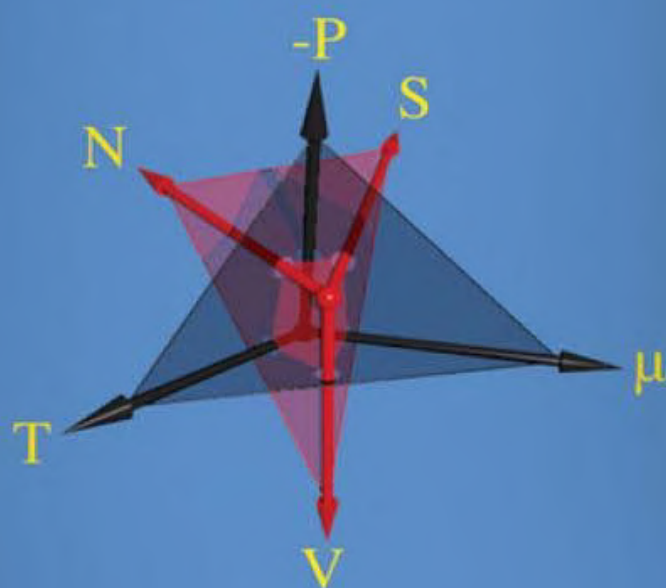


Classical and Geometrical Theory of
**CHEMICAL AND PHASE
THERMODYNAMICS**



Frank Weinhold

**CLASSICAL AND
GEOMETRICAL THEORY
OF CHEMICAL
AND PHASE
THERMODYNAMICS**

CLASSICAL AND GEOMETRICAL THEORY OF CHEMICAL AND PHASE THERMODYNAMICS

Frank Weinhold



WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2009 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

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, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

ISBN 978-0-470-40236-8

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

It was an act of desperation. For six years I had struggled with the blackbody theory. I knew the problem was fundamental, and I knew the answer. I had to find a theoretical explanation at any cost, except for the inviolability of the two laws of thermodynamics.

Max Planck (letter to R. W. Wood, 1931)

If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations. If it is found to be contradicted by observation—well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can offer you no hope; there is nothing for it but to collapse in deepest humiliation.

Sir Arthur Eddington (*The Nature of the Physical World*, 1929)

CONTENTS

PREFACE	xiii
PART I INDUCTIVE FOUNDATIONS OF CLASSICAL THERMODYNAMICS	1
1. Mathematical Preliminaries: Functions and Differentials	3
1.1 Physical Conception of Mathematical Functions and Differentials	3
1.2 Four Useful Identities	7
1.3 Exact and Inexact Differentials	10
1.4 Taylor Series	15
2. Thermodynamic Description of Simple Fluids	17
2.1 The Logic of Thermodynamics	17
2.2 Mechanical and Thermal Properties of Gases: Equations of State	18
2.3 Thermometry and the Temperature Concept	24
2.4 Real and Ideal Gases	30
2.4.1 Compressibility Factor and Ideal Gas Deviations	31
2.4.2 Van der Waals and Other Model Equations of State	36
2.4.3 The Virial Equation of State	44
2.5 Condensation and the Gas–Liquid Critical Point	47
2.6 Van der Waals Model of Condensation and Critical Behavior	50
2.7 The Principle of Corresponding States	54
2.8 Newtonian Dynamics in the Absence of Frictional Forces	56
2.9 Mechanical Energy and the Conservation Principle	58
2.10 Fundamental Definitions: System, Property, Macroscopic, State	60
2.10.1 System	60
2.10.2 Property	61
2.10.3 Macroscopic	63
2.10.4 State	64
2.11 The Nature of the Equilibrium Limit	65
3. General Energy Concept and the First Law	67
3.1 Historical Background of the First Law	67
3.2 Reversible and Irreversible Work	71
3.3 General Forms of Work	76
3.3.1 Pressure–Volume Work	76
3.3.2 Surface Tension Work	78
	vii

3.3.3	Elastic Work	79
3.3.4	Electrical (emf) Work	80
3.3.5	Electric Polarization Work	81
3.3.6	Magnetic Polarization Work	83
3.3.7	Overview of General Work Forms	84
3.4	Characterization and Measurement of Heat	85
3.5	General Statements of the First Law	87
3.6	Thermochemical Consequences of the First Law	89
3.6.1	Heat Capacity and the Enthalpy Function	89
3.6.2	Joule's Experiment	91
3.6.3	Joule–Thomson Porous Plug Experiment	93
3.6.4	Ideal Gas Thermodynamics	95
3.6.5	Thermochemistry: Enthalpies of Chemical Reactions	101
3.6.6	Temperature Dependence of Reaction Enthalpies	107
3.6.7	Heats of Solution	108
3.6.8	Other Aspects of Enthalpy Decompositions	112
4.	Engine Efficiency, Entropy, and the Second Law	117
4.1	Introduction: Heat Flow, Spontaneity, and Irreversibility	117
4.2	Heat Engines: Conversion of Heat to Work	122
4.3	Carnot's Analysis of Optimal Heat-Engine Efficiency	123
4.4	Theoretical Limits on Perpetual Motion: Kelvin's and Clausius' Principles	128
4.5	Kelvin's Temperature Scale	130
4.6	Carnot's Theorem and the Entropy of Clausius	134
4.7	Clausius' Formulation of the Second Law	139
4.8	Summary of the Inductive Basis of Thermodynamics	145
PART II	GIBBSIAN THERMODYNAMICS OF CHEMICAL AND PHASE EQUILIBRIA	147
5.	Analytical Criteria for Thermodynamic Equilibrium	149
5.1	The Gibbs Perspective	149
5.2	Analytical Formulation of the Gibbs Criterion for a System in Equilibrium	152
5.3	Alternative Expressions of the Gibbs Criterion	157
5.4	Duality of Fundamental Equations: Entropy Maximization versus Energy Minimization	160
5.5	Other Thermodynamic Potentials: Gibbs and Helmholtz Free Energy	162
5.6	Maxwell Relations	164
5.7	Gibbs Free Energy Changes in Laboratory Conditions	170
5.8	Post-Gibbsian Developments	180
5.8.1	The Fugacity Concept	181
5.8.2	The "Third Law" of Thermodynamics: A Critical Assessment	183

6. Thermodynamics of Homogeneous Chemical Mixtures	195
6.1 Chemical Potential in Multicomponent Systems	195
6.2 Partial Molar Quantities	197
6.3 The Gibbs–Duhem Equation	201
6.4 Physical Nature of Chemical Potential in Ideal and Real Gas Mixtures	204
7. Thermodynamics of Phase Equilibria	209
7.1 The Gibbs Phase Rule	211
7.2 Single-Component Systems	216
7.2.1 The Phase Diagram of Water	217
7.2.2 Clapeyron and Clausius–Clapeyron Equations for Phase Boundaries	219
7.2.3 Illustrative Phase Diagrams for Pure Substances	224
7.3 Binary Fluid Systems	233
7.3.1 Vapor–Pressure (P – x) Diagrams: Raoult and Henry Limits	237
7.3.2 The Lever Rule	241
7.3.3 Positive and Negative Deviations	243
7.3.4 Boiling-Point Diagrams: Theory of Distillation	247
7.3.5 Immiscibility and Consolute Behavior	250
7.3.6 Colligative Properties and Van’t Hoff Osmotic Equation	253
7.3.7 Activity and Activity Coefficients	260
7.4 Binary Solid–Liquid Equilibria	263
7.4.1 Eutectic Behavior	264
7.4.2 Congruent Melting	265
7.4.3 Incongruent Melting and Peritectics	266
7.4.4 Alloys and Partial Miscibility	266
7.4.5 Phase Boundaries and Gibbs Free Energy of Mixing	267
7.5 Ternary and Higher Systems	273
8. Thermodynamics of Chemical Reaction Equilibria	281
8.1 Analytical Formulation of Chemical Reactions in Terms of the Advancement Coordinate	281
8.2 Criterion of Chemical Equilibrium: The Equilibrium Constant	282
8.3 General Free Energy Changes: de Donder’s Affinity	285
8.4 Standard Free Energy of Formation	286
8.5 Temperature and Pressure Dependence of the Equilibrium Constant	288
8.5.1 Temperature Dependence: Van’t Hoff Equation	288
8.5.2 Pressure Dependence	289
8.6 Le Chatelier’s Principle	290
8.7 Thermodynamics of Electrochemical Cells	292

8.8	Ion Activities in Electrolyte Solutions	296
8.9	Concluding Synopsis of Gibbs' Theory	305
PART III METRIC GEOMETRY OF EQUILIBRIUM THERMODYNAMICS		311
9.	Introduction to Vector Geometry and Metric Spaces	313
9.1	Vector and Matrix Algebra	315
9.2	Dirac Notation	323
9.3	Metric Spaces	328
10.	Metric Geometry of Thermodynamic Responses	331
10.1	The Space of Thermodynamic Response Vectors	331
10.2	The Metric of Thermodynamic Response Space	333
10.3	Linear Dependence, Dimensionality, and Gibbs–Duhem Equations	337
11.	Geometrical Representation of Equilibrium Thermodynamics	345
11.1	Thermodynamic Vectors and Geometry	345
11.2	Conjugate Variables and Conjugate Vectors	348
11.3	Metric of a Homogeneous Fluid	353
11.4	General Transformation Theory in Thermodynamic Metric Space	357
11.5	Saturation Properties Along the Vapor-Pressure Curve	360
11.6	Self-Conjugate and Normal Response Modes	363
11.7	Geometrical Characterization of Common Fluids	366
11.8	Stability Conditions and the “Third Law” for Homogeneous Phases	376
11.9	The Critical Instability Limit	379
11.10	Critical Divergence and Exponents	384
11.11	Phase Heterogeneity and Criticality	386
12.	Geometrical Evaluation of Thermodynamic Derivatives	393
12.1	Thermodynamic Vectors and Derivatives	394
12.2	General Solution for Two Degrees of Freedom and Relationship to Jacobian Methods	401
12.3	General Partial Derivatives in Higher-Dimensional Systems	405
12.4	Phase-Boundary Derivatives in Multicomponent Systems	408
12.5	Stationary Points of Phase Diagrams: Gibbs–Konowalow Laws	414
12.6	Higher-Order Derivatives and State Changes	417
13.	Further Aspects of Thermodynamic Geometry	421
13.1	Reversible Changes of State: Riemannian Geometry	424
13.2	Near-Equilibrium Irreversible Thermodynamics: Diffusional Geometry	429

13.3	Quantum Statistical Thermodynamic Origins of Chemical and Phase Thermodynamics	439
13.3.1	Nonequilibrium Displacement Variables of Mayer and Co-workers	442
13.3.2	Quantum Statistical Thermodynamics and the Statistical Origins of Metric Geometry	445
13.3.3	Evaluation of Molecular Partition Functions for Reactive Mixtures	452
13.3.4	Quantum Cluster Equilibrium Theory of Phase Thermodynamics	455
Appendix: Units and Conversion Factors		465
AUTHOR INDEX		469
SUBJECT INDEX		473

PREFACE

This book has two primary aims. The first is to provide an accurate but accessible introduction to the theory of chemical and phase thermodynamics as first enunciated by J. Willard Gibbs. The second is to exhibit the transcendent beauty of the Gibbsian theory as expressed in the mathematical framework of Euclidean and Riemannian geometry.

Both aims may seem unrealistic within the pedagogical constraints of a textbook for undergraduates or beginning graduate students. However, the author believes that accurate and thorough grounding in the Gibbsian viewpoint is not only the best introduction to research-level thermodynamic applications, but also the low-barrier entryway to a remarkably simple and effective set of geometrical tools that make accurate thermodynamic reasoning accessible to students with only modest mathematical training.

In attempting this amalgamation of Gibbsian and geometric concepts, I have adhered closely in Parts I (Chapters 1–4) and II (Chapters 5–8) to the actual content of the first-semester physical chemistry course at the University of Wisconsin (Chem 561) for more than two decades. This includes the usual topics pertaining to the pre-Gibbsian historical development (Part I) and the final Gibbs synthesis of chemical and phase thermodynamics (Part II), expressed in the traditional language of partial differential calculus. Aside from certain subtle points of rigor and emphasis, the content of Chapters 1–8 can be closely mapped onto other introductory thermodynamics expositions, such as the venerable “Wisconsin” series of physical chemistry textbooks (as authored by Getman and Daniels in 1931, Daniels and Alberty in the author’s student days, and Silbey, Alberty, and Bawendi at present).

Part III (Chapters 9–13), in contrast, is quite novel, representing the first full textbook exposition of the metric geometry of equilibrium thermodynamics as originally formulated in a series of papers (1975–1978) by the author. Although this “thermodynamic geometry” has seen extensive research applications in such diverse areas as optimal process control and black hole thermodynamics, its many pedagogical and practical advantages have not been sufficiently exhibited for beginning students of physical chemistry.

In a sense, the material of Part III is far the easiest to master, even though it is logically equivalent to the traditional Gibbsian-based formalism outlined in Parts I and II. Indeed, it is conceivable that a motivated high school student with only basic skills in Euclidean geometry could reasonably begin with Part III, proceeding immediately to derive complex thermodynamic relationships with confidence and accuracy! (The only “trick” is to learn how to associate the geometrical distances or angles with measurable thermodynamic properties or equivalent partial differential expressions of Parts I and II, as illustrated in Fig. 11.2.) However, thoughtful students would undoubtedly find this short cut to be excessively “magical” if insufficiently supported by the historical and physical background of Parts I and II. Hence, Part III does not attempt to revisit all the topics of Parts I and II, as though this background were unfamiliar to the reader. Instead, the basic geometrical

isomorphism is established with traditional thermodynamic concepts of assumed familiarity, allowing students to carry out desired geometrical re-derivations of thermodynamic identities at their leisure (in analogy to the many examples provided in sidebars) while focusing primarily on novel extensions of the thermodynamic geometry, including many described here for the first time. Part III therefore assumes some familiarity with Parts I and II, but students with alternative physical chemistry backgrounds (e.g., the textbooks of Atkins, Engel–Ried, Levine, or Silbey–Alberty–Bawendi) should encounter little difficulty in picking up the thread.

I wish to express sincere gratitude to many teachers and colleagues, present and past, who have aided my understanding of thermodynamics and phase equilibria. These include Steve Berry, Bob Bird, Phil Certain, Dan Cornwell, Chuck Curtiss, Tom Farrar, John Ferry, Joop de Heer, Michael Fisher, Stan Gill, Joe Hirschfelder, Ed Jaynes, Fred Koenig, Arthur Lodge, Ralf Ludwig, Mike McBride, Gil Nathanson, John Perepezko, Tom Record, Peter Salamon, Jim Skinner, Laszlo Tisza, Worth Vaughan, Hyuk Yu, and John Wheeler.

I also wish to express my appreciation to David Strasfeld, Gil Nathanson, John Harriman, and (particularly) Bob Bird, who suggested helpful improvements to an early draft; to Mark Wendt, who prepared the rendered graphics for the cover and Figure 11.1; and to John Herbert, Phillip Thomas, and David Strasfeld (all former teaching assistants in Chem 561), who assembled problems and exercises to accompany the book.

Neither the writing of this book nor the original research on which it is based could have come about without the loving support of my family, for which I am deeply grateful.

FRANK WEINHOLD

Madison, Wisconsin

PART I

INDUCTIVE FOUNDATIONS OF CLASSICAL THERMODYNAMICS

Mathematical Preliminaries: Functions and Differentials

1.1 PHYSICAL CONCEPTION OF MATHEMATICAL FUNCTIONS AND DIFFERENTIALS

Science consists of interrogating nature by experimental means and expressing the underlying patterns and relationships between measured properties by theoretical means. Thermodynamics is the science of heat, work, and other energy-related phenomena.

An experiment may generally be represented by a set of stipulated control conditions, denoted $x_1, x_2, \dots, x_n$, that lead to a unique and reproducible experimental result, denoted z . Symbolically, the experiment may be represented as an input–output relationship,

$$\begin{array}{ccc} x_1, x_2, \dots, x_n, & \xrightarrow{\text{experiment}} & z \\ \text{(input)} & & \text{(output)} \end{array} \quad (1.1)$$

Mathematically, such relationships between independent ($x_1, x_2, \dots, x_n$) and dependent (z) variables are represented by *functions*

$$z = z(x_1, x_2, \dots, x_n) \quad (1.2)$$

We first wish to review some general mathematical aspects of functional relationships, prior to their specific application to experimental thermodynamic phenomena.

Two important aspects of any experimentally based functional relationship are (1) its *differential* dz , i.e., the *smallest sensible increment of change* that can arise from corresponding differential changes ($dx_1, dx_2, \dots, dx_n$) in the independent variables; and (2) its *degrees of freedom* n , i.e., the number of “control” variables needed to determine z uniquely. How “small” is the magnitude of dz (or any of the dx_i) is related to specifics of the experimental protocol, particularly the inherent experimental uncertainty that accompanies each variable in question.

For $n = 1$ (“ordinary” differential calculus), the dependent differential dz may be taken proportional to the differential dx of the independent variable,

$$dz = z' dx \quad (1.3)$$

where z' (the total *derivative* of z with respect to x) is evidently related to the differentials dz , dx by the ratio formula

$$z' = \frac{dz}{dx} \quad (1.4)$$

The validity of (1.3), i.e., the existence of the derivative dz/dx in (1.4), is an essential requisite for application of the formalism of differential calculus. It is therefore important that the magnitudes of differentials dz , dx be taken “sufficiently small” (but not “zero,” a meaningless and unphysical extrapolation in this context!) for the limiting ratio in (1.4) to have an experimentally well-defined value, within usual limits of experimental precision.

For the general case of n variables, the expression for dz must include corresponding “partial” contributions from each possible differential change dx_i . This is expressed by the important equation

$$dz = \sum_{i=1}^n \left(\frac{\partial z}{\partial x_i} \right)_{\underline{x}} dx_i = \sum_{i=1}^n z'_i dx_i \quad (1.5)$$

where

$$z'_i = \left(\frac{\partial z}{\partial x_i} \right)_{\underline{x}} \quad (1.6)$$

and where the subscript $\underline{x}$ denotes the list of all control variables held constant (i.e., all but the “active” variable dx_i). In general, each “partial” derivative $(\partial z / \partial x_i)_{\underline{x}}$ in (1.5) [like each ordinary derivative z' in (1.3)] is itself a function of all variables on which z depends. Equation (1.5) is referred to as the “chain rule” of partial differential calculus. It represents the most fundamental relationship between differential changes for a system with n degrees of freedom, and often forms the starting point for thermodynamic reasoning.

SIDEBAR 1.1: RECTANGLE EXERCISE

Exercise For a rectangle of sides x , y , find the function for area $A = A(x, y)$, its partial derivatives with respect to x and y , and its differential dA .

Solution The area function is $A(x, y) = xy$, so the partial derivatives are $(\partial A / \partial x)_y = y$ and $(\partial A / \partial y)_x = x$, and the differential is $dA = y dx + x dy$.

SIDEBAR 1.2: CIRCUMFERENCE EXERCISE

Exercise Suppose that the circumference of the Earth is snugly encircled with a band of 25,000-mile length. If the band is slightly *lengthened* by 10 ft, how high above the surface does it rise? (Does the Earth’s precise circumference matter?)

Solution Circumference C and radius R are related by $R = C/2\pi$. To determine the small radial change dR accompanying a change of circumference dC , we need $R' = dR/dC = 1/2\pi$. We can therefore approximate the radial increase ΔR as $\Delta R = R' \Delta C = (1/2\pi)(10 \text{ ft}) \cong 1.59 \text{ ft}$ (independent of C).

The important functional relationships of thermodynamic systems also permit second derivatives to be evaluated. For example, the derivative function $z'_i = z'_i(x_1, x_2, \dots, x_n)$ of (1.6) can be differentiated with respect to a second variable x_j to give the mixed second derivative of z with respect to x_i and x_j ,

$$z''_{ij} = \left(\frac{\partial z'_i}{\partial x_j} \right)_x = \frac{\partial^2 z}{\partial x_i \partial x_j} \quad (1.7)$$

As first shown by J. W. Gibbs, the analytical characterization of thermodynamic equilibrium states can be expressed completely in terms of such first and second derivatives of a certain “fundamental equation” (as described in Section 5.1).

Note that differentials (dz) have fundamentally different mathematical character than do functions (such as z, z', z''). The former are inherently “infinitesimal” (microscopic) in scale and carry multivariate dependence on all possible “directions” of change, whereas the latter carry only macroscopic numerical values. Thus, it is mathematically inconsistent to write equations of the form “differential = function” (or “differential = derivative”), just as it would be inconsistent to write equations of the form “vector = scalar” or “apples = oranges.” Careful attention to proper balance of thermodynamic equations with respect to differential or functional character will avert many logical errors.

The student of thermodynamics must learn to cope with the functional, differential, and derivative relationships in (1.2)–(1.7) from a variety of formulaic, graphical, and experimental aspects. Let us briefly discuss each in turn.

Formulaic Aspect The student should be familiar with analytical formulas for derivatives z' of common algebraic and transcendental functions z , such as

$$z = x^n, \quad z' = nx^{n-1}; \quad \text{or} \quad z = u^n, \quad z' = nu^{n-1}u' \quad (1.8a)$$

$$z = e^x, \quad z' = e^x; \quad \text{or} \quad z = e^u, \quad z' = e^u u' \quad (1.8b)$$

$$z = \ln x, \quad z' = \frac{1}{x}; \quad \text{or} \quad z = \ln u, \quad z' = \frac{u'}{u} \quad (1.8c)$$

These formulas are also generally sufficient for partial derivatives (because holding some terms *constant* in z can only simplify its differentiation!). Although such formulas may prove useful in certain contexts (such as homework problems based on assumed functional forms of forgiving mathematical simplicity), they are less useful than, for example, graphical or numerical techniques for dealing with realistic experimental data.

Graphical Aspect Functional relationships such as (1.1) and (1.2) can often be most effectively depicted in graphical (or geometric model) form. Innovative graphical methods were developed by Gibbs and others to display the complex thermodynamic relationships of single- and multicomponent chemical systems, as illustrated in Fig. 1.1. For thermodynamic purposes, the ability to “read” equations such as (1.2)–(1.5) through graphical visualization is more important than facility with analytical formulas such as (1.8a–c).

Graphical visualization of a function z or its derivative(s) is similar in the case of ordinary ($n = 1$) and multivariate systems, except that the latter necessarily requires additional dimensions. In a standard 2-dimensional graph, the *height* of the curve at given x_0

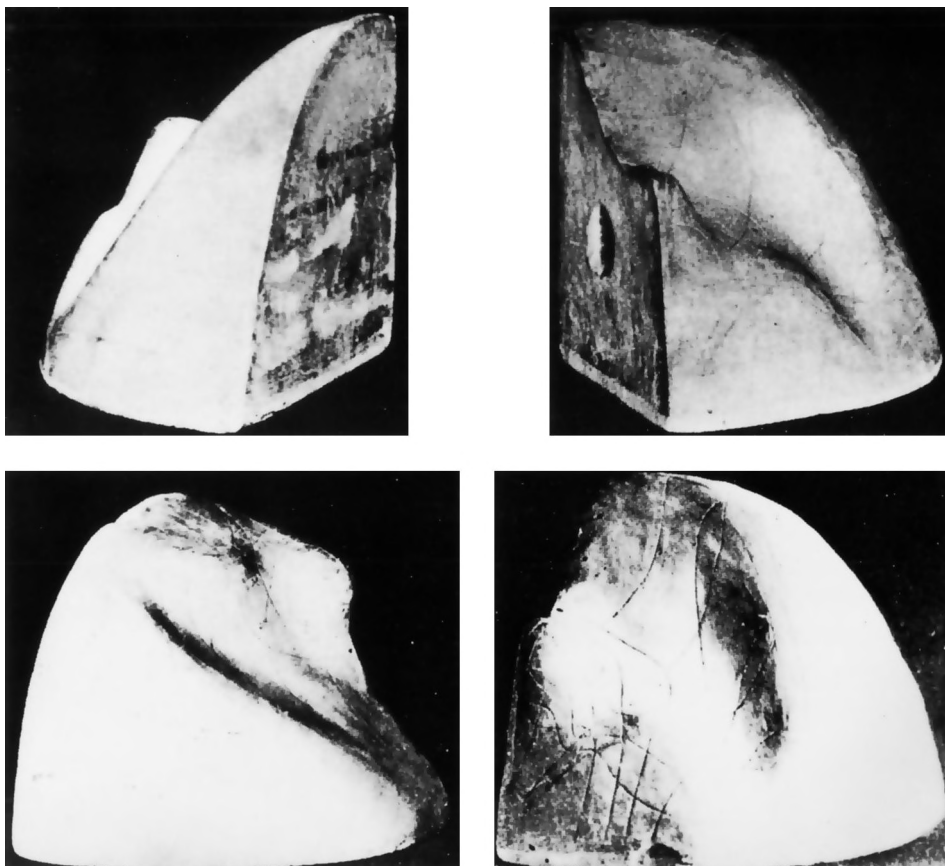
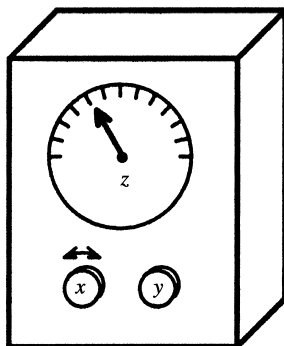


Figure 1.1 Geometrical model depicting thermodynamic properties of water in “Gibbs coordinates.” This plaster model, currently in the Beinecke Library at Yale University, was created by noted British physicist James Clark Maxwell as a gift to American thermodynamicist J. Willard Gibbs (see www.public.iastate.edu/~jolls/ for computer-generated representations by Professor K. R. Jolls).

represents the “strength” of $z = z(x_0)$, whereas the *slope* of the curve is the first derivative $z' = z'(x_0)$ and the *curvature* (variation of slope) is the second derivative $z'' = z''(x^{(0)})$. In a corresponding multidimensional graph, the slope $z'_i = (\partial z / \partial x_i)_{\underline{x}}$ of the surface generally depends on which “direction” dx_i is chosen (different slopes in different directions), and a similar remark applies to the curvature z''_{ij} for any chosen pair of directions dx_i, dx_j . In general, the slope or curvature in the x direction is independent of that in the y direction, so each partial derivative expresses independent information about the function. Of course, in the thermodynamic context, the partial derivatives generally correspond to experimental “response functions,” such as heat capacity or compressibility, that have no literal topographic character. However, it is useful to retain the intuitive topographic imagery (e.g., of a ski hill) to recognize that “slope” and “curvature” must generally depend on the “directions” chosen.

Experimental Aspect Experimental evaluation of a derivative z' (or z'_i in the multivariate case) might be envisioned with the following schematic “ z -meter” apparatus:



This apparatus, together with the usual mathematical expression for the limit ratio in (1.4), suggests the experimental protocol for measuring partial derivatives of z . Suppose that the effect of slightly tweaking the control x -dial about its initial setting $x^{(0)}$ by Δx is to give a slight deflection Δz of the z -needle from its initial position $z^{(0)}$. Then the derivative (1.4) can be evaluated as the limit

$$z'(x^{(0)}) = \left. \frac{dz}{dx} \right|_{x^{(0)}} = \lim_{\Delta x \rightarrow "0"} \frac{\Delta z}{\Delta x} = \lim_{\Delta x \rightarrow "0"} \frac{z(x^{(0)} + \Delta x) - z^{(0)}}{\Delta x} \quad (1.9)$$

Here the “0” of the limit means “sufficiently small for the limit to *exist*,” which is to be understood more precisely in the context of the experiment. A corresponding z -meter in the multivariable case would have n x_i -dials, each of which is tweaked in turn (holding the remaining $n - 1$ dials fixed) to determine the successive partial derivatives z_i , $i = 1, 2, \dots, n$. It is noteworthy that the multivariate dz carries sufficient information to evaluate *each* of its possible monovariate dx_i derivatives z'_i , confirming its status as a more powerful type of mathematical object.

We emphasize that mathematical limiting operations such as (1.9) must make *physical* sense in order to usefully serve thermodynamic applications. The student should always be prepared to make physical estimates of “how small” a sensible differential must be chosen for ratios such as (1.4) or (1.9) to have experimentally well-defined values. (For example, it makes no sense to measure the rainfall rate in a hurricane with a rainfall volume increment corresponding to one droplet, or one molecule, or smaller!) For physical purposes, a differential dz must be sufficiently small for onset of the linear regime expressed by (1.3) or (1.5), but never so small as to raise unjustified concerns about “dividing by zero” in equations such as (1.4) or (1.9).

1.2 FOUR USEFUL IDENTITIES

The special case of $n = 2$ degrees of freedom is often of particular interest. For this purpose, we write the function as $z = z(x, y)$, with the differential dz being given by the usual chain-rule expression

$$dz = \left(\frac{\partial z}{\partial x} \right) dx + \left(\frac{\partial z}{\partial y} \right) dy \quad (1.10)$$

This is the starting point for the four mathematical identities to be derived below.

(i) Reduction to $n = 1$ (Single Degree of Freedom) Suppose that the “independent” variables $x = x(u)$, $y = y(u)$ are both simple functions of a single variable u , so that $z = z(u)$ has only “ordinary” derivative dependence on u . What is dz/du ? To obtain this ratio, we can simply divide dz (1.10) by du to obtain

$$\frac{dz}{du} = \left(\frac{\partial z}{\partial x} \right)_y \frac{dx}{du} + \left(\frac{\partial z}{\partial y} \right)_x \frac{dy}{du} \quad (1.11)$$

Note closely the distinctions between ordinary (d) and partial (∂) derivatives throughout this formula.

Note also that we employ “physicist’s notation” for functions, in which both $z = z(u)$ and $z = z(x, y)$ express how z depends on the variables specified in parentheses (even though the mathematical formulas that express this dependence might be quite different in the two cases). Although somewhat “unmathematical,” the chosen notation better expresses the experimental relationship (1.1), in which control variables x_i might be chosen for convenience in many ways, but the target property z is independent of this choice. For example, the volume of a sphere could be equivalently expressed in terms of its measured diameter [$V = V(d) = \pi d^3/6$] or surface area [$V = V(A) = (\pi^{1/2}/6)A^{3/2}$], despite the fact that the mathematical dependence (i.e., whether there is a cubic or three-halves power in the chosen measurement argument) is different in the two cases.

(ii) Change of Differentiated Variable Suppose that we re-express $z = z(x, \theta)$ as a function of x and a new variable θ , where the “old” variable $y = y(x, \theta)$ is also expressible in the new independent variables (x, θ) . To find the expression for $(\partial z/\partial \theta)_x$ from the “old” differential expression (1.10), we merely divide (1.10) throughout by “ $d\theta$ at constant x ” [replacing the constrained ratio “ $dz/d\theta$ at constant x ” on the left-hand side by the proper partial derivative notation, $(\partial z/\partial \theta)_x$, and similarly for both ratios on the right-hand side]:

$$\left(\frac{\partial z}{\partial \theta} \right)_x = \left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial \theta} \right)_x + \left(\frac{\partial z}{\partial y} \right)_x \left(\frac{\partial y}{\partial \theta} \right)_x$$

However, the partial derivative $(\partial x/\partial \theta)_x = 0$ (because, at constant x , derivatives of x with respect to *anything* must vanish). The above equation thereby simplifies to

$$\left(\frac{\partial z}{\partial \theta} \right)_x = \left(\frac{\partial z}{\partial y} \right)_x \left(\frac{\partial y}{\partial \theta} \right)_x \quad (1.12)$$

Note how the right-hand side has the proper “balance” of differential terms, as though dy can be cancelled from numerator and denominator to give the desired partial derivative.

(iii) Change of Variable Held Constant Under the same change of variables $(x, y) \rightarrow (x, \theta)$, we can also obtain the partial derivative $(\partial z/\partial x)_\theta$ (with the new variable θ held constant). Starting again from (1.10), we “divide by dx at constant θ ” on both sides (using proper partial derivative notation for the constrained ratios) to obtain

$$\left(\frac{\partial z}{\partial x} \right)_\theta = \left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial x} \right)_\theta + \left(\frac{\partial z}{\partial y} \right)_x \left(\frac{\partial y}{\partial x} \right)_\theta$$

But $(\partial x/\partial x)_\theta = 1$ (since the variations of x with itself are *unity*, no matter what else is constant), so the equation becomes

$$\boxed{\left(\frac{\partial z}{\partial x}\right)_\theta = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_\theta} \quad (1.13)$$

Note that this identity clearly shows that $(\partial z/\partial x)_y \neq (\partial z/\partial x)_\theta$, i.e., that the variable held constant *matters* in these derivatives! (Strictly speaking, a lazy notation such as “ $\partial z/\partial x$ ” has *no meaning whatsoever*!) Although the inconvenient notation of partial derivatives makes it somewhat tedious to keep the inactive (constant) “background” variables in mind, it is important from a physical and pedagogical standpoint that this be done as carefully as possible. (The tedium of this notation is avoided in the geometrical thermodynamics to be presented in Part III.)

SIDEBAR 1.3: CHANGE-OF-VARIABLE EXERCISE

Exercise Suppose the rectangular area A in Sidebar 1.1 is expressed in terms of side x and perimeter P . What are $(\partial A/\partial P)_x$ and $(\partial A/\partial x)_P$?

Solution The new and old variables are related by $P = 2(x + y)$, or

$$y = \frac{1}{2}P - x$$

so that

$$\left(\frac{\partial y}{\partial x}\right)_P = -1, \quad \left(\frac{\partial y}{\partial P}\right)_x = \frac{1}{2}$$

From the identity (1.12), we obtain

$$\left(\frac{\partial A}{\partial P}\right)_x = \left(\frac{\partial A}{\partial y}\right)_x \left(\frac{\partial y}{\partial P}\right)_x = (x) \left(\frac{1}{2}\right) = \frac{1}{2}x$$

Similarly, from the identity (1.13), we obtain

$$\left(\frac{\partial A}{\partial x}\right)_P = \left(\frac{\partial A}{\partial x}\right)_y + \left(\frac{\partial A}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_P = y + (x)(-1) = \frac{1}{2}P - 2x$$

[Of course, in this case, it is also possible to solve explicitly for $A = A(x, P) = \frac{1}{2}Px - x^2$ and differentiate directly, but this “direct” route is often less practical than use of the identities (1.12), (1.13).]

(iv) Jacobi (Cyclic) Identity A provocative identity of great generality and usefulness for $n = 2$ is obtained by considering (1.10) under conditions of constant z (i.e., $dz = 0$). If we then “divide by dx at constant z ” (making the usual change of notation from ratio to

partial derivative), we obtain

$$0 = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_z$$

Noting that $(\partial z/\partial x)_y = 1/(\partial x/\partial z)_y$, we can rewrite the above equation as

$$\boxed{\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial z}{\partial x}\right)_y \left(\frac{\partial y}{\partial z}\right)_x = -1} \quad (1.14a)$$

Alternately, we can rewrite this identity as

$$\boxed{\left(\frac{\partial x}{\partial y}\right)_z = -\frac{(\partial z/\partial y)_x}{(\partial z/\partial x)_y}} \quad (1.14b)$$

As one can see in (1.14a), the variables (x, y, z) are “cycled” in the three derivatives, each appearing once in the numerator, once in the denominator, and once as the constant variable. The cyclic symmetry makes it easy (and advisable) to commit this identity to memory, even if it can be easily rederived from (1.10) for use as needed.

The identities (1.11)–(1.14) are among the most commonly employed in thermodynamic derivations, because *two* degrees of freedom underlie the important special case of “simple” substances (pure, homogeneous), as will be subsequently described.

1.3 EXACT AND INEXACT DIFFERENTIALS

While the existence of a functional relationship $z = z(x_1, x_2, \dots, x_n)$ allows its differential dz to be unambiguously determined, the reverse need not be the case. Differentials dz for which *no* corresponding function z exists are called *inexact* (or “imperfect,” often marked with a slash: $\bar{d}$), whereas those for which z exists are *exact* (or “perfect”). The basic distinction between exact (d -type) and inexact ($\bar{d}$ -type) differentials lies at the heart of thermodynamic usage of the differential concept, so we must understand clearly how the two cases can be mathematically distinguished. Differentials of heat, for example, are found to belong to the “imperfect” category, whereas those of energy are “perfect.”

It might seem that a suitable z (up to an arbitrary constant) could always be generated from a given differential form dz by merely evaluating the integral

$$z \stackrel{?}{=} \int dz$$

This is indeed always possible for a single variable $n = 1$ (ordinary calculus), where the distinction between exact and inexact differentials disappears. However, for $n > 1$, it is clear that integrals over dz must generally depend on the chosen *path* along which the integration is performed. Integrals of multivariate differentials are called *line integrals* (or path integrals) to indicate this distinction from ordinary (monovariate) integrals. For inexact $\bar{d}z$, the line integral $\int \bar{d}z$ is *path-dependent* (and therefore not uniquely defined), the signature

defect of inexactness. Only in the case of an *exact* differential dz does the indefinite integral $\int dz$ evaluate to a unique function z , independent of the chosen integration path.

Let us first consider this issue in the simple case $n = 2$, with independent variables x, y and dependent variable z . If a well-defined function $z(x, y)$ exists, then dz [of the form (1.10)] is certainly exact. Furthermore, if we evaluate the definite integral from initial (x_1, y_1) to final (x_2, y_2) , the result is simply

$$I = \int_{x_1, y_1}^{x_2, y_2} dz = z(x_2, y_2) - z(x_1, y_1) \quad (1.15)$$

The important point is that the final value of the integral depends only on the two endpoints, i.e., the value of the function z at (x_1, y_1) and (x_2, y_2) , but not the chosen path of integration (as illustrated in Sidebar 1.4). Moreover, in the special case of a *cyclic* integral (denoted $\oint$), where “initial” and “final” limits coincide, the integral (1.15) necessarily *vanishes* for an exact differential, independent of how the cyclic path is chosen. We can therefore state the following integral criterion for exactness:

$$\begin{aligned} \text{Integral criterion: } & \text{The differential } dz \text{ is exact if and only if} \\ & \oint dz = 0 \text{ for all possible paths} \end{aligned} \quad (1.16a)$$

A closely related criterion can be stated in graphical terms:

$$\begin{aligned} \text{Graphical criterion: } & \text{The differential } dz \text{ is exact if and only if its integral} \\ & z(x, y) = \int dz \text{ is graphable} \end{aligned} \quad (1.16b)$$

This criterion is rather self-evident, because the condition that $z = z(x, y)$ be “graphable” is merely that a *unique* z -value be given any chosen x, y , i.e., that $z = z(x, y)$ satisfies the requirements of a function. However, both criteria require global (integral) information that may be difficult to obtain from local measurements.

SIDEBAR 1.4: SUMMIT TRAIL PROBLEM

Problem On the coast of Hawaii, a sign points to a distant volcano with the information, “Summit: distance = 15.3 km, altitude = 4.2 km.” How can one determine which (if either) of the differential quantities dl (distance) or dh (altitude) is exact?

Solution By measuring (e.g., with ruler and altimeter) the differential changes dl, dh and integrating (summing up) their cumulative changes I_l, I_h from coast to summit,

$$I_l = \int_{\text{coast}}^{\text{summit}} dl, \quad I_h = \int_{\text{coast}}^{\text{summit}} dh$$

one could verify experimentally that I_h is independent of the path chosen to the summit, so that dh is exact, whereas I_l is path-dependent, so that dl is inexact by (1.16a).

As an alternative strategy, one might ask in a local bookshop for an “altitude map” and a “distance map” for Hawaii. A mathematically savvy shopkeeper may reply that the first (a “topo map”) is readily available, because altitude is easily graphable in topographic form, whereas the second is not, because distance is inherently a path-dependent, ungraphable quantity. This reply points, by (1.16b), to the same conclusion.

A more convenient differential criterion for exactness was established by Euler. Suppose that the differential dz consists as usual of contributions from dx and dy variations,

$$dz = M(x, y) dx + N(x, y) dy$$

where the respective coefficients $M = M(x, y)$ and $N = N(x, y)$ are stipulated functions of x and y . We can then state the Euler criterion as follows:

Euler criterion ($n = 2$): The differential $dz = M dx + N dy$ is exact if and only if

$$\left(\frac{\partial M}{\partial y}\right)_x = \left(\frac{\partial N}{\partial x}\right)_y \text{ at every point } x, y \quad (1.17)$$

It is easy to recognize that the Euler criterion will be satisfied if the integral or graphical criteria (1.16) are satisfied. Suppose that $z(x, y)$ indeed exists (e.g., displayed as a graph), so that (1.10) is assured. Comparison of (1.10) with the assumed form of the differential then shows that

$$M(x, y) = \left(\frac{\partial z}{\partial x}\right)_y, \quad N(x, y) = \left(\frac{\partial z}{\partial y}\right)_x \quad (1.18)$$

The M -derivative of the Euler criterion (1.17) can therefore be evaluated as

$$\left(\frac{\partial M}{\partial y}\right)_x = \left(\frac{\partial}{\partial y} \left(\frac{\partial z}{\partial x}\right)_y\right)_x = \frac{\partial^2 z}{\partial y \partial x} \quad (1.19a)$$

whereas the N -derivative is similarly

$$\left(\frac{\partial N}{\partial x}\right)_y = \left(\frac{\partial}{\partial x} \left(\frac{\partial z}{\partial y}\right)_x\right)_y = \frac{\partial^2 z}{\partial x \partial y} \quad (1.19b)$$

The Euler criterion is therefore equivalent to the familiar “mixed partials of a function are equal” rule of calculus. This cross-differentiation rule is also the condition for the function $z(x, y)$ to have well-defined (single-valued) first derivatives at each point, and thus to be graphable.

SIDEBAR 1.5: EXACT DIFFERENTIAL EXERCISES

Exercises Use the Euler criterion (1.17) to determine whether each of the following differentials dz is exact or inexact:

- (a) $dz = y dx + x dy$
- (b) $dz = y^2 dx + xy dy$
- (c) $dz = (y/x) dx + \ln(x) dy$
- (d) $dz = 2x^{-1/3}y^7(y dx + 12x dy)$

Solutions (a) exact; (b) inexact; (c) exact; (d) exact. To work out the solution of (d) in more detail, we note that

$$M = 2x^{-1/3}y^8, \quad N = 24x^{2/3}y^7$$

so that

$$\left(\frac{\partial M}{\partial y}\right)_x = 16x^{-1/3}y^7 = \left(\frac{\partial N}{\partial x}\right)_y$$

as required for exactness.

SIDEBAR 1.6: ILLUSTRATIVE LINE INTEGRALS

Let us examine the line integrals of two simple *inexact* differentials, namely,

$$\bar{d}z_1 = y dx, \quad \bar{d}z_2 = x dy \quad (\text{S1.6-1})$$

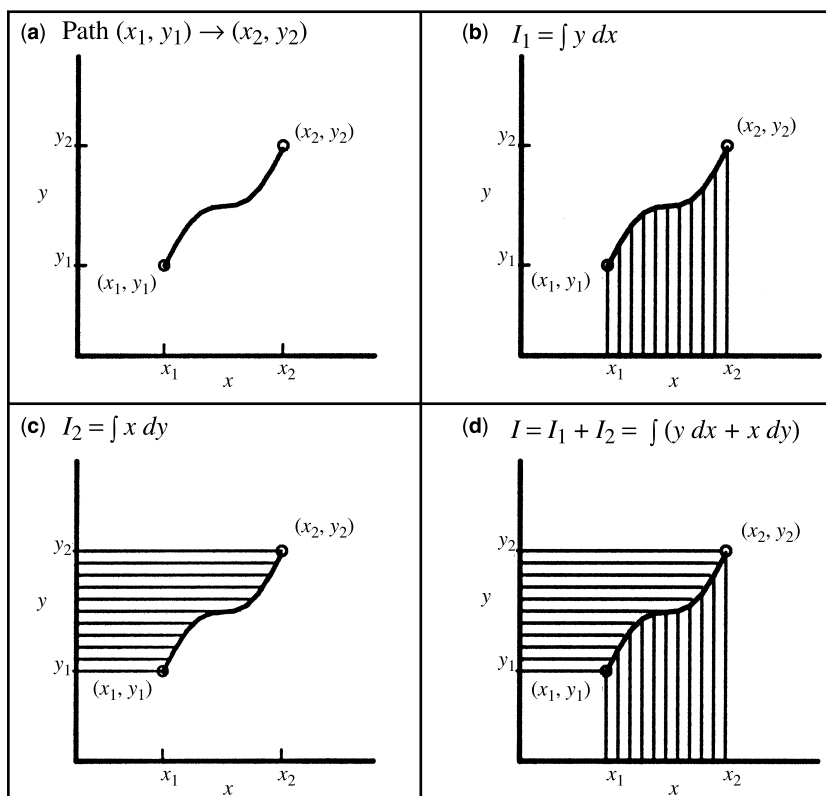
to see their explicit path dependence. We employ the path $y = y(x)$ shown in figure panel (a) to connect the initial point $P = (x_1, y_1)$ to the final point $Q = (x_2, y_2)$ in the definite integrals

$$I_1 = \int_P^Q \bar{d}z_1 = \int_P^Q y dx, \quad I_2 = \int_P^Q \bar{d}z_2 = \int_P^Q x dy \quad (\text{S1.6-2})$$

The first integral I_1 is just the area under the curve $y = y(x)$, as shown by the shaded region in panel (b). Similarly, the second integral I_2 is the area to the left of this curve, as shown by the shaded region in panel (c). Clearly, the values of both I_1 and I_2 are dependent on the chosen path of integration, confirming that $\bar{d}z_1$ and $\bar{d}z_2$ are inexact. However, the sum of these differentials, $dz = \bar{d}z_1 + \bar{d}z_2 = y dx + x dy$, is evidently exact [cf. part (a) of Sidebar 1.5]. By inspection, its integral

$$I = \int_P^Q dz = I_1 + I_2 \quad (\text{S1.6-3})$$

is the total area of the shaded L-shaped region in panel (d), which depends on the endpoints (x_1, y_1) , (x_2, y_2) but not the connecting path.



For arbitrary n , the more general statement of the Euler criterion can be formulated in terms of a general n -term differential form

$$dz = \sum_{i=1}^n R_i dx_i \quad (1.20)$$

with coefficients $R_i = R_i(x_1, x_2, \dots, x_n)$. The generalization of (1.17) is

$$\text{Euler criterion (general } n\text{): The differential } dz = R_1 dx_1 + R_2 dx_2 + \dots + R_n dx_n \quad (1.21)$$

is exact if and only if $\left(\frac{\partial R_i}{\partial x_j}\right)_{\underline{x}} = \left(\frac{\partial R_j}{\partial x_i}\right)_{\underline{x}}$ for all $i, j = 1, 2, \dots, n$

(i.e., mixed partial derivatives are equal for any chosen pair of variables x_i, x_j).

This fundamental relation underlies all thermodynamic descriptions of exact (conserved) differential quantities such as internal energy or entropy, as will be shown in subsequent chapters.

