T_0 S + p_0 V, \tag{B.5}$$

so that

$$\blacktriangleright \qquad dA = \mathrm{d}U - T_0 \, \mathrm{d}S + p_0 \, \mathrm{d}V. \tag{B.6}$$

We have

$$\mathrm{d}A = -T_0 \, \mathrm{d}_i S \leq 0. \tag{B.7}$$

For a given environment, the exergy is a state variable of the system. As a consequence of the second law, the exergy of a system in contact with its environment can only reduce. The system is at equilibrium with its environment if its exergy is minimum. So for a system in equilibrium with its environment we have

$$\mathrm{d}A = 0 \qquad \text{at equilibrium.} \tag{B.8}$$

This equation allows us to define conditions for equilibrium in a given environment.

Let us assume the system is in local thermodynamic equilibrium, although not necessarily in equilibrium with its environment. For such systems, the variation in internal energy is given by the first law

$$\mathrm{d}U = T \, \mathrm{d}S - p \, \mathrm{d}V. \tag{B.9}$$

---

[67] There are various synonyms for exergy. The most commonly encountered synonym is availability, hence the symbol A.

Combining this equation with the differential of the exergy, Eq. B.6, we find

$$\mathrm{d}A = (T - T_0)\,\mathrm{d}S - (p - p_0)\,\mathrm{d}V. \tag{B.10}$$

This expression is used to define equilibrium under four different conditions.

Case 1: the system is thermally isolated and has a fixed volume. Because in Eq. B.10 the two contributions to the variations in A are independent, both must vanish identically. The last term vanishes because we set $\mathrm{d}V = 0$. Because the system is thermally isolated, the temperature of the system is not bound to the temperature of the environment. For $\mathrm{d}A$ to vanish we therefore must have $\mathrm{d}S = 0$. In this case, the equilibrium condition $\mathrm{d}A = 0$ is equivalent to $\mathrm{d}U = 0$. The equilibrium conditions therefore are

$$\mathrm{d}S = 0, \qquad \mathrm{d}V = 0, \qquad \mathrm{d}U = 0. \tag{B.11}$$

Case 2: the system is in thermal contact with the environment but has fixed volume. The thermal contact leads to an equilibrium condition of $T = T_0$ and therefore $\mathrm{d}T = 0$. We also have $\mathrm{d}V = 0$. In this case, the equilibrium condition reduces to $\mathrm{d}A = \mathrm{d}U - T\,\mathrm{d}S = \mathrm{d}F = 0$, because T is constant (F is the Helmholtz free energy). The equilibrium conditions for this case are

$$\mathrm{d}T = 0, \qquad \mathrm{d}V = 0, \qquad dF = 0. \tag{B.12}$$

Case 3: the system is thermally isolated and remains isobaric. Thermal isolation requires $\mathrm{d}S = 0$, as in case 1. The isobaric condition requires that at equilibrium $p = p_0$, and therefore $\mathrm{d}p = 0$. The equilibrium condition becomes $\mathrm{d}A = \mathrm{d}U + p\,\mathrm{d}V = \mathrm{d}H = 0$, because p is constant (H is the enthalpy). The equilibrium conditions now are

$$\mathrm{d}S = 0, \qquad dp = 0, \qquad dH = 0. \tag{B.13}$$

Case 4: the system is in thermal contact with the environment and remains isobaric. For equilibrium we now need $T = T_0$ and $p = p_0$. The equilibrium condition is $\mathrm{d}A = \mathrm{d}U - T\,\mathrm{d}S + p\,\mathrm{d}V = \mathrm{d}G = 0$ (G is the Gibbs function). We now have

$$\mathrm{d}T = 0, \qquad dp = 0, \qquad dG = 0. \tag{B.14}$$

We find that each thermodynamic potential occurs with its own natural variables. The above analysis shows how to apply the maximum entropy condition under different external constraints.