Finally, we briefly mention the concept of an *integrating factor*, a multiplicative factor (L) that converts an inexact differential (df) to an exact differential (dg), namely,

$$Ldf = dg \quad (1.22)$$

Integrating factors L may or may not exist for a given df , and if they exist, they are generally non-unique (e.g., $L' = cL$ is also an integrating factor for any constant c). In simple cases, an integrating factor can be guessed “by inspection”; for example, it is easy to see that $L = 1/y$ is an integrating factor for the inexact differential in Sidebar 1.5(b). In more complex cases, the Euler condition (1.21) can be used to convert (1.22) into a differential equation for determining L . In the thermodynamic context, however, the most important integrating factor is that for the differential of heat, and this factor (namely, $L = 1/T$, the inverse temperature) will be obtained from *physical* considerations, rather than, for example, by solving a differential equation.

1.4 TAYLOR SERIES

A common situation in thermodynamics is that some property $z(x)$ and its lower derivatives (z' , z'' , z''' , ...) have been measured at a certain point x_0 , and one wishes to use this information to approximate the behavior of the function $z(x_0 + \Delta x)$ in the Δx -neighborhood of x_0 . For this purpose, the fundamental *Taylor series* (or MacLaurin series, the special case for $x_0 = 0$) yields approximations that are useful for sufficiently small Δx :

$$z(x_0 + \Delta x) \simeq z(x_0) + z'(x_0)\Delta x + \frac{1}{2!}z''(x_0)(\Delta x)^2 + \frac{1}{3!}z'''(x_0)(\Delta x)^3 + \dots \quad (1.23)$$

The student of thermodynamics should be able to generate such Taylor series expansions for common algebraic and trigonometric functions.

SIDEBAR 1.7: TAYLOR SERIES EXERCISES

Exercises Use the first few terms of the Taylor series expansion (1.23) to develop small- x approximations for the functions

- (a) $z(x) = (1 - x)^{-1}$
- (b) $z(x) = \ln(1 + x)$
- (c) $z(x) = [\cos(x)]^{-1/2}$
- (d) $z(x) = (1 + x^2)^{1/2}$

Solutions

- (a) $(1 - x)^{-1} \simeq 1 + x + x^2 + x^3 + \dots$
 - (b) $\ln(1 + x) \simeq x - x^2/2 + x^3/3 - \dots$
 - (c) $[\cos(x)]^{-1/2} \simeq 1 + x^2/4 + 7x^4/96 + \dots$
 - (d) $(1 + x^2)^{1/2} \simeq 1 + x^2/2 - x^4/8 + \dots$
-

Thermodynamic Description of Simple Fluids

2.1 THE LOGIC OF THERMODYNAMICS

Scientific theory relies on the logic of deductive reasoning. Each scientific deduction builds on one or more *premises* of the theory and leads to *conclusions* that can be tested experimentally.

Scientific theories themselves can be distinguished as “deductive” or “inductive” in nature, according to the underlying character of their premises. In a deductive theory, the fundamental premises are “axioms” or “postulates” that are neither questionable nor explainable within the theory itself. Outstanding examples of deductive theories include Euclidean geometry (based on Euclid’s five axioms) and quantum mechanics (based on Schrödinger’s prescription for converting classical trajectory equations into wave equations). An inductive theory, on the other hand, is based on *universal laws of experience* that express what has always been found to be true in the past, and may therefore be reasonably expected to hold in the future. Thermodynamics is the pre-eminent example of an inductive theory.

Although we cannot “prove” that an inductive law of experience will continue to hold tomorrow (any more than we can “prove” Euclid’s fifth axiom), every day’s continued success adds confidence to the predictions of the theory with respect to new phenomena. Every rigorous prediction of the theory is demanded to exhibit the same universality and infallibility as the underlying inductive laws themselves, further supporting our confidence in the generality and power of the theory. Few if any theories can now rival the confidence that scientists place in thermodynamics (*frontispiece*).

What are the empirical inductive laws on which thermodynamics rests? For future reference, Table 2.1 lists the six general statements IL-1–IL-6 of observational experience on which the present exposition will be based. Several of these require additional definitions or explanations before they can be properly understood. Each will be introduced explicitly in the text as its definitional basis is properly laid and its logical role in the formal construction of the theory becomes apparent.

Examples or illustrations of each law will be provided as they are introduced. However, if one merely understands “in equilibrium” to mean “at the same temperature,” one can

TABLE 2.1 The Inductive “Laws” 1–6 of Equilibrium Thermodynamics
[With Descriptive Alias in Brackets]

IL-1	Two degrees of freedom suffice to determine the equilibrium state of a fixed amount of a simple gas. [AKA: Simplest case of the “Gibbs phase rule”]
IL-2	The mutual relationships between the properties of systems at equilibrium are sufficiently well behaved to permit low-order (i.e., 1st and 2nd) derivatives to be determined by experimentally well-defined procedures. [AKA: Existence of mathematically well-behaved “equations of state”]
IL-3	Two bodies that are each in equilibrium with a third body are in equilibrium with one another. [AKA: “Zeroth law of thermodynamics”]
IL-4	All gases in mutual equilibrium satisfy $\lim_{p \rightarrow 0} PV_m = \text{universal constant}$ at sufficiently low pressure. [AKA: Universality of low-density “ideal” limit]
IL-5	There exists a macroscopic state property U (“internal energy”) whose infinitesimal changes in processes involving only differential absorption of heat dq or performance of work dw on the system are given by $dU = dq + dw$. [AKA: “First law of thermodynamics”]
IL-6	There exists a macroscopic state property S (“entropy”) that achieves the character of a <i>maximum</i> with respect to variations that do not alter the energy of an isolated system at equilibrium, and whose differential changes at equilibrium are given by $dS = dq/T$. [AKA: “Second law of thermodynamics”]

recognize the general consistency of the first four laws with gaseous equations of state as introduced in freshman chemistry.

2.2 MECHANICAL AND THERMAL PROPERTIES OF GASES: EQUATIONS OF STATE

Gases were among the earliest objects of systematic scientific study. Gases are also the simplest form of matter in many respects, so it is appropriate that a study of thermal and mechanical properties of matter begin with the gaseous state.

From the earliest studies of Robert Boyle (Sidebar 2.1) and others, the interdependence of certain properties of gases came to the fore, such as

- pressure P (“the spring of the gas”)
- volume V
- temperature T
- quantity, as measured by mass m or number of moles n

Indeed, this interdependence of all but a small number of independent properties (“degrees of freedom”) is observed to be one of the most striking and universal features of all known gaseous substances. If we consider a fixed quantity of a pure substance (“simple gas”) in the

quiescent “equilibrium” state in which properties such as P , V , T can be considered to have well-defined experimental values, we can summarize this observation, Inductive Law 1 (IL-1), as follows (cf. Table 2.1):

Observation IL-1: *Two* degrees of freedom suffice to determine the equilibrium state of a fixed amount of a simple gas.

This observation is the simplest case of the “Gibbs phase rule” (to be discussed in Section 7.1). It implies, for example, that pressure $P = P(V, T)$ is *uniquely* specified when V and T are chosen, and similarly, that $V = V(P, T)$ or $T = T(P, V)$ are uniquely determined when the remaining two independent variables are specified. Such functional relationships between PVT properties are called *equations of state*. We can also include the quantity of gas (as measured, for example, in moles n) to express the equation of state more generally as

$$f(P, V, T, n) = 0 \quad (2.1)$$

However, for simplicity, we often choose $n = 1$ (so that $V = V_m$ is the molar volume) to focus on PVT relationships only.

A particularly simple and well-known approximate equation of state is the “ideal gas” equation,

$$PV = nRT \quad (2.2)$$

where R , the “gas constant,” has the same numerical value ($R \simeq 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ in SI units; see Appendix) for all known gases. Equation (2.2) combines the results of many known “laws” discovered in the earliest studies of gas behavior, including “Boyle’s law” (or “Gay Lussac’s law”),

$$P \propto \frac{1}{V} \quad (\text{for fixed } T, n) \quad (2.3a)$$

“Charles’ law” (or “Gay-Lussac’s law”),

$$V \propto T \quad (\text{for fixed } P, n) \quad (2.3b)$$

“Avogadro’s hypothesis,”

$$V \propto n \quad (\text{for fixed } P, T) \quad (2.3c)$$

and “Dalton’s law of partial pressures,”

$$\text{“partial” pressure } P_i \propto n_i \quad (\text{for fixed } V, T) \quad (2.3d)$$

for each gas i of an ideal gas mixture with total pressure $P = \sum_i P_i$. The reader is assumed to be familiar with the ideal gas equation of state and its elementary applications.

SIDEBAR 2.1: ROBERT BOYLE (1627–1691)



Robert Boyle has been called the “Father of Chemistry” and the most influential scientist ever born in Ireland. He was born into fabulous wealth as fourteenth (and favorite) child of the Earl of Cork, said to be the largest landholder in Ireland and richest man in all Britain. The precocious young Robert learned Latin, Greek, and French while still a child, then, like his siblings, was sent away from home for early rearing and schooling, arriving finally (at age 8!) at Eton College near London to enjoy the privileges of the best education then available.

At age 12, Robert was sent (with his brother and an academic tutor) on a European tour that was to have a strong formative influence. His keen interest in mathematical studies was first kindled by travels in France and Switzerland. At age 13, he also experienced a violent thunderstorm that led to intense religious conversion and convictions that he maintained throughout life; in the words of a biographer,

... for him a God who could create a mechanical universe—who could create matter in motion, obeying certain laws out of which the universe as we know it could come into being in an orderly fashion—was far more to be admired and worshipped than a God who created a universe without scientific law.

At age 14, after learning Italian in preparation, Boyle continued on to Florence, where he came under the influence of Galileo’s writings and was greatly affected by the circumstances of Galileo’s house imprisonment and death in nearby Arcetri. He thereafter became a strong supporter of Galileo’s philosophy and approach to studying the world through mathematics and mechanics, a philosophy he would later extend into chemistry.

The turmoil of civil war at home left Robert stranded on the continent and somewhat embarrassed financially until age 17. Upon the death of his father, however, he scraped together sufficient funds (by selling his jewelry and borrowing from his tutor) to return home to England, where he had inherited a castle at Stalbridge and estates elsewhere. There, in the midst of ongoing civil strife, Boyle continued his studies in “fits and snatches,” maintaining written contacts with fellow-minded intellectuals of a “new philosophical college” or “Invisible College” that was later to become the nucleus of the Royal Society.

With the defeat of Charles I and ascendancy of Cromwell, Boyle was able to return to his estates in Ireland in 1652, becoming a very wealthy man who was thereafter able to devote himself entirely to science without concern for external support. He thereupon moved to Oxford to associate more closely with other Invisible College members, including leader John Wilkins (Warden of Wadham College), John Wallis (Professor of Geometry), and Christopher Wren (later Professor of Astronomy). Although he never held a university post at Oxford, Boyle equipped his personal rooms as a scientific laboratory and hired Robert Hooke as research assistant, embarking on the studies of gas properties for which he is best known, based on a novel vacuum pump (of Hooke’s design) that allowed

exploration of many aspects of gases and the vacuum state, which up to that time was widely believed to be “impossible.”

What is now known as “Boyle’s law” appeared in the 1662 Appendix to his 1660 work, *New Experiments Physio-Mechanicall, Touching the Spring of the Air and its Effects*. However, Boyle’s greater contribution was to liberate chemistry from the stultifying traditions of alchemy and Aristotelian scholasticism, instituting a modern research style based on the primacy of experiment and detailed publication of methods and results to allow duplication by others. Although he retained certain alchemical goals and beliefs, Boyle brought a decidedly new perspective to chemical studies, as enunciated in his famous book, *The Sceptical Chymist* (1661). As biographers J. J. O’Connor and E. F. Robertson [<http://www-history.mcs.st-andrews.ac.uk>] noted,

... although he did not develop any mathematical ideas himself, he was one of the first to argue that all science should be developed as an application of mathematics. Although others before him had applied mathematics to physics, Boyle was one of the first to extend the application of mathematics to chemistry, which he tried to develop as a science whose complex appearance was merely the result on simple mathematical laws applied to simple fundamental particles.

In 1668, Boyle left Oxford to live with his sister’s family in London, but two years later (at age 42), he suffered a debilitating stroke that left him paralyzed and only partially able to resume scientific interactions. He was a founding member of the Royal Society and was offered its presidency, but declined out of religious scruples concerning the necessary oaths. He continued to advocate strongly on behalf of a corpuscular (atomistic) and kinetic theory of gas properties and the nature of heat (in opposition, for example, to Newton and Lavoisier). Although he continued to produce a steady stream of publications, his health continued to decline until his death in 1691, one week after that of the sister with whom he had lived out the last decades of life.

Equation (2.2) is sometimes referred to as the “ideal gas law.” However, for our present purposes, we must recognize that this “law” [like those summarized in (2.3a–d)] is merely a crude approximation that never describes any real gas exactly, except in the idealized limit of “zero pressure” (to be discussed in Section 2.3). Hence, we must sharply distinguish between crude empirical “laws” (which are at most approximate rules of thumb) and true thermodynamic “laws” as summarized in Table 2.1. A difficulty for the beginning student of thermodynamics is to distinguish those equations that are based on the ideal gas approximation (and thus are practically *never* true) from those of rigorous thermodynamic quality. We shall often flag equations of the former type with “IG” (ideal gas), for example

$$PV \stackrel{\text{IG}}{=} nRT \quad (2.2')$$

to indicate that these hold true only in the limit $P \rightarrow 0$.

Pairwise relationships such as (2.3a–d) can be represented by simple 2-dimensional (2D) graphs, but the full *PVT* behavior of the equation of state requires a 3-dimensional (3D) representation (for fixed $n = 1$). Figure 2.1 illustrates some simple graphical representations of the ideal gas equation of state (2.2). Figure 2.1a illustrates Boyle’s law (2.3a) in

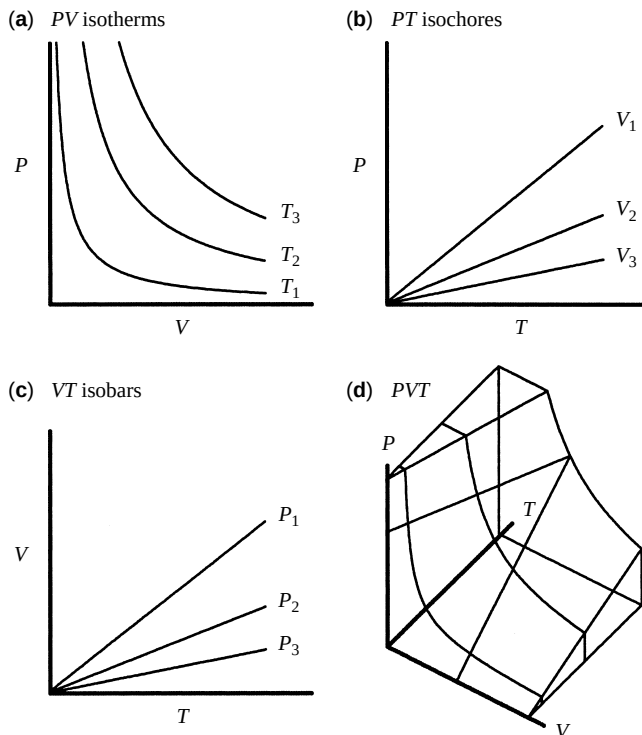


Figure 2.1 Graphical representation of isotherms (a), isochores (b), isobars (c), and the 3D *PVT* surface (d) for an ideal gas.

the *PV* diagram, depicting the *isotherms* (lines of constant T) for chosen temperatures T_1 , T_2 , T_3 . Figure 2.1b similarly shows *isochores* (lines of constant V) in the *PT* diagram, and Fig. 2.1c show *isobars* (lines of constant P) in the *VT* diagram. The composite *PVT* behavior is depicted in the 3D model of Fig. 2.1d, which can be viewed from different directions to see the various 2D projections shown in Fig. 2.1a–c.

For real gases, the graphical representations of equations of state naturally acquire more complex and significant details.

For any real substance, the equation of state (2.1) is always found to be sufficiently “smooth” that we can measure additional *derivative* properties, as summarized in the following Inductive Law 2 (cf. Table 2.1):

Observation IL-2: The mutual relationships between the properties of systems at equilibrium are sufficiently well-behaved to permit low-order (i.e., 1st and 2nd) derivatives to be determined by experimentally well-defined procedures.

While this observation permits many other thermal and mechanical properties to be obtained from the equation of state (2.1), this equation is not yet sufficient for a *complete* thermodynamic description, as will be discussed in Section 5.1.

From the equation of state in the form $V = V(P, T)$, we can now define two important derivative properties of the substance: the *coefficient of thermal expansion* α_P ,

$$\alpha_P \equiv \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P \quad (2.4)$$

and the *isothermal compressibility* β_T ,

$$\beta_T \equiv -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T \quad (2.5)$$

In words, α_P can be described as the fractional volume increase (dV/V) with respect to a temperature increase (dT) under isobaric conditions, while β_T is the corresponding fractional volume decrease ($-dV/V$) with respect to a pressure increase (dP) under isothermal conditions. Of course, both $\alpha_P = \alpha_P(P, T)$ and $\beta_T = \beta_T(P, T)$ vary with P, T , as do other thermodynamic properties. Numerical values of α_P, β_T (e.g., for 1 atm, 25°C) are often tabulated with other material properties, such as density, boiling point, or heat capacity, as unique “fingerprints” of a pure substance. [Throughout this book, experimental values are commonly drawn from standard sources, such as J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954) or any recent edition of the *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, FL).]

SIDEBAR 2.2: IDEAL GAS EXERCISE

Exercise Determine α_P and β_T for an ideal gas. How do these properties reflect the “unreal” character of such a gas?

Solution From $V = nRT/P$, we obtain the partial derivatives $(\partial V/\partial T)_P = nR/P$ and $(\partial V/\partial P)_T = -nRT/P^2$. From the definitions (2.4), (2.5), we therefore obtain

$$\alpha_P = \frac{1}{V} \left(\frac{nR}{P} \right) = \frac{1}{T}, \quad \beta_T = -\frac{1}{V} \left(-\frac{nRT}{P^2} \right) = \frac{1}{P} \quad (\text{S2.2-1})$$

Each property depends on only one (rather than the expected two) degrees of freedom, and each becomes pathological (divergent) in the limit of small T or P , respectively. For solids and liquids, α_P and β_T are rather insensitive to P, T variations, so low-order Taylor series approximations may be adequate. For gases, however, it is generally necessary to differentiate an accurate equation of state to obtain a realistic (P, T) dependence of α_P, β_T .

SIDEBAR 2.3: THERMOMETER PROBLEM

Problem For liquid Hg, $\alpha_P = 1.8 \times 10^{-4} \text{C}^{-1}$, $\beta_T = 3.9 \times 10^{-6} \text{atm}^{-1}$ (effectively constant over the range of laboratory conditions).

- What is $(\partial P/\partial T)_V$ for liquid Hg?
- What is the excess pressure that develops in a fixed-volume mercury thermometer when it is slightly overheated by 2°C?

Solution

- (a) For a pure fluid (two degrees of freedom), we can employ the chain rule for $V = V(T, P)$ to write

$$dV = \left(\frac{\partial V}{\partial T}\right)_P dT + \left(\frac{\partial V}{\partial P}\right)_T dP$$

or, in view of definitions (2.4), (2.5),

$$dV = (V\alpha_P) dT + (-V\beta_T) dP$$

Under constant-volume conditions ($dV = 0$), we can therefore obtain $(\partial P/\partial T)_V$ (i.e., the ratio “ dP/dT at constant V ”) as

$$\left(\frac{\partial P}{\partial T}\right)_V = -\frac{V\alpha_P}{-V\beta_T} = \frac{\alpha_P}{\beta_T}$$

[Alternatively, we could use the Jacobi cyclic identity (1.14b) to write directly

$$\left(\frac{\partial P}{\partial T}\right)_V = -\frac{(\partial V/\partial T)_P}{(\partial V/\partial P)_T}$$

which leads to the same result.] Substituting the numerical values of α_P and β_T , we obtain finally

$$\left(\frac{\partial P}{\partial T}\right)_V = \frac{\alpha_P}{\beta_T} = \frac{1.8 \times 10^{-4} \text{ }^\circ\text{C}^{-1}}{3.9 \times 10^{-6} \text{ atm}^{-1}} \simeq 46 \text{ atm } ^\circ\text{C}^{-1}$$

- (b) For the small overheating $\Delta T = 2^\circ\text{C}$, we can estimate the pressure increase ΔP as

$$\Delta P \simeq \left(\frac{\partial P}{\partial T}\right)_V \Delta T = (46 \text{ atm } ^\circ\text{C}^{-1})(2^\circ\text{C}) = 92 \text{ atm}$$

At such high internal pressure, a glass thermometer will predictably explode!

2.3 THERMOMETRY AND THE TEMPERATURE CONCEPT

Practical measurements of “temperature” long preceded the theory of this important concept. Thermodynamics clearly requires the temperature concept, but *thermometry* (the theory of temperature measurements) is so deeply intertwined with general thermodynamic theory that we must take care to avoid logical circularity.

Let us begin from common subjective perceptions of “hotness” or “coldness” of material bodies and the changes that result when bodies are brought into contact. We know from experience that when a “hotter” body comes into contact with a “colder” body, both bodies undergo changes (we say, “heat flows from the hotter to the colder body”) until they acquire the *same* “degree of hotness.” Thus, even before we can quantify this property, we can agree that two bodies share the *same* degree of hotness when there is *absence of change* (we say, “thermal equilibrium”) when they are brought into thermal contact.

Given this simple concept of thermal “sameness” or “equilibrium,” we can express the results of universal human observations in the following Inductive Law 3, also known as the *zeroth law of thermodynamics*:

Observation IL-3: Two bodies that are each in equilibrium with a third body are in equilibrium with one another.

Observation IL-3 expresses the “transitive” nature of thermal equilibrium, i.e., that if A shares this property with B, and B shares it with C, then A also shares it with C. This observation may seem such an obvious aspect of experience as not to warrant special mention, but it guarantees that we can consistently speak of a *definite property* that is shared by all bodies in thermal equilibrium. We call this property “temperature,” denoted (provisionally) by the symbol Θ .

Let us first attempt to establish an operational *mechanical temperature scale* for Θ that is based solely on mechanical concepts, such as pressure or volume, that are assumed to be well established. (Such a scale may be of little practical utility, but it satisfies the thermodynamicist’s penchant for orderly logic.) To this end, we recall from IL-1 (Table 2.1) that only *two* properties suffice to uniquely fix the value of Θ (as well as all other properties) of a simple gas. We may therefore choose P and V as these independent properties, and express Θ by the functional relationship

$$\Theta = \Theta(P, V) \quad (2.6)$$

Such a functional relationship indeed provides a feasible “mechanical thermometer” based only on measurements of the mechanical variables P , V .

To see how such a thermometer might work, we can suppose that a large array of “standard” systems $\Theta_1(P_1, V_1)$, $\Theta_2(P_2, V_2)$, $\dots$ are stored in a king’s vault as the “official temperature scale.” The Θ_1 standard might be kept as one mole of CO_2 gas at $P_1 = 1$ atm, $V_1 = 22.4$ L, the Θ_2 standard as one mole of water vapor at $P_2 = 2$ atm, $V_2 = 50$ L, and so forth. A sample $\Theta = \Theta(P, V)$ of unknown temperature can therefore be evaluated by bringing it into thermal contact with each member Θ_i of the standard array until a thermal “match” (i.e., equilibrium) is found. (This is similar to determining an unknown “color” by comparing it with all the paint samples at the hardware store until a match is found.) Unwieldy as this may seem, the functional relationship (2.6) yields a valid “mechanical temperature scale” that allows a consistent assignment of temperature Θ to any given system.

An obvious practical improvement can be achieved by choosing each “reference standard” Θ_i to employ the *same* gas and pressure (say, CO_2 at 1 atm), $\Theta_i = \Theta_i(P, V_i)$, so that only the variation of volume V_i is needed to assign Θ_i . This is the basis of *gas thermometry*, which underlies the current international temperature scale. However, it is even more convenient to recognize that the gas can be generally replaced by its corresponding liquid form, the two being related (as will be discussed in Section 2.5) by a *continuous* pathway that preserves the functional relationship (2.6) at every point on the path. For everyday purposes, it is adequate to store the liquid (e.g., water or alcohol) in a glass container under the approximately constant ambient pressure of the atmosphere, etched with markings (and assigned temperature values) that allow the volume variations of the fluid to be expressed as temperature readings. This is the principle of the earliest known instruments for measuring temperature.

However, it is clear that slight variations in vessel shape, etched markings, or external pressure can lead to disagreements as to which thermometer gives the “true” temperature. Moreover, the “reference points” chosen to standardize the readings between different thermometers could be subject to disagreements (see Sidebar 2.4), as could the choice of thermometric fluid (e.g., Hg vs. water, each of which has *different* values of α_p in different temperature ranges). Under these circumstances, the choice of the “true” temperature scale may become subject to non-scientific influences. We therefore seek a *universal* standard that avoids such arbitrary choices.

SIDEBAR 2.4: EARLY THERMOMETERS

Although “thermoscope” devices that could detect differences in temperature (but not assign magnitudes) were developed by Galileo and others in the 16th century, the first liquid-in-glass thermometer (water in a glass bulb, open to the atmosphere) was apparently developed by French physician Jean Rey in 1632. Its “reference points” are not recorded. Improvements led to alternative thermometric fluids (e.g., alcohol and liquid mercury) in sealed glass tubes to reduce the effects of atmospheric pressure variations.

From the 1640s onward, dozens of proposals were put forward for different temperature *scales*, based on two selected “reference points” and their assigned degree-values. Some leading examples are tabulated below.

Despite imaginative suggestions by Isaac Newton and others, the Celsius “centigrade” system became the international scientific standard, prevailing for about two centuries. However, in 1954 the Celsius scale was replaced (at the 10th General Conference on Weights and Measures) by the *single*-reference “ideal gas temperature scale” described in the text.

Date	Inventor	Reference Points			
		(°)	Lower Reference Point	(°)	Upper Reference Point
1641	Archduke Ferdinand of Tuscany	(?)	“Coldest winter cold”	(?)	“Hottest summer heat”
1688	J. Dalencé	(−10)	m.p. of snow	(+10)	m.p. of butter
1701	I. Newton	(0)	m.p. of ice	(+12)	Human body temperature
1708	O. Rømer	(0)	Ice–salt mix	(+60)	b.p. of H ₂ O
1742	D. G. Fahrenheit	(+32)	m.p. of ice	(+212)	b.p. of H ₂ O
1742	A. Celsius ^a	(0)	m.p. of ice	(+100)	b.p. of H ₂ O
1954	10th General Conference on Weights and Measures	(+273.16)	Triple point of H ₂ O	—	—

^aCelsius’s original 0–100 “centigrade” scale was “inverted” (100 = m.p. of ice, 0 = b.p. of H₂O), but was soon righted (on the suggestion of Linneaus) to its present form. Earlier proposals for water-based 0(m.p.)–100(b.p.) scales were apparently presented by Renaldi (1694) and Elvius (1710).

How can one rationally choose the “best” thermometer (or, equivalently, the units of the “true” temperature scale)? As expressed by Gibbs (see Sidebar 11.6), this question can be addressed in terms of “the conditions which it is most necessary for these units to fulfill for the convenience both of men of science and of the multitude.” From the scientific side, these necessary conditions include:

- (i) maximum *simplicity* and *consistency* with underlying theory (i.e., a “bad” thermometer makes thermodynamic calculations unnecessarily complex, whereas a “good” thermometer makes the theory as simple as possible)
- (ii) maximum *universality* (i.e., a “bad” thermometer is idiosyncratic and difficult to replicate, whereas a “good” thermometer is reproducible as needed throughout the world)
- (iii) minimum *conventions* (i.e., a “good” thermometer has the fewest possible “reference points” or other arbitrary conventions)

Let us now describe the conditions of simplicity, universality, and minimal reference points that motivated adoption of the currently accepted international temperature scale.

A rational temperature scale can be based on the limiting validity of the ideal gas equation of state (2.2) at sufficiently low pressure or density. This universal limiting behavior can be expressed in terms of the following Inductive Law 4:

Observation IL-4: All gases in mutual equilibrium satisfy

$$\lim_{P \rightarrow 0} (PV_m) = \text{universal constant}$$

at sufficiently low pressure.

From this observation, we are motivated to define temperature T on the “ideal gas temperature scale” as

$$T = \frac{1}{\text{constant}} \lim_{P \rightarrow 0} (PV_m) \quad (2.7a)$$

for *any* gas. Only one “constant” (the “gas constant” R) remains to be specified in (2.7a) by assigning the numerical value of T (in “kelvin” degrees, K) at a *single* reference point, chosen to be consistent with the experimental “universal constant” of IL-4:

$$T \equiv 273.16\text{K (exactly) at the H}_2\text{O triple point} \quad (2.7b)$$

More specifically, the H₂O triple point is the specific state in which the liquid, gaseous, and ice phases of H₂O (of terrestrial isotopic composition; in the absence of air) *coexist* in equilibrium. (Why such a triple point corresponds to a *unique* specification of temperature and other particulars will be discussed in Section 7.1.)

As defined by (2.7a, b), T is necessarily a non-negative “absolute” value (because P and V_m can only have non-negative values). The value “273.16” in (2.7b), although apparently arbitrary, is shrewdly chosen so that the *size* of a kelvin degree (K) agrees sensibly with that of the older Celsius degree (°C), the previous international standard. Thus, for “the multitude,” the change from centigrade (t) to ideal gas (T) standard appears to be merely a

numerical shift, virtually transparent in everyday usage. However, for “men of science,” the actual differences are profound.

How is the ideal scale (T) related to the previous Celsius (t) scale? The latter is a two-point scale whose reference points were formerly defined as

$$t = 0^\circ\text{C at the H}_2\text{O ice point} \quad (2.8a)$$

$$t = 100^\circ\text{C at the H}_2\text{O boiling point} \quad (2.8b)$$

(both at $P = 1$ atm, saturated with air). Even if pressure can be maintained at the required 1 atm with high precision, experimental replication of the H₂O ice point (2.8a) appears uncertain by about 0.02°C , an intolerable inaccuracy for scientific purposes. However, it was known that the water ice point is 0.010°C above the triple point (within experimental precision). Hence, the Celsius scale is now *defined* with respect to the Kelvin scale by

$$t(^{\circ}\text{C}) \equiv T(\text{K}) - 273.15 \text{ (exactly)} \quad (2.9)$$

This means that the melting point and boiling point of water are no longer *exactly* equal to 0 and 100°C (although this is still sensibly true within current experimental accuracy). Similarly, the value of the gas constant R (currently known to have the approximate value $R \simeq 8.314 \dots \text{J/mol}^{-1} \text{K}^{-1}$) is *determined* from the fundamental definition (2.7b).

The ideal scale, as defined by (2.7a, b), also has an entirely different (and quite surprising!) theoretical basis, related to the maximum efficiency of machines and the second law of thermodynamics. This alternative definition of T (suggested by Kelvin) will be discussed in Section 4.5. However, we can recognize at this point that such a *dual* connection to fundamental thermodynamic principles of great universality gives (2.7a, b) a double-justification to be considered the “true” temperature scale. We henceforth adopt this definition of T throughout this book.

SIDEBAR 2.5: “TEMPERATURE” FROM ENERGY DISTRIBUTIONS

Statistical mechanics (cf. Chapter 13) suggests an alternative way to extract temperature-like properties from molecular energy distributions. According to the classical Boltzmann distribution law, the number $N(E)$ of molecules having energy E is proportional *under equilibrium conditions* to the Boltzmann factor $e^{-E/kT}$,

$$N(E) \propto e^{-E/kT} \quad (\text{S2.5-1})$$

where $k = R/N_A$ (with N_A being Avogadro’s number) is Boltzmann’s constant. Accordingly, if we measure the populations $N(E)$ at two distinct energies E and $E + \Delta E$, we can obtain the ratio

$$\frac{N(E + \Delta E)}{N(E)} = e^{-\Delta E/kT} \quad (\text{S2.5-2})$$

This allows us to solve for the Boltzmann “temperature” T_{Boltz} :

$$T_{\text{Boltz}} = \frac{-\Delta E}{k \ln \left[\frac{N(E + \Delta E)}{N(E)} \right]} \quad (\text{S2.5-3})$$

The population ratio $N(E + \Delta E)/N(E)$ can often be inferred from the intensities of spectroscopic peaks, for example, the rotational lines of a microwave spectrum. This seemingly

allows one to determine the “temperature” of the molecular species from the intensities of a few spectral lines.

Although T_{Boltz} may be considered equivalent to thermodynamic temperature T when the molecules are *truly in equilibrium*, in numerous cases this cannot be safely assumed. Under nonequilibrium conditions, T_{Boltz} may be *negative* (as it is, for example, for a laser), or *different* for different energy modes (e.g., electronic vs. vibrational vs. rotational), or *inconsistent* from energy level to energy level within the same spectroscopic region. Common exceptions occur in the case of population inversions, or indeed in any case where the average energy level spacing ΔE fails to satisfy $kT \gg \Delta E$ (a condition implying blast-furnace temperatures even for vibrational stretching frequencies). In such cases, it is clearly invalid to consider T_{Boltz} as “temperature” in the thermodynamic sense. *Caveat emptor!*

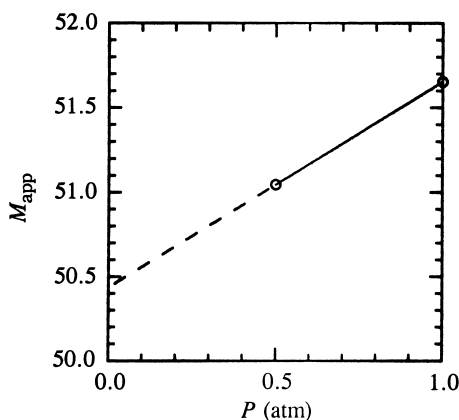
SIDEBAR 2.6: MOLECULAR WEIGHT PROBLEM

Problem From the ideal gas equation (2.2), the apparent molecular weight M_{app} can be estimated as

$$M_{\text{app}} = \frac{mRT}{PV}$$

where m is the mass (in grams) and R , P , V , T have their usual meaning. For an unknown gas, a student carefully measured the ratio mRT/PV : first at $P = 1.000$ atm, where the result was $M_{\text{app}} = 51.65_3$; then at $P = 0.500$ atm, where the result was $M_{\text{app}} = 51.04_5$. However, the student is told that the errors in both estimates are unacceptably large (1–2%), owing to the approximate nature of the ideal gas equation of state. How can she do better?

Solution From Observation IL-4, the student recalls that the ideal gas equation becomes increasingly exact in the limit $P \rightarrow 0$. Accordingly, she plots her measured values of M_{app} versus P and *extrapolates* toward the limit as $P \rightarrow 0$, as shown in the graph. Her new estimate, 50.44, is now within about 0.1% of the true molecular weight (50.49) of the unknown (CH_3Cl), and could be further improved, if needed, by additional measurement(s) at lower P .



2.4 REAL AND IDEAL GASES

The kinetic molecular theory (KMT; see Sidebar 2.7) of Bernoulli, Maxwell, and others provides deep insight into the molecular origin of thermodynamic gas properties. From the KMT viewpoint, pressure P arises merely from the innumerable *molecular collisions* with the walls of a container, whereas temperature T is proportional to the *average kinetic energy* of random molecular motions in the container of volume V . KMT starts from an ultrasimplified picture of each molecule as a mathematical *point* particle (i.e., with no “volume”) with mass m and average velocity $\bar{v}$, but no potential energy of interaction with other particles. From this purely “kinetic picture” of chaotic molecular motions and wall collisions, one deduces that the PVT relationships must be those of an ideal gas, (2.2). Hence, the inaccuracies of the ideal gas approximation can be attributed to the unrealistically oversimplified “noninteracting point mass” picture of molecules that underlies the KMT description.

Formal thermodynamics does not rest on KMT or other molecular assumptions (hence, their relegation to “sidebar” status in this book). Nevertheless, thermodynamic studies are highly valued for their ability to provide fundamental insights into the intermolecular forces that underlie chemical phenomena. Indeed, the most successful advances in thermodynamic theory and practice are often *inspired* by molecular insights, and the productive interplay between microscopic and macroscopic domains should be emphasized in a pedagogically useful presentation of thermodynamic principles. Accordingly, we discuss equations of state in terms of their ability to suggest *improvements* over the KMT ideal gas picture of intermolecular interactions.

SIDEBAR 2.7: KINETIC MOLECULAR THEORY OF IDEAL GASES

Let us consider N molecules (or n moles, with $N = nN_A$, N_A = Avogadro’s number), each a point particle of mass m and average velocity,

$$\bar{v} = (\bar{v}_x^2 + \bar{v}_y^2 + \bar{v}_z^2)^{1/2} \quad (\text{S2.7-1})$$

moving independently, randomly and incessantly in a cubic container of side L and volume $V = L^3$, and undergoing frequent collisions with the walls that are perceived as the “pressure” of the gas.

For simplicity, consider first a single molecule moving with speed $\bar{v}_x$ (and momentum $\bar{p}_x = m\bar{v}_x$) along the x direction. When this particle collides elastically with a wall, it undergoes an abrupt reversal of direction, with momentum change $\Delta p = 2m\bar{v}_x$. It then travels a distance $2L$ (in a time interval $\delta t = 2L/\bar{v}_x$) before returning to collide again with the same wall. By conservation of momentum, each collision imparts a force F on the wall, given by the momentum transfer rate

$$F = \frac{dp}{dt} = \frac{2m\bar{v}_x}{2L/\bar{v}_x} = \frac{m\bar{v}_x^2}{L} \quad (\text{S2.7-2})$$

The pressure P (= force/area) exerted on the wall by collisions due to this single molecule is therefore

$$P \text{ (per molecule)} = \frac{F}{L^2} = \frac{m\bar{v}_x^2}{L^3} = \frac{m\bar{v}_x^2}{V} \quad (\text{S2.7-3})$$

Because motion in each direction is equally likely,

$$\bar{v}_x^2 = \bar{v}_y^2 = \bar{v}_z^2 = \frac{1}{3}\bar{v}^2 \quad (\text{S2.7-4})$$

we can rewrite (S2.7-3) as

$$P(\text{per molecule}) = \frac{m\bar{v}^2}{3V} \quad (\text{S2.7-5})$$

The total pressure P (exerted equally on each wall of the container) is therefore

$$P = \frac{nN_A m\bar{v}^2}{3V} \quad (\text{S2.7-6})$$

which can be rewritten as

$$PV = n \left(\frac{1}{3} N_A m\bar{v}^2 \right) \quad (\text{S2.7-7})$$

According to the “energy equipartition” theorem of classical physics, the three translational kinetic energy modes each acquire average thermal energy $\frac{1}{2}kT$ (where $k = R/N_A$ is Boltzmann’s constant),

$$\frac{1}{2}m\bar{v}_x^2 = \frac{1}{2}m\bar{v}_y^2 = \frac{1}{2}m\bar{v}_z^2 = \frac{1}{2} \left(\frac{1}{3}m\bar{v}^2 \right) = \frac{1}{2}kT \quad (\text{S2.7-8})$$

or in molar terms, noting (S2.7-4),

$$N_A \left(\frac{1}{3}m\bar{v}^2 \right) = N_A kT = RT \quad (\text{S2.7-9})$$

so that (S2.7-7) becomes the ideal gas equation

$$PV = nRT \quad (\text{S2.7-10})$$

2.4.1 Compressibility Factor and Ideal Gas Deviations

As a first step toward improved description of real gases, we define the *compressibility factor* Z by the equation

$$Z \equiv \frac{(V_m)_{\text{obs}}}{(V_m)_{\text{ideal}}} \quad (2.10a)$$

Here $(V_m)_{\text{ideal}} = RT/P$ is the molar volume of an ideal gas and $(V_m)_{\text{obs}} = V/n$ is the actual (“observed”) molar volume; with these substitutions, Z can be defined equivalently as

$$Z \equiv \frac{PV}{nRT} \quad (2.10b)$$

TABLE 2.2 Variations of Observed $(PV_m)_{\text{obs}}$ and Compressibility Factor Z for CO_2 at 40°C^a

P (atm)	1	10	50	100	200	500	1100
$(PV_m)_{\text{obs}}$ (L atm mol $^{-1}$)	25.57	24.49	19.00	6.93	10.50	22.00	40.00
Z	0.99	0.95	0.74	0.27	0.41	0.87	1.56

^aNote that for an ideal gas, $(PV_m)_{\text{ideal}} = 25.70 \text{ L atm mol}^{-1}$ for this temperature.

Although Z is identically *one* for an ideal gas, its value generally deviates from this ideal limit for real gases. Like other measurable properties, Z is expected to depend on two degrees of freedom (for fixed $n = 1$), which can be taken, for example, as P and T ,

$$Z = Z(P, T) \quad (2.11)$$

(or other pairs of variables, according to convenience). We can recognize from (2.10b) that, for T defined as in (2.7), the low-pressure ideal gas limit (IL-4) can be equivalently expressed in terms of Z as

$$\lim_{P \rightarrow 0} Z = 1 \quad (2.12)$$

since $V_m \rightarrow (V_m)_{\text{ideal}}$ as $P \rightarrow 0$. Thus, measured values of Z carry information about deviations from ideal gas behavior at nonzero pressures.

Let us first examine the pressure dependence of the compressibility factor under isothermal conditions. Table 2.2 and Fig. 2.2 exhibit some representative P -dependent values of Z for gaseous CO_2 (at fixed temperature $40^\circ\text{C} = 313\text{K}$) in tabular and graphical form.

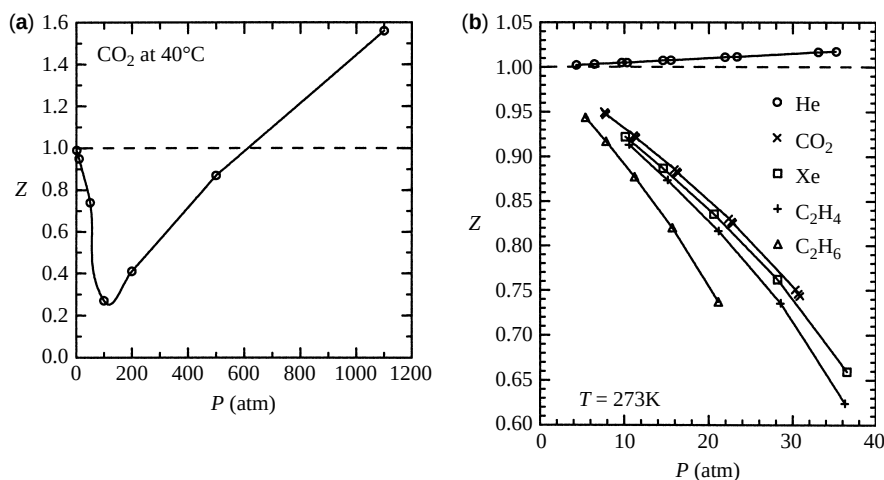


Figure 2.2 P -dependent compressibility factor Z for CO_2 (a) over a broad pressure range at 40°C , and (b) compared with other gases in a narrower low-pressure range at 0°C . (Interpolating curves connecting the data points are shown to aid visualization.)

The graph of Z versus P in Fig. 2.2 is rather typical, showing the following general features that should be noted:

- (i) Near the origin ($P \rightarrow 0$), Z approaches ideality ($Z \rightarrow 1$), as expected from (2.12) (a dashed line marks the “reference value” to aid visualization of deviations from ideality at other pressures).
- (ii) For low pressures (and densities), Z is generally less than unity ($Z < 1$), showing that the gas volume is *reduced* from its “ideal” value, presumably owing to *attractive forces* between molecules that continue to “draw molecules together” even at lower density (long-range separation).
- (iii) For high pressures, Z always becomes *greater than* unity ($Z > 1$), presumably owing to short-range *repulsive forces* that oppose the “squeezing together” of molecules at higher density.
- (iv) At some intermediate pressure (near 600 atm for CO_2), Z exhibits “cross-over” (near-ideal) behavior, presumably owing to *cancelling* effects of long-range attractive and short-range repulsive forces on gas volume (rather than, for example, total absence of such forces, as assumed in the KMT picture).

As shown in Fig. 2.2, the deviations from ideality appear rather negligible in the $P \simeq 1$ atm range of usual laboratory conditions. However, these errors become *strongly* negative (e.g., approximately 75% errors near 100 atm) or positive (e.g., apparently unbounded errors as P increases above 700 atm) in other pressure regions. Thus, the ideal gas approximation (2.2) is qualitatively unreasonable over a wide range of pressures.

It is also informative to examine the temperature dependence of Z . Figure 2.3 exhibits the schematic shifts of Z with T for a CO_2 -like gas (actually, the Van der Waals equation

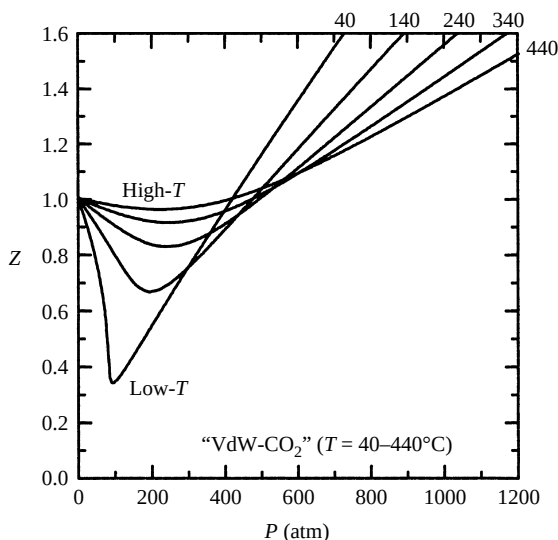


Figure 2.3 Schematic T -dependent shifts in Z for a CO_2 -like gas (i.e., Van der Waals approximation), showing the gradual shifts toward ideality at all pressures.

of state approximation for CO₂; Section 2.4.2). From this figure, we note the salient general features:

- (i) At low T , Z exhibits ever-larger negative deviations from $Z = 1$, culminating in near-zero Z values at condensation (Section 2.5), indicative of near-complete volume collapse under the dominant influence of attractive intermolecular forces.
- (ii) As T increases, Z increases at lower P and decreases at higher P , decreasing the deviations from ideality in both pressure ranges.
- (iii) At higher T , only the high- P positive deviations remain appreciable, reflecting the dominant influence of repulsive molecular forces in this limit.

The qualitative picture is that ideality is recovered at sufficiently high T , but nonideality becomes ever stronger (in the negative-deviations sense) as T decreases toward the gas–liquid condensation point.

Although the general behavior of $Z(P, T)$ shows parallels in all gases, the *magnitudes* of nonideal deviations (as well as their characteristic condensation temperatures) vary considerably from one gas to another, as shown in Fig. 2.2b. Most nearly ideal are low-boiling gases such as He and H₂, whereas less volatile or hydrogen-bonding gases such as NH₃ and H₂O exhibit dramatically larger deviations from ideality, even under near-ambient conditions. A quantitative description of $Z(P, T)$ must therefore take account of the strong variations of intermolecular forces (or, equivalently, differences in the intermolecular potential energy function; Sidebar 2.8) that appear to characterize different gases. Stated another way, $Z(P, T)$ contains rich information about hydrogen bonding and other intermolecular coordination phenomena of chemical importance.

SIDEBAR 2.8: THE INTERMOLECULAR POTENTIAL ENERGY FUNCTION

In classical mechanics, the force F acting between two molecules A, B separated by distance R is determined from the *potential energy function* $\mathcal{V}(R)$ by the equation

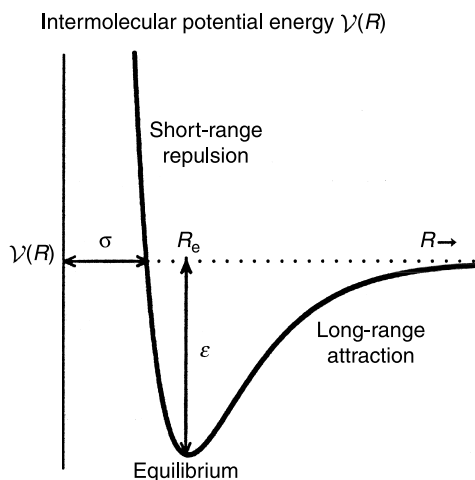
$$F(R) = -\frac{d\mathcal{V}}{dR} \quad (\text{S2.8-1})$$

[For more complex vector-field potentials depending on the relative orientation as well as the separation of the two particles, the corresponding vector expression is $\mathbf{F} = -\hat{\nabla}\mathcal{V}$, where $\hat{\nabla} = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$ is the gradient operator. Such vectorial aspects of intermolecular forces are obviously important for real molecules of nonspherical shape.]

Knowledge of $\mathcal{V}(R)$ is a “Holy Grail” that motivates and unifies many areas of modern and experimental physical chemistry research, including:

- (i) bulk thermodynamic measurements (as described in this chapter)
- (ii) molecular beam scattering (“quantum pinball,” using recoil patterns from molecular collisions to determine the shape of the scattering potential $\mathcal{V}$)
- (iii) spectroscopic measurements (using quantum energy levels determined from spectral transitions to determine the supporting potential $\mathcal{V}$)
- (iv) *ab initio* quantum chemistry computations (using direct solutions of Schrödinger’s equation to determine $\mathcal{V}$)

A schematic form of the intermolecular potential energy curve $\mathcal{V}(R)$ is shown below:



The horizontal dotted line marks the “zero” of interaction energy as $R \rightarrow \infty$, and the lowest potential energy below this line (at R_e) gives the attractive “well depth” ϵ of the interaction. At smaller R (particularly inside the “hard core diameter” σ), $\mathcal{V}(R)$ slopes steeply downward (i.e., rises steeply as R diminishes), so the force F is strongly positive (repulsive). At large R (well beyond R_e), $\mathcal{V}(R)$ slopes softly upward, so F is softly attractive. At the equilibrium distance $R = R_e$, the slope $d\mathcal{V}/dR$ vanishes and $F = 0$, corresponding to the well depth ϵ relative to the dissociation limit (dotted line) at $R \rightarrow \infty$.

The *Lennard-Jones* (“6–12”) potential $\mathcal{V}_{\text{LJ}}(R)$ provides a simple 2-parameter approximation to the intermolecular potential between closed-shell molecular species. This function can be written in the form (with empirical parameters a, b)

$$\mathcal{V}_{\text{LJ}}(R) = \frac{a}{R^{12}} - \frac{b}{R^6} \quad (\text{S2.8-2})$$

or, alternatively (in terms of hard-core diameter σ and equilibrium well depth ϵ) as

$$\mathcal{V}_{\text{LJ}}(R) = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right] \quad (\text{S2.8-3})$$

This potential exhibits a steeply repulsive R^{-12} term that dominates at small R and a weakly attractive R^{-6} term that dominates in the long-range limit, with empirical parameters a, b (or equivalently, ϵ, σ) that can be adjusted to best approximate the experimental behavior.

Problem Find the equations that express ϵ, σ in terms of a, b , or vice versa.

Solution From the graph above, we see that at $R = \sigma$,

$$\mathcal{V}(\sigma) = \frac{a}{\sigma^{12}} - \frac{b}{\sigma^6} = 0 \quad (\text{S2.8-4})$$

from which $\sigma = (a/b)^{1/6}$. To find ε , we need R_e , which is the point for which the derivative $d\mathcal{V}/dR$ vanishes,

$$\frac{d\mathcal{V}}{dR} = \frac{-12a}{R_e^{13}} + \frac{6b}{R_e^7} = 0 \quad (\text{S2.8-5})$$

from which $R_e = (2a/b)^{1/6} = 2^{1/6}\sigma$. Finally, from the potential value $(-\varepsilon)$ at $R = R_e$, we obtain

$$-\varepsilon = \mathcal{V}(R_e) = \mathcal{V}(2^{1/6}\sigma) = \mathcal{V}\left(\left(\frac{2a}{b}\right)^{1/6}\right) = \frac{a}{(2a/b)^2} - \frac{b}{(2a/b)} = \frac{-b^2}{4a} \quad (\text{S2.8-6})$$

from which $\varepsilon = b^2/4a$. The inverse relationships are $a = 4\varepsilon\sigma^{12}$, $b = 4\varepsilon\sigma^6$.

Exercises In terms of ε , σ , find expressions for

- (i) R_e
- (ii) the force at $R = \sigma$
- (iii) the value of $\mathcal{V}(R)$ at $R = 2R_e$
- (iv) the “curvature” (stretching force constant) $k = d^2\mathcal{V}/dR^2$ at R_e

Solutions (i) $2^{1/6}\sigma$; (ii) $24\varepsilon/\sigma$; (iii) $-\varepsilon(127/4096)$; (iv) $72\varepsilon/(2^{1/3}\sigma^2)$.

The Lennard-Jones potential is a rather crude representation of an actual intermolecular potential, chosen more often for computational simplicity than chemical accuracy. More accurate for many chemical purposes is the *Morse potential*

$$\mathcal{V}_{\text{Morse}}(R) = \varepsilon\{1 - \exp[-a(R - R_e)]\}^2 \quad (\text{S2.8-7})$$

In this expression, ε and R_e have the same physical meaning as discussed above, and the “width” parameter a is related to the bonding force constant k by

$$k = 2\varepsilon a^2 \quad (\text{S2.8-8})$$

It is an instructive student exercise to compare plots of $\mathcal{V}_{\text{LJ}}$ and $\mathcal{V}_{\text{Morse}}$ for fixed ε and R_e , noting the dramatic differences at both large and small R despite “matching” bond length and strength. One size does not fit all!

2.4.2 Van der Waals and Other Model Equations of State

In 1873, Dutch physicist J. D. Van der Waals (Sidebar 2.9) presented (in his doctoral thesis) the celebrated equation of state that now bears his name. The Van der Waals equation of state may be written in a form

$$\left[P + a\left(\frac{n}{V}\right)^2\right](V - nb) = nRT \quad (2.13)$$

that suggests how the ideal gas equation can be modified by including empirical parameters a , b (different for different gases) that partially correct for two of the most serious defects of the KMT ideal gas picture (Sidebar 2.7). Van der Waals' discovery of this equation was strongly motivated by his attempt to understand the essential *continuity* between gaseous and liquid states of matter, established a few years earlier by T. H. Andrews' discovery (1869) of the gas–liquid critical point (Section 2.4.3). The Van der Waals equation (and associated molecular viewpoint that underlies its derivation) marked a pivotal advance in the molecular theory of gases and liquids, recognized by the 1910 Nobel Prize in Physics.

SIDEBAR 2.9: J. D. VAN DER WAALS (1837–1923)



Johannes Diderik Van der Waals was born into the family of a Dutch carpenter in Leiden, the eldest of ten children. His working-class education did not include the classical languages then required to sit for admission to the university, so he pursued a career as an elementary school teacher, meanwhile using spare time to take advantage of a loophole that allowed outside students to attend a few courses in mathematics and science at the University of Leiden. In 1863, the government established a new tier of secondary-school education that allowed Van der Waals to begin teaching mathematics and science at Deventer and The Hague. Miraculously, another change of law then allowed the Minister of Education to grant special dispensations from the university

language requirement, and this loophole allowed Van der Waals to enter the doctoral program at Leiden University, where he completed (at age 36) a brilliant thesis, *On the Continuity of the Gaseous and Liquid State* (1873). This work, stimulated by the 1869 discovery of the liquid–gas critical point by Thomas Andrews, instantly propelled Van der Waals to scientific prominence and was to alter the course of physical chemistry; as Maxwell pronounced upon reading his first paper, “there can be no doubt that the name of Van der Waals will soon be among the foremost in molecular science.”

Van der Waals was soon invited to become the first Professor of Physics at the newly formed University of Amsterdam, which he helped raise to world prominence. Unfortunately, in 1881 (just after his formulation of the law of corresponding states), his young wife and mother of four children died of tuberculosis (at age 34), leaving Van der Waals so shaken that he never remarried, and published no work for about a decade. But, inspired by the work of Gibbs, he resumed work in the 1890s on binary solutions [for an engaging account, see J. M. H. Levelt Sengers, *How Fluids Unmix: Discoveries by the Schools of Van der Waals and Kamerlingh Onnes* (Royal Netherlands Academy of Arts and Sciences, Amsterdam, 2002)], capillarity, and other topics. Upon his retirement at age 70 in 1908, he was succeeded by his only son, Johannes Diderik Van der Waals, Jr. In 1910, he received the Nobel Prize in Physics.

Van der Waals' Nobel lecture (December 12, 1910) gives deep insight into his lifelong struggle to understand and extend his famous namesake equation, organized into four main points:

1. Van der Waals acknowledges the profound influence of Clausius' 1857 paper on the kinetic theory of heat, which provided the first correct proof of Maxwell's velocity distribution law and led to the

revelation for me. . . [that] this is obviously still so . . . in liquids, which can only be regarded as compressed gases at low temperature. Thus I conceived the idea that there is no essential difference between the gaseous and the liquid state of matter, [whereas] the crystalline state definitely behaves in a slightly different way.

Van der Waals clearly recognized the two factors (associated with his a , b parameters) that prevent exact agreement with Boyle's law, as well as the remaining discrepancy with experiment that "continually obsesses me, I can never free myself from it, it is with me even in my dreams."

2. Van der Waals emphasizes his conviction that the "constant" b actually carries significant V dependence, and his astonishment ("to my great joy") to learn that such dependence does not detract from essential consistency with the law of corresponding states.
3. Van der Waals recognizes the essential need to incorporate the additional physical effect of "pseudo-association":

. . . bluntly speaking, the result would be: an equation of state compatible with experimental data is totally impossible. No such equation is possible, unless something is added, namely that the molecules associate to form larger complexes . . . I have termed it "pseudo association" to differentiate it from the association which is of chemical origin. The possible formation of larger molecular complexes, particularly in the liquid state, has frequently been emphasized and the finding that the assumption is necessary to achieve agreement between the state equation and experiment will hence cause no surprise. Unfortunately, my examination is still incomplete. I have found it arduous.

Van der Waals also recognizes compelling evidence that "the attraction of the molecules decreases extremely quickly with distance, indeed that the attraction only has an appreciable value at distances close to the size of the molecules," and he recalls Debye's explanation of a prescient remark by Boltzmann: "Debye's remark implied that Boltzmann had predicted the formation of a complex." He further remarks that

If pseudo association exists in a substance, there are at least two types of molecules, namely simple and complex. I say at least two types because it cannot be assumed that all complexes are of equal size. But as a first step I have assumed only two types, i.e., simple molecules and n -fold molecules. For a really scientific treatment, of course, it would be necessary to assume all values of n as possible and to seek the law of distribution for these values.

In these remarks, one can see a pioneering suggestion of a *cluster mixture* theory of liquids with short-range (exchange-like) forces, along the lines of Mayer cluster theory (Sidebar 13.5) or quantum cluster equilibrium theory (Section 13.3.4).

4. Van der Waals acknowledges the unexpected difficulty in extending his equation to binary mixtures, and the key role of Gibbs' theory of free energy minimization in handling both binary mixtures and association complexes.

Finally, Van der Waals feels impelled to offer impassioned testimony to his personal belief in atoms and molecules, then still considered controversial among some physicists (see Sidebar 13.7):

... it does not seem to be superfluous, perhaps it is even necessary to make a general observation. It will be perfectly clear that in all my studies I was quite convinced of the real existence of molecules, that I never regarded them as a figment of my imagination, nor even as mere centres of force effects. I considered them to be the actual bodies ... When I began my studies I had the feeling that I was almost alone in holding that view ... Many of those who opposed it most have ultimately been won over, and my theory may have been a contributory factor. And precisely this, I feel, is a step forward. Anyone acquainted with the writings of Boltzmann and Willard Gibbs will admit that physicists carrying great authority believe that the complex phenomena of the heat theory can only be interpreted in this way.

Physical chemists of the current generation can hardly imagine how atomic and molecular concepts were once doubted by serious scientists, even into the early 20th century! Our current enlightenment reflects the debt that we owe to Boltzmann, Gibbs, Van der Waals, and other pioneer theorists.

Van der Waals sought to address two basic defects of the KMT “noninteracting point mass” picture: (i) neglect of the *finite molecular volume* that distinguishes molecules from mathematical “points”; and (ii) neglect of the *intermolecular attraction* that leads to condensation (liquid formation) at sufficiently low temperature. Whereas the ideal gas equation (2.2) exhibits no vestige of condensation phenomena, the Van der Waals equation (2.13) is intended to provide a unified description of gas–liquid (“fluid”) behavior, exhibiting the essential *commonality* that must be shared by these disparate forms of matter at the molecular level.

Let us first rewrite (2.2) in slightly different form

$$P_{\text{ideal}} V_{\text{ideal}} = nRT \quad (2.14)$$

to remind ourselves that P_{ideal} and V_{ideal} are associated with the unrealistic ideal gas KMT assumption of noninteracting mathematical point-particles of mass m and average speed $\bar{v}$ undergoing free translational motions throughout the gas container. We now consider how corrections (i) and (ii) modify the form of this equation, using Fig. 2.4 to illustrate these corrections.

(i) Let us suppose that each molecule can be represented by a hard sphere of diameter d , as shown in Fig. 2.4i. Because these hard spheres cannot interpenetrate, their centers of mass (shown as central dots) are prevented from approaching closer than distance d , thereby *reducing* the volume V_{ideal} that is available for free translational motions. (Think of a professional ice skater on a crowded public rink who is forced to restrict her performance glides to the small remaining portion of the ice surface that is not occupied by other skaters.) We can express this reduction as

$$V_{\text{ideal}} = V - V_{\text{excl}} \quad (2.15)$$

where V_{excl} is the “excluded” portion of the total volume V that is occupied by other molecules (confirming the intuition that real gas molecules indeed have finite volumes). We can estimate V_{excl} in a simple way, assuming that pair encounters are the primary collisional

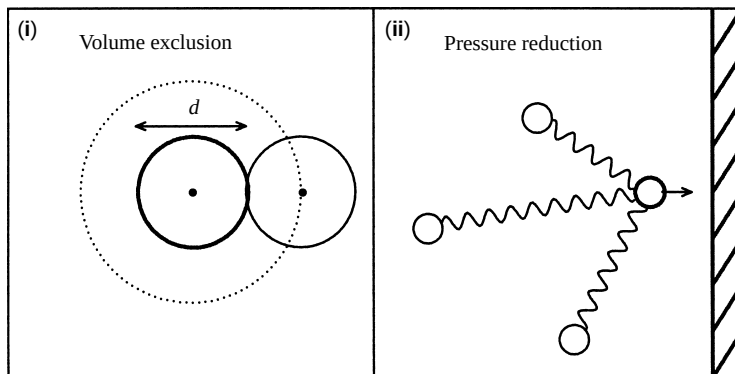


Figure 2.4 Schematic representation of Van der Waals corrections: (i) Dotted line showing the spherical excluded region (of volume $\frac{4}{3}\pi d^3$) surrounding a probe molecule (heavy circle) that is inaccessible to the center of mass of another molecule of diameter d . (ii) Wavy lines (molecular attractions) depicting the net “pulling” effect of attractions by surrounding molecules on a given probe molecule (heavy circle) about to strike the wall, thereby reducing the impact of collision and resulting pressure of wall collisions.

events that limit translational mobility. For any pair of adjacent molecules, we can recognize from Fig. 2.4i that each is surrounded by a sphere of volume $\frac{4}{3}\pi d^3$ that is excluded to the center of the other. To avoid overcounting, we assign *half* this volume to each member of the pair to obtain

$$V_{\text{excl}}(\text{per molecule}) = \frac{1}{2} \left(\frac{4}{3}\pi d^3 \right) \quad (2.16)$$

or, for the total system of $N = nN_A$ molecules,

$$V_{\text{excl}}(\text{total}) = nN_A \left(\frac{2}{3}\pi d^3 \right) \quad (2.17)$$

Because the molecular diameter d is unknown, we can rewrite (2.17) as

$$V_{\text{excl}} = nb \quad (2.18a)$$

where b is a constant, proportional to molecular volume,

$$b = \frac{2}{3} N_A \pi d^3 \quad (2.18b)$$

that can be treated as an empirical parameter. With the substitution (2.18a), we finally obtain from (2.15)

$$V_{\text{ideal}} = V - nb \quad (2.19)$$

to express the finite-volume correction (i) to (2.14).

(ii) Let us now consider the effect of intermolecular attractions, as pictured in Fig. 2.4ii. One can recognize from the figure that a molecule nearing a wall collision experiences anisotropic attractions from nearby molecules that all act to *reduce* its forward velocity toward the wall, and thus its contribution to pressure. We express this pressure

reduction P_{red} (with respect to the corresponding P_{ideal} of a noninteracting ideal gas) by the equation

$$P = P_{\text{ideal}} - P_{\text{red}} \quad (2.20)$$

The attractive forces, represented by wiggly lines in Fig. 2.4ii, may have complex R dependences that make it difficult to quantify the pressure reduction P_{red} . However, for a specified collider molecule, we can expect that the attractive force is proportional to the *number density* n/V of surrounding molecules, so that, for example, doubling the number of surrounding molecules at distance R also *doubles* their effect on collision impact and pressure. We express this proportionality in the form

$$P_{\text{red}}(\text{per collider molecule}) \propto \frac{n}{V} \quad (2.21)$$

Moreover, we recognize that the total number of collider molecules (or collisions) is also proportional to n/V , leading to an overall *squared* dependence on number density:

$$P_{\text{red}}(\text{total}) \propto \left(\frac{n}{V}\right)^2 \quad (2.22)$$

We introduce an empirical constant a to express this proportionality in the form

$$P_{\text{red}} = a\left(\frac{n}{V}\right)^2 \quad (2.23)$$

Equation (2.20) therefore becomes

$$P = P_{\text{ideal}} - a\left(\frac{n}{V}\right)^2 \quad (2.24)$$

which we can rewrite as

$$P_{\text{ideal}} = P + a\left(\frac{n}{V}\right)^2 \quad (2.25)$$

to express the effect of the molecular-attractions correction (ii).

It only remains to replace the uncorrected equation (2.14) by substituting the corrections expressed by (2.19) and (2.25). When these substitution are made, we obtain the Van der Waals equation (2.13), which can also be expressed in terms of molar volume $V_{\text{m}} = V/n$ as

$$\left(P + \frac{a}{V_{\text{m}}^2}\right)(V_{\text{m}} - b) = RT \quad (2.26)$$

Hence, this improved equation of state can be seen as a direct consequence of the improved incorporation of molecular volume and attraction effects as depicted in Fig. 2.4.

The relationship to the earlier discussion of real gas properties (Section 2.4.1) can be demonstrated by re-expressing the Van der Waals equation in terms of Z . To obtain Z_{vdw} for the Van der Waals gas, we first rewrite (2.13) in expanded form as

$$PV - nbP + \frac{an^2}{V} - \frac{abn^3}{V^3} = nRT \quad (2.27)$$

Dividing by nRT and using (2.10b), we obtain

$$Z_{\text{vdw}} = 1 + \frac{1}{RT} \left\{ \left[-a \left(\frac{n}{V} \right) \right] + \left[bP + ab \left(\frac{n}{V} \right)^2 \right] \right\} \quad (2.28)$$

From this equation, we can see that the total nonideality correction (in braces) contains a negative contribution (first bracketed term) that is indeed proportional to the “attractions” constant a , while the positive contribution (second bracketed term) is proportional to the finite-volume “repulsions” constant b , as was supposed in the interpretation of experimental Z behavior in Fig. 2.2. One can also see that the attractions term is *linearly* proportional to density n/V , whereas the repulsions term is proportional to *squared* density $(n/V)^2$, so that the former must always prevail at low density (low P) and the latter at high density (high P), as was shown in Fig. 2.2. Furthermore, one can recognize from the $1/RT$ prefactor that the entire nonideality correction must diminish with increasing T , as was noted in Fig. 2.3. Thus, regardless of the particular values chosen for a and b , the Van der Waals equation is expected to exhibit both pressure and temperature dependences that are qualitatively consistent with the observed $Z(P, T)$ behavior.

The Van der Waals constants a , b are generally chosen to represent gas–liquid critical properties, as will be described in Section 2.5. Table 2.3 lists values of these empirical constants for some selected gases. It can be seen from this table that the b values tend to increase (with notable exceptions, such as Ne compared with He!) in the expected manner with molecular size, as judged, for example, from standard molecular structure models. The a values similarly increase with qualitative measures of molecular attractions, as judged, for example, by their boiling points or other signatures of strong association such as hydrogen bonding.

The accuracy of the Van der Waals equation can be indicated by comparison with experimental compressibility factor data, as illustrated for a representative case (CO_2 at 40°C ; cf. Fig. 2.2) in Fig. 2.5. The improved description of real gas PVT behavior is immediately apparent. As the figure shows, the Van der Waals Z_{vdw} closely approximates the experimental Z through the low-pressure region of increasing negative deviations from ideality (up to about $P \simeq 100$ atm). However, Z_{vdw} rises too steeply at higher pressures (e.g., crossing the ideal line $Z = 1$ about 200 atm before the experimental crossing), grossly exaggerating the positive deviations from ideality at large P . Thus, the Van der Waals $Z_{\text{vdw}}(P)$ exhibits qualitatively reasonable shape (particularly in the low-pressure region) and is certainly a major improvement on the ideal gas description. Nevertheless, significant differences

TABLE 2.3 Van der Waals Constants a and b for Selected Gases

Gas	a (atm mol ² L ⁻²)	b (L mol ⁻¹)	Gas	a (atm mol ² L ⁻²)	b (L mol ⁻¹)
He	0.0341	0.02370	CO ₂	3.592	0.04267
Ne	0.2107	0.01709	NH ₃	4.170	0.03707
Ar	1.345	0.03219	H ₂ O	5.46	0.0305
H ₂	0.2444	0.02661	CH ₄	2.253	0.04278
O ₂	1.360	0.03183	CCl ₄	20.39	0.1383
N ₂	1.390	0.03913	Benzene	18.00	0.1154
Cl ₂	6.493	0.05622	<i>n</i> -Octane	37.32	0.2368

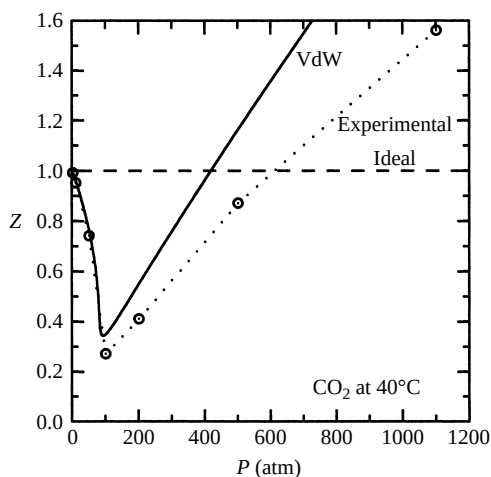


Figure 2.5 Compressibility factor $Z(P)$ for CO_2 at 40°C (cf. Fig. 2.2), comparing the Van der Waals approximation (solid line) with experimental values (circles, dotted line) and with the ideal gas approximation (dashed line).

from experiment are still evident, and one may conclude that the Van der Waals equation provides only the first tier of improvements to the ideal gas picture.

The success of the Van der Waals equation inspired many attempts to develop improved equations of state based on more complex functional forms and additional empirical fitting parameters. Sidebar 2.10 lists some modified equations of state (and associated empirical parameters) that have been proposed and found useful in engineering applications.

Although increasingly parametrized empirical equations of state can “fit” selected real gas properties with ever-increasing accuracy, the fundamental significance of these parameters is problematic. The parameters usually lack theoretical significance that would permit their numerical evaluation from deeper principles (i.e., quantum mechanics or other physical models and measurements lying outside the fitting procedure itself). For any given parametrization, the accuracy tends to vary from one PVT region to another, but if the accuracy is inadequate, there is no systematic way to improve the description. Thus, we seek an alternative formulation of the equation of state that is more directly connected to the underlying theory of intermolecular forces, is systematically improvable to any desired accuracy, and (at least in principle) requires no empirical “fitting” parameters to fix its form.

SIDEBAR 2.10: EMPIRICAL EQUATIONS OF STATE

A few of the simpler and more commonly known empirical equations of state are shown below for comparison with the Van der Waals equation $[(P + a/V^2)(V - b) = RT]$ of 1873. Many other such equations (some with 20 or more parameters!) have appeared in the literature. Note that empirical parameters of one equation are generally unrelated to

those in another equation, even if given the same symbol (a , b , etc.). All equations are expressed for $n = 1$ ($V = V_m$).

Clausius equation (1880)

$$\left[P + \frac{a}{T(V+c)} \right] (V-b) = RT \quad [\text{parameters (3): } a, b, c] \quad (\text{S2.10-1})$$

Dieterici equation (1899)

$$P(V-b) = RT \exp(-a/VRT) \quad [\text{parameters (2): } a, b] \quad (\text{S2.10-2})$$

Berthelot equation (1907)

$$\left(P + \frac{a}{TV^2} \right) (V-b) = RT \quad [\text{parameters (2): } a, b] \quad (\text{S2.10-3})$$

Redlich–Kwong equation (1949)

$$\left[P + \frac{a}{V(V+b)T^{1/2}} \right] (V-b) = RT \quad [\text{parameters (2): } a, b] \quad (\text{S2.10-4})$$

Beattie–Bridgeman equation (1927)

$$\left[P + \frac{A_0(1-a/V)}{V^2} \right] \left[\frac{V^2}{V-B_0(1-a/V)} \right] = RT \left(1 - \frac{C}{VT^2} \right) \\ [\text{parameters (5): } A_0, B_0, C, a, b] \quad (\text{S2.10-5})$$

Benedict–Webb–Rubin equation (1940)

$$P = RT/V + (BRT - A - C/T^2)/V^2 + (bRT - a - d/T)/V^3 \\ + \alpha(a + d/T)/V^6 + (c/T^2 V^3)(1 + \gamma/V^2)e^{-\gamma/V^2} \\ [\text{parameters (9): } A, B, C, a, b, c, d, \alpha, \gamma] \quad (\text{S2.10-6})$$

2.4.3 The Virial Equation of State

In 1901, H. Kamerlingh Onnes introduced a fundamentally new and improved description of real gas *PVT* properties in terms of the *virial equation of state*. [The word “virial,” deriving from the Latin word *viris* (“force”) was introduced into physics by R. Clausius, whom we shall meet later.] This equation expresses the compressibility factor $Z(V_m, T)$ in terms of a general *power series* expansion in inverse molar volume V_m . The starting point for the virial expansion is the ideal limiting behavior (2.12), which can also be expressed as

$$\lim_{(1/V_m) \rightarrow 0} Z = 1 \quad (2.29)$$

(since $V_m^{-1} \rightarrow 0$, or $V_m \rightarrow \infty$, as $P \rightarrow 0$). For fixed T , we can therefore employ the general Taylor series concept (1.23) to develop the V_m dependence of $Z(V_m, T)$ around the limit (2.29) as the infinite power series

$$Z(V_m, T) = 1 + B(T)\left(\frac{1}{V_m}\right) + C(T)\left(\frac{1}{V_m}\right)^2 + D(T)\left(\frac{1}{V_m}\right)^3 + \cdots \quad (2.30)$$

Here $B(T)$ is the *second virial coefficient*, $C(T)$ the *third virial coefficient*, and so forth. Formally, the virial coefficients can be defined as successive partial derivatives of Z with respect to inverse molar volume (density) under isothermal conditions; for example, $B(T)$ is given by

$$B(T) = \left(\frac{\partial Z}{\partial (1/V_m)} \right)_T \bigg|_{V_m \rightarrow \infty} \quad (2.31)$$

For low density (large V_m), the series (2.30) is expected to achieve useful accuracy with only a few terms. Higher densities within the domain of convergence require additional terms to achieve a desired accuracy. For some densities, the virial series may not converge at all.

The virial coefficients $B(T)$, $C(T)$, $D(T)$, $\dots$ are functions of temperature only. Although these coefficients might be treated simply as empirical fitting parameters, they are actually deeply connected to the theory of intermolecular *clustering*, as developed by J. E. Mayer (Sidebar 13.5). More specifically, the second virial coefficient $B(T)$ is related to the intermolecular potential for *pairs* of molecules, the third virial coefficient $C(T)$ to that for *triples* of molecules, and so forth. For example, knowledge of the intermolecular pair potential $\mathcal{V}(R)$ (see Sidebar 2.8) allows $B(T)$ to be explicitly evaluated by statistical mechanical methods as

$$B(T) = 2\pi N_A \int_0^\infty \left[1 - e^{-\mathcal{V}(R)/kT} \right] R^2 dR \quad (2.32)$$

Although we shall not pursue such specific molecular formulas in the present thermodynamic exposition, they form the basis for modern molecular-level understanding of gases and liquids [J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954)] and support the conviction that the virial equation (2.30) has deeper fundamental significance than other proposed equations of state.

The leading correction to ideality arises from the second virial coefficient $B(T)$, whose qualitative T dependence is shown in Fig. 2.6 (approximating the experimental data for CO_2). As shown in the figure, $B(T)$ rises from strongly negative values near the low- T condensation limit to weakly positive values at very high T . At an intermediate T known as the *Boyle temperature* T_{Boyle} , the second virial coefficient vanishes:

$$B(T_{\text{Boyle}}) = 0 \quad (2.33)$$

At this point, the gas behavior becomes anomalously “ideal,” owing to fortuitous cancellation (rather than absence) of attractive and repulsive potential energy contributions, and nonideality appears only in the higher-order density terms.

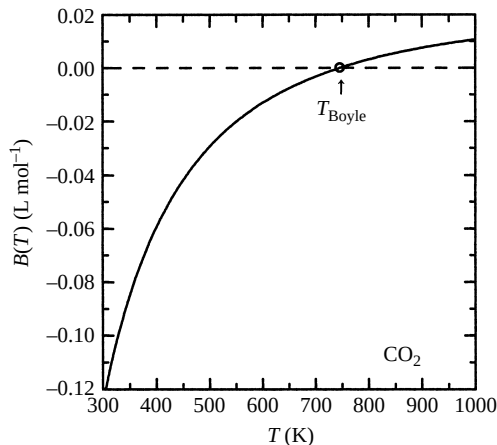


Figure 2.6 Representative T dependence of the second virial coefficient $B(T)$, showing the strong negative deviations from ideality at small T , the weak positive deviations at high T , and the Boyle temperature ($T_{\text{Boyle}} \simeq 750\text{K}$ for CO_2) where $B(T_{\text{Boyle}})$ vanishes.

The virial expansion (2.30) can also be used to expand empirical equations of state in power series form, allowing term-by-term comparisons of one equation with another. To illustrate this procedure, let us obtain the virial expansion for the Van der Waals equation (2.26), which can be rewritten as

$$P = \frac{RT}{V_m - b} - \frac{a}{V_m^2} \quad (2.34)$$

Multiplying through by V_m/RT to give Z on the left-hand side, we obtain

$$Z_{\text{vdW}} = \frac{V_m}{V_m - b} - \left(\frac{a}{RT}\right) \left(\frac{1}{V_m}\right) = \frac{1}{1 - b/V_m} - \left(\frac{a}{RT}\right) \left(\frac{1}{V_m}\right) \quad (2.35)$$

The final term on the right-hand side is already of desired $(1/V_m)^n$ power series form, but the first is not. However, we note that $x = b/V_m$ is a very small number ($x \ll 1$) and recall the general Taylor series expansion for $1/(1 - x)$ (Sidebar 1.7a):

$$\frac{1}{1 - b/V_m} = \frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots \quad (2.36)$$

Equation (2.35) then becomes

$$Z_{\text{vdW}} = \left[1 + \left(\frac{b}{V_m}\right) + \left(\frac{b}{V_m}\right)^2 + \left(\frac{b}{V_m}\right)^3 + \cdots \right] - \frac{a}{RT} \left(\frac{1}{V_m}\right) \quad (2.37)$$

which can be rearranged to standard virial series form

$$Z_{\text{vdW}} = 1 + \left[b - \frac{a}{RT}\right] \left(\frac{1}{V_m}\right) + [b^2] \left(\frac{1}{V_m}\right)^2 + [b^3] \left(\frac{1}{V_m}\right)^3 + \cdots \quad (2.38)$$

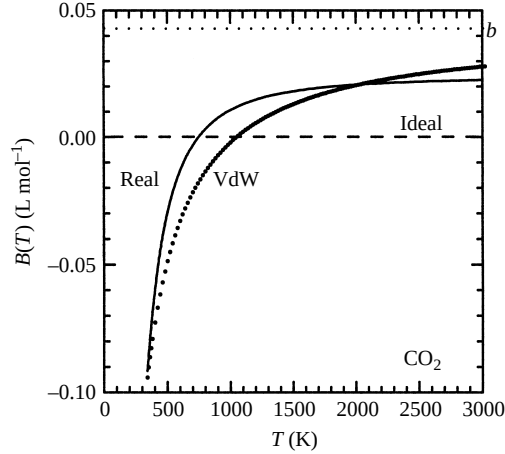


Figure 2.7 Comparison of Van der Waals $B_{\text{VdW}}(T)$ (heavy dotted line) with realistic $B(T)$ (solid line; cf. Fig. 2.6) for CO_2 . The light dotted line at $b = 0.04267$ marks the asymptotic limit of B_{VdW} as $T \rightarrow \infty$.

Comparing term-by-term with (2.30), we obtain the desired Van der Waals virial coefficients in the form

$$B_{\text{VdW}}(T) = b - \frac{a}{RT} \quad (2.39a)$$

$$C_{\text{VdW}}(T) = b^2 \quad (2.39b)$$

$$D_{\text{VdW}}(T) = b^3 \quad (2.39c)$$

and so forth.

The higher Van der Waals virial coefficients $C_{\text{VdW}}, D_{\text{VdW}}, \dots$ are rather unrealistic (temperature-independent), but the expression (2.39a) for $B_{\text{VdW}}(T)$ appears qualitatively reasonable. Figure 2.7 compares the plot of $B_{\text{VdW}}(T)$ (heavy dotted line) with the more realistic $B(T)$ (solid line) for CO_2 , showing the latter over a wider temperature range than in Fig. 2.6 in order to bring out the limiting high- T behavior. B_{VdW} is seen to approximate B rather closely in the low- T condensation limit, but it rises too slowly through the Boyle point, giving a T_{Boyle} estimate about 270K too high. In the high- T limit, $B_{\text{VdW}}(T)$ continues to rise toward the asymptotic value b (shown as the dotted horizontal line), that is more than twice as high as the most positive value of the actual $B(T)$. Thus, the Van der Waals approximation to the second virial coefficient $B(T)$ is qualitatively reasonable, but it exhibits significant errors (particularly at higher T) that indicate the need for further improvement.

2.5 CONDENSATION AND THE GAS-LIQUID CRITICAL POINT

Let us now examine more global aspects of the PVT behavior of a real substance, including regions beyond the domain of convergence of a virial series. Whereas an ideal gas can never undergo the phenomenon of condensation, it is known that all real substances undergo *phase transitions* to high-density liquid or solid forms at sufficiently low T or high P . In the ensuing, we generally consider $n = 1$ mole of substance ($V = V_m$) unless otherwise specified.

The abrupt volume collapse signaling condensation from gas to liquid phase is shown schematically in Fig. 2.8 (approximating the behavior of CO_2). The phase transition is depicted in a PV diagram (called the “indicator diagram,” for its relationship to the “Watt’s indicator” of early steam engines), showing the abrupt volume collapse (circles) near $P = 52 \text{ atm}$ on the $T = 280\text{K}$ isotherm. The horizontal plateau corresponds to 2-phase liquid–gas mixtures of variable proportions (modeled by the Maxwell equal-area construction; cf. Fig. 2.10), ranging from all liquid at the left ($V_L \simeq 81 \text{ mL}$) to all gas at the right ($V_G \simeq 265 \text{ mL}$). The dotted line in Fig. 2.8 marks the dome-like boundary of gas–liquid coexistence, showing how the gas–liquid volume difference ($V_G - V_L$) progressively narrows at higher temperatures or pressures approaching the top of the dome. Note that the horizontal segment of the $T = 280\text{K}$ isotherm does not depict the dispersed vapor bubbles of the seething liquid (or dispersed liquid droplets of the steamy vapor phase) that one “sees” when passing through the liquid–vapor boiling point at finite rate, but only the equilibrium proportions of liquid and vapor in the long-time limit when phase separation is complete.

Figure 2.9 shows the corresponding behavior in a PT diagram (called the “phase diagram,” for its special relationship to phase equilibria; Chapter 7). The circled point at $T = 280\text{K}$, $P = 52 \text{ atm}$ corresponds to the plateau in Fig. 2.8 (whose circled endpoint is now seen end-on in the PT projection). This point therefore falls on the gas–liquid *coexistence curve* (or “vapor pressure curve”), which extends from the “triple point” (triangle) to the “critical point” ($\times$). Two other coexistence curves also emanate from the triple point, corresponding to gas–solid and solid–liquid phase transitions that meet at the unique temperature and pressure for 3-phase (solid, liquid, gas) coexistence. The lines of the phase diagram thereby delineate *stability boundaries* for gas, liquid, or solid phases. Because the phases differ dramatically in mechanical, electrical, optical and other properties, the experimental locations of phase boundaries are of considerable importance to the materials scientist or engineer.

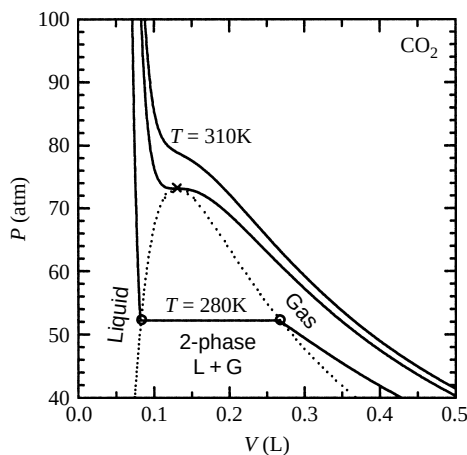


Figure 2.8 Representative supercritical ($T = 310\text{K}$) and subcritical ($T = 280\text{K}$) Van der Waals isotherms for CO_2 , showing the liquid–gas (L + G) condensation plateau ($P = 52 \text{ atm}$) for $T = 280\text{K}$, and outlining the 2-phase liquid–gas coexistence “dome” (dotted line) topped by the critical point ($\times$) at $T_c = 304\text{K}$, $P_c = 73 \text{ atm}$.

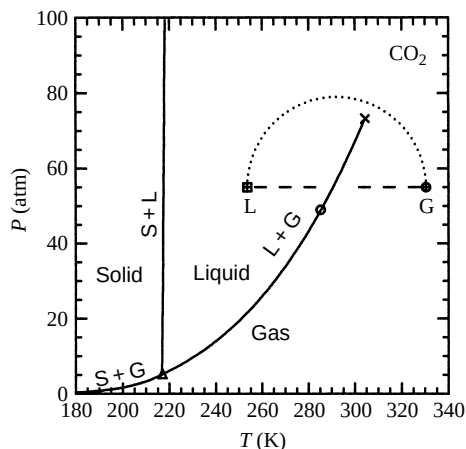


Figure 2.9 Phase diagram for CO_2 , showing solid–gas ($S + G$, “sublimation”), solid–liquid ($S + L$, “fusion”), and liquid–gas ($L + G$, “vaporization”) coexistence lines as PT boundaries of stable solid, liquid, or gaseous phases. The triple point (triangle), critical point ($\times$), and selected 280K isotherm of Fig. 2.8 (circle) are marked for identification. Note that the fusion curve tilts slightly forward (with slope 75 atm K^{-1}) and that the sublimation and vaporization curves meet with slightly discontinuous slopes (angle $< 180^\circ$) at the triple point. The dotted and dashed half-circle shows two possible paths between a “liquid” (cross-hair square) and a “gas” (cross-hair circle) state, one discontinuous (dashed) crossing the coexistence line, the other continuous (dotted) encircling the critical point (see text).

The “critical point” ($\times$) of the phase diagram is of special interest and importance. As shown in Fig. 2.9, the point (T_c, P_c) represents the terminus of the gas–liquid coexistence line, i.e., the point at which a distinction between “gas” and “liquid” forms is no longer meaningful. The same point occurs at the top of the dotted coexistence dome in Fig. 2.8. Below this point, the coexisting phases are distinguishable, for example, by the meniscus that separates the high-density liquid from the low-density gas form. However, at any $T > T_c$ or $P > P_c$, we can no longer observe a boiling point or other evidence of discontinuous change between distinct gas and liquid forms. In this supercritical region, we therefore speak only of a “fluid,” without attempting to distinguish whether it is gas or liquid (because this distinction is no longer meaningful). The existence of the gas–liquid critical point was first discovered in 1869 by Thomas Andrews, greatly influencing the subsequent work of Van der Waals and other liquid theorists.

The approach to the critical point, from above or below, is accompanied by spectacular changes in optical, thermal, and mechanical properties. These include critical opalescence (a bright milky shimmering flash, as incident light refracts through intense density fluctuations) and *infinite* values of heat capacity, thermal expansion coefficient α_P , isothermal compressibility β_T , and other properties. Truly, such a “confused” state of matter finds itself at a “critical” juncture as it transforms spontaneously from a uniform and isotropic form to a symmetry-broken (nonuniform and anisotropically separated) *pair* of distinct phases as (T_c, P_c) is approached from above. Similarly, as (T_c, P_c) is approached from below along the $L + G$ coexistence line, the densities and other phase properties are forced to become identical, erasing what appears to be a fundamental physical distinction between “liquid” and “gas” at all lower temperatures and pressures.

TABLE 2.4 Critical Constants T_c , P_c , V_c , and Compressibility Factor $Z_c = P_c V_c / RT_c$ for Selected Gases

Gas	T_c (K)	P_c (atm)	V_c (L mol ⁻¹)	Z_c	Gas	T_c (K)	P_c (atm)	V_c (L mol ⁻¹)	Z_c
He	5.2	2.26	0.0578	0.31	CO ₂	304.2	73.0	0.0957	0.280
Ne	44.4	26.9	0.0417	0.308	NH ₃	405.5	111.3	0.0724	0.242
Ar	150.7	48.0	0.0749	0.291	H ₂ O	647.1	217.7	0.0568	0.233
H ₂	33.2	12.8	0.0650	0.305	CH ₄	191.1	45.8	0.0993	0.290
O ₂	154.8	50.1	0.078	0.308	CCl ₄	556.4	45.0	0.276	0.272
N ₂	126.2	33.5	0.0899	0.291	Benzene	561.8	48.6	0.260	0.274
Cl ₂	417.1	76.1	0.124	0.276	<i>n</i> -Octane	569	24.6	0.488	0.257

Although liquids and gases appear superficially dissimilar under ordinary subcritical conditions, the fact that there is no separating boundary above the critical point indicates their essential “sameness.” Indeed, as found by Andrews, it is possible to pass *continuously* from “liquid” to “gas” by following a path of (T, P) changes that takes one *around* the critical point, thus bypassing any abrupt “transition” [as occurs at the boiling point if the connecting path remains below (T_c, P_c)]. Such a path is shown as a semicircular dotted line in Fig. 2.9, continuously connecting a “liquid” (L) and “gaseous” (G) state that would otherwise lead to a discontinuous boiling transition if connected by the direct isobaric heating path (horizontal dashed line). Note that no corresponding critical point is observed to terminate the liquid–solid or gas–solid phase boundaries for known substances. The critical point therefore establishes a deep-seated unity between gaseous and liquid phases that is not shared with a crystalline solid phase.

Table 2.4 displays critical constants T_c , P_c , V_c and critical compressibility factor Z_c for a number of common gases. (Accurate determination of the critical point is experimentally challenging, and quoted values are generally uncertain in the final decimal.) One can see from the table that many common “gases” (including N₂, O₂, and CH₄) are actually supercritical fluids (“permanent gases”) under ambient temperature conditions, incapable of liquefaction by any applied pressure whatsoever. (Aspects of cryogenic gas-liquefaction techniques are discussed in Section 3.6.3.)

2.6 VAN DER WAALS MODEL OF CONDENSATION AND CRITICAL BEHAVIOR

Despite its quantitative flaws, the Van der Waals equation (2.13) provides an intriguing description of condensation and critical behavior that mimics certain aspects of real fluid behavior. We therefore wish to examine some detailed aspects of the Van der Waals description of condensation and critical phenomena as a starting conceptual model.

Figure 2.10a displays details of some Van der Waals PV isotherms in the near-critical region. As shown in the figure, the isotherms below T_c exhibit pronounced oscillatory “loops,” quite unlike the experimental plateau-like behavior shown in Fig. 2.8. Indeed, no simple cubic equation could exhibit true plateau-like behavior, and the near-horizontal looping pattern in Fig. 2.10a is apparently the best that a cubic polynomial can do to represent such a “flattened” region. [The fact that the Van der Waals $P = P(V)$ isotherm is a cubic polynomial will be apparent from expanding (2.13).]