In an engineering context, the exergy is a measure of how much useful work a system can perform in a given environment.[68] This is the origin of the

---

[68] See Bejan, A. (2006) *Advanced Engineering Thermodynamics*, 3rd edn. J. Wiley & Sons, Hoboken.

synonym ‘availability’. Suppose the system can produce work in other ways than just expansion against the environment. The *useful work output* dL of a system then is defined as

$$\mathrm{d}L = -\mathrm{d}W - p_0\,\mathrm{d}V. \tag{B.15}$$

The useful work output is the total work output of the system $-\mathrm{d}W$ minus the work spent by expansion against the environment $p_0\,\mathrm{d}V$. Using the first law, Eq. B.1, the second law, Eq. B.2, and the definition of exergy, Eq. B.5 this can be rearranged to

$$\mathrm{d}L = -\mathrm{d}A - T_0\,\mathrm{d}_iS. \tag{B.16}$$

The first term on the right-hand side is the reversible part of the work output $\mathrm{d}L_{\mathrm{rev}} = -\mathrm{d}A$. Note that at equilibrium with the environment $\mathrm{d}A = 0$ so no useful work can be extracted. The above expression can be rewritten as the *Guoy–Stodola theorem*,

► $$dL_{\mathrm{lost}} = \mathrm{d}L_{\mathrm{rev}} - \mathrm{d}L = T_0\,\mathrm{d}_iS \geq 0. \tag{B.17}$$

The system destroys useful work at a rate which is proportional to its irreversible entropy production.

Appendix C

Thermodynamic diagrams

The thermodynamic state of a simple substance is determined by two variables. That means that a two-dimensional diagram with a particular thermodynamic variable on each of the two coordinate axes will be sufficient to describe any thermodynamic state of the substance. The *pV* diagrams in Chapter 1 and 2 are examples. As we saw in Chapter 2, the area on a *pV* diagram corresponds to (work) energy. The area inside a closed cycle *C* on a *pV* diagram is

$$\text{area} = \oint_C p\, \mathrm{d}V, \tag{C.1}$$

which also equals the total work performed by an air parcel over the cycle. We tend to reserve the phrase 'thermodynamic diagram' for two-dimensional diagrams where the area corresponds to an energy. So, for example, a diagram with T and S on the axes would also be a thermodynamic diagram with the integral of $T\,\mathrm{d}S$ being the heat input for an air parcel over a cycle.

We can construct alternative thermodynamic diagrams with area conserving transformations of *pV* diagrams or *TS* diagrams. Here we will not describe in general how such transformations may be constructed; they can range from fairly trivial ones, such as rotations or translations, to more convoluted ones, such as area preserving deformations.

In atmospheric science we often use *tephigrams*, which are rotated *Ts* diagrams (we use specific entropy here, so area corresponds to energy per unit mass), and so-called *skewT–log p diagrams*, where the coordinate axes are essentially skewed versions of T and $\ln p$. Tephigrams derive their name from the combination 'T-ϕ', with ϕ an early twentieth century notation for specific entropy. The tephigram is now the most commonly used thermodynamic diagram in Europe; in the US the skew*T*–log*p* diagram is currently more fashionable. Some local meteorological services use their own variants of thermodynamic diagrams.

The tephigram is a rotated and translated rectangular subsection of the *Ts* diagram. Figure C.1 illustrates how a tephigram maps onto a *Ts* diagram. The