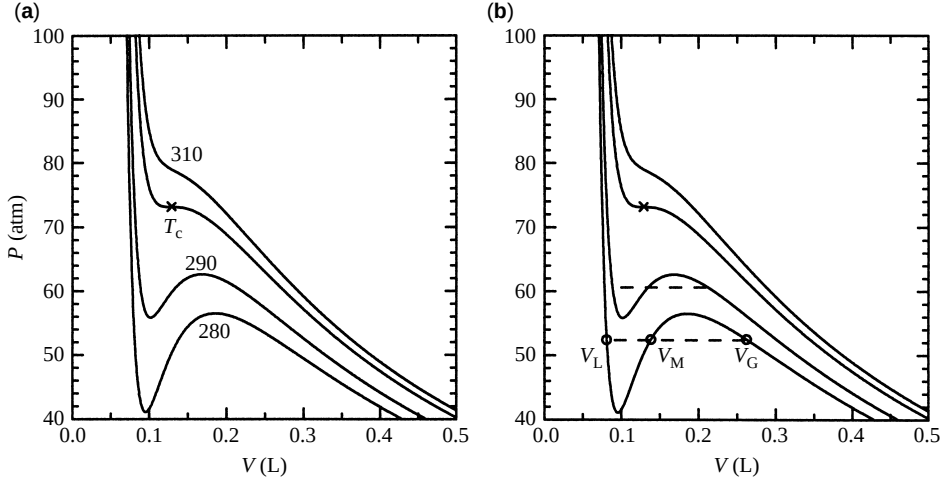


Figure 2.10 Representative Van der Waals PV isotherms for CO_2 near the critical point ($\times$), showing (a) contrasting monotonic behavior above T_c (at $T = 310\text{K}$) compared with oscillatory “loops” below T_c (at $T = 280$ and 290K); (b) Maxwell’s equal-area construction for finding pressure P_0 (horizontal dashed line) that cuts off equal areas in the upper loop (between V_M and V_G) and the lower loop (between V_L and V_M); $P_0 = 52.2\text{ atm}$ for $T = 280\text{K}$, 60.4 atm for 290K .

It was shown by J. C. Maxwell that a horizontal line can be drawn through the Van der Waals loop region in such a way that the area enclosed *above* the line in the upward loop exactly matches that enclosed *below* the line in the downward loop (“Maxwell’s equal-area construction”). As shown in Fig. 2.10b, this horizontal line (say, at pressure P_0) can be taken as the Van der Waals approximation to the actual condensation plateau, bounded on the left by the steeply sloping “liquid” branch, and on the right by the more gently sloping “gaseous” branch of the isotherm. The three points where this horizontal line $P = P_0$ crosses the Van der Waals isotherm may be obtained as the *roots* of the cubic polynomial $P = P(V)$ for $P = P_0$, i.e., as solutions of the equation

$$V^3 - \left[b + \frac{RT}{P_0} \right] V^2 + \left[\frac{a}{P_0} \right] V - \left[\frac{ab}{P_0} \right] = 0 \quad (2.40)$$

According to the general theory of cubic equations, this equation will have three distinct real roots V_1, V_2, V_3 whenever the discriminant D ,

$$D = \left[\frac{3a - 2bRT}{P_0} - \left(\frac{RT}{P_0} \right)^2 - b^2 \right]^3 + \frac{1}{4} \left[9a \left(\frac{2b}{P_0} - \frac{RT}{P_0^2} \right) + 2 \left(b + \frac{RT}{P_0} \right)^3 \right]^2 \quad (2.41)$$

is positive, $D > 0$. If the roots are ordered $V_1 < V_2 < V_3$ we can identify $V_1 = V_L$ as the “liquid” root, $V_3 = V_G$ as the “gas” root, and $V_2 = V_M$ as the “middle” (unphysical) root. That the behavior is profoundly unphysical near V_M can be seen from the fact that $(\partial P / \partial V)_T$ is *positive* in this region, i.e., volume *increases* with increased P (in violation of the 2nd law of thermodynamics; see Section 5.2). Moreover, the second derivative

$(\partial^2 P / \partial V^2)_T$ is necessarily changing sign near V_M (from positive at lower V to negative at higher V). These loop anomalies characterize all subcritical isotherms $T < T_c$, whereas $(\partial P / \partial V)_T < 0$ for all $T > T_c$.

The critical point (T_c, P_c) can therefore be uniquely identified with the *inflection point* that occurs at the top of the coexistence dome, i.e., the point at which both first and second derivatives of the isotherm vanish,

$$\left(\frac{\partial P}{\partial V}\right)_T = \left(\frac{\partial^2 P}{\partial V^2}\right)_T = 0 \quad \text{at } P_c, V_c, T_c \quad (2.42)$$

This analytic characterization allows us to find V_c, T_c (and therefore P_c) as solutions of the two differential equations in (2.42). (A noncalculus method for finding these same solutions is sketched in Sidebar 2.11 below.)

Let us solve the inflection equations (2.42). Starting from the Van der Waals equation in the form

$$P = \frac{RT}{V-b} - \frac{a}{V^2} \quad (2.43)$$

we obtain from the first criterion, $(\partial P / \partial V)_T = 0$,

$$\left(\frac{\partial P}{\partial V}\right)_T = -\frac{RT}{(V-b)^2} + \frac{2a}{V^3} = 0 \quad (2.44a)$$

which is rewritten as

$$\frac{2a}{V^3} = \frac{RT}{(V-b)^2} \quad \text{at } T_c, V_c \quad (2.44b)$$

The second criterion, $(\partial^2 P / \partial V^2)_T = 0$, leads to

$$\left(\frac{\partial^2 P}{\partial V^2}\right)_T = \frac{2RT}{(V-b)^3} - \frac{6a}{V^4} = 0 \quad (2.45a)$$

which can be rewritten as

$$\frac{2RT}{(V-b)^3} = \frac{6a}{V^4} = \frac{3}{V} \left(\frac{RT}{(V-b)^2} \right) \quad \text{at } T_c, V_c \quad (2.45b)$$

where we have used (2.44b) to eliminate a in the final term. Removing the common factors $RT/(V-b)^2$ from both sides of (2.45b), and inserting the value V_c at which the inflection conditions hold, we obtain $\frac{2}{3} V_c = V_c - b$, or

$$V_c = 3b \quad (2.46)$$

When this value of V_c is substituted back into (2.44b), we obtain the expression for T_c as

$$T_c = \frac{8a}{27Rb} \quad (2.47)$$

Finally, substituting these values of T_c , V_c back into the original Van der Waals equation (2.43) gives P_c ,

$$P_c = \frac{RT_c}{V_c - b} - \frac{a}{V_c^2} = \frac{(8a/27b)}{2b} - \frac{a}{9b^2}$$

which simplifies to

$$P_c = \frac{a}{27b^2} \quad (2.48)$$

Equations (2.46)–(2.48) give the Van der Waals prediction of the critical state (P_c , V_c , T_c) for chosen empirical parameters a , b .

SIDEBAR 2.11: VAN DER WAALS CRITICAL PROPERTIES WITHOUT DERIVATIVES

By factoring the Van der Waals polynomial into its roots, we can rewrite (2.40) in the form

$$(V - V_L)(V - V_M)(V - V_G) = 0 \quad (S2.11-1)$$

As is apparent from Fig. 2.10b, the three roots necessarily coalesce at the critical point

$$V_L = V_M = V_G = V_c \quad (S2.11-2)$$

so (S2.11-1) can be rewritten at this point as

$$(V - V_c)^3 = V^3 - [3V_c]V^2 + [3V_c^2]V - [V_c^3] = 0 \quad (S2.11-3)$$

At the same point, (2.40) is

$$V^3 - \left[b + \frac{RT_c}{P_c} \right] V^2 + \left[\frac{a}{P_c} \right] V - \left[\frac{ab}{P_c} \right] = 0 \quad (S2.11-4)$$

The coefficients of each power V^n in (S2.11-3), (S2.11-4) can therefore be equated term-by-term to give

$$3V_c = b + \frac{RT_c}{P_c} \quad (S2.11-5)$$

$$3V_c^2 = \frac{a}{P_c} \quad (S2.11-6)$$

$$V_c^3 = \frac{ab}{P_c} \quad (S2.11-7)$$

Solution of these three equations for the three unknowns T_c , P_c , V_c leads, as before, to (2.46)–(2.48). (Note, however, that this “trick” of equating algebraic coefficients does not work for more general equations of state, so the calculus-based derivation is clearly preferred.)

The expressions (2.46)–(2.48) for the three critical-point parameters can be inverted in many possible ways (depending on which pair of T_c , P_c , V_c are chosen as independent variables) to obtain the two Van der Waals parameters a , b . The most commonly chosen solutions are

$$a = \frac{9}{8}RT_cV_c = 3P_cV_c^3 \quad (2.49)$$

$$b = \frac{1}{3}V_c \quad (2.50)$$

but many other expressions are equally valid [e.g., $a = (3/4)^3(RT_c)^2/P_c$, $b = RT_c/8P_c$ in terms of T_c , P_c]. If we choose R to be a third “unknown” for determinate solution of (2.46)–(2.48), we obtain the rather dubious consequence of Van der Waals theory that

$$R \text{ “=” } \frac{8}{3} \frac{P_cV_c}{T_c} \quad (2.51)$$

Tests of this “prediction” against experimental critical-point data of Table 2.4 reveal large deviations (e.g., an approximately 20% error even in the most favorable case of He) that reflect serious quantitative defects of the Van der Waals description. This is but one of many indications that the Van der Waals equation, although a distinct improvement over the ideal gas equation, is still a significantly flawed representation of real fluid properties.

2.7 THE PRINCIPLE OF CORRESPONDING STATES

Even though the Van der Waals equation is quantitatively unreliable, its qualitative mathematical form suggests certain deeper truths. Particularly striking is the fact that the critical state (P_c , V_c , T_c) seems to form the “reference point” or “origin” from which the Van der Waals equation can be expressed in an elegant *universal* form for all gases.

To see how this universal form arises, let us first substitute the expressions (2.49)–(2.51) for the empirical constants a , b , R into the Van der Waals equation (2.34) to obtain

$$P = \frac{RT}{V-b} - \frac{a}{V^2} = \frac{8}{3} \frac{P_cV_c}{T_c} \left(\frac{T}{V - 1/3V_c} \right) - \frac{3P_cV_c^2}{V^2} \quad (2.52)$$

Dividing by P_c and rearranging gives

$$\left[\frac{P}{P_c} + \frac{3}{(V/V_c)^2} \right] \left(\frac{V}{V_c} - \frac{1}{3} \right) = \frac{8}{3} \left(\frac{T}{T_c} \right) \quad (2.53)$$

We now define the “reduced” variables P_r , V_r , T_r , which are *dimensionless* pure numbers representing “fractional proximity” to the critical point,

$$P_r \equiv \frac{P}{P_c}, \quad V_r \equiv \frac{V}{V_c}, \quad T_r \equiv \frac{T}{T_c} \quad (2.54)$$

Gases having the same numerical values of P_r , V_r , T_r are said to be in “corresponding states,” because each stands in the same proportionate relationship to its own critical point. With the substitutions (2.54), the Van der Waals equations become expressed in pure numbers

$$\left(P_r + \frac{3}{V_r^2} \right) \left(V_r - \frac{1}{3} \right) = \frac{8}{3} T_r \quad (2.55)$$

independent of empirical parameters, physical constants, or choice of units.

In effect, the substitutions (2.54) have allowed us to write the Van der Waals equation of state in a universal functional form

$$f_{\text{vdW}}(P_r, V_r, T_r) = 0 \quad (2.56)$$

that is mathematically identical for all gases. The elegant simplicity of (2.55) indicates that all gases “look the same” when expressed in reduced-variables form. Could a similar conclusion hold true for more accurate equations of state?

Van der Waals himself originally conjectured that universal functional relationships of the form (2.56) remain valid beyond the Van der Waals approximation. This conjecture, the principle (often “law”) of corresponding states, can be stated as follows:

Principle of Corresponding States: For real gases, there exists a universal equation of state of the form

$$f(P_r, V_r, T_r) = 0 \quad (2.57)$$

in terms of reduced variables P_r , V_r , T_r .

This principle asserts that gases in the same corresponding state are, in effect, “identical” with respect to their thermodynamic *PVT* relationships.

The validity of the principle of corresponding states has been carefully investigated experimentally, particularly by Guggenheim [E. A. Guggenheim, *J. Chem. Phys.* **13**, 253 (1945)]. Available experimental evidence indicates that numerous real gases satisfy this principle to high accuracy, but not exactly. One obvious consequence of this principle is that the critical compressibility factor $Z_c = P_c V_c / RT_c$ should be a universal constant for all substances, but one can readily judge from experimental critical constants (Table 2.4) that the actual values of Z_c vary significantly (e.g., from 0.23 for H_2O to 0.31 for He). Theoretical criteria [see, e.g., K. S. Pitzer, *Quantum Chemistry* (Prentice-Hall, Englewood Cliffs, NJ, 1953), pp. 344–9] also identify cases where the corresponding states hypothesis must fail. Although (2.57) falls short of being a rigorous inductive “law” of thermodynamics, it often serves as a useful guide for constructing more accurate equations of state. It may also be useful in estimating inaccessible *PVT* properties of one gas in terms of accessible properties of another gas, as illustrated in Sidebar 2.12.

SIDEBAR 2.12: CORRESPONDING STATES PROBLEM

Problem You are asked to estimate the molar volume of CCl_4 at $T = 500\text{K}$, $P = 50\text{ atm}$, but available laboratory facilities do not permit such high temperatures in the high-pressure container. How can you proceed?

Solution Choose a different gas, such as CO_2 , whose critical temperature is in a more accessible region. From the critical properties in Table 2.4, the desired state of CCl_4 has the reduced variables

$$T_r = \frac{500}{556.4} = 0.899, \quad P_r = \frac{50}{45.0} = 1.11$$

The corresponding state of CO_2 is therefore at the more convenient experimental values

$$T = 0.899(304.2) = 273\text{K}, \quad P = 1.11(73) = 81 \text{ atm}$$

If the molar volume of CO_2 at (273K, 81 atm) is measured to be $V(\text{CO}_2)$, the corresponding molar volume of CCl_4 at (500K, 50 atm) can be estimated by equating the reduced volumes

$$\frac{V(\text{CCl}_4)}{V_c(\text{CCl}_4)} = \frac{V(\text{CO}_2)}{V_c(\text{CO}_2)}$$

to give the estimate

$$V(\text{CCl}_4) \simeq V(\text{CO}_2) \frac{0.276}{0.0957}$$

whose accuracy should exceed that obtained directly from the Van der Waals equation.

2.8 NEWTONIAN DYNAMICS IN THE ABSENCE OF FRICTIONAL FORCES

“Dynamics” is the study of matter in motion. The starting point for classical dynamics is Newton’s equation of motion; for a particle of mass m at position $\mathbf{r} = \mathbf{r}(t)$, the force $\mathbf{F}$ leads to motion described by

$$\mathbf{F} = m\mathbf{a} = \frac{d\mathbf{p}}{dt} \quad (2.58)$$

where $\mathbf{a} = d^2\mathbf{r}/dt^2$ is the acceleration, $\mathbf{p} = m\mathbf{v}$ is the momentum, and $\mathbf{v} = d\mathbf{r}/dt$ is the velocity of the particle. For many purposes, the terms “mass” m (the quantity of matter) and “weight” $W = W(\mathbf{r})$ (the gravitational attraction force at point $\mathbf{r}$) may be considered synonymous, because the values of these properties are simply proportional,

$$\frac{m_1}{m_2} = \frac{W_1}{W_2} \quad (2.59)$$

for objects on a laboratory balance.

An important concept of dynamics is the “work” w performed by an external force acting through a distance. If force $\mathbf{F}$ is applied through differential distance $d\mathbf{r}$, the differential work is defined as

$$dw = \mathbf{F} \cdot d\mathbf{r} \quad (2.60)$$

The dot product between vectors $\mathbf{F} = (F_x, F_y, F_z)$ and $d\mathbf{r} = (dx, dy, dz)$ can be written more explicitly in terms of Cartesian components as

$$dw = F_x dx + F_y dy + F_z dz \quad (2.61a)$$

or in terms of the lengths $F = (\mathbf{F} \cdot \mathbf{F})^{1/2}$, $dr = (d\mathbf{r} \cdot d\mathbf{r})^{1/2}$ and angle θ between vectors,

$$dw = F \cos \theta dr \quad (2.61b)$$

Note from (2.58) that force $\mathbf{F}$ has the physical units of (mass)(length)(time) $^{-2}$ [SI: newton (N)], while work w has units of (mass)(length) ^{2} (time) $^{-2}$ [SI: joule (J)]. Attention to dimensional units of a physical quantity, as well as facility in switching from one set of units to another, will facilitate understanding of thermodynamics and its applications in diverse fields of knowledge.

As illustrative examples of the work concept, let us develop expressions for two well-known types of work, using the general definition (2.60):

1. *Weight-lifting.* In the Earth's gravitational field (with gravitational force constant $g \simeq 9.8 \text{ m s}^{-2}$), the gravitational force on mass m is $F_{\text{grav}} = mg$, directed vertically. The associated gravitational work w_{grav} in lifting mass m from h_1 to h_2 is therefore

$$w_{\text{grav}} = \int_{h_1}^{h_2} F_{\text{grav}} dh = \int_{h_1}^{h_2} mg dh = mg(h_2 - h_1) \quad (2.62)$$

2. *Extending a spring.* For a simple “Hooke’s law” spring, the force of extending the spring increases in proportion to extension distance r (with proportionality “force constant” κ), $F_{\text{spring}} = \kappa r$. In this case,

$$dw_{\text{spring}} = F_{\text{spring}} dr = \kappa r dr \quad (2.63a)$$

and the integrated work of stretching from 0 to r is

$$w_{\text{spring}} = \int_0^r \kappa r dr = \kappa \left(\frac{r^2}{2} \right) \quad (2.63b)$$

In the examples (2.62), (2.63), the force vector $\mathbf{F}$ was parallel to the displacement vector $d\mathbf{r}$ [i.e., $\theta = 0$ in (2.61b)]. In these cases, the vectorial aspects of the work expression (2.60) could be ignored, because the integration path from initial to final r has only a 1-dimensional character. More generally, the integration path may vary in 3-dimensional (nonlinear) manner from initial point A (at $\mathbf{r}_A$) to final point B (at $\mathbf{r}_B$), and the value of the integral is not uniquely determined until the exact A $\rightarrow$ B *path* is specified. In this case, the work $w_{A \rightarrow B}$ must be evaluated as a *line integral* (or *path integral*; cf. Sidebar 1.6),

$$w_{A \rightarrow B} = \int_{\substack{A \\ \text{path}}}^B \mathbf{F} \cdot d\mathbf{r} \quad (2.64)$$

whose value depends on details of the chosen path (not just the endpoints A, B). The path dependence can be averted by restricting $\mathbf{F}$ to be a “conservative force,” for which (by definition) the integral in (2.64) depends on the endpoints only. This restriction is ordinarily assumed in elementary classical dynamics, but it excludes the frictional forces that are of primary interest in a theory of heat effects. Thus, the Newtonian dynamics of conservative

forces provides only an important first step toward a comprehensive “thermodynamic” theory of thermal and mechanical phenomena.

2.9 MECHANICAL ENERGY AND THE CONSERVATION PRINCIPLE

What is “energy,” and why is it a “conserved” quantity?

The concept of energy conservation emerged haltingly over more than two centuries. The difficulty of this concept is associated with the malleable quality of energy itself, which appears in a bewildering multiplicity of interconvertible “forms” (mechanical, thermal, electromagnetic, chemical, nuclear, . . .). Each discovery of a new energy form raised challenges to a general formulation of the energy concept and the associated conservation principle.

The concept of “conservation” of a dynamical property was first enunciated by R. Descartes (1644) as a keystone of natural philosophy. According to Descartes, the “overall sum of all products of *quantitas materiae* (m) and velocity in the world is constant,” and the constancy of this quantity (which would now be identified as linear momentum) called attention to its primary importance as a descriptor of dynamical behavior. C. Huygens (1669) subsequently enunciated the conservation principle of “*vis viva*,” defined as the product of mass and velocity squared (now recognizable as proportional to kinetic energy $\frac{1}{2}mv^2$), in force-free elastic collisions. By 1738, D. Bernoulli was employing the more general principle of conservation of “*vis viva plus vis mortua*” (kinetic plus potential energy) in all friction-free collisions. The term “energy” was first applied in this context by T. Young in 1787, but not pervasively established in the scientific literature for another half-century. In modern terms, the principle of conservation of total mechanical energy E is expressed in terms of kinetic energy (“energy of motion”) and potential energy $\mathcal{V}(\mathbf{r})$ (“energy of position”), with force $\mathbf{F} = -d\mathcal{V}/d\mathbf{r}$. Early disputes were often due to unclear or conflicting definitions of force, work, and other energy-related dynamical properties.

The importance of energy lies in its conservation. The units of energy, $(\text{mass})(\text{length})^2(\text{time})^{-2}$, are the same as those of work, yet energy and work are *not* equivalent properties. We say instead that work is only one form of energy, interconvertible with others. Similarly, kinetic and potential energy are distinct but interconvertible contributions to the total conserved quantity. To the question “What is energy?” we reply that “It is the conserved quantity (among quantities of this type).” To the question “Why is energy conserved?” we reply that “Fulfillment of the conservation principle is what identifies and distinguishes energy.” Sidebar 2.13 further suggests the symbiotic relationship between a conservation principle and the quantity conserved. At a deeper level, Emmy Noether’s theorem [E. Noether, *Nachr. d. Königl. Gesellsch. d. Wiss. zu Göttingen, Math.-phys. Klasse* 235–57 (1918)] reveals how conservation principles for energy and linear or angular momentum are associated with general space–time symmetry properties (homogeneity, isotropy) in classical mechanics.

[Nowadays, it is recognized that energy serves as the quantum-mechanical generator of time evolution, insuring its essential invariance in dynamical processes. However, it is also recognized that precise determination of energy is limited by Heisenberg’s uncertainty principle ($\Delta E \Delta t > \hbar$), so that, strictly speaking, energy conservation can only be verified within this inherent quantal imprecision. Such limits on quantal energy conservation are of no practical significance for macroscopic objects.]

SIDEBAR 2.13: PARABLE OF THE BLOCKS (AFTER FEYNMAN)

[The following parable (a story that serves to teach a difficult concept or lesson) is modeled after that first presented by master physicist R. P. Feynman: R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*, Vol. I (Addison-Wesley, Reading MA, 1963), pp. 4–1 ff.]

A young mother (of scientific inclination) notes that her son Junior's favorite playthings are a set of 37 building blocks. Part of each day's task is gathering up the blocks that Junior has strewn about the nursery (perhaps hidden under the bed or behind the closet door). Although the blocks are of many markings and colorations, and sometimes deviously hidden, she recognizes a useful "block-conservation principle"

$$N_{\text{blocks}} = 37 \quad (\text{S2.13-1})$$

that establishes her task is done when all 37 blocks are accounted for. [On one occasion, only 36 blocks were found, but she noticed that a window was left open! Sure enough, the missing block was found in the garden below where Junior had thrown it, so she carefully nailed shut the windows, sealed keyholes, etc., to prevent losses to the surroundings.]

One day, she is surprised to find that only 36 blocks are present. However, she notices that Junior has been playing with a saw from his tool set, and she observes an uncommon quantity of sawdust strewn about! She gathers and weighs the sawdust, and is delighted to find that it matches the weight of a test block. Thus, she modifies the block-conservation principle to read

$$N_{\text{blocks}} + \frac{\Delta(\text{sawdust})}{k_{\text{sawdust}}} = 37 \quad (\text{S2.13-2})$$

where k_{sawdust} (mass of an intact block) is a fixed constant. Thereafter, her daily task includes sweeping up and weighing sawdust to verify that the result indeed satisfies (S2.13-2).

On another day, her confidence in the conservation principle is again shaken when she finds a tally of only 36. However, she notices that the level of soapy dishwater in the nursery sink is slightly higher than usual! With a ruler, she measures the change in water level as she submerges a test block in the sudsy water, and is delighted to find that the change accounts for the missing block. So she again modifies the block-conservation principle to read

$$N_{\text{blocks}} + \frac{\Delta(\text{sawdust})}{k_{\text{sawdust}}} + \frac{\Delta(\text{sinkwater})}{k_{\text{sinkwater}}} = 37 \quad (\text{S2.13-3})$$

where $k_{\text{sinkwater}}$ (change of sinkwater depth per block) is another constant. Thereafter, her daily task includes measuring the sinkwater level with a ruler to verify that the result satisfies (S2.13-3).

On still another day, she is again disheartened to discover that the left-hand side of (S2.13-3) gives only 36. However, she notices that Junior has been playing with matches, and she observes that the room temperature is slightly warmer than usual! So she carefully measures the heat of combustion of a test block and calculates (from the known heat capacity of the nursery contents) the expected room temperature change,

delighted to find that it matches the observed thermometer change. Accordingly, she again modifies the conservation principle to read

$$N_{\text{blocks}} + \frac{\Delta(\text{sawdust})}{k_{\text{sawdust}}} + \frac{\Delta(\text{sinkwater})}{k_{\text{sinkwater}}} + \frac{\Delta(\text{temperature})}{k_{\text{temp}}} = 37 \quad (\text{S2.13-4})$$

and includes temperature measurement in her daily verification protocol.

Now, you must imagine that the first term is absent, and the block-conservation principle reads simply

$$\frac{\Delta(\text{sawdust})}{k_{\text{sawdust}}} + \frac{\Delta(\text{sinkwater})}{k_{\text{sinkwater}}} + \frac{\Delta(\text{temperature})}{k_{\text{temp}}} = 37 \quad (\text{S2.13-5})$$

The mother wishes to demonstrate this wonderful principle to her husband. She weighs the sawdust, measures the sinkwater, and reads the thermometer to evaluate the successive terms in (S2.13-5), then announces proudly that the result is once again “37 blocks!” “But,” asks the husband, “what exactly *are* the ‘blocks’ you are talking about?”

Such is the puzzling nature of the energy concept.

2.10 FUNDAMENTAL DEFINITIONS: SYSTEM, PROPERTY, MACROSCOPIC, STATE

The rigor and power of equilibrium thermodynamics is purchased at the price of precise operational definitions. In this section, we wish to carefully define four of the most important thermodynamic terms: system, property, macroscopic, and state. Although each term has an everyday meaning, it is important to understand the more rigorous and precise aspects of their usage in the thermodynamic context.

2.10.1 System

We begin with the time-honored definition of “the system” that is familiar to all students of thermodynamics:

Definition: The *system* is the part of the universe of special interest, separated from its *surroundings* (everything else) by definite *boundaries*.

The first task of the student of thermodynamics is to clearly recognize the boundary “wall” (real or virtual) that separates and distinguishes the system of interest from everything else in the universe. Possible examples of a thermodynamic system include:

- a quantity of gas in a piston (the boundaries are the walls of the cylinder)
- a rubber band (“virtual” boundaries are the outer surface of the rubber band, separating the rubbery material from the surrounding air or vacuum)
- the nuclear spins in a nuclear magnetic resonance spectrometer cavity (intermingled with “surrounding” electronic material in the field region as well as everything outside the cavity)
- the entire universe (no boundaries and no surroundings)

As schematized in Fig. 2.11, a chosen initial system (a) might later be enlarged to include a work reservoir, a heat reservoir, or a mass reservoir, thus giving a larger system (b).

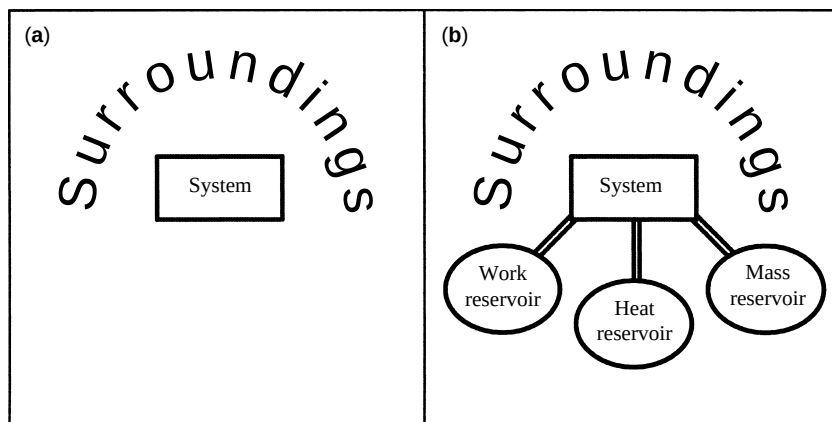


Figure 2.11 Schematic representation of (a) a simple and (b) an enlarged system, separated by boundaries (heavy lines) from surroundings.

Partitions or boundaries (even imagined ones!) play a critical role in virtually every aspect of thermodynamic theory. Because the boundary effectively defines the system, we often classify the system according to physical properties of its boundaries. Simple examples include:

- “*open*” system: able to exchange mass with surroundings
- “*closed*” system: unable to exchange mass (but may exchange heat or work) with surroundings
- “*isolated*” system: not affected by *any* change in surroundings; boundaries do not permit exchange of heat, work, mass (i.e., *any* quantity) with surroundings

We shall later deal with “adiabatic,” “diathermal,” “semipermeable,” and other types of partitions.

2.10.2 Property

The “property” concept may appear so generic as not to require special attention. However, for thermodynamic purposes, the considered set of properties differs significantly from that in other areas of science. Allowed thermodynamic properties are closely tied to specific experimental circumstances of the chosen system, including its quiescence and stability, and a number of the variables that are commonly assumed in a Newtonian mechanical description (such as position and momentum) play no thermodynamic role. The following definition may be adopted:

Definition: For thermodynamic purposes, a *property* of the system is the reproducible result^b of a definite measurement^a on the system, verifiable by an independent observer^c and capable of change.^d

Let us now comment on specific terms or phrases in this definition and the implied restrictions on what are considered valid properties from the thermodynamic viewpoint:

- (a) “definite measurement”: An *operational* definition of the specific experimental protocol is implicitly required to define the property. If there is no imaginable

experimental apparatus that could yield the numerical value of the property, then there is no “property” in the thermodynamic sense.

- (b) “reproducible result”: Successive measurements (repeated at leisure) must yield the *same* numerical value, within the inherent experimental uncertainty of the instrument. Therefore, “transients” and “fluctuations” are not considered properties for thermodynamic purposes, because the system is not yet in the quiescent (equilibrium) condition required for reproducibility. Furthermore, “miracles” or other one-time phenomena are not considered properties for thermodynamic purposes.
- (c) “verifiable by an independent observer”: Repetition and verification of the measurement must be possible both for the original investigator as well as for any independent observer. Here, “independent” means that the alternative observer
 - (i) has no knowledge of the preparation of the system
 - (ii) makes no special reference to the surroundings of the system
 Restriction (i) implies that “history” of the system is not a relevant thermodynamic property. Restriction (ii) implies that “position” or “orientation” of the system are not considered thermodynamic properties, because different observers must be free to select their own preferred laboratory coordinate systems. (Note that omission of position $\mathbf{r}$ as a relevant “property” strongly distinguishes thermodynamics from classical dynamics, where spatial location $\mathbf{r}$ of the center of mass is a prominent variable of the system.)
- (d) “capable of change”: A proposed property must be *capable* of detectable variations under the experimental conditions and time scale. Some properties are therefore *ignorable* in the context of a particular thermodynamic investigation, such as
 - optical rotation of the sample (if no racemization reactions are active)
 - magnetization (if the magnetic field is constant and uniform)
 - radioactive decay rate (if negligible compared with experimental uncertainties)
 - viscoelastic properties (if stress–strain relaxation rates are negligible)

Remarks (a)–(d) sharply restrict the number of imagined “properties” that are relevant to thermodynamic description (Sidebar 2.14).

SIDEBAR 2.14: FLUID PROPERTIES?

Problem The wine steward brings a sealed bottle containing the system of interest. He extols the many “properties” of the fluid that justify its lofty price, including

- (a) pressure P
- (b) volume V
- (c) temperature T
- (d) chemical formula/composition
- (e) restaurant location
- (f) vintage
- (g) “character,” “impudence”
- (h) energy U of the bulk fluid
- (i) energy of an individual molecule in the fluid
- (j) electric polarization (dipole moment) of the bulk fluid

Decide which of these attributes is (Y) or is not (N) a “property” in the thermodynamic sense.

Solution

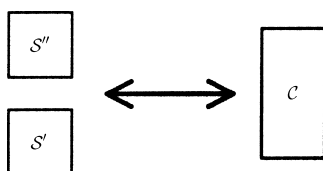
- (a) Y
- (b) Y
- (c) Y
- (d) N (unless chemical reactions are active, e.g., due to cork leak)
- (e) N (position/orientation unimportant)
- (f) N (“history”)
- (g) N (impudence-meter?)
- (h) Y
- (i) N (fluctuation; molecular collisions rapidly redistribute energy)
- (j) N (ignorable, unless variable electric field is present)

2.10.3 Macroscopic

It is generally recognized that thermodynamics applies to bulk macroscopic (“big”) systems, but not to microscopic (“little”) systems. Precisely what is “big” and what is “little” in this context? Can we sensibly apply thermodynamics to 1 cm^3 or 1 mm^3 or $1\text{ }\mu\text{m}^3$ of water? Could a flea do thermodynamics?

It is simplistic to suppose that only an “infinite system” warrants macroscopic designation. We suspect that certain quite small systems (e.g., 1 mm^3 of He gas) might be satisfactorily macroscopic for thermodynamic purposes. Let us frame the definition of “macroscopic” in more precise and operational terms that allow a realistic finite limit to be established.

Definition: Suppose that identical copies^a of a system S are combined to form a new composite system C ,



and that C can be partitioned^b to restore the original systems S' , S'' . We define two special classes of properties:

- Those (R_i ; *intensive*) whose values in any composite C are numerically equal^c to their values in each S ($R_C = R_{S'} = R_{S''}$)
- Those (X_i ; *extensive*) whose values in any composite C are simple *additives* of the value in S , according to the number of copies in the composite ($X_C = X_{S'} + X_{S''}$)

If *all* independent properties of S belong to one class or the other^d, then S (and all its composites) may be called *macroscopic*.

The following comments apply to specific terms or phrases in this definition:

- (a) “identical copies”: This implies that S can be reproduced in arbitrary numbers of copies, all having identical properties (i.e., one can make “ensembles”).

- (b) “can be partitioned . . .”: This implies that system S is *stable* with respect to contact with copies of itself, so that partitions can be added and removed without change. (This restriction would exclude, for example, a near-critical mass of ^{235}U .)
- (c) “numerically equal”: Such “equality” is always understood to mean “within the inherent experimental uncertainty.”
- (d) “one class or the other”: Note that the definition provides an *operational* criterion for classifying any property as “intensive,” “extensive,” or “other” class.

The distinction between intensive (R_i) and extensive (X_i) (vs. “other”) properties is quite important. Qualitatively, one can say that intensive properties R_i are *uniform everywhere in the system*, whereas extensive properties X_i are *additive in subsystems* (“scale with the size of the system”).

According to this definition, the failure to qualify as macroscopic can be traced to one or more properties that *lack* intensive or extensive character. This failure can usually be attributed to surface effects, i.e., to properties that still depend on the *shape* or *size* of the system in a nonextensive manner. We know that, for sufficiently large systems, the surface effects (which scale as r^2) eventually become negligible compared with the bulk volume effects (which scale as r^3). (Although the preceding remarks are specific to ordinary 3-dimensional systems, similar remarks apply to “edge effects” in 2-dimensional systems or “end effects” in 1-dimensional systems.) Empirically, we know that macroscopic character, once established, persists for all larger sizes. Hence, the operational definition above allows us to determine the *smallest* sample of S that still qualifies as “macroscopic” in a particular experimental framework.

It should be emphasized that the criterion for macroscopic character is based on *independent* properties only. (The importance of properly enumerating the number of independent intensive properties will become apparent in the discussion of the Gibbs phase rule, Section 5.1). For example, from two independent extensive variables such as mass m and volume V , one can obviously form the ratio m/V (density ρ), which is *neither* extensive nor intensive, nor independent of m and V . (That density cannot fulfill the “uniform value throughout” criterion for intensive character will be apparent from consideration of any 2-phase system, where ρ certainly varies from one phase region to another.) Of course, for many thermodynamic purposes, we are free to choose a *different* set of independent properties (perhaps including, for example, ρ or other ratio-type properties), rather than the base set of intensive and extensive properties that are used to assess macroscopic character. But considerable conceptual and formal simplifications result from choosing properties of pure intensive (R_i) or extensive (X_i) character as independent arguments of thermodynamic state functions, and it is important to realize that this “pure” choice is always possible if (and only if) the system is macroscopic.

2.10.4 State

With the preceding three definitions established, we can now address the important concept of “state” (or “thermodynamic state,” or “thermodynamic equilibrium state”), the central concept of equilibrium thermodynamics:

Definition: Two macroscopic systems having all the same numerical values of the independent intensive properties are said to be in the same *state* (regarded as identical for thermodynamic purposes).

The thermodynamic state is therefore considered equivalent to specification of the *complete set of independent intensive properties* $\{R_1, R_2, \dots, R_n\}$. The fact that “state” can be specified without reference to extensive properties is a direct consequence of the macroscopic character of the thermodynamic system, for once this character is established, we can safely assume that *system size does not matter* except as a trivial overall scale factor. For example, it is of no thermodynamic consequence whether we choose a cup-full or a bucket-full as “sample size” for a thermodynamic investigation of the normal boiling-point state of water, because thermodynamic properties of the two systems are trivially related.

2.11 THE NATURE OF THE EQUILIBRIUM LIMIT

Let us first summarize the principal features of the thermodynamic equilibrium states that are the principal focus of a thermodynamic description. According to the definitions of Section 2.10, such states are:

- reproducible (i.e., they can be arbitrarily copied to make ensembles)
- time-independent (on the relevant time-scale of the experiments)
- stable with respect to contact with copies of themselves (i.e., they can be partitioned in arbitrary ways and rejoined without change)
- independent of the mode of preparation or condition of surroundings
- independent of size or extent (because macroscopic character renders this aspect of the system superfluous)
- equivalent to complete specification of the independent intensive properties $\{R_1, R_2, \dots, R_n\}$

The number n of independent intensities $\{R_i\}$ remains to be clearly specified (Section 5.1). However, we can say (e.g., on the basis of IL-1, Table 2.1) that this number is remarkably *small*—for example, $n = 2$ for a pure substance!

The small number of variables needed for thermodynamic state description is certainly surprising from a microscopic molecular dynamic viewpoint. For the complete molecular-level description of an arbitrary “state” (phase-space configuration) of the order of 10^{23} particles, we should expect to require an enormously complex nonequilibrium function $\tilde{S}_{\text{non-eq}}$ with a long list (of the order of Avogadro’s number) of independent variables (i.e., positions r_i and velocities $\dot{r}_i$), $\tilde{S}_{\text{non-eq}}(r_1, r_2, \dots, r_{10^{23}}, \dots)$. However, if we allow time evolution until equilibrium is achieved, we find that a vastly simpler description is possible for the resulting equilibrium state $\mathcal{S}_{\text{eq}}$, which depends only on a *few* state properties $R_1, R_2, \dots$, i.e., for a pure substance,

$$\tilde{S}_{\text{non-eq}}(r_1, r_2, \dots, r_{10^{23}}, \dots) \xrightarrow[\text{evolution}]{\text{time}} \mathcal{S}_{\text{eq}}(R_1, R_2) \quad (2.65)$$

The enormous reduction in complexity (2.65) is one of the most remarkable and characteristic features of the equilibrium limit, and it justifies extreme care that the rigorous criteria for this remarkable limit are truly satisfied.

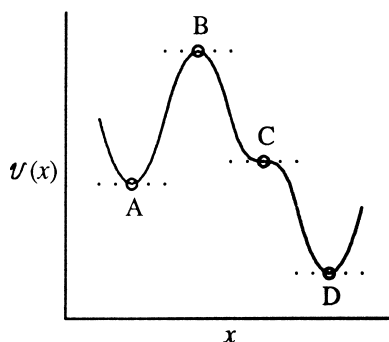
Much of the pre-Gibbsian development of thermodynamics (Chapters 3 and 4) is formulated in terms of nonequilibrium processes, involving the left-hand side of (2.65). This pre-Gibbsian terrain must be traversed in order to understand the historical roots,

terminology, and experimental basis of modern chemical thermodynamics. However, the Gibbs formulation (Chapters 5–8) is firmly set in the vastly simpler domain of equilibrium states, as represented by the right-hand side of (2.65). It is this Gibbsian domain that we shall find to be the proper mathematical framework for a comprehensive description of chemical and phase equilibria.

Thermodynamic equilibrium states also exhibit certain formal analogies to mechanical equilibrium states (Sidebar 2.15). The thermodynamic “potentials” that underlie these analogies will be discussed in Chapter 5.

SIDEBAR 2.15: MECHANICAL EQUILIBRIUM STATES

A mechanical system with potential energy function $\mathcal{V}(x)$ is said to be “in equilibrium” at point x if the forces acting at that point are zero, i.e., $d\mathcal{V}/dx = 0$. This condition is satisfied at the *stationary points* of the potential, such as points A, B, C, D in the potential sketched below:



These stationary points are further classified as to *stability* of the equilibrium by examining the second derivative:

Point(s)	Classification	2nd Derivative
A, D	Stable equilibrium	$d^2\mathcal{V}/dx^2 > 0$
B	Unstable equilibrium	$d^2\mathcal{V}/dx^2 < 0$
C	Critical point	$d^2\mathcal{V}/dx^2 = 0$

Of the two stable equilibrium points, D may be further classified as “absolutely stable” (deepest minimum) and A as “metastable” (local minimum). The unstable point B is also called a “transition state.” Critical point C is a “marginally stable state” that may be further classified as stable or unstable according to the sign of the lowest nonvanishing derivative.

Thermodynamic equilibrium states correspond only to the stable or marginally stable states ($d^2\mathcal{V}/dx^2 \geq 0$) of the mechanical analog. The first law of thermodynamics establishes the thermodynamic potential, while the second law of thermodynamics establishes the stability condition, as discussed in Chapter 5.

General Energy Concept and the First Law

3.1 HISTORICAL BACKGROUND OF THE FIRST LAW

Credit for discovering the first law of thermodynamics is attributable to several individuals. We shall single out four key contributors for special mention: Rumford, Mayer, Joule, and Helmholtz. A brief account of their contributions allows us to better appreciate the difficult concepts embodied in the first law (interconvertibility of heat and work, and conservation of their total). It also illustrates the vital mix of inspiration, experimental verification, and rigorous theoretical formulation that may be required to bring ideas from the domain of creative speculation to well-established scientific law.

A first clear statement of the interconvertibility of work and heat, together with a qualitative numerical estimate of the “mechanical equivalent of heat” was given in 1798 by Count Rumford (born Benjamin Thompson; see Sidebar 3.1). At the time, Rumford was supervising the boring of cannons in the military arsenal at Munich (now in Germany, then part of the Holy Roman Empire). Horses were employed to perform the heavy work of turning the drill stem to ream out the cannon bore, and soldiers were periodically summoned to pour buckets of water over the cannon to quench the intense frictional heat produced by the drilling process. From close observation of this process, Rumford was able to estimate the “heat produced by a horse working for an hour”; in modern terms, his estimate equates to a value of about 5.5 J cal^{-1} (not far from the accepted conversion factor 4.184 J cal^{-1} , which relates the joule and calorie in modern SI terms).

Rumford’s studies (along with those of Humphrey Davy; see Section 3.4) contributed to gradual decline of the caloric theory of heat and its replacement by the modern kinetic molecular theory. By about 1840, the interconversion of heat and work was clearly understood, as well as the association of heat with molecular motion. However, there was as yet no clear statement of the conservation principle for the total heat plus work.

Credit for the first recognizable statement of the principle of conservation of energy (heat plus work) apparently belongs to J. Robert Mayer (Sidebar 3.2), who published such a statement in 1842. Mayer also obtained a (slightly) improved estimate, approximately 3.56 J cal^{-1} , for the mechanical equivalent of heat. Mayer had actually submitted his first paper on the energy-conservation principle two years earlier, but his treatment of the concepts of force, momentum, work, and energy was so confused that the paper was rejected. By 1842, Mayer had sufficiently straightened out his ideas to win publication,

and his claim to discovery of the first law of thermodynamics rests on this work. However, the entire subject was still hedged with considerable uncertainty (as witness Mayer's rather indifferent value of J/q). Although a recognizable statement of the first law was claimed by Mayer, there remained ample room to doubt the quantitative accuracy or generality of this claim.

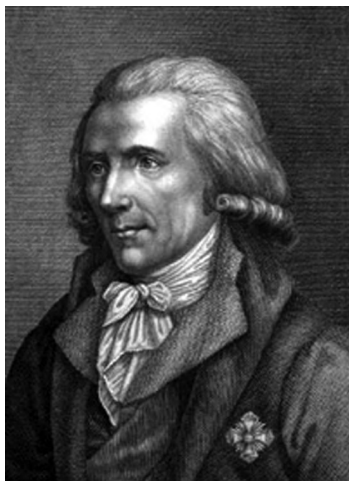
That doubt was emphatically removed by the precise experimental work of James P. Joule (Sidebar 3.3). In the years 1840–49, Joule carried out a classic series of studies of the interconversion of work and heat in multiple forms, replacing vague speculations with a firm experimental basis for the first law. Specifically, Joule studied the heating effects of

- electric current (“Joule heat” $q = I^2R$ of current I passing through resistance R in unit time)
- rotation of paddle wheels in liquids (studied in both H_2O and Hg to demonstrate the independence from liquid identity)
- compression of gases
- forcing liquids through fine tubes

Joule obtained a much improved estimate of the “Joule’s equivalent” of heat ($J/q = 4.15 \text{ J cal}^{-1}$, within 1% of the modern value) and demonstrated its quantitative consistency for all these effects. Thus, Joule’s name is rightly attached to the SI unit of energy, and he deserves to be considered the scientist most responsible for quantitative establishment of the first law of thermodynamics.

On the theoretical side, it was Hermann von Helmholtz (Sidebar 3.4) who first presented a clear and comprehensive mathematical formulation of energy conservation as a principle of universal validity, applicable to all natural phenomena. Helmholtz’s landmark paper of 1847, “Über die Erhaltung der Kraft,” reflected some lingering ambiguities of the force (*Kraft*) concept, but exhibited the deep integration of the first law into analytical dynamics in a clear and modern way. Helmholtz deserves to be counted the scientist most responsible for rigorous mathematical formulation of the first law.

SIDEBAR 3.1: COUNT RUMFORD (1753–1814)



“Count Rumford” was born as Benjamin Thompson in 1753 in Woburn, Massachusetts. Although American by birth, his loyalties were wholly with the British Tories, for whom he served as a spy in the prerevolutionary turmoil of New England. He had risen from modest circumstances as a young man by marrying a rich widow (a strategy he was to repeat much later when he married Lavoisier’s widow after the famous chemist was guillotined during the French Revolution). When the Yankees prevailed in the revolution of 1776, Thompson was forced to flee as a traitor, first to Canada, then to England, where he became a knight in the court of King George III.

Rumford’s fortunes eventually led him to Bavaria, where he rose to prominence as a Count

of the Holy Roman Empire. He was said to be the second most powerful political figure in all Bavaria, second only to the Elector. At one time, he served simultaneously as minister of war, minister of police, major general of the army, Chamberlain of the Bavarian court, and State Councillor to the Elector (and all the while a British spy!).

Rumford's inventiveness was legendary. In addition to his scientific inventions (including the combustion calorimeter and comparative photometer), he is credited with inventing (among other things) the modern cookstove or kitchen range, the pressure cooker, the drip percolator coffee maker, central steam heating, the baking oven, thermal underwear, and (for bureaucrats) the form-in-triplicate. In his duties of maintaining the army, Rumford carried out pioneering nutritional studies (on the minimum rations needed to sustain a soldier), as well as the aforementioned studies of work and heat leading to the first law of thermodynamics. Rumford was also “der Englischer” who designed the beautiful English Gardens (Englischer Garten) of Munich. Credited as founder of the Royal Institution of London, Rumford maintained close ties to English science and society throughout his life.

Rumford's approach to science was pragmatic rather than academic. He once observed that “a habit of keeping the eyes open to every thing that is going on in the ordinary course of the business of life has oftener led to . . . sensible schemes for investigation and improvement than all the more intense meditations of philosophers in the hours expressly set aside for study.”

Although Rumford achieved considerable eminence and power, he did not inspire affection. A biographer observed that he was “utterly without scruples or principles, . . . caustic and treacherous to his peers, and tyrannical to his subordinates.”

[Sanborn C. Brown. *Count Rumford, Physicist Extraordinary* (Doubleday, Garden City, NY, 1962).]

SIDEBAR 3.2: J. R. MAYER (1814–78)



Julius Robert Mayer was born in 1814 in Heilbronn, Germany. Though originally hoping to become a scientist, Mayer was a mediocre student, and he eventually became a physician in 1840. Upon completing medical studies at Tübingen, Munich, and Paris, he followed the European tradition of embarking on an adventure trip prior to beginning his medical career. For this purpose, he signed on as a ship's doctor on a Dutch vessel bound for Java in the tropics. One day, a sailor was badly injured, and Mayer was called into the ship's operating room. There he was startled to see the bright red color of the blood spurting from the sailor's wounds, quite unlike anything he had seen in Germany. After expressing his surprise, he was informed that

this was a well-known phenomenon in the tropics, a consequence of the reduced blood oxygen required to maintain body temperature in equatorial regions compared with the colder northern climes.

Mayer became intrigued with this phenomenon, realizing that the reduced demand for body heat might allow a greater amount of physical work to be performed for given food intake. He reasoned that a fixed amount of food (energy) was used both to preserve body heat and to perform physical work, and that the heat and work output might appear in different proportions according to circumstances, but with constant overall total. He became obsessed with the idea of such constancy and began working feverishly on his theory. As described in the text, his first paper announcing the conservation of heat plus work was rejected, but he performed additional experiments to better estimate the J/q equivalence factor and solicited help in improving the mathematical presentation of his paper, which was finally accepted and published in *Liebig's Annalen der Chemie* in 1842.

Mayer resumed his medical career in Heilbronn, but was much involved in controversies over the priority for his scientific claims, which had attracted scant attention. Distressed in addition by matters of domestic grief, Mayer attempted suicide and was placed in an asylum in 1851. Although he was later released, Mayer's mind never fully recovered, and his influence on subsequent events in the history of thermodynamics was minimal.

SIDEBAR 3.3: J. P. JOULE (1818–99)



James Prescott Joule (pronounced “Jool,” not “Jowl”) was born in 1814 in Salford (near Manchester), England, the son of a wealthy brewer. He was a gifted student, home-schooled for a time, and then personally tutored by noted chemist and atomic-theory propounder John Dalton. However, when his father fell ill, Joule was compelled to abandon university studies to take over the family business affairs. From the proceeds of the brewery, he was able to construct an adjacent laboratory, said to be the finest in all England. There he continued his private investigations, self-financing his research on the heating effects of electricity and other frictional work forms while maintaining close contacts with leading academic scientists. He became a lifelong friend of William Thomson (Lord Kelvin) of Glasgow University, with whom he collaborated in studying the “Joule–Thomson

effect” of gaseous expansion and cooling under adiabatic conditions. He was a devout Quaker, shy and temperate in manner, and he suffered throughout his life from a spinal disorder, although this never diminished his passionate pursuit of scientific truth. After his wife's death in 1854, leaving their two young children, Joule participated in selling the family brewing business, then continued to largely finance his own researches until his funds ran out about 1875, after which he became a Civil List pensioner until his death in 1899.

SIDEBAR 3.4: H. VON HELMHOLTZ (1821–94)



Hermann Ludwig Ferdinand von Helmholtz was born in 1821 in Potsdam, Germany. Helmholtz was one of the towering figures of 19th-century European science, remarkable in the breadth of his contributions. Although he originally hoped to study physics, his father persuaded him to switch to medical studies, for which government scholarships were available. He became a practicing physician and later a professor of physiology at Königsberg, Bonn, and Heidelberg. But his interests in physics, mathematics, and philosophy consumed his spare time, and his work in these areas culminated in appointment to the professorship of physics at Berlin in 1871. His outstanding contributions to science and mathematics included fundamental work in electrodynamics and the theory of electricity (including the Helmholtz coil), meteorology, hydrodynamics, non-Euclidean geometry, and abstract principles of dynamics (such as those

that led to his mathematical formulation of the first law of thermodynamics). His contributions to anatomy and physiology included pioneering studies of the optics of human vision, physiological acoustics and the mechanism of the inner ear, and the theory of tone perception.

3.2 REVERSIBLE AND IRREVERSIBLE WORK

The distinction between “reversible” and “irreversible” work is one of the most important in thermodynamics. We shall first illustrate this distinction by means of a specific numerical example, in which a specified system undergoes a certain change of state by three distinct *paths* approaching the idealized reversible limit. Later, we introduce a formal definition for reversible work that summarizes and generalizes what has been learned from the path dependence in the three cases. In each case, we shall evaluate the integrated work $w_{1\rightarrow 2}$ from the basic path integral,

$$w_{1\rightarrow 2} = \int_1^2 dw = \int_1^2 \mathbf{F} \cdot d\mathbf{r} \quad (3.1)$$

passing from initial state “1” to final state “2” by a specified path. We shall show that

- work depends on the chosen path $1 \rightarrow 2$
- *maximum* useful work is performed as the path approaches the idealized reversible limit

The system chosen for this example consists of a spring and attached platform (Fig. 3.1). The surroundings consist of an assortment of weights (totaling 1000 g) and a vertical

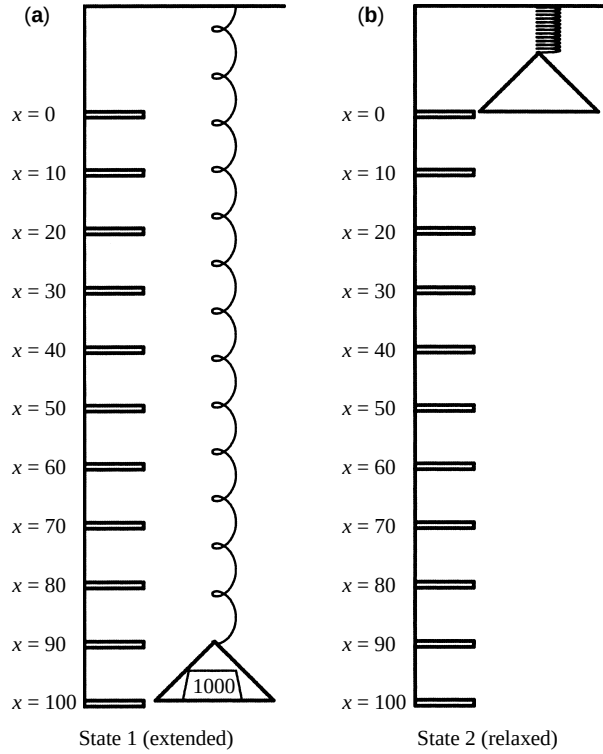


Figure 3.1 (a) Initial extended state 1 and (b) final relaxed state 2 of the spring + platform system discussed in the text.

array of 10 shelves at 10 cm intervals alongside the suspended spring. The initial state “1” is the fully extended spring, with the platform held at $x = 100$ cm extension (Fig. 3.1a) by a 1000 g weight. The final state “2” is the fully relaxed spring, with the platform at $x = 0$ (Fig. 3.1b). The spring is assumed to satisfy Hooke’s law,

$$F_{\text{spring}} = -\kappa x \quad (3.2a)$$

with force constant κ chosen as

$$\kappa = 10g \quad (3.2b)$$

where $g = 980 \text{ cm s}^{-2}$ is the acceleration due to gravity (note the subtle typographical distinction between g and g). With this convenient choice, each 10 g weight added to the platform will cause the spring to extend 10 cm. As usual, we shall assume a massless spring and platform as well as frictionless weights and platforms to simplify the calculations.

We now consider the three distinct paths (a)–(c) depicted schematically in Fig. 3.2:

- (a) In the first path (Fig. 3.2a), the 1000 g weight is dismounted from the platform onto the shelf at $x = 100$ cm. The spring recoils violently and, after furious rattling and reverberations, comes to rest at relaxed-state equilibrium position 2 at $x = 0$.

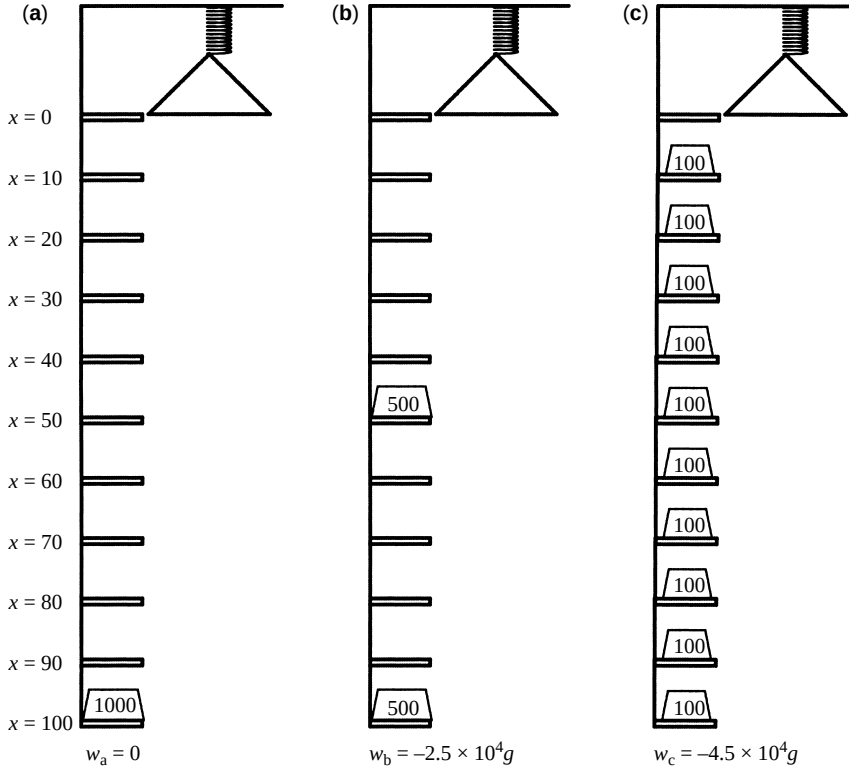


Figure 3.2 Final state of system and surroundings for the three paths (a)–(c) discussed in the text.

(b) In the second path (Fig. 3.2b), a 500 g weight is dismounted onto the shelf at $x = 100$ cm. The spring recoils (although not so furiously as before) until it stabilizes at its new equilibrium position $x = 50$ cm. Then the remaining 500 g weight is dismounted onto the adjacent shelf at $x = 50$ cm, and again the spring recoils to final equilibrium state 2 at $x = 0$.

(c) In the third path (Fig. 3.2c), a 100 g weight is dismounted at $x = 100$ cm and the spring recoils (now rather quietly) to new equilibrium position $x = 90$ cm. Then another 100 g weight is dismounted at $x = 90$ cm, another at $x = 80$ cm, and so forth, each recoil carrying the platform to the next higher shelf to dismount another 100 g weight. Finally, the last 100 g weight is dismounted at $x = 10$ cm, and the spring recoils to the final equilibrium state 2 at $x = 0$.

Let us now assess the work performed in each path. In path (a), no weights were lifted, and therefore no work was performed:

$$w_a = 0 \quad (3.3a)$$

In path (b), a single 500 g weight was lifted by 50 cm, so the gravitational work performed by the system was $mg\Delta x = (500 \text{ g})g(50 \text{ cm})$, or

$$w_b = -2.5 \times 10^4 g \quad (3.3b)$$

in multiples of the gravitational force constant g . Finally, in path (c), one 100 g weight was raised by 10 cm, a second 100 g weight was raised by 20 cm, . . . , and finally a ninth 100 g weight was raised by 90 cm. The total useful work performed on the surroundings is therefore $(100\text{ g})g(10 + 20 + \cdots + 90\text{ cm})$, or

$$w_c = -4.5 \times 10^4 g \quad (3.3c)$$

Note that a *minus* sign is attached to each work quantity in (3.3b, c), because work was formally “lost” from the system to the surroundings. This is in accord with the general *acquisitive convention* for heat and work that will be adopted throughout this book. If heat is expelled to the surroundings (exothermic heat), or work is performed on the surroundings (“useful” work), both are considered “lost” from the system to the surroundings and assigned *negative* values. Conversely, heat absorbed from the surroundings, or work performed by the surroundings on the system, are both considered “gained” by the system and assigned *positive* values. (Note that earlier books often used the opposite convention for the sign of work.) Because the acquisitive convention leads us, paradoxically, to assign the *minimum* (most negative) numerical value to the *maximum useful* work quantity, we shall often intend the colloquial phrase “maximum useful work” to refer to maximum $|w|$ (absolute value), with “useful” implying a negative sign when necessary.

From (3.3a–c), we see that the work performed is indeed path-dependent, and the useful work increases in the order $|w_a| < |w_b| < |w_c|$. Each path in Fig. 3.2 is said to be *irreversible*, because one cannot restore the system and surroundings to their original state 1 (Fig. 3.1a) without expenditure of additional work [most for path (a), least for path (c)]. The increased work output and reduced irreversibility in path (c) is clearly associated with the greater number of points at which the gravitational and spring forces were in perfect balance (i.e., at each shelf level). We can envision a *limiting* “reversible” path that uses an unlimited number of closely spaced shelves and arbitrarily small weights, so that only infinitesimal mass transfers are required to move the spring one way or the other and maintain the balance of gravitational and spring forces at every point. More generally, we define such a path as follows:

Definition: A *reversible path* is a limiting process in which every point of the path is an equilibrium point, so that only an infinitesimal amount of work is sufficient to reverse the path.

For the irreversible paths (3.3a–c), the system was out of equilibrium ($F_{\text{spring}} > F_{\text{gravity}}$) at many points, whereas the limiting reversible path is characterized by perfect force balance

$$F_{\text{spring}} = F_{\text{gravity}} \quad (3.4)$$

at every point. It is clear that the useful work w_{rev} performed in the limiting reversible path must be greater (i.e., more negative) than that of any irreversible path.

We can evaluate the reversible work w_{rev} by substituting the equilibrium condition (3.4) into the general work integral (3.1). Let $m(h)$ be the mass raised to height $h = 100 - x$. The equilibrium condition (3.4) requires that $m(h)$ be chosen such that

$$m(h) = kx/g = (10g)(100 - h)/g = 1000 - 10h \quad (3.5)$$

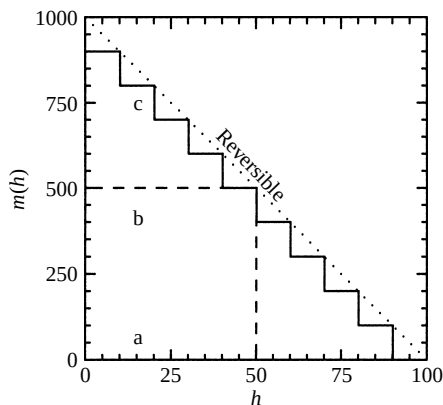


Figure 3.3 Plot of mass elevated to height h , $m(h)$, versus h for path a (baseline), path b (dashed line), path c (solid line), and reversible path (dotted line). The work of each path is proportional to the area under the path.

Noting that $dx = -dh$, we obtain for the reversible work integral

$$w_{\text{rev}} = - \int_{h=0}^{h=100} (1000 - 10h)g \, dh = -5 \times 10^4 g \quad (3.6)$$

As expected, this value exceeds (in absolute value) the useful work obtained from any irreversible path (3.3a–c).

We can compare w_{rev} with w_a , w_b , w_c in more direct graphical form, as shown in Fig. 3.3. This figure shows the plot of $m(h)$ versus h for each path (except the first, which flatlines on the abscissa). The gravitational work is obtained (with proportionality factor g) as the integral $\int m(h) \, dh$, i.e., as the area under the $m(h)$ versus h curve for each path. It is evident that the triangular area under the reversible path exceeds that of any irreversible path within the triangle. The triangular area can be evaluated as

$$\text{area} = \frac{1}{2}(\text{base})(\text{height}) = \frac{1}{2}(100)(1000) = 5 \times 10^4 \quad (3.7)$$

in agreement with (3.6).

We conclude by summarizing in Table 3.1 some key distinctions between reversible and irreversible processes, taking as an example the expansion of a gas against a piston, with external pressure P_{ext} :



Note that an irreversible process may also be called a “real” or “finite-time” or “dissipative” process, because all real processes (i.e., those that occur in finite time) are to some extent irreversible and dissipative in nature.

TABLE 3.1 Comparison of Reversible and Irreversible Gas Expansion with Respect to Equilibration, Reversibility, Rate, and Work Capacity

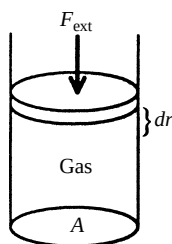
	Reversible Expansion	Irreversible Expansion
Equilibrium	$P_{\text{int}} = P_{\text{ext}}$ at every point; always in equilibrium	Inhomogeneities; turbulence; “sticky” piston; no well-defined P_{int} ; not in equilibrium at every point
Reversibility	Can be retraced without change (restoring surroundings to original)	Cannot be restored to original state without <i>some</i> change in surroundings
Rate	Infinitely slow; idealized limiting case	Real finite-time process
Work	Maximum possible useful work: $ w_{\text{rev}} = w_{\text{max}} $	Dissipates some of the available work as heat, $ w_{\text{irrev}} < w_{\text{max}} $

3.3 GENERAL FORMS OF WORK

Elementary mechanical work forms were considered in Section 2.8. In the present section, we present a broader overview of the varieties of work that are commonly encountered in thermodynamic investigations. The goal is to introduce experimental techniques and operational terminology that underlie the definition and measurement of each work type. We also draw attention to formal patterns among the different forms of work that anticipate their unification with heat in a generalized energy-conservation principle.

3.3.1 Pressure–Volume Work

Let us begin by considering the process of gas compression in a cylinder of cross-sectional area A , with external force F_{ext} driving the piston downward through a differential distance dr , as shown in the following diagram:



Since the force and displacement are parallel, we can write the general differential work expression (2.60) for PV work in the simpler form

$$dw_{PV} = F_{\text{ext}} dr \quad (3.8)$$

By dividing one factor and multiplying the other by the cross-sectional area A , we can rewrite (3.8) as

$$dw_{PV} = \left(\frac{F_{\text{ext}}}{A} \right) (A dr) \quad (3.9)$$

where we recognize that F_{ext}/A , the force per unit area, is the external pressure P_{ext} , while $A dr$ is the (negative) volume change ($-dV$) of the gas. The general PV -work differential is therefore

$$dw_{PV} = -P_{\text{ext}} dV \quad (3.10)$$

Note that the sign in (3.10) conforms to the acquisitive convention; if the gas (system) is compressed ($dV < 0$), work is performed *on* the gas ($dw_{PV} > 0$), whereas for gas expansion ($dV > 0$), the gas performs useful work on the surroundings ($dw_{PV} < 0$).

For a general process $A \rightarrow B$, the PV work is evaluated as the path integral

$$w(A \rightarrow B) = - \int_A^B P_{\text{ext}} dV \quad (3.11)$$

For the special case of a reversible process, where internal $P_{\text{int}} = P_{\text{ext}} = P$ at every point of the path, we obtain the simpler expression

$$w_{\text{rev}}(A \rightarrow B) = - \int_A^B P dV \quad (3.12)$$

where P is the (equilibrium) internal gas pressure.

SIDEBAR 3.5: REVERSIBLE ISOTHERMAL EXPANSION OF AN IDEAL GAS

Let us evaluate the general integral (3.12) for reversible change of an ideal gas from initial state “1” (P_1, V_1, T_1) to final state “2” (P_1, V_1, T_1) under isothermal conditions ($T_1 = T_2$). Recalling that $P = nRT/V$ for an ideal gas, we obtain (with n, R, T all constant)

$$w_{\text{rev}} = - \int_1^2 P dV = - \int_1^2 \frac{nRT}{V} dV = -nRT \int_1^2 d \ln V = -nRT(\ln V_2 - \ln V_1)$$

so that

$$w_{\text{rev}} \stackrel{\text{IG}}{=} -nRT \ln \left(\frac{V_2}{V_1} \right)$$

Note that $w < 0$ if the gas expands ($V_2 > V_1$) to perform useful work on the surroundings, and $w > 0$ if the gas was compressed ($V_1 > V_2$) by the surroundings, in accordance with the acquisitive convention.

The PV -work integral (3.11) can also be displayed graphically in the “indicator diagram” (PV diagram) of Watt. In this diagram (Fig. 3.4), the external pressure P_{ext} is plotted as a function of volume V , and the hatched *area* under this curve (between initial V_A and final V_B) is the integrated work w . It is easy to see from such a graph that the area depends on the shape of the $P_{\text{ext}}(V)$ curve, and thus that w is path-dependent. [“Watt’s Indicator” was a mechanical device on early steam engines that displayed the $P(V)$ curve of the engine cycle, allowing the steam engineer to maximize work output by maximizing the area traced out by the indicator needle.]

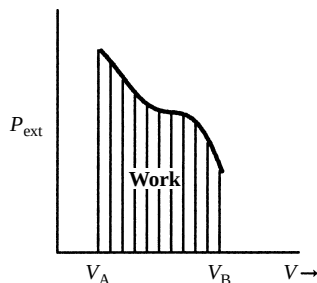


Figure 3.4 Indicator (PV) diagram, showing work performed as the hatched area beneath the curve of P_{ext} versus V (heavy line).

Let us also characterize PV work in more descriptive verbal terms as a *quantity-transfer* process, using gravitational work (2.62) as a model. In the latter case, one can recognize that the essential work process is transfer of a quantity of *mass* (an extensive quantity) through a difference in *gravitational field strength* (an intensive property, having uniform value throughout a laboratory system). (Note that gravitational field strength is usually not a controllable laboratory variable, so gravitational work terms are “ignorable” in the context of thermodynamic investigations of laboratory-scale systems.) In analogous fashion, we can characterize PV work as follows:

Pressure–volume work: The transfer of a quantity of *volume* (extensive) through a difference in external *pressure* (intensive).

This description evokes an image reminiscent of a *waterfall* as a work source (powering, for example, a waterwheel or turbine generator). Just as a quantity of water falls from a higher to a lower gravitational field to produce waterfall work, so does a quantity of volume “fall” from higher to lower pressure to produce PV work. Such parallel images can also help to suggest the formal mathematical parallels that relate various work forms, as we shall demonstrate in Sections 3.3.2–3.3.7.

Because of its pervasive involvement in energy changes in terrestrial laboratory conditions, PV work is often the primary (or only) work term included in a thermodynamic description. Any volume change in a terrestrial system requires displacement (lifting) of the atmospheric air column pressing down on the laboratory object at approximately constant pressure $P_{\text{ext}} \simeq 1 \text{ atm}$. PV work is particularly important in any system involving gases, especially chemical reactions that consume or produce gaseous species. Even for condensed liquid and solid materials, where thermal volume changes are typically smaller, the importance of PV -work terms increases sharply under the high-pressure conditions of, for example, building foundations, suboceanic environments, and planetary interiors.

3.3.2 Surface Tension Work

A two-dimensional analog of PV work can be recognized in fluid films that exhibit *surface tension* (tendency of the film surface to contract against an opposing spreading force). The surface tension work w_{surf} (of, for example, a soap film) can be measured by a rectangular wire-frame device with moveable edge, as shown in Fig. 3.5.

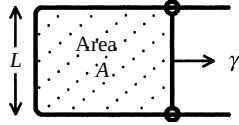


Figure 3.5 Wire-frame device with sliding edge (length L) for measuring the surface tension work w_{surf} of a soap film of area A and surface tension γ .

The film surface tension $\gamma = F_{\text{surf}}/L$ is the force per unit length exerted by the film on the slide-bar. If the slide-bar is extended by distance dr to stretch the film, the work required is

$$dw_{\text{surf}} = F_{\text{surf}} dr = (F_{\text{surf}}/L)(L dr) \quad (3.13)$$

or, since $L dr = dA$,

$$dw_{\text{surf}} = \gamma dA \quad (3.14)$$

This is clearly the two-dimensional analog of PV work. Note that the sign differs in (3.10) and (3.14) because work is performed on the system by reducing the volume in (3.10), but by expanding the area in (3.14).

Surface tension work can also be characterized analogously:

Surface tension work: The transfer of a quantity of *area* (extensive) through a difference in *surface tension* (intensive).

Note that “extensive” is used in the sense of “2D thermodynamics,” where 3D (volume-type) contributions are absent; in 3D thermodynamic systems, surface-type contributions such as (3.14) must be negligible in order that macroscopic 3D extensivity be satisfied (Section 2.10).

3.3.3 Elastic Work

One can also consider the 1-dimensional (1D) analog of PV (3D) or surface tension (2D) work. In this case, consider a rubber band of length L that undergoes differential stretching dL against the “tension” (force) t exerted by the rubber band, as shown in Fig. 3.6. The elastic work performed on the system (the rubber band) is evidently

$$dw_{\text{elastic}} = t dL \quad (3.15)$$

and can be characterized as follows:

Elastic work (1D): A quantity of *length* (extensive) moved through a difference in *tension* (intensive).

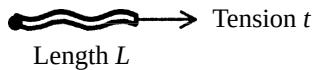


Figure 3.6 Schematic depiction of stretching a rubber band of length L against the tension (force) t to perform elastic work.

The analogous case of an elastic spring satisfying Hooke's law ($t = \kappa r$) was treated previously in (2.63a).

3.3.4 Electrical (emf) Work

The flow of electric charge Q through a difference in voltage ("electromotive force," emf, E) gives rise to electrical emf work w_{emf} , whose differential form can be expressed as

$$dw_{\text{emf}} = E dQ \quad (3.16)$$

Parallel to other work forms, this can be expressed as follows:

Electrical (emf) work: The transfer of a quantity of *electric charge* (extensive) through a difference in *electromotive force* (intensive).

Electrical work is usually performed in an electrical circuit, as shown schematically in Fig. 3.7. The circuit consists of a voltage source (e.g., a battery) E connected through a resistance R with circulating current I . These quantities are related by Ohm's law,

$$I = E/R \quad (3.17)$$

where, in SI units, emf E is in volts (V), resistance R in ohms (Ω), and current I in amperes (A). The current I expresses the rate of charge flow per unit time,

$$I = \frac{dQ}{dt} \quad (3.18)$$

which, by convention, is taken *opposite* to the physical direction of electron flow.

From (3.18), the differential charge dQ (in coulombs, $C = A s^{-1}$) is

$$dQ = I dt \quad (3.19)$$

and from (3.17) the emf E is

$$E = IR \quad (3.20)$$

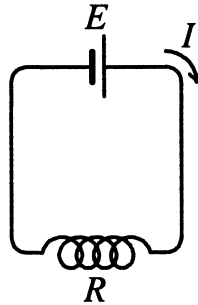


Figure 3.7 Schematic electrical circuit, with voltage source E , electric current I , and resistance R .

The electrical work (3.16) can therefore be expressed in alternative form as

$$dw_{\text{emf}} = E dQ = (IR)(dQ) = I^2 R dt \quad (3.21)$$

This agrees with the expression for “Joule heat” (Section 3.1), which played a prominent role in the quantitative establishment of the first law of thermodynamics.

3.3.5 Electric Polarization Work

Another type of electrical work, called “electric polarization” (or “capacitance”) work w_{pol} , can be carried out in a parallel-plate capacitor (“condenser”), as depicted schematically in Fig. 3.8. As its name implies, the capacitor consists of two large parallel plates, of area A and separation distance δ , enclosing a dielectric medium (the system). In the presence of the dielectric (insulating) medium, opposite charges $\pm Q$ are loaded onto the two plates to create a potential difference E between the plates. (If the medium were a conductive material such as metallic copper, the charges would simply self-neutralize through the medium, allowing no potential difference E to be maintained.) We shall assume that the plates are so large, and the separation so small, that the electric field strength $\mathcal{E}$ is uniform between the plates (no “edge effects”). Under these conditions, the electric field strength (the gradient of electric potential) is defined simply as

$$\mathcal{E} = \frac{E}{\delta} \quad (3.22)$$

where E is the voltage difference between the plates.

Given a particular substance in the capacitor, the quantity measured is the *capacitance* $\mathcal{C}$,

$$\mathcal{C} = \frac{|Q|}{E} \quad (3.23)$$

the ratio of charge supported to potential difference (with units of farads, $\text{F} = \text{C V}^{-1}$). However, the supported charge Q (extensive) is proportional to plate area A and (for fixed potential) inversely proportional to plate separation δ , so that $\mathcal{C}$ carries an incidental

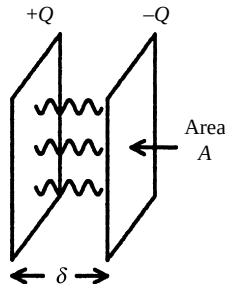


Figure 3.8 Schematic parallel-plate capacitor, showing plate area A , separation δ , and net charge $\pm Q$ on each plate surrounding the dielectric medium.

dependence on the geometrical factors of the capacitor ($C \propto A/\delta$) as well as the important dependence on the physical medium. To separate these factors, we define

$$C = \frac{A}{\delta} \varepsilon \quad (3.24)$$

where ε is the “permittivity” of the medium between the plates. We then compare the measured permittivity (ε) when the medium is present with the corresponding “permittivity of the vacuum” (ε_0) when the capacitor is *empty* (with capacitance C_0). The dimensionless ratio of these quantities is called the “dielectric constant” κ of the medium,

$$\kappa = \frac{\varepsilon}{\varepsilon_0} = \frac{C}{C_0} \quad (3.25)$$

which is the fundamental material property obtained in a capacitance measurement.

Physically, κ expresses the ability of the substance to withstand opposite charges at small distances without dielectric breakdown (neutralization), or, alternatively, the ability to “screen” charges from one another by reducing the Coulombic potential E between charges. From (3.25), one can recognize that $\kappa \geq 1$ for any real substance (i.e., compared with a vacuum, the medium *increases* capacitance by reducing the potential E between fixed charges Q). For gases, $\kappa \simeq 1$ remains close to the vacuum limit ($\kappa = 1$), whereas for liquids and solids, values vary widely over the range $\kappa \simeq 2 - 100$ (e.g., $\kappa \simeq 78$ for H_2O). Higher values of κ require increased work to charge up the condenser to fixed E , or, alternatively, increased drain on batteries to hold the potential difference constant.

Let us now evaluate the work of charging the condenser. For fixed potential E , we envision transfer of infinitesimal charge dQ from the negative to the positive plate, with differential work $dw = E dQ$ as in (3.16). From the definitions (3.22), (3.23), we can write $E = \mathcal{E}\delta$ and $Q = C\mathcal{E}\delta$, so that

$$E dQ = (\mathcal{E}\delta) d[C\mathcal{E}\delta] = \mathcal{E} d[C\mathcal{E}\delta^2] \quad (3.26a)$$

We now use (3.24) to replace $C = (A/\delta)\kappa\varepsilon_0$ in the brackets, recognizing that $A\delta = V$ is the condenser volume,

$$d[C\mathcal{E}\delta^2] = d[(A/\delta)\kappa\varepsilon_0\mathcal{E}\delta^2] = d[\kappa\varepsilon_0\mathcal{E}V] \quad (3.26b)$$

which allows us to rewrite (3.26a) as

$$E dQ = \mathcal{E} d[\kappa\varepsilon_0 V \mathcal{E}] \quad (3.26c)$$

However, for thermodynamic purposes, the quantity dw_{pol} that we seek is the work performed on the *medium* (the system). This is the *difference* between the work performed in charging up the condenser with and without the medium:

$$dw_{\text{pol}} = (E dQ)_{\text{medium}} - (E dQ)_{\text{vacuum}} \quad (3.27)$$

We evaluate this difference by using (3.26c) for both terms (in the vacuum case, with $\kappa = 1$) to obtain

$$dw_{\text{pol}} = \mathcal{E} d[(\kappa - 1)\varepsilon_0 V \mathcal{E}] \quad (3.28)$$

The quantity in brackets in (3.28) is called the “polarization” M ,

$$M = (\kappa - 1)\epsilon_0 V \mathcal{E} \quad (3.29)$$

that is, the total *dipole moment* of the medium. As expected, M is proportional to system volume V (i.e., it is an extensive property) as well as to the strength of the electric field $\mathcal{E}$ that induces the polarization. We therefore obtain the final expression for electrical polarization work in the form

$$dw_{\text{pol}} = \mathcal{E} dM \quad (3.30)$$

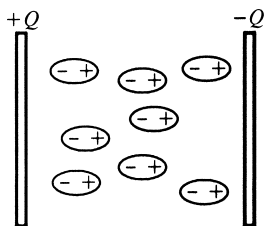
or, in words:

Electrical polarization work: The transfer of a quantity of *dipole moment* (extensive) through a difference in *electric field strength* (intensive).

Sidebar 3.6 describes some molecular aspects of the electrical polarization phenomenon.

SIDEBAR 3.6: MOLECULAR VIEW OF ELECTRIC POLARIZATION

From the molecular viewpoint, the medium acquires macroscopic polarization M because the molecules possess dipole moments μ that preferentially align along the electric field of the charged plates, as depicted schematically below:



Two polarization mechanisms are possible. If the molecules possess a permanent electric dipole moment μ_{perm} , each molecule can align its moment with the field direction by *reorientation*, producing a macroscopic dipole moment. Even if $\mu_{\text{perm}} = 0$ in the field-free limit, each molecule can achieve a field-dependent dipole moment μ_{ind} by *induction*. The induced dipole moment is proportional to field strength, $\mu_{\text{ind}} = \alpha \mathcal{E}$, where α is the electric polarizability of the molecule. In both cases, work must be performed on the system to achieve the macroscopic polarization. Molecules with large permanent dipole moments correspond to high κ .

3.3.6 Magnetic Polarization Work

As might be anticipated from the Maxwellian symmetry of electric and magnetic phenomena, a completely parallel magnetic work form dw_{mag} arises from magnetic polarization in

the presence of a magnetic field, namely,

$$dw_{\text{mag}} = \mathcal{H} d\mathcal{M} \quad (3.31)$$

Here, $\mathcal{H}$ is the magnetic field strength, and $\mathcal{M}$ is the “magnetization” (total magnetic moment) of the substance. In words:

Magnetic work: The transfer of a quantity of *magnetization* (extensive) through a difference in *magnetic field strength* (intensive).

Magnetic work is usually negligible for ordinary (“diamagnetic”) substances with no unpaired spins, but becomes significant for paramagnetic or ferromagnetic substances (e.g., gaseous O_2 , metallic Fe) in the presence of external magnetic fields. In a field-free environment ($\mathcal{H} = 0$) the internal molecular magnetic moments of, for example, gaseous O_2 molecules will tend to be randomly oriented, with no resultant macroscopic magnetization ($\mathcal{M} = 0$). However, in a finite field ($\mathcal{H} \neq 0$), these internal magnetic moments become macroscopically aligned to produce net magnetization ($\mathcal{M} \neq 0$), requiring macroscopic work as given by (3.31).

3.3.7 Overview of General Work Forms

Let us recap what has been learned from the assortment of work forms considered in Sections 3.3.1–3.3.6. In each case, the differential work form dw_i of type i can be expressed in terms of an intensive property R_i and the complementary extensive property X_i in the form

$$dw_i = R_i dX_i \quad (3.32)$$

Table 3.2 summarizes the intensive “field” R_i , extensive “quantity” X_i , and work element $dw_i = R_i dX_i$ for the various work types that have been considered.

The total work dw_{total} is generally a summation of contributing work types,

$$dw_{\text{total}} = -P dV + \gamma dA + \cdots + \mathcal{H} d\mathcal{M} = \sum_{i=1}^{\text{types}} R_i dX_i \quad (3.33)$$

running over all relevant pairs (R_i, X_i) of “conjugate” intensive and extensive properties. This quantity is generally conserved if frictional dissipation is absent, i.e., if “heat

TABLE 3.2 Summary of Work Types, with Associated R_i (Intensive) and X_i (Extensive) Properties and Differential Work Element dw_i

Type i	Intensive R_i	Extensive X_i	$dw_i = R_i dX_i$
Gravitational	Gravitational field strength g	Mass displacement mh	$dw_{\text{grav}} = g d(mh)$
Pressure–volume	Negative pressure $-P$	Volume V	$dw_{PV} = -P dV$
Surface tension	Surface tension γ	Area A	$dw_{\text{surf}} = \gamma dA$
Elastic	Tension t	Length L	$dw_{\text{elastic}} = t dL$
Emf	Electromotive force E	Charge Q	$dw_{\text{emf}} = E dQ$
Polarization	Electric field strength $\mathcal{E}$	Electric dipole moment M	$dw_{\text{pol}} = \mathcal{E} dM$
Magnetization	Magnetic field strength $\mathcal{H}$	Magnetic dipole moment $\mathcal{M}$	$dw_{\text{mag}} = \mathcal{H} d\mathcal{M}$

effects” are missing. Based on the strong mathematical pattern seen in Table 3.1 and (3.33), we might conjecture that the “missing” thermal term needed to restore the energy conservation principle to full generality is also of the form

$$d(\text{thermal}) \stackrel{?}{=} R_{\text{thermal}} dX_{\text{thermal}} \quad (3.34)$$

where “ R_{thermal} ” is the controlling intensive variable (temperature?) and “ X_{thermal} ” is the conjugate extensive variable (entropy?) pertaining to thermal heating effects. However, confirmation of this conjecture must await proper definition of “heat,” “entropy,” and other terms yet to be discussed.

3.4 CHARACTERIZATION AND MEASUREMENT OF HEAT

Before about 1800, heat was widely considered to be a material substance, called “caloric” (listed as such by Lavoisier in his first “Table of Elements”!). Caloric was supposed to be a weightless, invisible fluid that could penetrate (“dissolve into”) any object, but could then be extracted (“squeezed out”) by friction. The fuzzy imagery of heat as a fluid, based on a naive but appealing analogy, presented a serious impediment to development of a rational theory of heat.

Around 1800, experimental challenges to caloric theory were being presented by Count Rumford (cannon boring) and Humphrey Davy (melting of ice by friction). It became apparent that heat could be produced from a body in *unlimited* quantity by friction, further stretching its credibility as a “substance.” By about 1840, caloric theory was overturned by the modern kinetic molecular theory of heat (Sidebar 2.7), which identified heat with the energy of random molecular motions.

Heat q and temperature T are related but distinct concepts. Temperature T can be identified as “degree of hotness” (Section 2.3), whereas heat q is a “quantity of thermal energy.” The same quantity of heat might be stored under different conditions, for example, as “high-temperature heat” or “low-temperature heat” in heat reservoirs of different temperature. Further aspects of how the temperature of a heat quantity affects its usefulness (e.g., for conversion to useful work) will be discussed in Chapter 4.

Useful quantitation of heat q as a quantity of energy can be traced to the studies of Joseph Black around 1803. Black recognized that different substances vary in their capacity to absorb heat, and he undertook systematic measurements of the “heat capacity” C (the ratio of heat absorbed to temperature increase) for many substances. He recognized that a fixed quantity of any pure substance (e.g., 1 g of water) has a unique value of C , which can be chosen as a calorimetric standard for defining “quantity of heat” in a convenient way. In this manner, he introduced the “calorie” as a unit of heat:

Calorie: the quantity of heat required to raise the temperature of 1 g of H_2O by 1°C (now 4.184 J, by definition).

With this definition, the quantity of heat q (in calories) is given by

$$q = c_{\text{water}} \Delta T \quad (3.35)$$

where $c_{\text{water}} = 1 \text{ cal g}^{-1} \text{ K}^{-1}$ is the “specific heat” (on a per-gram basis) of H_2O . [Note that K^{-1} and $^{\circ}\text{C}^{-1}$ are equivalent as dimensional units, because the degree size is identical in K and $^{\circ}\text{C}$ scales. Note also that c_{water} varies slightly with temperature and pressure, so that (3.35) is only a provisional definition.]

Although the definition (3.35) allows practical progress, it rests on other concepts of temperature and thermal capacity that border on circular reasoning. Accordingly, we shall first attempt to formulate an alternative *mechanical definition of heat* that is of no practical significance, but satisfies the thermodynamicist’s penchant for logical order. As in the case of temperature (Section 2.3), we attempt to characterize heat in terms of mechanical variables only (e.g., P , V), which are well defined in a pre-thermodynamic context.

To do so, we first define the concept of “adiabatic” walls (and associated adiabatic systems) that *only* permit PV changes, as illustrated in Fig. 3.9a. A system enclosed in an adiabatic container is only subject to PV -work changes, but not to any other type of perturbation (or, as we should say colloquially, not to “heat effects”). [Experimentally, we can find container walls that adequately satisfy this criterion by sealing a vacuum within the walls (to prevent convection) and silvering the outer surface to high reflectivity (to prevent radiation), as in, for example, a Dewar flask.] We can similarly define “diathermal” walls as *nonadiabatic*, i.e., those permitting thermal equilibration with the surroundings, as illustrated in Fig. 3.9b.

Let us now carry out a change of state ($A \rightarrow B$) on the system with each type of container and measure the mechanical work ($w_{\text{adiabatic}}$, $w_{\text{diathermal}}$) associated with each. From these two measurements, we can define heat $q = q_{A \rightarrow B}$ absorbed in this process as

$$q \equiv w_{\text{adiabatic}} - w_{\text{diathermal}} \quad (3.36)$$

It may not be initially apparent why this definition coincides with (3.35), but the right-hand side of (3.36) clearly depends only on PV measurements, and thus meets the criterion of a purely mechanical definition of heat. [We shall see in the following section that $w_{\text{adiabatic}}$ is identical to the internal energy change, so the equivalence of (3.35) and (3.36) is rather trivial.]

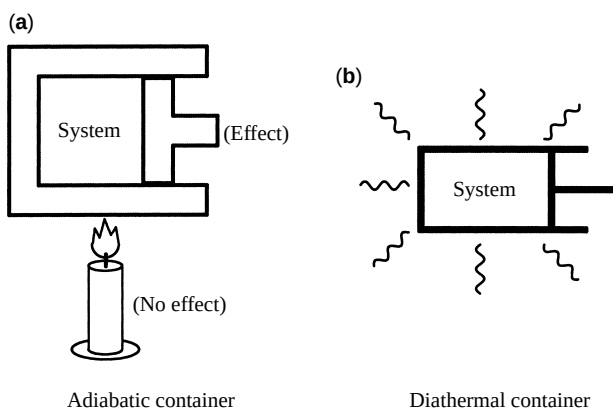


Figure 3.9 Schematic representation of (a) an adiabatic container, allowing PV work by movement of a piston, but unaffected by other changes in the surroundings; (b) a diathermal (nonadiabatic) container, allowing thermal equilibration with the surroundings.

3.5 GENERAL STATEMENTS OF THE FIRST LAW

With definitions of work w and heat q established, we proceed to formal statement(s) of the first law of thermodynamics (cf. IL-5; Table 2.1). Although the first law is sometimes stated colloquially as “Energy is conserved” (or, somewhat more satisfactorily, “Energy is conserved if heat is taken into account”), a proper statement requires the introduction of a new quantity, *internal energy* U , that can be distinguished from “energy” as used in the mechanical framework:

First Law (IL-5): There exists a macroscopic state property U (“internal energy”) whose change, in a process $A \rightarrow B$ involving only the absorption of heat q or performance of work w on the system, is given by

$$\Delta U = U_B - U_A = q + w \quad (3.37a)$$

or, in differential form,

$$dU = dq + dw \quad (3.37b)$$

From (3.37a), we can recognize that in an adiabatic process ($q = 0$),

$$(\Delta U_{A \rightarrow B})_{\text{adiabatic}} = w_{\text{adiabatic}} \quad (3.38a)$$

whereas for the same change of state carried out under diathermal conditions,

$$(\Delta U_{A \rightarrow B})_{\text{diathermal}} = q + w_{\text{diathermal}} \quad (3.38b)$$

However, the change in state property U must be independent of the path from A to B, $(\Delta U_{A \rightarrow B})_{\text{adiabatic}} = (\Delta U_{A \rightarrow B})_{\text{diathermal}}$, so we deduce from (3.38a, b) that $q = w_{\text{adiabatic}} - w_{\text{diathermal}}$, as stated in (3.36).

The essential content of the first law is that $dU = dq + dw$ is the (exact) differential of a *state* property, and hence independent of the path from A to B. The path integral over dU from A to B therefore depends only on the values of internal energy (U_A , U_B) at the two endpoints

$$\int_A^B dU = U_B - U_A \quad (3.39)$$

whereas the corresponding path integrals over dq or dw are path-dependent. The fact that U returns to the same value (i.e., is “conserved”) whenever the system returns to the same state, whereas q and w do not, expresses the deeper significance of U compared with its individual contributions from diverse forms of heat or work.

As stated above, the first law refers only to closed systems that allow no exchange of mass with the surroundings. To obtain an alternative expression of the first law in words, we might therefore consider *the entire universe* (which perforce has no surroundings) as the “system,” which leads to the somewhat melodramatic pronouncement:

First Law: The internal energy of the universe is a constant.

Far more useful are statements that establish the mathematical nature of the internal energy function. Any of the following statements can be considered fully equivalent to the first law, once U is defined in terms of heat and work:

$$U \text{ is a } \textit{state} \text{ property} \quad (3.40a)$$

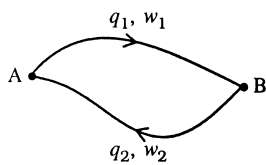
$$U \text{ is a } \textit{conserved} \text{ quantity} \quad (3.40b)$$

$$dU \text{ is an } \textit{exact} \text{ differential} \quad (3.40c)$$

$$\int_A^B dU = U_B - U_A \quad \text{for any path from A to B} \quad (3.40d)$$

$$\oint dU = 0 \quad \text{for any cyclic path} \quad (3.40e)$$

We can also derive certain provocative consequences of the first law that make no direct reference to the internal energy function. Consider, for example, a cyclic process that involves passing from A to B by path “1” (with heat q_1 and work w_1), then returning from B to A by alternative path “2” (with heat q_2 and work w_2):



According to (3.40e), the ΔU changes on these paths must be equal in magnitude and opposite in sign:

$$\Delta U_1 = q_1 + w_1 = -\Delta U_2 = -(q_2 + w_2) \quad (3.41a)$$

which can be rearranged to read

$$q_1 + q_2 = -(w_1 + w_2) \quad (3.41b)$$

The left-hand side of (3.41b) is the net heat absorbed from the surroundings, while the right-hand side is the net work performed on the surroundings. From the mandatory balance of these terms, we can assert:

Consequence of First Law: It is impossible to construct a machine (“perpetual motion machine of the first kind”) that produces useful work in a cyclic process without a compensating change in the surroundings.