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

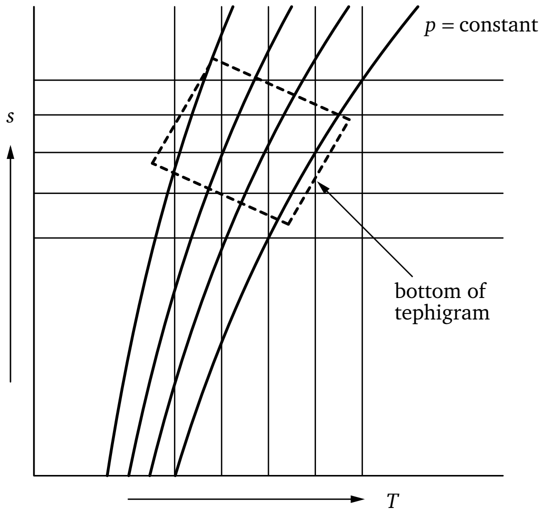


FIGURE C.1 The dashed rectangle illustrates the location of a typical atmospheric tephigram on a *Ts* diagram. The solid black lines represent isolines of the pressure. The rotation is such that the 1000 hPa pressure isoline is quasi horizontal, close to the bottom of the tephigram. The thin lines represent isolines of temperature and potential temperature, respectively.

rotation is chosen so as to make the isolines of pressure quasi-horizontal, see Figure C.2a. The surface pressure is near the bottom of the tephigram and upper tropospheric pressures are near the top.

The coordinate axes of the tephigram are most easily defined as

$$x = T_0 s + c_p T, \tag{C.2a}$$

$$y = T_0 s - c_p T, \tag{C.2b}$$

with T_0 a relevant reference temperature and c_p the ideal gas heat capacity. With this transformation, on lines parallel to the x-axis, defined by $\mathrm{d}y = 0$, the temperature and entropy satisfy

$$T_0 \,\mathrm{d}s = c_p \,\mathrm{d}T. \tag{C.3}$$

But according to Eq. 3.40 this is the differential of s for an ideal gas at constant p and when the temperature is T_0. In other words, isolines of pressure are horizontal in the tephigram when the temperature is T_0. Another advantage of this mapping is that at fixed temperature the y-coordinate is linear in $\ln(p_0/p)$, with p_0 a reference pressure. So to a good degree of accuracy, the y-coordinate corresponds to the geometric height of a parcel.

The skew*T*–log*p* diagram is strictly speaking only a thermodynamic diagram for ideal gases. A skew*T*–log*p* diagram is located similarly on a *Ts* diagram as a tephigram but it introduces an additional deformation to the rectangle so as to make the pressure isolines straight and parallel to the