Such a statement is called an “impossibility axiom” or “principle of impotence.” Although the history of science contains numerous examples of inventors claiming perpetual motion devices, closer examination has always exposed the “compensating change” (e.g., heat absorption from the surroundings, or a string through the table leg, manipulated by an accomplice beneath the stage) that brings the device into compliance with the first law.

The inductive generality and validity of the first law is therefore upheld by every day that passes without successful challenge to this principle.

3.6 THERMOCHEMICAL CONSEQUENCES OF THE FIRST LAW

The first law of thermodynamics leads to a broad array of physical and chemical consequences. In the following Sections 3.6.1–3.6.8, we describe the formal theory of heat capacity and the enthalpy function, the measurements of heating effects that clarified the energy and enthalpy changes in real and ideal gases under isothermal or adiabatic conditions, and the general first-law principles that underlie the theory and practice of “thermochemistry,” the measurement of heat effects in chemical reactions.

The first law is formulated as a completely rigorous, general, and exact statement of inductive experience. Logical consequences of the first law, if derived with sufficient mathematical care, inherit the same rigor, generality, and exactness as the first law itself. Accordingly, we shall pay special attention to the logic of mathematical derivations, insuring that no logical loopholes intrude to weaken the thermodynamic conclusions. As previously mentioned, it is important in this regard to identify and distinguish relationships based on ideal gas assumptions or other approximations (marked, e.g., by $\simeq$ or $\stackrel{\text{IG}}{=}$ signs) from those that enjoy the full force and authority of the first law.

3.6.1 Heat Capacity and the Enthalpy Function

The provisional characterization (Section 3.4) of heat capacity C as the ratio of heat absorbed to the temperature increase corresponds to a differential ratio of the form (on a per-mole basis, unless otherwise specified)

$$“C = \frac{dq}{dT}” \quad (3.42)$$

However, the “definition” (3.42) leaves heat capacity ill-determined, because the imperfect differential dq harbors a dependence on the (unspecified) path along which dq is measured (“path” referring to how the remaining non- T degree of freedom is specified).

Two common experimental paths for heat measurement are conditions of constant volume (q_V) or constant pressure (q_P). These conditions lead to the corresponding constant-volume (C_V) or constant-pressure (C_P) heat capacities

$$C_V = \frac{dq_V}{dT} \quad (3.43a)$$

$$C_P = \frac{dq_P}{dT} \quad (3.43b)$$

The definitions (3.43), although improving on (3.42), are still somewhat unsatisfactory owing to the presence of the nonstate property q .

To remedy this defect for C_V , we can rewrite the first law (3.37a) for constant-volume conditions as

$$\Delta U = q_V + w_V \quad (3.44)$$

However, it is evident from (3.10) that $w_V = 0$ under the usual conditions where only PV work is allowed. Under these conditions, $q_V = \Delta U$ (or $dq_V = dU$), and (3.43a) becomes

$$C_V = \left(\frac{\partial U}{\partial T} \right)_V \quad (3.45)$$

defined entirely in terms of *state properties*. This definition makes it clear that C_V becomes a well-defined property of a substance when the *state* is well-specified.

Let us now attempt to find a similar expression for $C_P = dq_P/dT$ in terms of state properties only. Under constant- P conditions, the first law (3.37a) becomes

$$\Delta U = q_P + w_P \quad (3.46)$$

If only PV work is allowed, we obtain from (3.10) or (3.12), at constant $P = P_{\text{ext}}$,

$$w_P = -P\Delta V = -P(V_2 - V_1) \quad (3.47)$$

Under these conditions, (3.46) can be rewritten as

$$\Delta U = U_2 - U_1 = q_P - P(V_2 - V_1) \quad (3.48a)$$

or, equivalently,

$$q_P = U_2 - U_1 + P(V_2 - V_1) = (U + PV)_2 - (U + PV)_1 \quad (3.48b)$$

The right-hand side of (3.48b) has the appearance of a difference in a *state function* ($U + PV$) between the two states 1, 2. This state function is identified as the “enthalpy” (symbolized by H), with fundamental definition

$$\boxed{H \equiv U + PV} \quad (3.49)$$

According to (3.48b), enthalpy H can be described as “constant-pressure heat”

$$\Delta H = q_P \quad (3.50)$$

We can therefore rewrite the expression (3.43b) for C_P in final partial differential form in terms of state properties only:

$$C_P \equiv \left(\frac{\partial H}{\partial T} \right)_P \quad (3.51)$$

According to (3.51), C_P is the heat capacity measured under ordinary laboratory conditions (i.e., with open laboratory vessels exposed to the constant pressure of the atmosphere), where the measured heat changes are *enthalpy* changes ΔH . According to (3.45), C_V is the corresponding heat capacity under constant- V conditions (i.e., in a bomb calorimeter), where the measured heat changes are *internal energy* changes ΔU .

What is the difference between C_P and C_V ? We first express this difference in terms of a standard *thermodynamic identity* for 1 mole of a simple gas:

$$C_P - C_V = \left[P + \left(\frac{\partial U}{\partial V} \right)_T \right] \left(\frac{\partial V}{\partial T} \right)_P \quad (3.52)$$

To prove this identity, we first write from the definitions (3.45), (3.51),

$$C_P - C_V = \left(\frac{\partial H}{\partial T} \right)_P - \left(\frac{\partial U}{\partial T} \right)_V \quad (3.53)$$

where, from (3.49), under constant- P conditions,

$$\left(\frac{\partial H}{\partial T} \right)_P = \left(\frac{\partial(U + PV)}{\partial T} \right)_P = \left(\frac{\partial U}{\partial T} \right)_P + P \left(\frac{\partial V}{\partial T} \right)_P \quad (3.54)$$

We now use the mathematical identity (1.13) to change the variable held constant from P to V :

$$\left(\frac{\partial U}{\partial T} \right)_P = \left(\frac{\partial U}{\partial T} \right)_V + \left(\frac{\partial U}{\partial V} \right)_T \left(\frac{\partial V}{\partial T} \right)_P \quad (3.55)$$

On substituting (3.55) into the right-hand side of (3.54), we obtain

$$\left(\frac{\partial H}{\partial T} \right)_P = \left(\frac{\partial U}{\partial T} \right)_V + \left[\left(\frac{\partial U}{\partial V} \right)_T + P \right] \left(\frac{\partial V}{\partial T} \right)_P$$

which, inserted into (3.53), gives (3.52)—*QED*. We shall later prove (Sidebar 5.5)

$$C_P - C_V = TV\alpha_P^2/\beta_T \quad (3.56)$$

which gives a more explicit relationship to the commonly tabulated properties α_P , β_T .

The partial derivative $(\partial V/\partial T)_P$ in (3.52) can be evaluated from the ordinary equation of state, but what is $(\partial U/\partial V)_T$? We next describe two classic experiments that provide useful information about this quantity.

3.6.2 Joule's Experiment

A crude estimate of $(\partial U/\partial V)_T$ was obtained by Joule in 1845 with the experimental apparatus sketched schematically in Fig. 3.10. A known quantity of gas at initial T , P occupies one bulb of a double flask, connected by a stopcock to a vacuum in the other bulb, the whole being submerged in a water bath of known heat capacity c_{bath} . When the stopcock is opened, the gas expands irreversibly into the vacuum, performing no useful work on the surroundings ($w = 0$). Under these work-free conditions, $\Delta U = q = c_{\text{bath}}\Delta T$, according to the first law. Experimentally, the measured temperature change ΔT is negligibly small

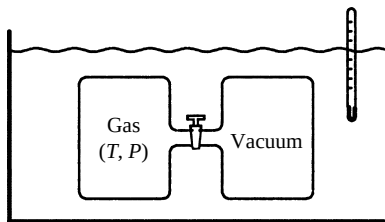


Figure 3.10 Schematic of Joule's experiment for irreversible expansion of a gas into an evacuated chamber in a water bath.

(isothermal conditions). One therefore concludes that $\Delta U \simeq 0$ for the volume expansion $\Delta V \neq 0$ under isothermal conditions, or

$$\left(\frac{\partial U}{\partial V}\right)_T \simeq 0 \quad (3.57)$$

In fact, Joule's experiment is not sufficiently accurate to detect the actual (small) value of $(\partial U/\partial V)_T$ for real gases at near-ambient conditions. However, it will later be proven (Sidebar 5.6) that the vanishing of $(\partial U/\partial V)_T$ becomes exact in the ideal gas limit,

$$\left(\frac{\partial U}{\partial V}\right)_T^{\text{IG}} \equiv 0 \quad (3.58)$$

and may be taken as a defining characteristic of the ideal gas approximation. Equation (3.58) asserts that *only temperature changes can affect the internal energy of an ideal gas*, i.e.,

$$U \stackrel{\text{IG}}{=} U(T) \text{ only} \quad (3.59)$$

Because $H = U + PV$, where $PV = nRT$ depends only on T for an ideal gas, the enthalpy also depends only on T :

$$H \stackrel{\text{IG}}{=} H(T) \text{ only} \quad (3.60)$$

The monovariate dependences in (3.59), (3.60) warn of the nonphysical character of the ideal gas approximation.

SIDEBAR 3.7: IDEAL GAS HEAT CAPACITIES

Problem Show that $C_P - C_V = nR$ for an ideal gas.

Solution From the general identity (3.52) and the ideal gas relationship (3.58), we obtain

$$C_P - C_V = \left[P + \left(\frac{\partial U}{\partial V}\right)_T \right] \left(\frac{\partial V}{\partial T}\right)_P = P \left(\frac{\partial V}{\partial T}\right)_P$$

For the ideal gas, with $V = nRT/P$, we obtain $(\partial V/\partial T)_P = nR/P$, so that, finally,

$$C_P - C_V \stackrel{\text{IG}}{=} nR$$

3.6.3 Joule–Thomson Porous Plug Experiment

In the period 1852–62, J. P. Joule and W. Thomson (later Lord Kelvin) perfected a clever method for measuring the *isenthalpic* property $(\partial T/\partial P)_H$, which has come to be called the *Joule–Thomson coefficient*, symbolized μ_{JT} :

$$\mu_{JT} \equiv \left(\frac{\partial T}{\partial P} \right)_H \quad (3.61)$$

The Joule–Thomson experiment can be described as “adiabatic expansion in a pipe through a porous plug,” as pictured schematically in Fig. 3.11.

The porous plug (traditionally meerschaum, but equally well a silk handkerchief) separates the adiabatic chamber with opposing pistons held at *fixed* pressures P_i , P_f (with $P_i > P_f$). The gas is initially in the left-hand chamber (at V_i , T_i) under the high pressure P_i , but oozes reversibly through the porous plug to expand into the right-hand chamber (at V_f , T_f) under the low pressure P_f . The experiment consists of measuring the final temperature difference $\Delta T = T_f - T_i$ for fixed pressure difference $\Delta P = P_f - P_i$, and evaluating the ratio in the small- ΔP limit:

$$\mu_{JT} = \lim_{\Delta P \rightarrow 0} \left(\frac{\Delta T}{\Delta P} \right)_{\text{“JT”}} \quad (3.62)$$

where subscript “JT” represents the (as yet incompletely specified) conditions of the experiment.

Let us now analyze the conditions of the Joule–Thomson expansion in more detail. From the adiabatic character ($q = 0$) of the expansion, the first law tells us that

$$\Delta U = w = \text{net work performed on the gas} \quad (3.63)$$

Because the pressures P_i , P_f are fixed constants, we can see that work $w_L = P_i(V_i - 0)$ was performed *on* the gas in the left chamber, while work $w_R = P_f(0 - V_f)$ was performed *by* the gas in the right chamber, leading to net work w given by

$$w = P_i V_i - P_f V_f = -\Delta(PV) \quad (3.64)$$

From (3.63) and (3.64), we recognize that $\Delta U = -\Delta(PV)$, i.e., that Joule–Thomson expansion occurs under conditions of *constant enthalpy*:

$$\Delta U + \Delta(PV) = \Delta H = 0 \quad (3.65)$$

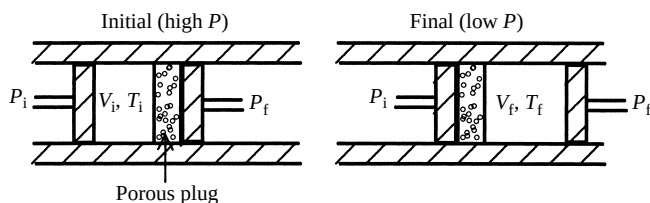


Figure 3.11 Joule–Thomson porous-plug experiment, showing the initial state P_i , V_i , T_i (left) and final state P_f , V_f , T_f (right) of the gas as it passes reversibly through the porous plug under fixed pressures P_i , P_f and adiabatic conditions.

The measured Joule–Thomson coefficient (3.62) can therefore be identified more precisely as $\mu_{JT} = (\partial T / \partial P)_H$, as claimed in (3.57).

The measured Joule–Thomson coefficient μ_{JT} provides valuable information about how the enthalpy of real gases depends on variables other than temperature. To obtain information about the P dependence of H , we can employ the Jacobi cyclic identity (1.14b) to rewrite the Joule–Thomson coefficient as

$$\mu_{JT} = \left(\frac{\partial T}{\partial P} \right)_H = - \frac{(\partial H / \partial P)_T}{(\partial H / \partial T)_P} \quad (3.66)$$

Recognizing $C_P = (\partial H / \partial T)_P$ in the denominator, we can rewrite (3.66) as

$$\left(\frac{\partial H}{\partial P} \right)_T = -\mu_{JT} C_P \quad (3.67)$$

Measured values of μ_{JT} and C_P therefore allow us to obtain the values of $(\partial H / \partial P)_T$ for real gases. Other derivatives such as $(\partial H / \partial V)_T$, $(\partial U / \partial P)_T$, or $(\partial U / \partial V)_T$ then follow readily.

As expected for any property of a real gas, $\mu_{JT} = \mu_{JT}(T, P)$ varies with the chosen temperature and pressure. Figure 3.12 illustrates some aspects of the (T, P) variations of μ_{JT} for CO_2 . As can be seen, $\mu_{JT}(\text{CO}_2)$ is positive throughout the near-ambient region, but trends toward zero or negative values at high T and P , typical of many gases.

Positive values $\mu_{JT} > 0$ are the “usual” low- T case for most common gases (i.e., all except He and H_2 at room temperature). In this case, the gas *cools* on expansion under adiabatic conditions, indicative of the dominance of *attractions* between molecules. The contrary high- T case of $\mu_{JT} < 0$ (e.g., for H_2 above 193K) leads to the gas *warming* on adiabatic expansion, indicative of the dominance of intermolecular *repulsions*. The crossover from positive to negative values of μ_{JT} occurs at the “Joule–Thomson inversion temperature” T_I , where

$$\mu_{JT} = 0 \quad \text{at } T = T_I \quad (3.68)$$

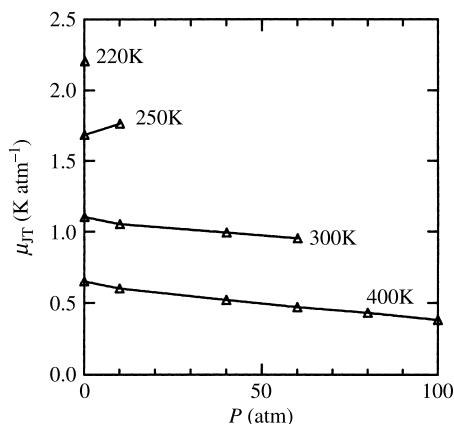


Figure 3.12 Qualitative temperature and pressure dependence of the Joule–Thomson coefficient $\mu_{JT}(T, P)$ for CO_2 .

Joule–Thomson inversion typically occurs at temperatures far above the critical temperature ($T_1 > T_c$). We shall later prove (Sidebar 5.5) the general thermodynamic identity

$$\mu_{JT} = \frac{TV\alpha_p - V}{C_p} \quad (3.69)$$

from which we can readily infer that

$$T_1 = \alpha_p^{-1} \quad (3.70)$$

showing that Joule–Thomson inversion points can be predicted from the coefficient of thermal expansion (2.4).

The fact that a gas can be cooled ($\mu_{JT} > 0$) or warmed ($\mu_{JT} < 0$) by merely expanding under adiabatic ($q = 0$) conditions may seem quite counterintuitive. How can we change the temperature of a gas without adding or subtracting *heat*? The answer follows quite directly from the first law. Under adiabatic conditions, $\Delta U = w$, so the work performed by the gas in reversible adiabatic expansion must be compensated by the change ΔU in internal energy, that is, by a *temperature* change (since heat capacity is nonzero).

The ability to cool (and eventually liquefy) gases by adiabatic expansion underlies industrial gas liquefaction processes. Adiabatic cooling of gaseous nozzle-jet expansions is also an important technique in modern molecular beam and mass spectrometric research. Thermodynamicist John Fenn, winner of the 2002 Nobel Prize in Chemistry, pioneered many of the techniques of adiabatic nozzle-beam cooling.

3.6.4 Ideal Gas Thermodynamics

In this section we summarize some leading thermodynamic properties of the fictitious “ideal gas” system, with defining characteristics

$$PV = nRT \quad (3.71a)$$

$$\left(\frac{\partial U}{\partial V}\right)_T = 0 \quad (3.71b)$$

[Because practically *all* the equations of this section are tainted by the errors of the ideal gas approximation (3.71a, b), we do not explicitly mark each one with the “IG” symbol as done elsewhere in this book; however, the reader should mentally flag these equations as approximations.]

Some aspects of the kinetic molecular theory (KMT) of ideal gases were outlined in Sidebar 2.7. The simplest form of KMT refers to monatomic ideal gases, for which the internal energy U and enthalpy $H = U + PV = U + nRT$ can be written explicitly as

$$U = \frac{3}{2} nRT \quad (3.72a)$$

$$H = \frac{5}{2} nRT \quad (3.72b)$$

From (3.72), the heat capacities $C_V = (\partial U / \partial T)_V$ and $C_P = (\partial H / \partial T)_P$ can be evaluated as

$$C_V = \frac{3}{2} nR \quad (3.73a)$$

$$C_P = \frac{5}{2} nR \quad (3.73b)$$

The expressions in (3.72) and (3.73) are valid only for monatomic ideal gases such as He or Ar, and must be replaced by somewhat different expressions for diatomic or polyatomic molecules (Sidebar 3.8). However, the classical expressions for polyatomic heat capacity exhibit serious errors (except at high temperatures) due to the important effects of quantum mechanics. (The failure of classical mechanics to describe the heat capacities of polyatomic species motivated Einstein's pioneering application of Planck's quantum theory to molecular vibrational phenomena.) For present purposes, we may envision taking more accurate heat capacity data from experiment [e.g., in equations such as (3.84a)] if polyatomic species are to be considered. The term "perfect gas" is sometimes employed to distinguish the monatomic case [for which (3.72), (3.73) are satisfactory] from more general polyatomic ideal gases with $C_V > \frac{3}{2} nR$.

SIDEBAR 3.8: CLASSICAL THEORY OF MONATOMIC AND POLYATOMIC IDEAL GASES

The original kinetic molecular theory assumes that the constituents of an ideal gas are point particles of mass m and average velocity v , without internal structure. In this case, internal energy U consists only of kinetic energy in the three independent directions (x , y , z) of free translational motion, each contributing $\frac{1}{2} nRT$ of average thermal energy to give (3.72a).

More generally, the classical "equipartition of energy" theorem asserts that *each* independent kinetic-type or potential-type thermal "mode" of a molecule acquires $\frac{1}{2} nRT$ of average thermal energy. Thus, to determine U , it is only necessary to count the number of independent translational, rotational, and vibrational thermal modes. For a general non-linear molecule of N atoms, the available thermal modes arise from three independent translations (contributing kinetic energy only), three independent rotations (contributing kinetic energy only), and $3N - 6$ independent vibrations (contributing one each of kinetic and potential type). [For the special case of a diatomic or linear polyatomic, only *two* rotational modes are meaningful (because no "rotation" occurs about the molecule's own axis) and only $3N - 5$ independent vibrations are present.]

As an example, we consider the benzene molecule (C_6H_6) with $N = 12$ atoms. In this case there are three kinetic modes of translation, three kinetic modes of rotation, and a kinetic and potential energy mode from each of the $3N - 6 = 30$ molecular vibrations, making 66 thermal modes in all. According to the equipartition theorem, each thermal mode acquires $\frac{1}{2} nRT$ of average thermal energy, giving $U = 33 nRT$ as the total internal energy and $C_V = 33R$ (or $C_P = 34R \simeq 67.6 \text{ cal mol}^{-1}$) as the molar heat capacity of benzene in the ideal gas approximation. The measured values for benzene, either in the liquid ($C_P = 31.67 \text{ cal mol}^{-1}$ at 20°C) or gaseous range ($C_P = 28.86 \text{ cal mol}^{-1}$ at 220°C), are far short of the classical equipartition value, due to the important effects of quantum mechanics.

Only for the monatomic case ($N = 1$) is the classical equipartition estimate qualitatively reasonable.

From the starting expressions (3.71)–(3.73), we now consider various simple processes involving an ideal gas.

Changes in U and H The key fact to recall is that only T changes can affect U or H of an ideal gas. Accordingly, we can write, for temperature change $\Delta T = T_2 - T_1$,

$$\Delta U = \int_{T_1}^{T_2} C_V dT = C_V \Delta T \quad (3.74a)$$

$$\Delta H = \int_{T_1}^{T_2} C_P dT = C_P \Delta T \quad (3.74b)$$

Reversible Isothermal Expansion Let us consider the heat and work of ideal gas expansion from V_1 to V_2 under isothermal conditions ($\Delta T = 0$). We recognize from (3.74a) that

$$\Delta U = q + w = 0 \quad (3.75)$$

so that $q = -w$ (or $dq = -dw$). Under reversible conditions (with $P_{\text{ext}} = P_{\text{int}} = nRT/V$) the heat and work differentials are

$$dq = -dw = P dV = nRT \frac{dV}{V} \quad (3.76)$$

The integrated heat and work of the process are therefore

$$q = -w = \int_{V_1}^{V_2} nRT \frac{dV}{V} = nRT \int_{V_1}^{V_2} d \ln V = nRT \ln \left(\frac{V_2}{V_1} \right) \quad (3.77)$$

Note from (3.71a) that isothermal P and V changes are related by

$$PV = \text{constant} \quad (3.78)$$

Reversible Adiabatic Expansion Under reversible adiabatic conditions ($q = 0$), the first law and (3.74a) establish that

$$\Delta U = dw = -P dV = C_V dT \quad (3.79)$$

so that

$$0 = C_V dT + P dV \quad (3.80)$$

For an ideal gas, we can rewrite dT through the identity

$$dT = d \left(\frac{PV}{nR} \right) = \frac{P dV + V dP}{nR} \quad (3.81)$$

and thereby obtain from (3.80)

$$0 = \frac{C_V}{nR}(P dV + V dP) + P dV = \frac{(C_V + nR)P dV + (C_V)V dP}{nR} \quad (3.82)$$

Multiplying (3.82) through by nR/PV , and recognizing from (3.73a, b) (cf. Sidebar 3.7) that $C_V + nR = C_P$, we obtain

$$0 = C_P \frac{dV}{V} + C_V \frac{dP}{P} = C_P d \ln V + C_V d \ln P \quad (3.83)$$

We now define the dimensionless *heat capacity ratio* γ as

$$\gamma \equiv \frac{C_P}{C_V} \quad (3.84a)$$

where, from (3.73a, b), for a monatomic ideal gas

$$\gamma = \frac{5}{3} \quad (3.84b)$$

In terms of this ratio, we can rewrite (3.83) as

$$0 = \gamma d \ln V + d \ln P = d \ln(PV^\gamma) \quad (3.85)$$

Equation (3.85) establishes that $\ln(PV^\gamma)$ is a constant, hence also that

$$\boxed{PV^\gamma = \text{constant}} \quad (3.86)$$

in a reversible adiabatic process. Equation (3.86) is called the *adiabatic equation of state* to contrast it with the corresponding equation of state (3.78) for isothermal conditions. By using the ideal gas equation of state, (3.86) can be expressed equivalently in terms of variables T , V as

$$TV^{\gamma-1} = \text{constant} \quad (3.87)$$

or in terms of T , P as

$$TP^{(1-\gamma)/\gamma} = \text{constant} \quad (3.88)$$

(Of course, the “constant” has a different numerical value in each of these equations.) For a monatomic ideal gas, these equations take the more specific form corresponding to (3.84b):

$$PV^{5/3} = \text{constant} \quad (3.89a)$$

$$TV^{2/3} = \text{constant} \quad (3.89b)$$

$$TP^{-2/5} = \text{constant} \quad (3.89c)$$

Figure 3.13 compares isothermal and adiabatic expansions of a monatomic ideal gas from a common starting point. As seen in the figure, the fall of pressure $P \propto V^{-5/3}$

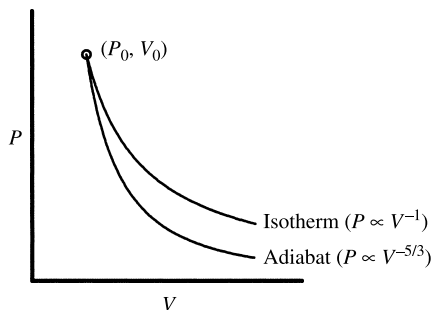


Figure 3.13 Comparison of isothermal and adiabatic expansions of a monatomic ideal gas from a common starting point P_0 , V_0 , showing the steeper fall of pressure (and temperature) in the adiabatic case.

under adiabatic conditions must be *faster* than that ($P \propto V^{-1}$) under isothermal conditions of gas expansion. (It will later be proven more generally that $\gamma \geq 1$ is a general thermodynamic stability condition.) The faster pressure drop can also be associated with the gas *cooling* that is expected under adiabatic conditions (since $P \propto T$ for any given volume of ideal gas). Sidebar 3.9 illustrates typical problems involving reversible and irreversible ideal gas expansion under isothermal and adiabatic conditions.

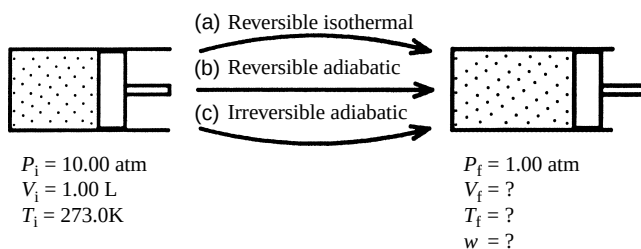
SIDEBAR 3.9: IDEAL GAS EXPANSION PROBLEMS

Problem Consider 1.00 L of a monatomic ideal gas (e.g., Ne), initially at 273.0 K and 10.00 atm pressure in a piston. The gas is allowed to expand to a final pressure of 1.00 atm by three different paths:

- reversible isothermal expansion
- reversible adiabatic expansion
- irreversible adiabatic expansion (against constant external $P_{\text{ext}} = 1 \text{ atm}$)

Calculate the final volume V_f , temperature T_f , and work w performed in each process.

Solution A solution often begins with a diagram outlining the problem:



Note that $n \neq 1$ mole for the problem as given.

- (a) *Reversible isothermal expansion* Under isothermal conditions,

$$T_f = 273.0\text{K}$$

and $PV = \text{constant}$, so that

$$V_f = (P_i V_i)/P_f = 10.0\text{ L}$$

From (3.77), the work w is therefore

$$w = -nRT \ln(V_f/V_i) = -P_i V_i \ln(V_f/V_i)$$

which evaluates to

$$w = -(10.00)(1.00) \ln\left(\frac{10.00}{1.00}\right) = -23.03\text{ L atm}$$

(or $w = -2.33\text{ kJ}$). In evaluating the expression, we have used the fact that $nRT = P_i V_i$ (instead of solving for n) as a short-cut. Note also that it is generally advantageous to write quantities as *dimensionless ratios* whenever possible, reserving the introduction of physical units until last.

- (b) *Reversible adiabatic expansion* For this case, we use (3.86) to write

$$1 = \frac{P_i}{P_f} \left(\frac{V_i}{V_f}\right)^{5/3}$$

from which

$$V_f = (10.00)^{3/5} = 3.98\text{ L}$$

From the ideal gas law, we can write

$$\frac{T_f}{T_i} = \frac{P_f V_f}{P_i V_i} = \frac{(1.00)(3.98)}{(10.00)(1.00)} = 0.398$$

from which

$$T_f = 0.398(273.0) = 108.7\text{K}$$

Finally, under adiabatic conditions ($q = 0$), we can use the first law and (3.74a) to express w as

$$w = \Delta U = C_V \Delta T = \frac{3}{2} nR \Delta T = \frac{3}{2} (P_f V_f - P_i V_i)$$

which evaluates to

$$w = \frac{3}{2} [(1.00)(3.98) - (10.00)(1.00)] = -9.03\text{ L atm}$$

(or $w = -9.14\text{ kJ}$). Note that one might have also evaluated $w = \int P dV = \text{constant} \times \int V^{-\gamma} dV$, but this is more tedious than the equations used above.

- (c) *Irreversible adiabatic expansion* In this case, the gas expands under constant $P_{\text{ext}} = P_f = 1.00 \text{ atm}$, so the work performed (“lifting the external atmosphere”) is

$$w = -P_f \Delta V = -P_f(V_f - V_i)$$

However, under adiabatic conditions ($q = 0$) we can also write $w = \Delta U = C_V \Delta T$, or

$$w = C_V \Delta T = \frac{3}{2} nR(T_f - T_i) = \frac{3}{2} (P_f V_f - P_i V_i)$$

Equating these two expressions for w , we obtain the equation

$$-P_f(V_f - V_i) = \frac{3}{2} (P_f V_f - P_i V_i)$$

and solve for the lone remaining unknown, V_f , to obtain

$$V_f = \frac{32}{5} V_i = 6.40 \text{ L}$$

The work w can then be obtained from either equation above as

$$w = -5.40 \text{ L atm}$$

(or $w = -5.47 \text{ kJ}$), and the final temperature $T_f = (P_f V_f / P_i V_i) T_i$ as

$$T_f = 175 \text{ K}$$

The following table summarizes V_f , T_f , and w for each path:

Process	V_f (L)	T_f (K)	w (L atm)
(a) Reversible isothermal	10.00	273.0	−23.03
(b) Reversible adiabatic	3.98	108.7	−9.03
(c) Irreversible adiabatic	6.40	175.0	−5.40

Note that the reversible isothermal path produces maximum useful work, while the reversible adiabatic path produces maximum cooling.

3.6.5 Thermochemistry: Enthalpies of Chemical Reactions

Chemical reactions are the changes (Δ) of greatest interest to chemists. The heat liberated or absorbed in chemical reactions (i.e., reaction enthalpy ΔH_{rxn} , under the usual conditions of open laboratory vessels) has been the subject of intense interest and quantitative calorimetric study from the dawn of the modern chemical era. In the present section, we merely wish to sketch how first-law principles underlie the entire theory and practice of modern thermochemistry, without entering the domain of practical applications, which are usually discussed in introductory chemistry textbooks.

Our principal emphasis is on constant-pressure heat or enthalpy changes,

$$q_P = \Delta H = \Delta U + \Delta(PV) \quad (3.90)$$

which must be carefully distinguished from the corresponding internal energy changes ($\Delta U = q_V$) as measured in bomb-calorimeter conditions. The numerical difference, $\Delta(PV)$, between ΔH and ΔU reaction heats is rather small for reactions involving only liquids or solids, but quite appreciable for reactions involving changes (Δn_{gas}) in the number of moles of *gaseous* species:

$$\Delta(PV) \simeq \begin{cases} 0 & \text{for liquids, solids} \\ RT\Delta n_{\text{gas}} & \text{for gases} \end{cases} \quad (3.91)$$

The formal discussion of ΔU_{rxn} values would parallel the present treatment of ΔH_{rxn} values in all essential respects.

The formal treatment of ΔH_{rxn} is facilitated by a convenient general notation for chemical reactions. It is well known that a balanced chemical reaction of generic form



has aspects of a balanced algebraic equation, with equality of atom numbers on reactant and product sides of the “chemical equation.” To further emphasize this similarity, we might rewrite (3.92a) by taking all terms to the right-hand side of the equals sign as

$$0 = cC + dD - aA - bB \quad (3.92b)$$

or, more generally,

$$\boxed{0 = \sum_i \nu_i A_i} \quad (3.93)$$

In (3.93), the A_i are chemical species and the ν_i are the corresponding stoichiometric coefficients of the balanced reaction, with the convention

$$\text{stoichiometric coefficients} \begin{cases} \nu_i > 0 & \text{for products} \\ \nu_i < 0 & \text{for reactants} \end{cases} \quad (3.94)$$

as illustrated in (3.92b).

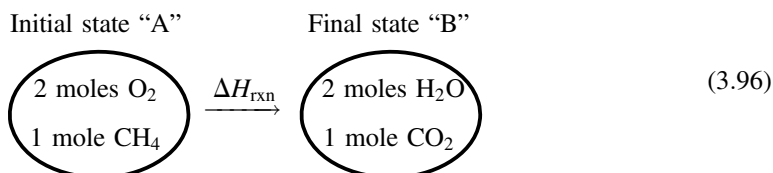
For example, the balanced chemical reaction



could be written with ν_i , A_i , $i = 1, \dots, 4$, as follows:

$$\begin{array}{c|cccc} i & 1 & 2 & 3 & 4 \\ \hline A_i & \text{CH}_4 & \text{O}_2 & \text{CO}_2 & \text{H}_2\text{O} \\ \nu_i & -1 & -2 & +1 & +2 \end{array} \quad (3.95b)$$

Let us now consider the “change of state” $A \rightarrow B$ represented by a chemical equation. For (3.95), this could be represented by



[Of course, the actual chemical reaction may not “go to completion” as represented by (3.96), but that is an aspect of chemical equilibrium that will be treated in Chapter 8.] If H_A represents the enthalpy of reactant species and H_B that of product species in (3.96), the ΔH for this process is

$$\Delta H = H_B - H_A = H_{\text{products}} - H_{\text{reactants}} = \Delta H_{\text{rxn}} \quad (3.97)$$

To characterize the ΔH_{rxn} for the chemical process in (3.96), we recall the properties of H previously established in the first law.

(1) H is a State Property

The general criterion (1.16a) for a state property (conserved quantity, exact differential) can be expressed as

$$\oint dH \quad \text{for any cycle} \quad (3.98)$$

A first consequence of (3.98) is evident from envisioning the reaction in terms of forward and backward reaction paths between reactant state A and product state B:



As shown in (3.99), the forward and reverse reactions must have *equal and opposite* ΔH_{rxn} values, so that if, for example, $\Delta H_{A \rightarrow B} < 0$ (exothermic), then $\Delta H_{B \rightarrow A} > 0$ (endothermic). ΔH_{rxn} must therefore be specified for the reaction as written.

A second consequence follows from rewriting the original transformation $A \rightarrow B$ in terms of alternative chemical reaction pathways that involve the same initial and final states. As depicted in (3.100a), the indirect reaction sequence $A \rightarrow X \rightarrow Y \rightarrow Z \rightarrow B$ must have exactly the *same* overall ΔH as the direct reaction $A \rightarrow B$,



which is expressed by the conservation equation

$$\Delta H_{A \rightarrow B} = \Delta H_{A \rightarrow X} + \Delta H_{X \rightarrow Y} + \Delta H_{Y \rightarrow Z} + \Delta H_{Z \rightarrow B} \quad (3.100b)$$

Cyclic enthalpy conservation relationships such as (3.100) are summarized by

Hess' law ("law of constant heat summation"): Reaction enthalpies add together as do the associated chemical reactions.

which is merely the thermochemical consequence of the first law. The vast literature of precise thermochemical measurements bears eloquent testimony to the exactness and generality of first-law relationships in chemical reactions.

(2) $H = H(T, P)$

Like all state properties, H is a function of temperature and pressure. Meaningful values of ΔH_{rxn} must therefore include implicit or explicit specification of T, P at which the chemical reaction was performed. Furthermore, it is necessary that all reaction products be brought back to the *same* temperature and pressure before the final $\Delta H_{\text{rxn}}(T, P)$ heat change is tallied up.

Conventional notation for $\Delta H(T, P)$ is often based on a selected "standard state" (such as $T = 298\text{K}$, $P = 1\text{ atm}$), which is designated by a "degree" circle (ΔH°). Alternatively, the T, P values can be explicitly specified by subscript and superscript values (ΔH_T^P). Because T is often the more important variation, a hybrid notation such as illustrated in (3.101a–c) is common (with, e.g., 1 atm as standard-state pressure):

$$\Delta H^\circ: \quad P = 1\text{ atm}, T = 298\text{K} \quad (3.101a)$$

$$\Delta H_{273}^\circ: \quad P = 1\text{ atm}, T = 273\text{K} \quad (3.101b)$$

$$\Delta H_0^\circ: \quad P = 1\text{ atm}, T = 0\text{K} \quad (3.101c)$$

Different choices of "standard" state are often adopted by different reference sources (such as 1 bar instead of 1 atm as the standard pressure), so it is important to check that "°" has a consistent meaning when using $\Delta H_{\text{rxn}}^\circ$ values from different sources.

(3) H is Extensive

The extensive property implies that ΔH_{rxn} is additive in the various components and multiplicative in the amounts (e.g., mole numbers) of each component. As a consequence, we can write

$$\Delta H_{\text{rxn}} = \sum_i \nu_i \bar{H}_i \quad (3.102)$$

where $\bar{H}_i$ is the molar enthalpy of species i and the stoichiometric coefficients ν_i (3.94) express the "Δ" (gain for products, loss for reactants) in the reaction change of state.

However, the absolute enthalpies $\bar{H}_i$ of individual species are generally unknown (indeterminate up to an arbitrary "zero" of the energy scale). Accordingly, to make use

of (3.102) we must replace each $\bar{H}_i$ by an enthalpy *difference*,

$$\Delta\bar{H}_i = \bar{H}_i - \bar{H}_i^{\text{ref}} \quad (3.103)$$

where the chosen reference state “ $\bar{H}_i^{\text{ref}}$ ” cancels out of any chemical reaction:

$$0 = \sum_i \nu_i \bar{H}_i^{\text{ref}} \quad (3.104)$$

It is convenient to choose $\bar{H}_i$ with respect to a “standard form” of the *elements* of which A_i is composed, since the elements must indeed “cancel out” in any balanced chemical reaction, as required by (3.104). Accordingly, we replace

$$\bar{H}_i \rightarrow \Delta H_f^\circ [A_i] \quad (3.105)$$

where $\Delta H_f^\circ [A_i]$ is the standard (molar) enthalpy of *formation* of A_i from its elements in their standard form. “Standard form” is taken as the most stable form of the element at 1 atm, 25°C [e.g., gaseous diatomic $\text{H}_2(\text{g})$ for hydrogen, solid graphite $\text{C}(\text{s})$ for carbon, rhombic solid $\text{S}_8(\text{s})$ for sulfur, and so forth]. With this replacement, (3.102) becomes

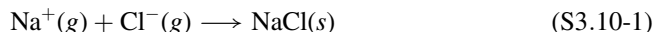
$$\Delta H_{\text{rxn}}^\circ = \sum_i \nu_i \Delta H_f^\circ [A_i] \quad (3.106)$$

Equation (3.106) is the well-known starting point for thermochemical calculations as described in many elementary textbooks. Standard enthalpies of formation $\Delta H_f^\circ [A_i]$ have been measured and tabulated (see, e.g., the NIST website: <http://webbook.nist.gov/chemistry>) for a vast number of chemical compounds A_i , so (3.106) makes it rather easy to obtain $\Delta H_{\text{rxn}}^\circ$ values under standard state conditions for virtually any chemical reaction of interest.

Thermochemical cycles extend to much less routine applications than those associated with (3.106). As an illustrative example, Sidebar 3.10 summarizes the “Born–Haber cycle,” by which a key quantity of ionic lattice theory is obtained from fiendishly indirect thermochemical measurements.

SIDEBAR 3.10: BORN–HABER CYCLE

In the general theory of ionic crystals (such as table salt, NaCl), a key physical quantity is the “cohesive energy” E_{xtal} of forming the solid crystal from its constituent ions. For sodium chloride, for example, this is the energy lowering in the reaction



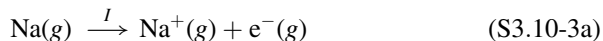
equivalent (except for a sign change) to the reaction enthalpy in (S3.10-1),

$$E_{\text{xtal}} = -\Delta H_{\text{rxn}} \quad (\text{S3.10-2})$$

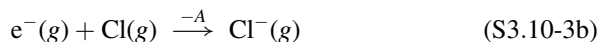
However, this energy change cannot be measured directly (e.g., by heating the crystal), because the crystal dissociates to gaseous $\text{NaCl}(\text{g})$ *molecules*, not to ions.

Born and Haber suggested a clever method of determining E_{xtal} by an alternative reaction cycle for which experimental thermochemical data are readily available. The required data are:

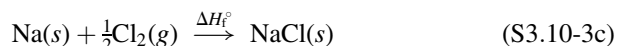
- (1) I = ionization potential of Na, determined from the reaction



- (2) A = electron affinity of Cl, determined (with change of sign) from



- (3) $\Delta H_f^\circ [\text{NaCl}(s)]$, determined from the standard formation reaction



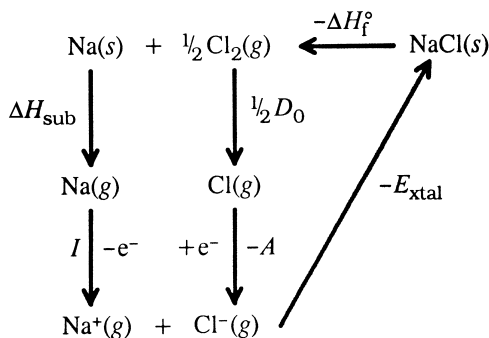
- (4) ΔH_{sub} , the sublimation energy for metallic solid Na(s), from



- (5) D_0 , the dissociation energy of Cl_2 to atoms, from



The following diagram shows how the product species $\text{NaCl}(s)$ can be connected to reactant ions $\text{Na}^+(g)$, $\text{Cl}^-(g)$ either by the direct path (S3.10-1) or by the succession of reactions (S3.10-3a–e):



From the diagram, one can directly read the Born–Haber relationship between the two paths:

$$E_{\text{xtal}} = -\Delta H_f^\circ + \Delta H_{\text{sub}} + \frac{1}{2}D_0 + I - A \quad (\text{S3.10-4})$$

Although surprisingly circuitous, this expression for the cohesive energy of an ionic crystal carries the full authority of the first law.

3.6.6 Temperature Dependence of Reaction Enthalpies

From equations such as (3.106) one can readily determine $\Delta H_{\text{rxn}}(T_1)$ at the “standard” temperature T_1 of thermochemical compilations. How can we find the value $\Delta H_{\text{rxn}}(T_2)$ at other temperatures T_2 of interest? This is the subject of Kirchhoff’s equation, which we shall derive as a simple consequence of the first law.

To evaluate the formal temperature derivative of $\Delta H = \Delta H_{\text{rxn}}$ (under isobaric conditions, assumed throughout this section), we note that (3.106) is of the form [cf. (3.103)]

$$\Delta H = \sum_i \nu_i (\bar{H}_i - \text{constant}) \quad (3.107)$$

Differentiating (3.107) with respect to T (at constant P), we obtain Kirchhoff’s equation in the form

$$\left(\frac{\partial(\Delta H)}{\partial T} \right)_P = \sum_i \nu_i \left(\frac{\partial \bar{H}_i}{\partial T} \right)_P = \sum_i \nu_i \bar{C}_{P,i} = \Delta C_P \quad (3.108)$$

or in simpler form (with constant P understood),

$$\boxed{\frac{d(\Delta H)}{dT} = \Delta C_P} \quad (3.109)$$

Here $\bar{C}_{P,i}$ is the molar heat capacity of species A_i at constant pressure and ΔC_P is the overall change in heat capacity (products minus reactants)

$$\Delta C_P = C_P^{\text{prod}} - C_P^{\text{react}} \quad (3.110)$$

with the usual convention (3.94) for stoichiometric coefficients.

We can integrate Kirchhoff’s equation (3.109) from T_1 to T_2 to obtain the desired $\Delta H(T_2)$ as

$$\Delta H(T_2) = \Delta H(T_1) + \int_{T_1}^{T_2} (\Delta C_P) dT \quad [+ \Delta H_{\text{latent}}] \quad (3.111)$$

The term in brackets, ΔH_{latent} , refers to the latent heat of *phase change* (if any) that occurs between T_1 and T_2 , e.g.,

$$\Delta H_{\text{latent}} = \begin{cases} \Delta H_{\text{vap}}, & \text{vaporization} \\ \Delta H_{\text{fusion}}, & \text{fusion (melting)} \\ \Delta H_{\text{sub}}, & \text{sublimation} \end{cases} \quad (3.112)$$

The latent-heat terms (3.112) become necessary whenever the integrand ΔC_P undergoes discontinuous change at a phase transition, with accompanying release of “hidden” ΔH . [The latent heat contribution is automatically included if one understands $\int (\Delta C_P) dT$ as *Lebesgue integration*.] For numerical evaluation of the integral in (3.111), power series

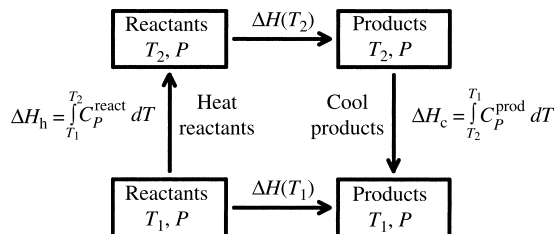


Figure 3.14 Graphical depiction of Kirchhoff's equation (3.111), showing the direct path for reaction at T_1 compared with the indirect path for reaction at T_2 , with heating of initial reactants (ΔH_h) and cooling of final products (ΔH_c) to complete the thermochemical cycle: $\Delta H(T_1) = \Delta H_h + \Delta H(T_2) + \Delta H_c$.

approximations are often useful in representing the T -dependent heat capacity $C_p(T)$ of each contributing reagent:

$$C_p(T) \simeq a + bT + cT^2 + \dots \quad (3.113)$$

The validity of the Kirchhoff integration formula (3.111) can be verified graphically by consideration of the thermochemical cycle shown in Fig. 3.14. As shown in the figure, the enthalpy change $\Delta H(T_1)$ for the direct reaction path at T_1 must match the total enthalpy change for the alternative path of

- (1) heating the reactants to T_2 (ΔH_h)
- (2) performing the reaction at T_2 [$\Delta H(T_2)$]
- (3) cooling the products to T_1 (ΔH_c)

The terms ΔH_h or ΔH_c involve integrals of the total heat capacities of reactants or products over the temperature range T_1 to T_2 (with reverse direction for products),

$$\Delta H_h + \Delta H_c = \int_{T_1}^{T_2} C_p^{\text{react}} dT + \int_{T_2}^{T_1} C_p^{\text{prod}} dT = \int_{T_1}^{T_2} (C_p^{\text{react}} - C_p^{\text{prod}}) dT \quad (3.114)$$

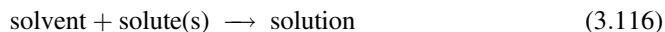
(including any latent heats that may appear in this temperature range for any reactant or product). The first-law conservation of enthalpy then requires that

$$\Delta H(T_1) = \Delta H_h + \Delta H_c + \Delta H(T_2) \quad (3.115)$$

which is equivalent to (3.111).

3.6.7 Heats of Solution

A special type of "reaction" is the formation of a solution from its components,



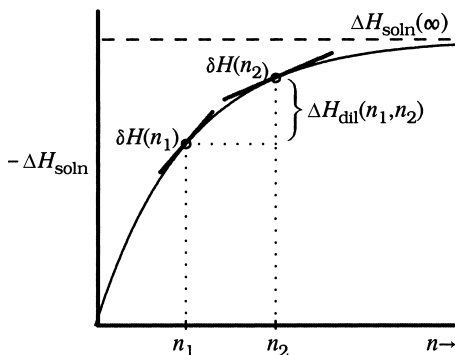
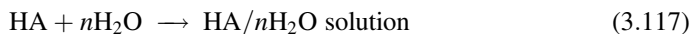


Figure 3.15 Plot of integral heat of solution $\Delta H_{\text{soln}}(n)$ versus n (= moles H_2O /moles acid), showing the infinite-dilution limit $\Delta H_{\text{soln}}(\infty)$, the heat of dilution $\Delta H_{\text{dil}}(n_1, n_2)$ from n_1 to n_2 , and the differential heat of solution (slope of tangent line) $\delta H(n_1)$, $\delta H(n_2)$ for representative concentrations n_1 , n_2 .

for which $\Delta H_{\text{rxn}} = \Delta H_{\text{soln}}$ is the “heat of solution.” A familiar example is addition of acid HA to water to form an aqueous solution $\text{HA}(aq)$,



for which the enthalpy change $\Delta H = \Delta H_{\text{soln}}(n)$ is called the “integral heat of solution.” Such reactions are found to exhibit high exothermicity for small n (= moles H_2O /moles HA) but ever-smaller increments as n increases indefinitely, where one is essentially adding pure water to nearly pure water. Figure 3.15 illustrates the general form of this dependence, showing the strongly negative value of $\Delta H_{\text{soln}}(n)$ as $n \rightarrow 0$, and the asymptotic approach to the infinite-dilution limit

$$\Delta H_{\text{soln}}(\infty) = “\Delta H_{\text{soln}}” = \text{integral heat of solution at infinite dilution} \quad (3.118)$$

at large n . [Unless otherwise stated, “aqueous” species $\text{HA}(aq)$ refer to this infinite-dilution limit.]

As shown in Fig. 3.15, the increment in $\Delta H_{\text{soln}}(n)$ between concentrations n_1 and n_2 is called the “heat of dilution” $\Delta H_{\text{dil}}(n_1, n_2)$

$$\Delta H_{\text{dil}}(n_1, n_2) = \Delta H_{\text{soln}}(n_2) - \Delta H_{\text{soln}}(n_1) \quad (3.119)$$

Furthermore, the *slope* of the ΔH_{soln} curve at $n = n_1$ is called the “differential heat of solution” $\delta H(n_1)$:

$$\delta H(n_1) = \left. \frac{\partial(\Delta H_{\text{soln}})}{\partial n} \right|_{n_1} \quad (3.120)$$

A similar expression applies to the corresponding quantity for the solute HA.

Let us consider a more general A/B solution of given concentration (in mole numbers n_A , n_B). If δH_A denotes the differential heat for the “solvent” A, and δH_B that for the

“solute” B, then the integral heat of solution is given by the simple additivity relation

$$\Delta H_{\text{soln}} = n_A \delta H_A + n_B \delta H_B \quad (3.121)$$

We can prove (3.121), starting from the general chain-rule expression for $d(\Delta H_{\text{soln}})$ at constant T and P :

$$d(\Delta H_{\text{soln}}) = \left(\frac{\partial(\Delta H_{\text{soln}})}{\partial n_A} \right) dn_A + \left(\frac{\partial(\Delta H_{\text{soln}})}{\partial n_B} \right) dn_B = \delta H_A dn_A + \delta H_B dn_B \quad (3.122)$$

Integrating this expression from “droplet” to full size at *constant* concentration (surely not the way the chemist prepares the solution, but an imaginable mathematical operation), we obtain

$$\Delta H_{\text{soln}} = \int_0^{n_A, n_B} d(\Delta H_{\text{soln}}) = \int_0^{n_A} \delta H_A dn_A + \int_0^{n_B} \delta H_B dn_B \quad (3.123)$$

Under these conditions δH_A and δH_B are *constants*, so that (3.123) immediately reduces to (3.121). Further aspects of such additivity relations involving “partial molar” quantities will be explored in Section 6.2.

From the earliest systematic studies it was observed that ΔH_{soln} values tend to be correlated with empirical solubility. Thus, a “highly soluble” salt such as Na_2SO_4 exhibits high exothermicity in aqueous solution ($\Delta H_{\text{soln}} \ll 0$). Conversely, the addition of NaCl to benzene is found to be highly endothermic ($\Delta H_{\text{soln}} \gg 0$), corresponding to the low solubility of polar salts in apolar organic solvents (“like dissolves like”). The propensity for solution formation to be related to its exothermicity is mirrored in other types of reactions, where the degree of spontaneity (or “vigor”) of the reaction often corresponds to high exothermicity. Such propensity led to the attempt by Berthelot to formulate a general principle of spontaneity as follows:

Berthelot’s principle: Chemical reactions proceed spontaneously if $\Delta H < 0$.

However, such an “enthalpic criterion of spontaneity” is undoubtedly erroneous. This can be seen from as simple an experiment as adding such highly soluble salts as NaCl or NH_4NO_3 to water; both processes will be found to be *endothermic* (the latter startlingly so).

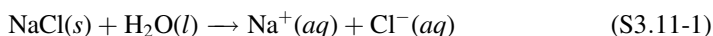
Although empirical solubility rules and ΔH_{soln} values show *some* degree of correlation with polarity or other attributes of solute and solvent, the situation at the molecular level can be rather complex (Sidebar 3.11). Heat evolution or absorption is thus a deep clue to the microscopic nature of solution formation, indicating its possible relationship to chemical reaction phenomena



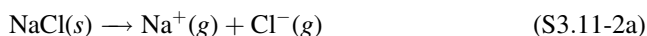
Further aspects of the molecular description of aqueous solvation phenomena will be considered in Chapters 6, 7, and elsewhere in this book.

SIDEBAR 3.11: MOLECULAR ASPECTS OF AQUEOUS SOLUBILITY

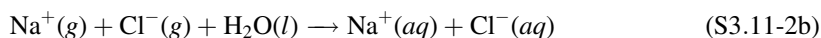
Formation of an aqueous solution of a strong electrolyte (shown here for NaCl),



can be pictured in terms of an alternative thermochemical pathway involving initial dissociation of the crystalline salt into gas-phase ions,



followed by aqueous solvation of the ions to form the final solution,



The initial step (S3.11-2a) is highly endothermic, corresponding to the *cohesive energy* of the crystal as evaluated by the Born–Haber cycle (Sidebar 3.10):

$$\Delta H_{2a} \gg 0$$

The final step (S3.11-2b), corresponding to the *solvation energy* of the ions, is highly exothermic:

$$\Delta H_{2b} \ll 0$$

For NaCl, the net result of (3.11-2a, b) is slightly *endothermic*:

$$\Delta H_1 = \Delta H_{2a} + \Delta H_{2b} = \Delta H_{\text{soln}} = 13.35 \text{ kcal mol}^{-1}$$

reflecting the slightly greater suitability of Na^+ , Cl^- ions for forming an ionic crystal than for accommodation in water.

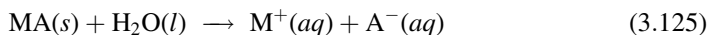
Solvation of ions by water involves formation of hydrogen bonds, competing with the hydrogen-bond network of water itself. The number and strength of solvent hydrogen bonds to a given ion depends sensitively on the relative size of the ion compared with the solvent molecule, as well as its net charge, electronegativity, and other electronic properties. In effect, the most favorable solvation energy arises from ion–water clusters of high enthalpic and entropic stability, corresponding to properly “matched” hydrogen-bonding properties. Similarly, the cohesive energy E_{xtal} in the initial step (S3.11-2a) depends on the crystal structure (different, e.g., for NaCl and CsCl), as well as favorable cation–anion matching conditions involving size, charge, polarizability, and other electronic properties of the ions.

Because both steps (S3.11-2a, b) involve rather subtle dependences on *chemical* properties of the ions (especially their complementary match to other species), the net enthalpy change ΔH_{soln} in (S3.11-1) can vary surprisingly from one electrolyte to another, even for

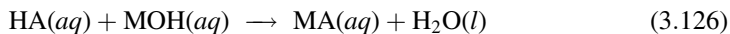
ions of the same charge and chemical family. Perhaps most mysterious in this respect is the hydrogen-bonded structure of water itself. Without proper understanding of hydrogen bonding in water, one can scarcely hope to understand the molecular-level nature of aqueous solvation for either electrolytes or nonelectrolytes. Fortunately, the thermodynamic description is independent of the current state of molecular-level understanding of aqueous solvation phenomena.

3.6.8 Other Aspects of Enthalpy Decompositions

As remarked in Sidebar 3.11, soluble salts MA (M = cation, A = anion) often behave as strong electrolytes in aqueous solutions, dissociating completely into ionic species as expressed by



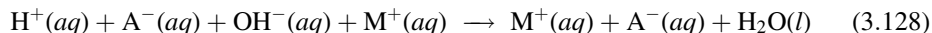
Moreover, ΔH_{rxn} for the neutralization reaction between the corresponding aqueous acid (HA) and base (MOH),



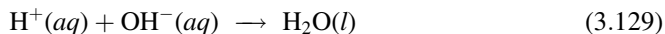
is found to be independent of the identity of M and A:

$$\Delta H_{\text{rxn}} = -13.345 \text{ kcal mol}^{-1} \quad (\text{all strong electrolytes MA}) \quad (3.127)$$

This suggests that we can also write the left-hand side of the reaction (3.126) in completely dissociated ionic form, analogous to (3.125):



The “spectator” species $\text{M}^+(aq)$, $\text{A}^-(aq)$ can therefore be cancelled from both sides of this equation to give the net ionic equation



which is indeed independent of M and A.

Reactions such as (3.129) suggest the utility of extending (3.106) to include reactions involving ions as reactants or products, for example,

$$\Delta H_{\text{rxn}}(3.129) = \Delta H_f^\circ [\text{H}_2\text{O}(l)] - \Delta H_f^\circ [\text{H}^+(aq)] - \Delta H_f^\circ [\text{OH}^-(aq)] \quad (3.130)$$

Of course, we cannot define ionic $\Delta H_f^\circ [\text{ion}^\pm(aq)]$ values by the formation reaction used for neutral compounds, because the standard elemental forms are neutral. However, we can create a consistent table of standard ionic enthalpies of formation by arbitrarily choosing the ΔH_f° value for a single ion (e.g., H^+), then determining the values for other ions by thermochemical consistency conditions. It is convenient to arbitrarily assign to $\text{H}^+(aq)$ a zero enthalpy of formation:

$$\Delta H_f^\circ [\text{H}^+(aq)] \equiv 0 \quad (3.131)$$

From the known formation enthalpy of liquid water, $\Delta H_f^\circ[\text{H}_2\text{O}(l)] = -68.315 \text{ kcal mol}^{-1}$, and the measured value (3.127), we obtain the remaining unknown $\Delta H_f^\circ[\text{OH}^-(aq)]$ of (3.130) as

$$\Delta H_f^\circ[\text{OH}^-(aq)] = -54.970 \text{ kcal mol}^{-1} \quad (3.132)$$

From the starting values (3.131), (3.132), the value of any desired $\Delta H_f^\circ[\text{M}^+(aq)]$ or $\Delta H_f^\circ[\text{A}^-(aq)]$ can be readily obtained from the known formation enthalpy of the corresponding $\text{MOH}(aq)$ or $\text{HA}(aq)$, respectively. The enthalpies of strong electrolytes can therefore be decomposed into constituent ionic enthalpies

$$\Delta H_f^\circ[\text{MA}(aq)] = \Delta H_f^\circ[\text{M}^+(aq)] + \Delta H_f^\circ[\text{A}^-(aq)] \quad (3.133)$$

in a convenient, general, and rigorous manner.

Is there a corresponding decomposition for nonelectrolytes? One might suppose that the analogous “building blocks” of a nonelectrolyte compound are its *chemical bonds*. Based on the known similarity and transferability of particular bond types from one molecule to another, one could then attempt to assign each bond a specific “bond enthalpy” (denoted $DH^\circ[\text{bond}]$) such that the overall sum of DH° s is related to (the negative of) ΔH_f° by

$$\sum_{\text{bonds}} DH^\circ[\text{bond}] \stackrel{?}{=} -\Delta H_f^\circ[\text{compound}] + \Delta H_{\text{atom}} \quad (3.134)$$

where ΔH_{atom} expresses the overall change of reference state from “standard” elemental form (for ΔH_f°) to gaseous atoms (for DH°). Empirical average $DH^\circ[\text{bond}]$ values have been tabulated [see, e.g., L. Pauling, *Nature of the Chemical Bond*, 3rd edn (Cornell University Press, Ithaca, 1960)] that allow simple estimates (approximately $\pm 5\%$) of ΔH_f° for many organic compounds (Sidebar 3.12). However, it is important to recognize that bond-enthalpy decompositions such as (3.134) are *approximations*, useful for qualitative purposes only and unjustified by any rigorous thermodynamic principles. (The non-transferability and nonadditivity of bond enthalpies arise from electronic delocalization and other coupling effects that vary from one molecular environment to another.) Although motivated by thermodynamics, expressions such as (3.134) are of irreducibly

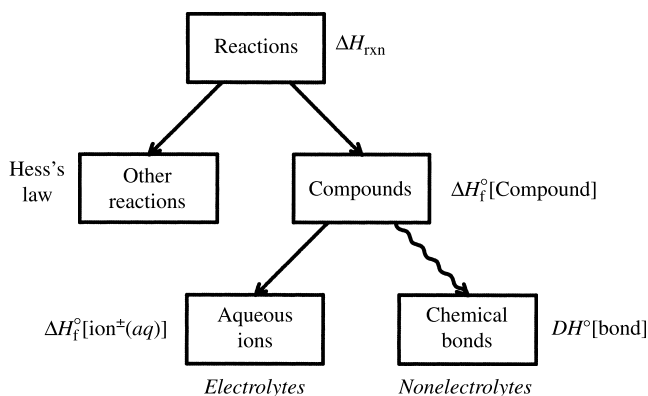


Figure 3.16 Schematic representation of enthalpy decompositions having exact (straight arrows) or approximate (wavy arrow) consistency with the first law of thermodynamics.

inexact and empirical character, and their accuracy varies considerably across blocks and bonding types of the periodic table.

Figure 3.16 summarizes various enthalpy decomposition schemes that are justified by the first law of thermodynamics. The results of innumerable thermochemical measurements based on these decompositions provide eloquent testimony to the accuracy and generality of the first law.

SIDEBAR 3.12: BOND ENTHALPIES

A chemically reasonable assumption is that the exothermicity $-\Delta H_{\text{rxn}}$ of a chemical reaction is related to the number and types of bonds made ($\nu_{\text{bond}} > 0$) or broken ($\nu_{\text{bond}} < 0$) in the reaction

$$-\Delta H_{\text{rxn}} \simeq \sum_{\text{bonds}} \nu_{\text{bond}} DH^\circ[\text{bond}] \quad (\text{S3.12-1})$$

where $DH^\circ[\text{bond}]$ is the “bond enthalpy” of given bond type. Such an assumption is based on the well-known localized character of chemical bonding and near-transferability of chemical bonds from one molecule to another, leading to near-additivity of bond contributions to molecular properties.

$DH^\circ[\text{A—B}]$ of a specific A—B bond may be expressed as the enthalpy required to dissociate a bonded A—B(g) species into gaseous fragments:

$$DH^\circ[\text{A—B}] = H[\text{A(g)}] + H[\text{B(g)}] - H[\text{A—B(g)}] \quad (\text{S3.12-2})$$

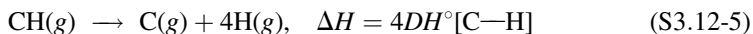
The sum of bond enthalpies is therefore related (with change of sign) to ΔH_f° of the gaseous compound, but with an altered reference state of gaseous atoms El(g) [rather than the “standard” form El(std)] for each element “El,”

$$\sum_{\text{bonds}} DH^\circ[\text{bond}] = -\Delta H_f^\circ [\text{compound(g)}] + \Delta H_{\text{atom}} \quad (\text{S3.12-3})$$

where the total “atomization energy” ΔH_{atom} expresses the change of elemental reference states from standard to gaseous-atom form:

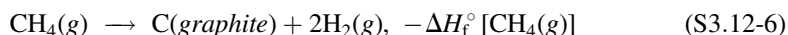
$$\Delta H_{\text{atom}} = \sum_{\text{El}} \{H[\text{El(g)}] - H[\text{El(std)}]\} = \sum_{\text{El}} \Delta H_{\text{atom}}[\text{El}] \quad (\text{S3.12-4})$$

For example, the usual structural description of a methane molecule (CH_4) in terms of four C—H single bonds suggests that ΔH_{rxn} for the total dissociation to gaseous atoms is simply four times the bond enthalpy $DH^\circ[\text{C—H}]$ of a “typical” C—H bond:



We can relate the constituent bond enthalpies $DH^\circ[\text{C—H}]$ to $\Delta H_f^\circ[\text{CH}_4(\text{g})]$ by decomposing the above reaction into separate steps:

- (1) Conversion of CH_4 to its elements in their standard form (the reverse of the standard formation reaction)



- (2) Atomization of each standard elemental form into gaseous atoms, with associated enthalpy change ΔH_{atom}

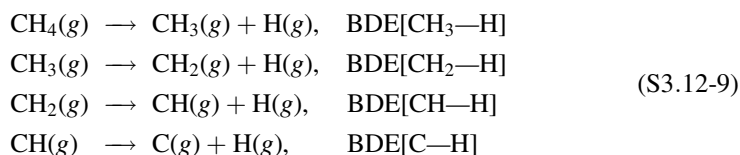


From the enthalpic equivalence of these alternative reaction paths, we obtain

$$4DH^\circ[\text{C—H}] = -\Delta H_f^\circ\{\text{CH}_4(\text{g})\} + \Delta H_{\text{atom}}[\text{C}] + 4\Delta H_{\text{atom}}[\text{H}] \quad (\text{S3.12-8})$$

as a special case of (3.134).

We could also envision dissociating $\text{CH}_4(\text{g})$ one bond at a time, with associated “bond dissociation energies” $\text{BDE}[\text{bond}]$ for each successive bond:



The BDEs are well-defined thermodynamic quantities that can be employed in rigorous first-law calculations, but their values strongly *differ* in each of the above steps. Only the *average* of the four $\text{BDE}[\text{CH}_n\text{—H}]$ values could be taken as an estimate of bond enthalpy $DH^\circ[\text{C—H}]$,

$$\begin{aligned} DH^\circ[\text{C—H}] &\simeq \frac{1}{4} \{ \text{BDE}[\text{CH}_3\text{—H}] + \text{BDE}[\text{CH}_2\text{—H}] + \text{BDE}[\text{CH—H}] \\ &\quad + \text{BDE}[\text{C—H}] \} \end{aligned} \quad (\text{S3.12-10})$$

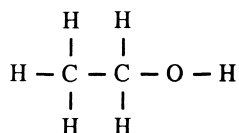
Useful empirical values of $DH^\circ[\text{bond}]$ can be obtained from such average BDE values, averaged over many representative molecules. The following table lists Pauling’s “best” single-bond enthalpies and associated atomization energies ΔH_{atom} for some common first-row atoms:

$DH^\circ[\text{A—B}] \text{ (kcal mol}^{-1}\text{)}$					
Atom	H	C	N	O	F
H	104				
C	99	83			
N	93	70	38		
O	111	84	—	33	
F	135	105	65	44	37
ΔH_{atom}	52	171	113	60	19

These are useful for “quick and dirty” estimates of ΔH_f° , as illustrated in the following problem.

Problem Estimate $\Delta H_f^\circ[\text{C}_2\text{H}_5\text{OH}(\text{g})]$ from the bond enthalpies tabulated above.

Solution From the standard Lewis structure representation



ethanol consists of five C—H bonds, one C—C bond, one C—O bond, and one O—H bond, with total bond-enthalpy sum

$$\sum_{\text{bonds}} DH^\circ[\text{bond}] = 5(99) + 83 + 84 + 111 = 773 \text{ kcal mol}^{-1} \quad (\text{S3.12-11})$$

For the total atomization energy $\Delta H_{\text{atom}} = 2 \Delta H_{\text{atom}}[\text{C}] + 6 \Delta H_{\text{atom}}[\text{H}] + \Delta H_{\text{atom}}[\text{O}]$, we obtain

$$\Delta H_{\text{atom}} = 2(171) + 6(52) + 60 = 714 \text{ kcal mol}^{-1} \quad (\text{S3.12-12})$$

The estimated ΔH_f° is therefore

$$\Delta H_f^\circ [\text{C}_2\text{H}_5\text{OH}(g)] \simeq -773 + 714 = -59 \text{ kcal mol}^{-1} \quad (\text{S3.12-13})$$

This estimate compares fairly well with the measured value, $\Delta H_f^\circ [\text{C}_2\text{H}_5\text{OH}(g)] = -55.795 \text{ kcal mol}^{-1}$.

Engine Efficiency, Entropy, and the Second Law

4.1 INTRODUCTION: HEAT FLOW, SPONTANEITY, AND IRREVERSIBILITY

The first law of thermodynamics seems to allow many thermal phenomena that are not observed experimentally. Indeed, the high symmetry of first-law-type relationships such as (3.99) would seem to suggest that the time sequence between “initial” and “final” states of a thermal process could be arbitrarily *reversed* with no special consequence (other than the expected sign change of ΔH). Yet experience says that this is not so.

Consider three elementary examples:

1. *Reverse-Rumford*. As noted in Sidebar 3.1, Count Rumford observed that the work performed *by* a horse on the cannon-bore was exactly compensated by the heat expelled *from* the cannon-bore to the surroundings, demonstrating the heat/work equivalence demanded by the first law. Imaginably, the process might have occurred in the *opposite* direction, with heat added *to* the cannon-bore (e.g., from a blow-torch) appearing as equivalent work performed *on* the horse (e.g., driving it backward in its hoofprints), in equal harmony with the first law. Yet such a “reverse-Rumford” process has never been observed.
2. *Reverse Heat Flow*. Joseph Black’s observations (Section 3.4) establish that two samples S_1 , S_2 of equal heat capacity but different initial temperatures T_1 , T_2 can exchange *equal* quantities of heat q to achieve equal and opposite heating/cooling effects to final $T_{av} = \frac{1}{2}(T_1 + T_2)$, with no change in the surroundings:

$$S_1(T_1) + S_2(T_2) \longrightarrow S_1(T_{av}) + S_2(T_{av}), \quad \Delta H = 0$$

Evidently, the exact reverse of this process, in which two equal- T samples spontaneously exchange q to warm one sample and cool the other by a compensating amount, would be equally in compliance with the first law. Yet such “reverse heat flow” has never been observed.

3. *Reverse Solvation*. As described in Section 3.6.7, when a solid crystal of a $\text{Na}_2\text{SO}_4(s)$ is placed in water, the crystal spontaneously dissolves with emission of heat q :

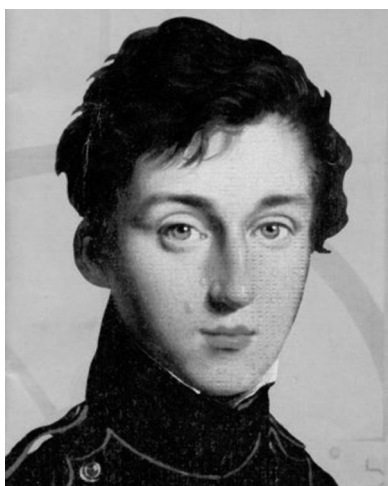


The reverse of this process, in which the same quantity of heat is *added* to the aqueous solution with spontaneous re-emergence (precipitation) of the salt crystal, is also fully compliant with the first law. However, such thermally induced “reverse solvation” has never been observed.

These three examples (and many others that might be imagined) indicate that the first law is inadequate to provide a complete picture of the intrinsic “natural” time-ordering or directionality of spontaneous thermal processes. As discussed in Section 3.2 (see Table 3.1), the irreversibility of spontaneous natural events (“time’s arrow”) is deeply tied to dissipative heating effects that underlie thermodynamic theory. Proper characterization of spontaneity and irreversibility in thermal processes therefore requires a further extension of the inductive basis of thermodynamic theory: the second law of thermodynamics.

The second law is more subtle and difficult to comprehend than the first. The full scope of the second law only became clear after an extended period of time in which (as expressed by Gibbs) “truth and error were in a confusing state of mixture.” In the present chapter, we focus primarily on the work of Carnot (Sidebar 4.1), Thomson (Sidebar 4.2), and Clausius (Sidebar 4.3), which culminated in Clausius’ clear enunciation of the second law in terms of the entropy function. This in turn led to the masterful reformulation by J. W. Gibbs, which underlies the modern theory of chemical and phase thermodynamics and is introduced in Chapter 5.

SIDEBAR 4.1: SADI CARNOT (1796–1832)



Sadi Carnot (full name Nicolas Léonard Sadi Carnot, “Sadi” after a Persian poet) was born into one of the most erudite and influential families of the turbulent Napoleonic period. Sadi’s father, Lazare Carnot, was a leading scientist and mathematician of his time, as well as a noted military commander who achieved high ministerial office under Napoleon. The father’s profound intellectual influence on Sadi is apparent from parallels between Lazare’s 1803 treatise, *Fundamental Principles of Equilibrium and Movement*, and Sadi’s famous 1824 monograph, *Reflections on the Motive Power of Fire* (*Réflexions sur la puissance motrice du feu*), which applied similarly general and abstract analysis to purely mechanical and thermomechanical devices, respectively. Among other accomplishments of this remarkable family, Sadi’s younger brother, Hippolyte,

became a noted writer and statesman, and the latter’s eldest son, Marie Francois Sadi Carnot, later became a president of the Third Republic.

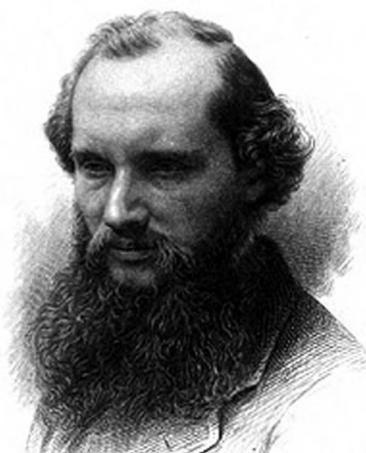
Under his father’s tutelage until age 16, Sadi entered the École Polytechnique (which his father had helped to found) to pursue a career in military engineering. However, the continued political turmoil, including his father’s exile after the Battle of Waterloo, brought considerable disappointment and frustration to the self-effacing young military officer. He thereupon took retirement from active military service in 1818 to pursue personal studies in Paris, with roommate Hippolyte. Sadi’s broad interests in mathematics, physics, chemistry, natural history, literature, music, and athletics were combined with

intense concentration on analysis of heat engines, culminating in his remarkable 118-page pamphlet, *Réflexions sur la puissance motrice du feu*, published in the year following his father's death in 1823. He was recalled to active military service in 1826, but again resigned in 1828 to resume his scientific studies. In 1832, at the height of his creative powers and apparently exhausted by overwork, Sadi developed an inflammation of the lungs that was followed by scarlet fever, and thereafter by a fatal cholera attack in the epidemic that struck Paris in August of that year. The fruits of his later researches (including clear recognition of the equivalence of heat and work) were not revealed until long after his death, when selected portions of his notes were published by Hippolyte in 1878 as an appendix to the second edition of the *Réflexions*, by then widely recognized as a landmark of thermodynamics.

Carnot's prescient pamphlet, apparently distributed only among a small circle of friends, remained unknown in the scientific literature for about a decade. Fortunately, its content and value were recognized by fellow French engineer Émile Clapeyron, who built on its concepts and extended its methods, including, for example, the first graphical PV representation of the Carnot cycle. Clapeyron's 1834 paper was the means by which Carnot's discoveries first became known to William Thomson, who made them the centerpiece of his own later work on thermodynamic theory.

Although Carnot's remarkable conclusions were essentially correct, it should be remarked for the sake of historical accuracy that present-day expositions of Carnot's analysis benefit from the refinements of more than a century of historical hindsight. Indeed, Carnot's work was set firmly in the then-current framework of caloric theory (Section 3.4), decades before the mechanical theory of heat and clear recognition of the first law of thermodynamics emerged. Thus, it is all the more remarkable that Carnot achieved such profoundly correct insights from what were at best rudimentary, incomplete, and incorrect premises for a theory of heat processes. His guiding principle (as for his father Lazare in the case of purely mechanical phenomena) was recognition of the impossibility of certain types of perpetual-motion devices. The later work of Thomson and Clausius allowed Carnot's insights to be expressed in more nearly the form they are taught today.

SIDEBAR 4.2: WILLIAM THOMSON, LORD KELVIN (1824–1907)



William Thomson (later Lord Kelvin) was born in Ireland in the year of Carnot's *Réflexions*, and became known as one of the most commanding scientific personalities of his era. Both his father (James, later Professor of Mathematics at the University of Glasgow) and his elder brother (also James) were notable scientists as well. Like Carnot, Thomson and his brother were home-schooled by their father, who was widowed when William was only six years old. Both boys proved to be prodigies, and William was first enrolled in the University of Glasgow when only ten years old. Among other accomplishments, William taught himself French by reading Laplace and Fourier, and the latter's analysis of heat diffusion had a formative influence on Thomson's interest in thermodynamic questions.

William entered Cambridge at the age of 16, earning high honors in mathematics. Early in 1846, just prior to his permanent appointment as Professor of Natural Philosophy at Glasgow (at age 22!), he spent a few months in the laboratory of H. G. Regnault in Paris. There he apparently learned of Clapeyron's 1834 paper and, indirectly, of Carnot's *Réflexions*, which he was unable to find in Paris. Thomson was finally able to obtain and read Carnot's work in 1848, and it became the acknowledged basis for his initial 1848 proposal of the absolute temperature scale (corrected in 1854 to that used today).

The following decade marked Thomson's most concentrated and productive period of research in thermodynamics, including his notable collaboration with Joule. Thereafter, Thomson turned increasingly to technology and invention (particularly involving electrical and magnetic phenomena), although he remained in close contact with Joule, Helmholtz, and other thermodynamic leaders and continued to publish in this area throughout his career. Thomson's theory of telegraphic signaling led him to deep involvement in the enterprise to lay a transatlantic cable, which was completed in 1866 to great acclaim after a decade-long effort. This success, together with other inventions and practical improvements, led to considerable public renown and wealth, including a personal yacht in which he travelled widely. Thomson ascended to the peerage as Baron Kelvin of Largs in 1892, taking this name from the small river Kelvin that winds through the Glasgow campus and the coastal town of Largs in which he maintained an impressive estate. He was known for his self-confidence and adoption of increasingly controversial and ill-judged positions on such questions as the nonexistence of atoms, Darwin's theory, the age of the Earth, the impossibility of flight, and alternative mechanical models of electromagnetism. (A biographer uncharitably remarked that during the first half of Thomson's career he seemed incapable of being wrong, while during the second half of his career he seemed incapable of being right.) He remained prominent in the public eye to the end of his career, and is buried in Westminster Abbey.

SIDEBAR 4.3: RUDOLF CLAUSIUS (1822–88)



Rudolf Julius Emmanuel Clausius was born into the large family of a church minister in Köslin, in the Pomeranian portion of Prussia (now Poland) near the Baltic Sea. He studied at the University of Berlin, taught Gymnasium science classes for a time, then completed his doctorate at the University of Halle in 1847 and gained an instructorship at Berlin in 1850. His first publications stemmed from his thesis work on atmospheric light scattering, but his attention soon turned sharply toward thermodynamics. His first publication in this area was the monumental 1850 paper [Über die bewegende Kraft der Wärme, and die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen. *Ann. Phys. (Leipzig)* **79**, 368–97, 500–24 (1850); English translation: On the moving force of heat, and the laws regarding the nature of

heat itself which are deducible therefrom. *Phil. Mag.* **2**, 1–21, 102–19 (1851)], in which he successfully reformulated Carnot’s analysis in a rigorous and convincing manner. His incisive papers of the next 15 years gave ever more succinct and general expression to the first and second laws (as Clausius identified them), culminating in his 1865 paper in which entropy was formally identified and used to express thermodynamic relationships in essentially modern form.

Clausius’ renown quickly led to a prestigious *Ordinarius* appointment at Zürich in 1855. However, he yearned to return to his German homeland, and in 1867 accepted appointment to the physics professorship at Würzburg, then later (1869) moved to the chair at Bonn, where he remained for the remainder of his career.

Clausius was an ardent nationalist and patriot. In the Franco-German war of 1870, he took leave of his university duties to enlist (nearing age 50!) for duties at the front, earning the Iron Cross in 1871. He commanded an ambulance corps of students recruited from Bonn that removed wounded soldiers from front-line action. In one of these engagements, he suffered a leg injury that left him disabled and in pain for life. His passionate advocacy of Germanic culture later led to unfortunate disputes with Thomson, Tait, Maxwell, and others over personal or nationalistic priority for the mechanical theory of heat, the first law, and other discoveries.

Clausius’ great paper of 1850 can be recognized as a landmark in the development of thermodynamics. As remarked by Thomson in 1851, “the merit of first establishing [Carnot’s theorem] upon correct principles is entirely due to Clausius.” In his 1889 eulogy of Clausius, Gibbs praised the 1850 paper in the following terms:

This memoir marks an epoch in the history of physics. If we say, in the words used by Maxwell some years ago, that thermodynamics is “a science with secure foundations, clear definitions, and distinct boundaries,” and ask when those foundations were laid, those definitions fixed, and those boundaries traced, there can be but one answer. Certainly not before the publication of that memoir. The materials indeed existed for such a science, as Clausius showed by constructing it from such materials, substantially, as had for years been the common property of physicists. But truth and error were in a confusing state of mixture. Neither in France, nor in Germany, nor in Great Britain, can we find the answer to the question [of ultimate heat conversion to work]. The case was worse than this, for wrong answers were confidently urged by the highest authorities. That question was completely answered, on its theoretical side, in the memoir of Clausius, and the science of thermodynamics came into existence . . . The constructive power thus exhibited, this ability to bring order out of confusion, this breadth of view which could apprehend one truth without losing sight of another, this nice discrimination to separate truth from error—these are qualities which place the possessor in the first rank of scientific men.

Clausius’ key insight was to recognize that in a Carnot cycle only a portion of the heat is transferred to the low-temperature reservoir, while the remainder appears as work in accordance with the first law. This required explicit renunciation of the caloric concept (which seemingly made heat, as a material substance, the conserved quantity) as well as accurate quantification of heat through the first law ($dq = dU - dw$). This explains the curious chronology of the second law, the essence of which originates in Carnot’s 1824 monograph, but the foundations of which are finally laid decades later in Clausius’ 1850 paper, only after the intervening development of the first law and the mechanical theory of heat.

It is noteworthy that Clausius also had to rely on second-hand information about Carnot’s *Réflexions*; as he states in his introduction: “I have not been able to obtain a copy of this book, and am acquainted with it only through the work of Clapeyron and Thomson, from the latter of whom are quoted the extracts afterwards given.”

4.2 HEAT ENGINES: CONVERSION OF HEAT TO WORK

The impetus to studies leading to the second law was the discovery of the steam engine (by James Watt and others) around 1769. The steam engine provided a novel source of mechanical work from heated steam that made earlier human- and animal-based forms of work obsolete. Such devices fundamentally devalued many forms of human manual labor in the industrial workplace, resulting in the industrial revolution that profoundly reshaped the post-18th-century world. At the time of its discovery, no proper scientific foundation existed for understanding how a steam engine works. Each new refinement yielded improved quantities of work for the fuel input, promising ever-cheaper sources of mechanical power production. Were there any *limits* to such improvements in efficiency? What were the *principles* that governed or constrained such machines? Might one even dream of a “perpetual motion machine,” the ultimate enticement of shady investment schemes? Such questions strongly motivated scientific studies over more than a century, culminating finally in the second law. As aptly remarked by science historian L. J. Henderson in 1917, “Science owes more to the steam engine than the steam engine owes to science.”

The rudimentary elements of a steam engine are shown schematically in Fig. 4.1. The engine consists of a cylinder and piston linked to a flywheel, with valves allowing steam input from the boiler (at temperature t_h) and expulsion through the exhaust pipe to the surroundings (at temperature t_c). Each engine cycle can be broken into four steps:

1. Hot steam from the boiler (t_h) enters the cylinders.
2. The steam expands (and cools), driving the piston back and turning the flywheel in the power stroke.
3. The expanded (and cooled) gas is vented at the exhaust pipe (t_c).
4. The momentum of the flywheel returns the piston to the starting position.

Although noncyclic variations are possible, we shall consider a generic “heat engine” to be such a device, operating in a cycle, that produces useful work from a quantity of heat input in each cycle. (In principle, the spent steam could be recovered from the exhaust pipe and recycled to the boiler, so no net input is required except boiler heat.)

How can we quantify the efficiency $\mathcal{E}$ of the steam engine? The desired output is the useful work produced by the engine ($-w$), while the input cost is the heat input q_h to the boiler (i.e., the cost of coal to maintain the boiler at t_h). We therefore define engine efficiency $\mathcal{E}$ as

$$\mathcal{E} \equiv \frac{\text{useful work}}{\text{input heat to boiler}} = \frac{-w}{q_h} \quad (4.1)$$

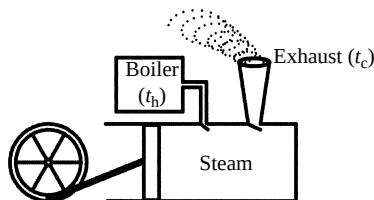


Figure 4.1 Schematic steam engine, showing boiler input (at temperature t_h), exhaust output (at temperature t_c), and connected flywheel for piston and steam cylinder.

From the first law, one can readily see that

$$-w \leq q_h \quad (4.2a)$$

which sets the upper limit

$$\mathcal{E} \leq 1 \quad (\text{first law}) \quad (4.2b)$$

i.e., no device can produce more useful work than is input as heat. However, we suspect that the first-law limit (4.2b) must be further tightened to express the real-world limitations on heat to work conversion.

4.3 CARNOT'S ANALYSIS OF OPTIMAL HEAT-ENGINE EFFICIENCY

The question of possible limits on the production of work from heat engines was taken up brilliantly by a young French military engineer, Sadi Carnot (Sidebar 4.1). Carnot's monograph of 1824, *Reflections on the Motive Power of Fire*, pointed to the answers to these questions in a remarkably bold and incisive (if abstract) way. Carnot introduces the question of the "motive power" (ability to cause movement) of "fire" (heat) in its most general terms:

Every one knows that heat can produce motion. That it possesses vast motive-power no one can doubt, in these days when the steam-engine is everywhere so well known . . . The question has often been raised whether the motive power of heat is unbounded, whether the possible improvements in steam-engines have an assignable limit—a limit which the nature of things will not allow to be passed by any means whatever; or whether, on the contrary, these improvements may be carried on indefinitely . . . We propose now to submit these questions to a deliberate examination.

Carnot's ensuing analysis of the steam engine culminated in an idealized "engine" of highest possible efficiency that could be represented as an abstract mathematical "Carnot cycle" in a PV diagram. Understanding the logic of this supreme thermodynamic abstraction is our first task.

Carnot recognized that the essence of a "heat engine" (a device able to convert heat to work) is a *working fluid operating between two temperature reservoirs*. As shown schematically in Fig. 4.2, the "working fluid" (e.g., steam) operates between a high-temperature reservoir t_h (e.g., the boiler) and a low-temperature reservoir t_c (e.g., the exhaust pipe to the surroundings), absorbing a quantity of "high-temperature heat" q_h from the hot reservoir and expelling a (smaller) quantity of "low-temperature heat" q_c to the cold reservoir, with production of work w . According to the first law, the difference in heat energy $|q_h| - |q_c|$ absorbed and expelled at the two reservoirs must equal the work produced by the engine:

$$-w = |w| = |q_h| - |q_c| = q_h + q_c \quad (4.3)$$

Because of the somewhat confusing way in which the algebraic signs of w , q_h , q_c (as assigned by the acquisitive convention) conflict with convenient verbal or graphical description of heat and work exchanges, we often take the liberty of using these as symbols for *magnitudes* only (invisible absolute-value bars), leaving it to descriptive words or arrow directions (as in Fig. 4.2) to depict the direction of transfer with respect to system and surroundings.

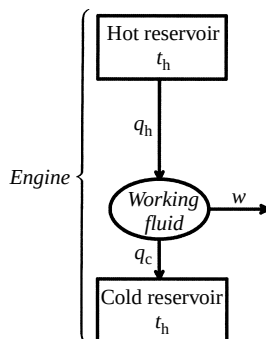


Figure 4.2 Schematic “engine” (as envisioned by Carnot), consisting of a working fluid that absorbs heat q_h from a hot reservoir (temperature t_h) and expels a smaller quantity of heat q_c to a cold reservoir (temperature t_c), with performance of work w ($= |q_h| - |q_c|$) on the surroundings. Arrows denote the magnitude and direction of heat or work flow.

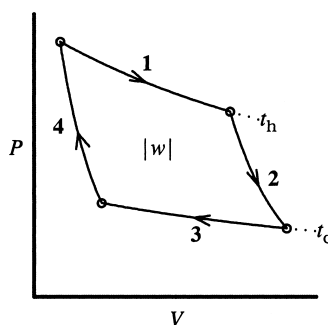


Figure 4.3 Reversible Carnot cycle, showing steps: (1) reversible isothermal expansion at t_h ; (2) reversible adiabatic expansion and cooling from t_h to t_c ; (3) reversible isothermal compression at t_c ; (4) reversible adiabatic compression and heating back to the original starting point. The total area of the Carnot cycle, $\oint P dV$, is the net useful work $|w|$ performed in the cyclic process (see text).

Carnot realized that the basic functions of the idealized engine of Fig. 4.2 might be achieved by subjecting the working fluid (e.g., a chosen gas) to a simple cyclic sequence of expansions and compressions under alternating isothermal and adiabatic conditions. Such a sequence, now known as a “Carnot cycle,” has the following four *reversible* steps 1–4, which are conveniently represented in a PV diagram (“indicator diagram”), as shown in Fig. 4.3:

1. The gas first undergoes reversible isothermal expansion at t_h , with energy change (as required by the first law)

$$\Delta U_1 = q_1 + w_1 \quad (4.4a)$$

where q_1 can be identified with q_h (the heat absorption from the high-temperature reservoir) in Fig. 4.2.

2. The gas is further expanded under reversible adiabatic conditions ($q_2 = 0$), with energy change

$$\Delta U_2 = w_2 \quad (4.4b)$$

and resultant adiabatic cooling from t_h to t_c .

3. The gas is now reversibly compressed under the isothermal conditions (t_c) of the low-temperature reservoir, with energy change

$$\Delta U_3 = q_3 + w_3 \quad (4.4c)$$

where the expelled heat q_3 is identified with the low-temperature q_c in Fig. 4.2.

4. Finally, the gas is reversibly compressed under adiabatic conditions ($q_4 = 0$),

$$\Delta U_4 = w_4 \quad (4.4d)$$

with resultant adiabatic heating from t_c to t_h .

Completion of these four steps returns the system to its initial state, ready for another cycle.

Let us now determine the overall heat, work, and energy changes in the reversible Carnot cycle. From (4.4a–d), we can evaluate the net work w as

$$w = w_1 + w_2 + w_3 + w_4 \quad (4.5)$$

It is easy to recognize that $|w| = \oint P dV$ is simply the total *area* enclosed by the Carnot cycle in the PV diagram, with greater area corresponding to greater useful work. (Indeed, the PV “indicator” diagram is so-called because it emulates the original “Watt’s indicator,” a mechanical device with a PV -sensitive needle that allowed the steam engineer to optimize the working conditions by visually maximizing the area swept out by the needle.) From the conservation of energy in a cyclic process, we also obtain

$$\Delta U = 0 = \Delta U_1 + \Delta U_2 + \Delta U_3 + \Delta U_4 = q_1 + q_3 + w = q_h + q_c + w \quad (4.6)$$

from which we conclude that

$$-w = q_h + q_c \quad (4.7)$$

The definition (4.1) then allows us to evaluate the efficiency $\mathcal{E}_{\text{Carnot}}$ of the reversible Carnot cycle as

$$\mathcal{E}_{\text{Carnot}} \equiv \frac{-w}{q_h} = \frac{q_h + q_c}{q_h} \quad (4.8)$$

or (since $q_c < 0$, $q_h > 0$),

$$\mathcal{E}_{\text{Carnot}} = 1 - \frac{|q_c|}{|q_h|} \quad (4.9)$$

Note that Carnot’s efficiency (4.9) is definitely more restrictive than the first-law efficiency (4.2b), constrained by the ratio of heat expelled (at t_c) to heat absorbed (at t_h).

Carnot's cycle of Fig. 4.3 is certainly a remarkably simple realization of a heat engine. Can any other engine do better? Surprisingly, Carnot concluded that the answer must be "No"! Carnot's general conclusion can be summarized in the following statement:

Carnot's principle:

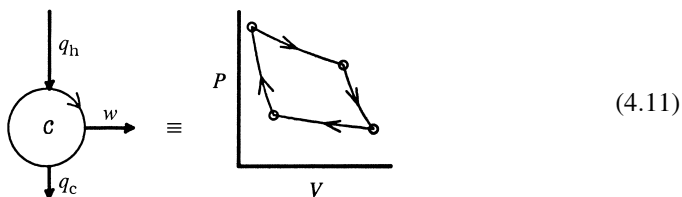
- (i) The efficiency of the reversible Carnot cycle is the *maximum* achievable for any engine operating between the same two temperatures t_h, t_c ;
- (ii) *All* reversible heat engines operating between the same two heat reservoirs have the *same* efficiency (independent of working fluid).