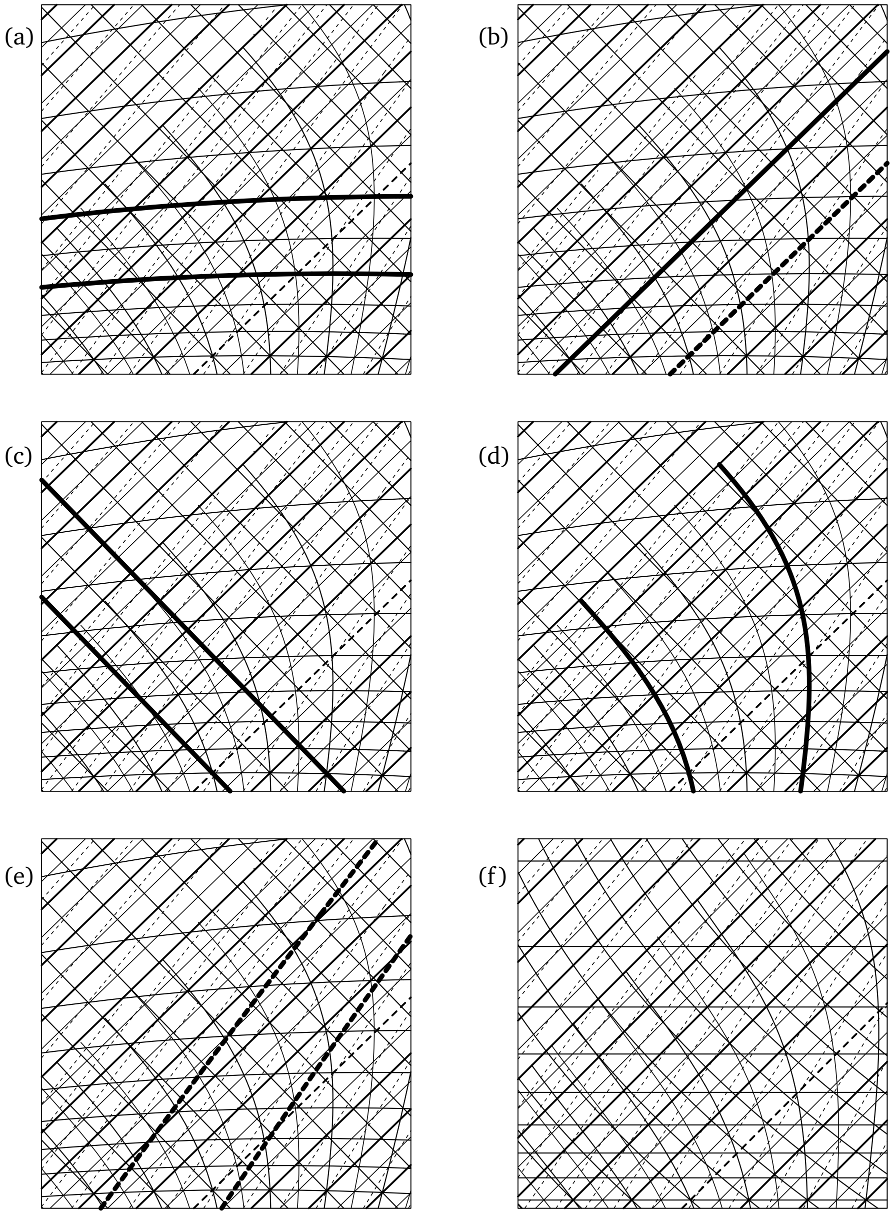


FIGURE C.2 Tephigrams with the bold lines highlighting isolines of (a) pressure, (b) temperature (0°C dashed), (c) potential temperature, (d) equivalent potential temperature, and (e) saturated water vapour mixing ratio. Panel (f) shows a skew$T$–log$p$ diagram with the corresponding isolines.

horizontal boundaries of the deformed rectangle. The coordinate axes on a skew$T$–log$p$ diagram are defined as

$$x = T/T_0 + \ln(p_0/p), \tag{C.4a}$$

$$y = \ln(p_0/p), \tag{C.4b}$$

with T_0 a fixed reference temperature which determines the skew of different isolines in the diagram and p_0 a reference pressure. Lines parallel to the x-axis, defined by $\mathrm{d}y = 0$, are isolines of pressure. Also, to a good degree of accuracy, the y-coordinate is proportional to the geometric height of a parcel. We can now see that for an ideal gas

$$x\,\mathrm{d}y = \frac{v\,\mathrm{d}p}{RT_0} + \ln(p_0/p)\,\frac{\mathrm{d}p}{p}. \tag{C.5}$$

Any integral over a closed cycle in the xy plane of the second term on the right-hand side will vanish because it is a function of p alone. Thus the integral over a closed cycle in the xy plane (that is, on a skew$T$–log$p$ diagram) will be proportional to the work performed on an ideal gas parcel (the contour integral of $v\,\mathrm{d}p$ equals the contour integral of $-p\,\mathrm{d}v$).

Tephigrams and skew$T$–log$p$ diagrams look fairly similar, see Fig C.2, and can be used interchangeably. In tephigrams, entropy isolines are straight and pressure isolines are curved; in skew$T$–log$p$ diagrams this is the other way around.

The isolines of specific entropy are by definition the dry adiabats; in a tephigram the isolines of specific entropy are normally labelled by their potential temperature, with the potential temperature corresponding to the actual temperature at the reference level of $p_0 = 1000\,\mathrm{hPa}$. The dry adiabats are highlighted in Figure C.2c.

We can use Eq. 6.12 to calculate the location of the pseudo-adiabats. These then produce a set of curved isolines on the tephigram, see Figure C.2d. Pseudo-adiabats are isolines of saturated pseudo-equivalent potential temperature θ_e.[69] The pseudo-adiabats are labeled with the wet-bulb potential temperature, θ_w, which is the temperature where the pseudo-adiabat crosses the reference pressure level of p_0.

Tephigrams furthermore show isolines of saturated mixing ratio r_{vs}, see Figure C.2e.

Figure C.3 shows a larger scale tephigram with all the isolines labeled.

So a tephigram or a skew$T$–log$p$ diagram shows isolines of: temperature T (isotherms), pressure p (isobars), potential temperature θ (dry adiabats), pseudo-equivalent potential temperature θ_e (pseudo-adiabats), and saturated water vapour mixing ratio r_{vs}. Tephigrams usually show several kinds of ad-

---

[69] Accurate equations used to construct pseudo-adiabats are given in Bolton, D. (1980) *Mon. Wea. Rev.* **108**, 1046–1053.

ditional information, but the above set of five isolines form the basis of any tephigram or its variants.

The real power of a tephigram is that it can be used to perform calculations graphically. As explained in Chapter 6, the calculation of, for example, the lifting condensation level or the convective available potential energy can be straightforwardly performed using a tephigram. In this sense a tephigram is a *nomogram,* a graphical calculation chart.

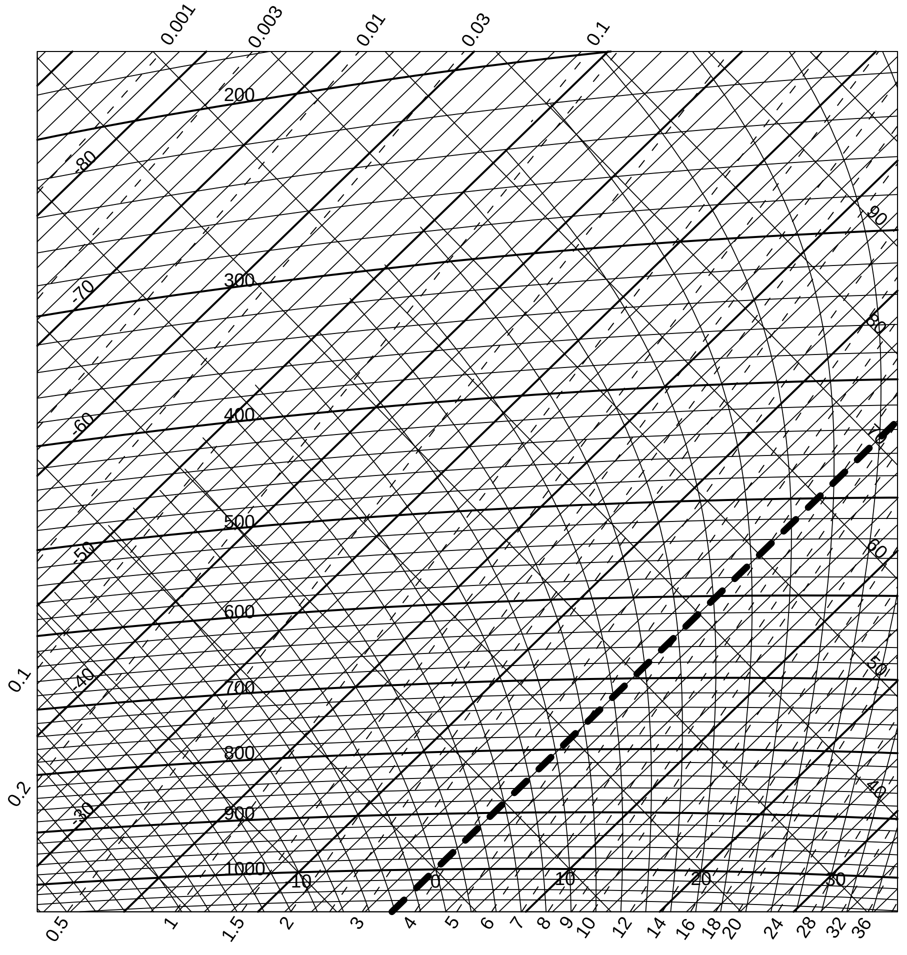


FIGURE C.3 A blank tephigram. Pressures are in hPa, temperatures in °C, and mixing ratios in g kg^{-1}.

Appendix D
Relationship between energy density and energy flux

Following the arguments leading to Eq. 1.11, we can deduce that the number of particles N with positive velocity $\mathscr{U}$ hitting the wall per unit time and per unit area equals

$$N_{\mathscr{U}} = \tilde{n}_{\mathscr{U}} \mathscr{U}, \tag{D.1}$$

with $\tilde{n}_{\mathscr{U}}\,\mathrm{d}\mathscr{U}$ the volumetric number density of particles with x velocity between $\mathscr{U}$ and $\mathscr{U} + \mathrm{d}\mathscr{U}$. Integrating this equation over positive $\mathscr{U}$ we find the total number of particles N hitting the wall per unit time and per unit area,

$$N = \int_{\mathscr{U}>0} \tilde{n}_{\mathscr{U}} \mathscr{U}\,\mathrm{d}\mathscr{U}. \tag{D.2}$$

We can write the particle density $\tilde{n}_{\mathscr{U}}$ in terms of the particle density $\tilde{n}_{\boldsymbol{U}}$ of particles with vector velocity $\boldsymbol{U}$ if we integrate out the dependence on the other velocity components, $\mathscr{V}$ and $\mathscr{W}$,

$$\tilde{n}_{\mathscr{U}} = \iint \tilde{n}_{\boldsymbol{U}}\,\mathrm{d}\mathscr{V}\,\mathrm{d}\mathscr{W}. \tag{D.3}$$

We can now write N as

$$N = \frac{1}{2} \iiint \tilde{n}_{\boldsymbol{U}} |\mathscr{U}|\,\mathrm{d}\mathscr{U}\,\mathrm{d}\mathscr{V}\,\mathrm{d}\mathscr{W}, \tag{D.4}$$

where the absolute value function allows us to include negative $\mathscr{U}$ and the resulting double counting is compensated by the factor 1/2. The particle density is a function only of particle speed, say $C = |\boldsymbol{U}|$, not its direction. We therefore find

$$N = \frac{1}{2} \int_0^\infty \tilde{n}_C \langle|\mathscr{U}|\rangle_C\,\mathrm{d}C, \tag{D.5}$$

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1

where $\langle|\mathscr{U}|\rangle_C$ is the average of $|\mathscr{U}|$ if the total speed is C. On geometric grounds it follows that

$$\langle|\mathscr{U}|\rangle_C = C/2 \tag{D.6}$$

(this is equivalent to the statement that the average z coordinate of a hemispherical dome on the xy-plane equals $\langle z\rangle = 1/2$) and we conclude that

$$N = \frac{1}{2}\int_0^\infty \tilde{n}_C \frac{C}{2}\,\mathrm{d}C = \frac{1}{4}\tilde{n}\langle C\rangle. \tag{D.7}$$

If the particles carry an average energy e, then the above equation can be used to relate the volumetric energy $\tilde{u} = e\tilde{n}$ to the total energy flux J_u. We find

► $$J_u = \frac{1}{4}\tilde{u}\langle C\rangle. \tag{D.8}$$

This relationship is used to find a relationship between the energy density and energy flux for the thermal radiation field where all the particles, the photons, have speed c.

Index

Thermal Physics of the Atmosphere Maarten H. P. Ambaum

ISBN: 978-0-470-74515-1