Carnot's principle can be summarized as an inequality that limits the possible efficiency $\mathcal{E}_{\text{real}}$ of any real engine:

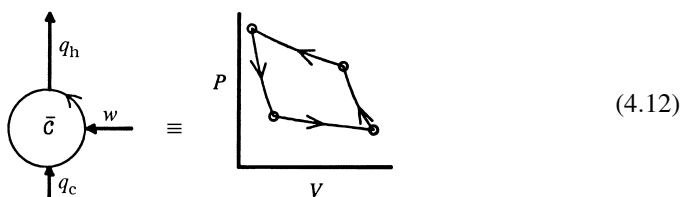
$$\mathcal{E}_{\text{real}} \leq 1 - \frac{|q_c|}{|q_h|} = \mathcal{E}_{\text{Carnot}} \quad (4.10)$$

The simple inequality (4.10) captures the essence of the second law. Its general consistency with universal inductive experience will be established in Section 4.4, and its further consequences (culminating in the final form of the second law as expressed by Clausius) will be developed in Sections 4.5–4.7. Thus, Carnot's remarkable principle provides virtually complete answers to the questions posed at the beginning of this chapter, although the relationship of (4.10) to these broader issues will certainly not become obvious until the following section.

To explore the rich consequences of Carnot's principle (4.10), let us begin by adopting the following alternative schematic representation of a Carnot cycle $\mathcal{C}$:



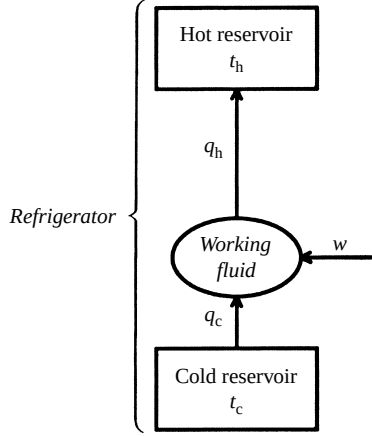
The clockwise direction in $\mathcal{C}$ corresponds to the clockwise direction in the Carnot cycle, with heat and work input/output as shown in Fig. 4.2. We can similarly envision a *reverse Carnot engine* ("heat pump") $\bar{\mathcal{C}}$, which is obtained by reversing the directions of heat and work arrows and traversing the Carnot cycle in counterclockwise direction:



$\bar{\mathcal{C}}$ corresponds to an ideal Carnot *refrigerator* (Sidebar 4.4), which removes heat q_c from the cold reservoir and expels heat q_h at the high-temperature reservoir, with input of work w from the surroundings.

SIDEBAR 4.4: CARNOT REFRIGERATOR

The reverse of a heat engine can be described as a “heat pump” or *refrigerator*, as shown in the following diagram:



As before, the essence of the refrigeration device is a working fluid (such as Freon or other “refrigerant”). In this case, the working fluid removes heat q_c from the low-temperature chamber t_c and expels a larger quantity of heat q_h to the external room surroundings t_h , with input of work $w (= |q_h| - |q_c|)$ from the compressor motor.

However, in this case the measure of refrigerator efficiency $\bar{\mathcal{E}}$ must be modified (because the function is different). Highest refrigeration efficiency is associated with greatest possible heat removal q_c from the cold chamber and smallest possible work input w , i.e., by the ratio

$$\bar{\mathcal{E}} \equiv \frac{\text{heat removed at } t_c}{\text{work input}} = \frac{q_c}{w} \quad (\text{S4.4-1})$$

(Note that q_c and w are both positive quantities according to the acquisitive convention, because “the system” is still the working fluid, not the cold chamber.)

Just as the Carnot cycle $\mathcal{C}$ of Fig. 4.3 can be claimed to be the most efficient possible heat engine ($\mathcal{E}_{\text{real}} \leq \mathcal{E}_{\text{Carnot}} = 1 - |q_c/q_h|$), so too can the reverse Carnot cycle $\bar{\mathcal{C}}$ be claimed to be the most efficient possible refrigerator:

$$\bar{\mathcal{E}}_{\text{real}} \leq \bar{\mathcal{E}}_{\text{Carnot}} = \frac{q_c}{w} = \frac{q_c}{-(q_h + q_c)} = \frac{1}{|q_h/q_c| - 1} \quad (\text{S4.4-2})$$

Thus, the development of the second law of thermodynamics might be based just as rigorously on the efficiency of refrigeration devices.

4.4 THEORETICAL LIMITS ON PERPETUAL MOTION: KELVIN'S AND CLAUSIUS' PRINCIPLES

Carnot's principle (4.10) may not seem particularly compelling from experience. However, we can easily derive some consequences from (4.10) that are indeed more obvious statements about the irreversibility of natural events, and hence provide compelling inductive proof of the truth of Carnot's principle. These derivative principles were first obtained by Thomson (Kelvin) and Clausius.

The derivations presented below illustrate the logical technique of “proof by contradiction.” In this method of proof, we begin by assuming that Carnot's principle is *untrue*, then demonstrate that we could easily produce “crazy” consequences that contradict experience if this assumption were valid. That is, we conclude that Carnot's principle must be true, because the contrary assumption leads to inconsistencies with inductive experience.

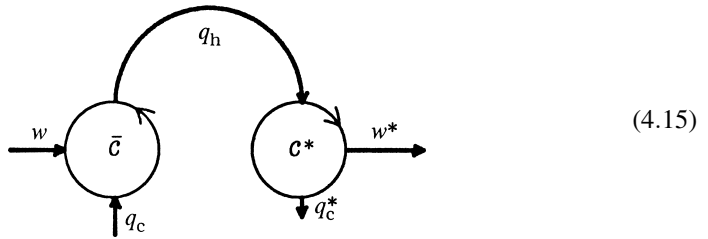
Let us therefore begin by assuming that Carnot's principle is *false*, i.e., that there exists some “new and improved model” $\mathcal{C}^*$ whose efficiency exceeds that of the reversible Carnot cycle. The hypothetical $\mathcal{C}^*$ engine can be represented as



with

$$\mathcal{E}^* = \frac{-w^*}{q_h^*} > \mathcal{E}_{\text{Carnot}} = \frac{-w}{q_h} \quad (4.14)$$

With the improved $\mathcal{C}^*$ in hand, we can now envision operating the old Carnot cycle as a heat pump $\bar{\mathcal{C}}$, then coupling this to $\mathcal{C}^*$ as shown in (4.15), using the heat output q_h from heat pump $\bar{\mathcal{C}}$, to drive the improved heat engine $\mathcal{C}^*$ (i.e., with $|q_h| = |q_h^*|$):



The composite $\bar{\mathcal{C}}\mathcal{C}^*$ machine (4.15) has astonishing properties! According to (4.14) (with $q_h = q_h^*$, now “hidden” within the machine), we must have net work *output* w^{**}

$$w^{**} = w + w^* = |w| - |w^*| < 0 \quad (4.16a)$$

which is exactly compensated by the net heat *input* q^{**} from the single remaining reservoir (t_c) of the surroundings,

$$q^{**} = q_c + q_c^* = |q_c| - |q_c^*| > 0 \quad (4.16b)$$

(because all components still satisfy the first law). Hence, the $\bar{C}C^*$ device has the miraculous property of removing heat $|q_c + q_c^*|$ from a reservoir and converting it *entirely* to useful work $|w + w^*|$, as represented by

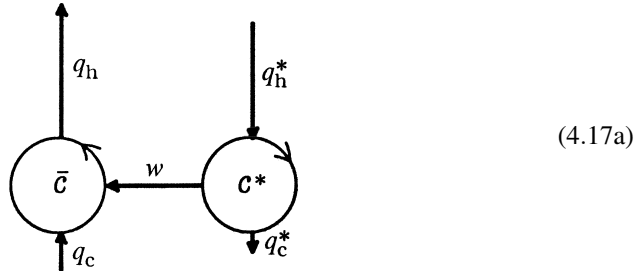


As concluded by Thomson, such a machine is impossible!

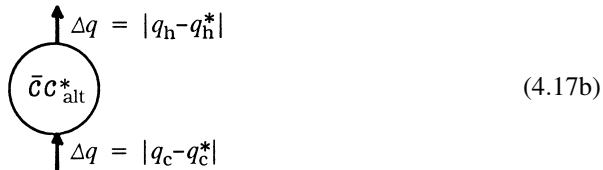
Principle of Thomson (Kelvin): *It is impossible to devise an engine that, working in a cycle, would produce no effect other than the extraction of heat from a reservoir and the performance of an equal amount of work (“perpetual motion of the second kind”).*

According to Thomson’s principle, no machine can convert heat to useful work unless *some* of that heat is transferred to a colder reservoir. This implies that *no* quantity of heat can be converted to work at a single temperature. It also implies that heat is a less useful form of energy, because some of it must always be “thrown away” by transfer to a lower temperature. Because the falsity of Carnot’s principle would imply easy violations of Thomson’s principle (which are never observed), we conclude that Carnot’s principle must be valid.

Still another contradiction with experience can be deduced from the assumed existence of any C^* device that falsifies Carnot’s principle. Let us again suppose that the old Carnot cycle is operated as a heat pump $\bar{C}$, now coupled to the improved C^* device through the work ($w = w^*$), as follows:



As can be seen, this alternative $\bar{C}C^*_{\text{alt}}$ coupling results in a composite device that can be represented as



The astonishing $\bar{C}C^*_{\text{alt}}$ device is seen to have the effect of transferring a quantity of heat Δq from a cold reservoir to a hot reservoir with no input of work or other

influence from the surroundings: heat runs uphill! As asserted by Clausius, such a device is impossible:

Principle of Clausius: It is impossible to devise an engine that, working in a cycle, shall produce no effect other than the transfer of heat from a colder to a hotter body.

Again we conclude that Carnot's principle must be true, because devices that contradict the principle of Clausius are never observed.

Let us conclude this section by quoting some alternative statements of the second law in colloquial form:

Second Law Statements

1. Carnot's cycle is the most efficient possible heat engine.
2. $\mathcal{E}_{\text{real}} \leq \frac{q_h + q_c}{q_h}$
3. No possible machine can convert heat to work at constant T without some compensating change in the surroundings.
4. Heat doesn't spontaneously flow "uphill."

These capture various aspects of the more general and comprehensive statements of Carnot, Clausius, and Gibbs that are still to follow.

4.5 KELVIN'S TEMPERATURE SCALE

Kelvin first suggested how the Carnot efficiency (4.9) might be used to define an "absolute" temperature scale. As Carnot's principle asserts, the efficiency

$$\mathcal{E}_{\text{rev}} = \mathcal{E}_{\text{Carnot}} = \frac{q_h + q_c}{q_h} = 1 - \left| \frac{q_c}{q_h} \right| \quad (4.18)$$

is characteristic of *any* reversible engine, independent of the working fluid. Hence, the only remaining physical properties of the device are the temperatures of the two heat reservoirs. Knowledge of the Carnot efficiency of a reversible heat engine must therefore be equivalent to knowledge of the temperature at which the engine is operating. In effect, if we ask the steam engineer "How efficiently are your best possible heat engines working today?", we can convert the answer into an assigned "Kelvin temperature" T^K .

According to Kelvin's definition, absolute Kelvin temperature T^K can be consistently assigned in terms of the efficiency of the reversible engine operating between the two reservoir temperatures as

$$\mathcal{E}_{\text{rev}} \equiv 1 - \frac{T_c^K}{T_h^K} \quad (4.19)$$

or, from comparison of (4.18) and (4.19),

$$\left| \frac{q_c}{q_h} \right| = \frac{T_c^K}{T_h^K} \quad (4.20)$$

Equation (4.19) could also be expressed as

$$T^K \equiv T_{\text{ref}}^K (1 - \mathcal{E}_{\text{rev}}) \quad (4.21)$$

where $T_{\text{ref}}^K = T_h^K$ is a single chosen fixed point of the Kelvin scale (e.g., the triple point of H_2O set to 273.16 K).

As defined by (4.19) or (4.21), it is easy to recognize that T^K is an “absolute” (strictly non-negative) quantity. Furthermore, one can see from (4.19) that the *highest* possible efficiency ($\mathcal{E} \rightarrow 1$) is achievable only at the absolute zero of the Kelvin scale ($T_c^K \rightarrow 0$). In addition, the *lowest* efficiency of converting heat to work ($\mathcal{E} \rightarrow 0$) occurs when the two reservoirs approach the same temperature ($T_h^K \rightarrow T_c^K$), consistent with the statement of Kelvin’s principle in Section 4.4. Such limits on engine efficiency can be used to paraphrase the three laws of thermodynamics in somewhat whimsical form as follows (the ultimate formulation of the “no free lunch” principle):

Paraphrase of the Three Laws of Thermodynamics

- (I) $\mathcal{E} \leq 1$: “You can’t get something for nothing; the best you can hope for is to break even.”
- (II) $\mathcal{E} \rightarrow 1$ only if $T \rightarrow 0$: “You can’t break even except at absolute zero.”
- (III) $T > 0$: “You can’t get to absolute zero!”

Carnot’s principle in Kelvin form (4.19) makes clear that the usefulness of a certain quantity of heat energy q depends on its temperature. Thus, a quantity of “high- T heat” intrinsically carries greater *work capacity* than the equivalent quantity of “low- T heat.” Even if the first law tells us that a quantity of heat q and work w are energetically equivalent, the second law restricts what *fraction* can actually be extracted from q as useful work, depending on its temperature.

Introduction of the Kelvin temperature scale T^K may seem to be in conflict with the internationally accepted ideal temperature scale T^I , as defined in Section 2.3. Satisfyingly, this is not so, as proved in Sidebar 4.5. The fact that these two temperature scales are *identical* serves as a second theoretical justification for adopting $T^K = T^I = T$ as the “true” temperature scale, and allows us to rewrite (4.19) in simpler form

$$\mathcal{E}_{\text{rev}} = 1 - \frac{T_c}{T_h} \quad (4.22)$$

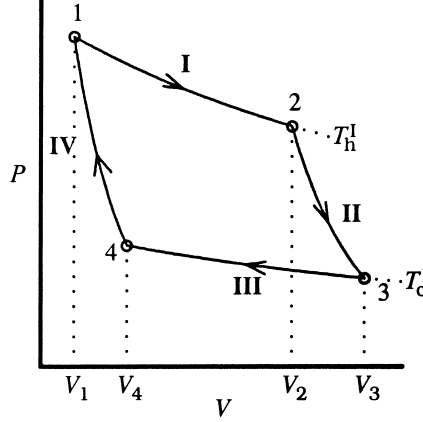
without qualifying superscripts. Sidebar 4.6 illustrates a simple use of the Kelvin relationship (4.22) in solving problems.

SIDEBAR 4.5: EQUIVALENCE OF KELVIN AND IDEAL TEMPERATURE SCALES

Claim The Kelvin scale T^K as defined by (4.19) *agrees* with the “ideal” scale T^I based on the ideal gas limit.

Proof Because the Carnot efficiency is independent of working fluid, we can assume one mole of an ideal gas for this purpose. Let us label the four stopping points of the

cycle as 1, ..., 4 (with volumes $V_1, \dots, V_4$) and the four steps as **I**, ..., **IV**, as shown in the figure below. The temperatures for the isothermal steps **I** and **III** are labeled by the respective values T_h^I, T_c^I on the standard ideal scale.



Let us begin by proving a simple identity between the ideal gas volumes $V_1, \dots, V_4$. For adiabatic step **II** ($2 \rightarrow 3$), the adiabatic equation of state (3.86) requires that $P_2 V_2^\gamma = P_3 V_3^\gamma$, while for adiabatic step **IV** ($4 \rightarrow 1$) the corresponding relationship is $P_1 V_1^\gamma = P_4 V_4^\gamma$. Dividing the latter equation by the former gives the relationship

$$\frac{P_1}{P_2} \left(\frac{V_1}{V_2} \right)^\gamma = \frac{P_4}{P_3} \left(\frac{V_4}{V_3} \right)^\gamma \quad (\text{S4.5-1})$$

For isothermal step **I** ($1 \rightarrow 2$), we can infer $P_1/P_2 = V_2/V_1$, and for isothermal step **III** ($3 \rightarrow 4$), we similarly infer $P_4/P_3 = V_3/V_4$. When these ratios are substituted into the above equation to give a relationship between volumes only, we obtain the desired identity

$$\frac{V_2}{V_1} = \frac{V_3}{V_4} \quad (\text{S4.5-2})$$

Let us now return to the main claim by evaluating the work and heat in each step **I**, ..., **IV**:

I. For isothermal expansion $1 \rightarrow 2$ of an ideal gas ($\Delta U_I = 0$), (3.77) gives

$$-w_I = q_I = RT_h^I \ln(V_2/V_1) \quad (\text{S4.5-3})$$

II. For adiabatic expansion $2 \rightarrow 3$ ($q_{II} = 0$), (3.79) gives

$$\Delta U_{II} = w_{II} = \int_{T_h^I}^{T_c^I} C_V dT \quad (\text{S4.5-4})$$

III. For isothermal compression $3 \rightarrow 4$, (3.77) gives

$$-w_{III} = q_{III} = RT_c^I \ln(V_4/V_3) = -RT_c^I \ln(V_2/V_1) \quad (\text{S4.5-5})$$

where we have used the identity $V_2/V_1 = V_3/V_4$ to rewrite the final form.

IV. For adiabatic compression $4 \rightarrow 1$, (3.79) gives

$$\Delta U_{IV} = w_{IV} = \int_{T_c^I}^{T_h^I} C_V dT = -w_{II} \quad (\text{S4.5-6})$$

From the sum of the work terms, we obtain

$$w = w_I + w_{II} + w_{III} + w_{IV} = -R(T_h^I - T_c^I) \ln(V_2/V_1) \quad (\text{S4.5-7})$$

and from the initial isothermal step I, we obtain

$$q_I = q_h = RT_h^I \ln(V_2/V_1) \quad (\text{S4.5-8})$$

The efficiency $\mathcal{E}_{\text{Carnot}}$ can therefore be evaluated as

$$\mathcal{E}_{\text{Carnot}} = \frac{-w}{q_h} = \frac{R(T_h^I - T_c^I) \ln(V_2/V_1)}{RT_h^I \ln(V_2/V_1)} = \frac{T_h^I - T_c^I}{T_h^I} = 1 - \frac{T_c^I}{T_h^I} \quad (\text{S4.5-9})$$

Comparison with (4.19) shows that this can only be true if T^I and T^K values are *identical* (assuming their match at a single “reference” point to insure the same degree size):

$$T^I = T^K = T \quad (\text{S4.5-10})$$

QED

SIDEBAR 4.6: T-DEPENDENT ENGINE PERFORMANCE

Problem A steam engine operates with a boiler temperature of 200°C . What is the maximum theoretical efficiency of the engine if operated on a warm summer day ($+30^\circ\text{C}$) compared with a cold winter day (-30°C)? How much more fuel will be required to perform the same work in the former case?

Solution With boiler temperature $T_h = 200^\circ\text{C} = 473\text{K}$, the maximum (reversible) efficiency $\mathcal{E}_{\text{warm}}$ on a warm day with $T_c = 30^\circ\text{C} = 303\text{K}$ is determined from (4.22) to be

$$\mathcal{E}_{\text{warm}} = 1 - \frac{303}{473} = 0.36$$

whereas the corresponding $\mathcal{E}_{\text{cold}}$ on a cold day with $T_c = -30^\circ\text{C} = 243\text{K}$ is

$$\mathcal{E}_{\text{cold}} = 1 - \frac{243}{473} = 0.49$$

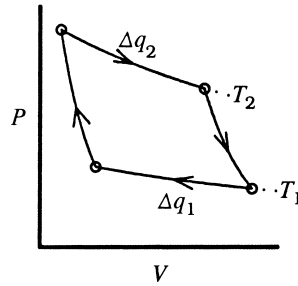
Thus, to perform equivalent work at the reduced efficiency $\mathcal{E}_{\text{warm}}$, the steam engineer must provide additional boiler heat q_h , corresponding to increased fuel costs of

$$\frac{0.49 - 0.36}{0.36}(100) = 36\%$$

While the actual efficiency of any real steam engine falls below the theoretical optimum values calculated above, the improved fuel economy of operating the engine with higher boiler temperature T_h or reduced exhaust temperature T_c is well known to steam engineers.

4.6 CARNOT'S THEOREM AND THE ENTROPY OF CLAUSIUS

We can redraw the Carnot cycle in a slightly simplified form:



to emphasize the fact that the only heat exchanges (Δq_1 , Δq_2) occur on the two isothermal limbs of the cycle, at respective temperatures T_1 , T_2 . The efficiency of this reversible Carnot cycle is

$$\mathcal{E}_{\text{rev}} = 1 + \frac{\Delta q_1}{\Delta q_2} = 1 - \frac{T_1}{T_2} \quad (4.23)$$

which can be rewritten as

$$\frac{\Delta q_1}{T_1} + \frac{\Delta q_2}{T_2} = 0 \quad (4.24a)$$

Without loss of generality, this summation could be extended to all four limbs ($i = 1, \dots, 4$) of the Carnot cycle (because adiabatic steps contribute nothing to the sum),

$$\sum_{i=1}^4 \frac{\Delta q_i}{T_i} = 0 \quad (4.24b)$$

or, in integrated form,

$$\oint_{\text{Carnot}} \frac{dq_{\text{rev}}}{T} = 0 \quad (4.24c)$$

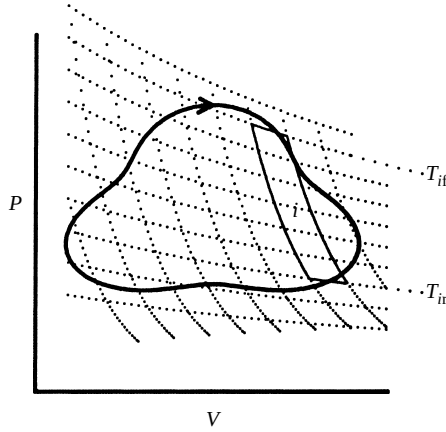
It was recognized by Carnot that cyclic integrals of the form (4.24c) must actually vanish for *all* reversible cycles:

$$\text{Carnot's Theorem: } \oint \frac{dq_{\text{rev}}}{T} = 0 \quad \text{for any reversible cycle} \quad (4.25)$$

The proof of this important theorem is sketched in Sidebar 4.7.

SIDEBAR 4.7: PROOF OF CARNOT'S THEOREM

Given the validity of (4.25b) for any Carnot cycle, we can extend the result to a *general* cyclic path, as shown in the following diagram:



As shown in the diagram, we can embed the path (heavy solid line) in a *grid* of isotherms and adiabats (dotted lines), dividing the path area into a grid of mini-Carnot cycles of infinitesimal width. The *i*th mini-Carnot cycle (light solid line) is shown delimited by isotherms T_{if} (forward direction) and T_{ir} (reverse direction), with associated differential heat exchanges dq_{if} , dq_{ir} satisfying

$$\text{ith cycle : } \frac{dq_{if}}{T_{if}} + \frac{dq_{ir}}{T_{ir}} = 0 \quad (S4.7-1)$$

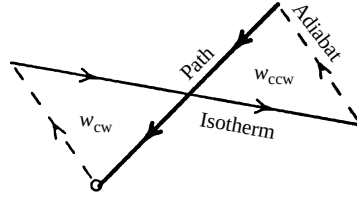
as established in (4.24a). Because the adiabats give no contributions to the heat exchange with the surroundings, the issue is whether we can sufficiently approximate the path integral of dq_{rev}/T around the actual path in terms of the sum over the grid of mini-Carnot cycles, i.e., whether

$$\oint \frac{dq_{\text{rev}}}{T} \stackrel{?}{=} \sum_i \left(\frac{dq_{if}}{T_{if}} + \frac{dq_{ir}}{T_{ir}} \right) \quad (S4.7-2)$$

We can achieve the desired equality along each small path segment (with reversible heat exchange dq_{path}) by choosing isotherm T_i of the mini-Carnot cycle i (with reversible heat exchange dq_{isotherm}) such that

$$dq_{\text{path}} \stackrel{?}{=} dq_{\text{isotherm}} \quad (\text{S4.7-3})$$

To achieve this goal (and thus establish the theorem), we shall choose the isotherm T_i that *divides* the infinitesimal path segment, with flanking adiabats chosen such that the path also divides the isotherm of mini-Carnot cycle i . In the limit when path segment and grid spacings become infinitesimal, the intersecting path segment and mini-isotherm can be approximated as straight lines (up to second order), as shown in the following diagram (approximating the “return” segment at T_{ir} in the diagram above):



Let us now consider the thermodynamic mini-cycle that follows the arrows in a figure-of-eight loop along each adiabat (dashed line), isotherm (thin solid line), and path segment (heavy solid line) of the intersection region, returning to a starting point (circle) along the path. For this mini-cycle, we must have (by the first law)

$$\Delta U_{\text{mini}} = \Delta q_{\text{mini}} + \Delta w_{\text{mini}} = 0 \quad (\text{S4.7-4})$$

By construction, the “clockwise” and “counter-clockwise” triangles are congruent, so the work performed in the clockwise loop (w_{cw}) must be equal in magnitude and opposite in sign to that in the counter-clockwise loop (w_{ccw}), leading to zero net work

$$\Delta w_{\text{mini}} = w_{\text{cw}} + w_{\text{ccw}} = 0 \quad (\text{S4.7-5})$$

This implies that the remaining term, Δq_{mini} , is also vanishing (to second order),

$$\Delta q_{\text{mini}} = dq_{\text{path}} - dq_{\text{isotherm}} = 0 \quad (\text{S4.7-6})$$

which establishes the matching condition

$$dq_{\text{path}} = dq_{\text{isotherm}} \quad (\text{S4.7-7})$$

and proves the desired result—*QED*.

It was initially appreciated by R. Clausius that Carnot’s theorem (4.25) allows the second law to be reformulated in a profoundly improved form. Clausius recognized that (4.25) is nothing more than the *exactness condition* (1.16a) for the differential dq_{rev}/T , i.e., that $L = 1/T$ is an *integrating factor* for the inexact differential $\bar{d}q_{\text{rev}}$ [cf. (1.22)]. Accordingly, we can conclude that dq_{rev}/T is the differential of a *state property*, a conserved quantity that

evidently plays a central thermodynamic role. Clausius proposed the name “entropy” for this state property (from the Greek $\tau\rho\epsilon\pi\iota\nu$, “to give a direction”), with symbol S :

$$dS \equiv \frac{dq_{\text{rev}}}{T} \quad (4.26)$$

Clausius’ recognition of the thermodynamic entropy S was a crucial watershed in thermodynamic theory, allowing the subsequent development of the theory to proceed in a far more accurate and coherent manner.

Let us employ (4.26) to rewrite the first law in improved form. Under reversible conditions, where PV work can be expressed as $dw_{\text{rev}} = -P dV$ [cf. (3.12)], we can now use (4.26) to write dq_{rev} correspondingly as

$$dq_{\text{rev}} = T dS \quad (4.27)$$

Note that S is an *extensive* property, so that (4.27) conforms to the general pattern expected from the parallel work contributions to the first law [cf. (3.34), with $R_{\text{thermal}} = T$, $X_{\text{thermal}} = S$]. With (4.27), we can rewrite (3.37b) as

$$dU = dq_{\text{rev}} + dw_{\text{rev}} = T dS - P dV \quad (4.28)$$

which can be called the “combined first and second law” (for closed systems), expressed entirely in terms of state properties.

The entropy function S immediately *simplifies* thermodynamic theory in important respects. From (4.28), we can recognize that S and V are the basic variables of the energy function U ,

$$U = U(S, V) \quad (4.29)$$

which allow U to be *graphed* [cf. (1.16b)], as required by its conserved property. The three basic variables (U, S, V) of (4.29) span “Gibbs space,” which allows thermodynamic relationships to be represented in a particularly incisive graphical form (see Fig. 1.1).

The entropy function S also simplifies the graphical depiction of the Carnot cycle. Consider, for example, the form of the Carnot cycle shown in the PV diagram of Fig. 4.4a. The corresponding ST diagram for the same Carnot cycle is shown in Fig. 4.4b. As can be seen, the ST representation of the Carnot cycle is a simple rectangle whose

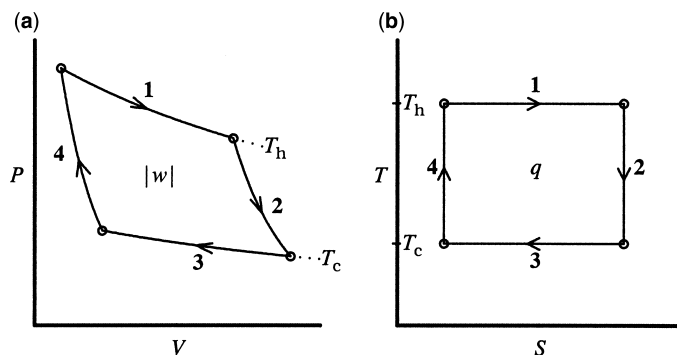


Figure 4.4 Reversible Carnot cycle, shown in (a) PV diagram, (b) ST diagram.

area ($q = \oint T dS$) is easily evaluated, whereas the equivalent area in the PV representation ($|w| = \oint P dV = q$, by the first law) involves a more difficult calculation.

Let us briefly summarize the mathematical formalism of thermodynamics to this point. A mathematical formalism begins with identification of the proper variables. We began with the standard *mechanical* variables of classical mechanics: P , V , and N (the quantity of matter, held fixed for the present). To these have now been added three important *thermal* variables:

T (defined by the zeroth law)

U (defined by the first law)

S (defined by the second law)

All these variables qualify as *state* properties of a thermodynamic system. From the fundamental relationship (4.29), we infer from the chain rule that

$$dU = \left(\frac{\partial U}{\partial S} \right)_V dS + \left(\frac{\partial U}{\partial V} \right)_S dV \quad (4.30)$$

Comparison of the combined first/second law (4.28) with (4.30) leads to the more general and rigorous thermodynamic definitions for the intensive properties T , $-P$ respectively “conjugate” to the extensive properties S , V :

$$T \equiv \left(\frac{\partial U}{\partial S} \right)_V \quad (4.31a)$$

$$-P \equiv \left(\frac{\partial U}{\partial V} \right)_S \quad (4.31b)$$

valid for closed systems where only PV work is allowed.

More generally, we can recognize that the “natural” arguments of the internal energy function U are a fundamental set of extensive properties $X_1, X_2, \dots, X_i, \dots$ (whose number remains to be established),

$$U = U(X_1, X_2, \dots, X_i, \dots) \quad (4.32)$$

From the fundamental functional relationship (4.32), one obtains the *conjugate* intensive property R_i of each extensive X_i as the partial derivative

$$R_i \equiv \left(\frac{\partial U}{\partial X_i} \right)_{X_{j(\neq i)}} \quad (4.33)$$

The total differential of U can therefore be expressed as

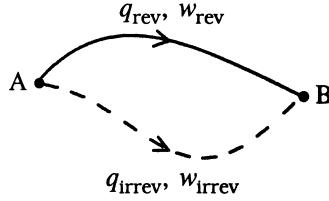
$$dU = \sum_i R_i dX_i \quad (4.34)$$

The conjugacy pattern was previously recognized for general work forms (Section 3.3.7), but is now seen to extend to *all* contributions to internal energy. Equations (4.32)–(4.34) generalize the corresponding equations of classical dynamics to incorporate the “thermo-”dynamic effects of friction and irreversibility (Sections 2.8–2.9), with internal energy U playing the role of the potential energy function $\mathcal{V}$. With Clausius’ recognition of the entropy, the long-sought extension of mechanics to the domain of thermal phenomena was successfully achieved.

4.7 CLAUSIUS' FORMULATION OF THE SECOND LAW

Clausius proceeded to demonstrate the power of entropy to express the deep consequences of the second law. We begin by introducing the *inequality of Clausius*, which complements Carnot's theorem (4.25) for the irreversible case.

For this purpose, consider a change of state $A \rightarrow B$, which can be achieved by either a reversible path (with heat q_{rev} and work w_{rev}) or an irreversible path (with heat q_{irrev} and work w_{irrev}), as shown schematically below:



As shown in Section 3.2, the reversible path yields *maximum* useful work:

$$-w_{\text{rev}} < -w_{\text{irrev}} \quad (4.35)$$

From the first law, the internal energy change $\Delta U(A \rightarrow B)$ must be independent of the path between states:

$$\Delta U_{\text{rev}} = q_{\text{rev}} + w_{\text{rev}} = \Delta U_{\text{irrev}}(A \rightarrow B) = q_{\text{irrev}} + w_{\text{irrev}} \quad (4.36)$$

From (4.35) and (4.36), we therefore conclude that, for any $A \rightarrow B$,

$$q_{\text{rev}} > q_{\text{irrev}} \quad (4.37a)$$

or, if $A \rightarrow B$ is a differential process,

$$dq_{\text{rev}} > dq_{\text{irrev}} \quad (4.37b)$$

The sense of this inequality is preserved if we divide both sides by $T (> 0)$:

$$\frac{dq_{\text{rev}}}{T} > \frac{dq_{\text{irrev}}}{T} \quad (4.38)^a$$

If we now integrate both sides of inequality (4.38) over any cyclic path, we obtain

$$\oint \frac{dq_{\text{rev}}}{T} > \oint \frac{dq_{\text{irrev}}}{T}, \quad \text{any cyclic path} \quad (4.39)$$

However, we recognize from Carnot's theorem (4.25) that the left-hand-side of this inequality is *zero*, from which we conclude:

$$\text{Inequality of Clausius: } \oint \frac{dq_{\text{irrev}}}{T} < 0$$

(4.40)

^aAs always, we should be prepared to define the discretized quantities of (4.38) in *operational* terms by employing, for example, a sequence of “stop-flow” operations (analogous to those of Fig. 3.1) to obtain the T for each incremental dq_{irrev} along the chosen pathway, thereby allowing dq_{rev} to be obtained for alternative *reversible* isothermal passage between the same two stopping points.

The inequality (4.40) merely expresses the fact that an irreversible process is always *dissipative*, evolving heat to the surroundings and *losing* some of the useful work that was potentially available.

By combining (4.25) and (4.40), we can write more generally

$$\oint \frac{dq}{T} \leq 0 \quad \begin{cases} < \text{irrev} \\ = \text{rev} \end{cases} \quad (4.41)$$

where the inequality refers to an irreversible (spontaneous) process and the equality to a reversible process. From the definition of entropy (4.26), we can also rewrite the inequality (4.38) as

$$dS > \frac{dq_{\text{irrev}}}{T} \quad (4.42a)$$

or more generally as

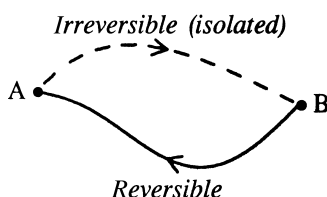
$$dS \geq \frac{dq}{T} \quad \begin{cases} > \text{irrev} \\ = \text{rev} \end{cases} \quad (4.42b)$$

For future reference, we also rewrite (4.42b) in the alternative form

$$dq \leq T dS \quad (4.43)$$

with the understanding that equality is reserved for the reversible (equilibrium) limit, whereas the inequality expresses the spontaneous (irreversible) relationship.

Let us now consider the most celebrated statement of the second law. For this purpose, consider the following cyclic process:



In this process, the system undergoes cyclic changes between two equilibrium states A, B:

1. The system is first allowed to undergo a *spontaneous* (irreversible) change $A \rightarrow B$ under *isolated* conditions (i.e., in which no heat, work, or other forms of energy are exchanged with the surroundings; Section 2.10).
2. The system is then returned ($B \rightarrow A$) to its original starting point under *reversible* conditions (i.e., carefully controlled by suitable coupling to work or heat reservoirs in the environment).

Because this cycle includes an irreversible step (1), we know from the Clausius inequality (4.40) that

$$\oint_{A \rightarrow B} \frac{dq}{T} < 0 \quad (4.44)$$

We now divide the cyclic integral (4.44) into its outbound irreversible ($A \rightarrow B$) and inbound reversible ($B \rightarrow A$) steps:

$$0 > \int_A^B \frac{dq_{\text{irrev(iso)}}}{T} + \int_B^A \frac{dq_{\text{rev}}}{T} \quad (4.45)$$

The first term in (4.45) is *zero*, because no heat exchange with the surroundings occurs under isolated conditions ($dq_{\text{irrev(iso)}} = 0$):

$$\int_A^B \frac{dq_{\text{irrev(iso)}}}{T} = 0 \quad (4.46)$$

The second term in (4.45) is merely $S_A - S_B$, the negative of the entropy change $\Delta S_{A \rightarrow B} = \Delta S_{\text{irrev(iso)}}$ in the spontaneous step from A to B:

$$\int_A^B \frac{dq_{\text{rev}}}{T} = S_A - S_B = -\Delta S_{A \rightarrow B} = -\Delta S_{\text{irrev(iso)}} \quad (4.47)$$

(Recall that entropy is a state property, so $\Delta S_{A \rightarrow B}$ is independent of whether a reversible or irreversible path was followed.) From (4.45)–(4.47), we conclude finally that

$$\boxed{\Delta S_{\text{irrev(iso)}} > 0} \quad (4.48)$$

i.e., that entropy *increases* in any spontaneous process in an isolated system. Sidebar 4.8 illustrates this entropy increase for two specific spontaneous processes: spontaneous heat flow through a temperature difference and spontaneous volume increase through a pressure difference.

SIDEBAR 4.8: TWO SPONTANEOUS PROCESSES IN ISOLATED SYSTEMS

Let us consider two specific spontaneous processes $A \rightarrow B$:

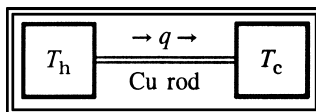
- (i) spontaneous heat flow from a hotter (T_h) to a colder (T_c) temperature
- (ii) spontaneous volume flow from a higher (P_h) to a lower (P_l) pressure

both of which are allowed to occur under conditions of total isolation from the surroundings (i.e., no exchange of heat, work, or other form of energy with the surroundings). In each case, the associated entropy change $\Delta S_{A \rightarrow B}$ will be evaluated by performing the *same* change of state under reversible conditions (i.e., by coupling the original system to suitable heat or work reservoirs that maintain the system in equilibrium at each infinitesimal step).

- (i) **Spontaneous Heat Flow** We consider the transfer of a quantity of heat q from a hot reservoir (T_h) to a cold reservoir (T_c), first (a) by the spontaneous (irreversible) process in isolation, then (b) by a controlled reversible process. The two temperature

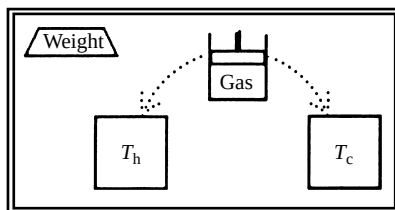
reservoirs are assumed very large (“infinite heat capacity”) so that their temperatures are unaffected by the small heat transfer q .

- (a) *Irreversible Heat Transfer in Isolation.* As shown in the following diagram, the system consists of two reservoirs in isolation, initially in equilibrium state A:



The reservoirs are brought into thermal contact (e.g., by a connecting copper rod), allowing heat q to flow irreversibly from T_h to T_c . The contact is removed and the reservoirs allowed to re-equilibrate to final state B. What is $\Delta S_{A \rightarrow B} = \Delta S_{\text{reservoirs}}$?

- (b) *Reversible Heat Transfer.* To answer the above question, we must transfer the same quantity of heat q by a reversible process. Consider therefore the enlarged system shown in the following diagram:



This includes an ideal gas in a piston, weights, and other necessary paraphernalia to carry out the following reversible operations. First, the gas piston is brought into thermal contact with the T_h reservoir, and the gas undergoes *reversible isothermal expansion* at T_h (lifting weights as necessary) to withdraw heat q from this reservoir, where

$$q = q_{\text{rev}} = T_h \Delta S_h$$

and ΔS_h is the entropy change of the T_h reservoir. Next, the piston is detached from the T_h reservoir and a *reversible adiabatic expansion* is performed to cool the gas to T_c . Finally, the gas piston is brought into thermal contact with the T_c reservoir, and the gas is subjected to *reversible isothermal compression* at T_c (lowering the weights as necessary) to deposit heat q into this reservoir, where

$$q = q_{\text{rev}} = T_c \Delta S_c$$

and ΔS_c is the entropy change in the T_c reservoir. The overall entropy change for the system (reservoirs) is therefore

$$\Delta S_{\text{reservoirs}} = \Delta S_h + \Delta S_c = -\frac{|q|}{T_h} + \frac{|q|}{T_c} > 0$$

since $T_h > T_c$. Thus, we can conclude that

$$\Delta S_{A \rightarrow B} > 0$$

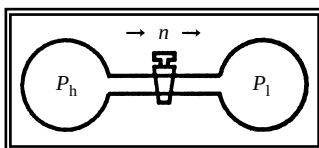
i.e., the entropy of the system *increased* in the isolated spontaneous process. (In the reversible process, the *total* entropy change was

$$\Delta S_{\text{total}} = \Delta S_{\text{reservoirs}} + \Delta S_{\text{gas}} = 0$$

but only the term $\Delta S_{\text{reservoirs}}$ corresponds to $\Delta S_{A \rightarrow B}$ for the irreversible step.)

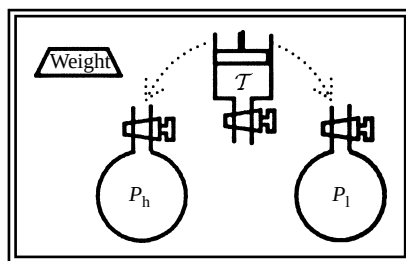
(ii) **Spontaneous Volume Flow** We consider the isothermal transfer of n moles of ideal gas from a high-pressure (P_h) to a low-pressure (P_l) reservoir, both reservoirs being so large as to experience no appreciable pressure change from the transfer.

(a) *Irreversible Volume Transfer in Isolation.* As shown in the following diagram, the reservoirs are initially in equilibrium state A, connected by a closed stopcock:



The stopcock is opened, allowing spontaneous transfer of n moles of gas from the P_h to the P_l reservoir. The stopcock is then closed and the reservoirs are allowed to re-equilibrate to final state B. What is $\Delta S_{A \rightarrow B} = \Delta S_{\text{reservoirs}}$?

(b) *Reversible Volume Transfer.* To answer the question, we must transfer the same quantity of gas by a reversible process. Consider therefore the enlarged system shown in the following diagram:



This includes a “transporter” piston T , weights, etc., to carry out the following reversible operations, all under isothermal conditions. Attach the empty transporter piston T to the high-pressure P_h reservoir. Reversibly remove an aliquot of n moles into T at constant P_h (doing no work), up to volume V_h . Detach T from the P_h reservoir and allow *reversible isothermal expansion* of the transporter gas from V_h to V_l , performing work

$$-w_{\text{rev}} = nRT \ln(V_l/V_h) = nRT \ln(P_h/P_l)$$

Now attach T to the P_l reservoir and reversibly transfer the aliquot of n moles of gas at constant P_l (with no work) into the reservoir to reach the final equilibrium state B of the reservoirs.

Because the entire process was performed isothermally with ideal gases, the internal energy change (whether for the transporter gas or for the reservoirs as a

whole) is zero:

$$\Delta U_{\text{total}} = q_{\text{total}} + w_{\text{total}}$$

where $w_{\text{total}} = w_{\text{rev}}$ is given above. Since $P_h > P_l$ (and n , R , T are all positive constants), we may conclude that $q_{\text{rev}} = -w_{\text{rev}} > 0$,

$$q_{\text{total}} = q_{\text{rev}} = T\Delta S_{\text{reservoirs}} = -w_{\text{rev}} = nRT \ln(P_h/P_l) > 0$$

and therefore that

$$\Delta S_{\text{reservoirs}} = \Delta S_{A \rightarrow B} > 0$$

which establishes that entropy *increased* in the original isolated spontaneous process.

The general inequality (4.48) leads to the famous Clausius formulation of the second law:

The second law of thermodynamics (Clausius' formulation): In isolated systems, spontaneous changes are always accompanied by a net *increase* in entropy.

Because the *universe* (perforce lacking “surroundings”) may be considered the ultimate “isolated system,” Clausius himself re-expressed the first and second laws in the famous couplet:

Die Energie der Welt ist constant. Die Entropie der Welt strebt einem Maximum zu.
(The energy of the universe is constant. The entropy of the universe strives toward a maximum.)

The latter statement evokes the image of inexorable entropy increase as the ultimate “progress variable” of the universe. Entropy presumably evolves toward an eventual equilibrium limit that marks the end of spontaneous change in our universe: “heat death” (*Wärmetod*). Sidebar 4.9 warns against common conceptual errors that result from superficial application of the entropy-increase principle (4.48).

SIDEBAR 4.9: COMMON MISCONCEPTIONS

Problem Explain the purported “violation of the second law” in each of the following statements:

1. A dog dish full of water is placed outside and left undisturbed on a cold Wisconsin night. The water spontaneously freezes. Clearly, the final state (ordered crystalline ice) has lower entropy than the initial state (disordered liquid water), so $\Delta S < 0$, proving that the second law is invalid.

2. A pot of water, heated gradually on a stove, is seen to develop highly structured convection patterns (“Bénard cells”), resembling a checkerboard of ascending and descending columns. Clearly, such “order out of chaos” represents a process with $\Delta S < 0$, proving that the second law is invalid.
3. According to current theories, the Earth’s original state (disordered primordial ooze) has evolved spontaneously to ever more elaborate and organized life forms (ordered). Clearly, the entropy of the Earth has diminished, $\Delta S < 0$, proving that the second law is invalid.

Solution

1. The dog dish is not an isolated system; hence, there is no violation.
 2. The pot of water is not an isolated system; hence, there is no violation.
 3. The Earth is not an isolated system; hence, there is no violation.
-

4.8 SUMMARY OF THE INDUCTIVE BASIS OF THERMODYNAMICS

The second law represents the final entry to the list of inductive laws 1–6 (Table 2.1) that constitute the basis of the formal theory of equilibrium thermodynamics. All further thermodynamic relationships to be derived in this book rest on this inductive basis

It should be noted that the concept of first and second laws (*Hauptsätze*) was introduced into thermodynamic theory by Clausius in a rather belated and ad hoc manner. The designation is chronologically scrambled, inasmuch as Carnot’s formulation of the second law preceded full recognition of the first law by about a quarter-century. The designation is also unsatisfactory if it is understood to represent logical precedence in an axiomatic formulation (analogous, for example, to Newton’s three laws of dynamics or Euclid’s five axioms of geometry). Indeed, the logical necessity of other preceding laws is explicitly recognized in the “zeroth law” [R. H. Fowler and E. A. Guggenheim. *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1939): “Thermal equilibrium is a transitive relationship”] or the “minus first law” of Brown and Uffink [H. R. Brown and J. Uffink. *Stud. Hist. Phil. Mod. Phys.* **32B**, 525–38 (2001): “Thermal equilibration happens”]. The logical situation was further confused by later introduction of the “third law,” which apparently has no authentic role in the inductive or axiomatic foundations of equilibrium thermodynamics (Sections 5.8.2 and 11.8).

In an important sense, the familiar statements of thermodynamic laws (as they evolved from the work of Carnot, Thomson, Clausius, and others) are ill-suited either for formal axiomatization or for practical recognition of thermodynamic relationships in heterogeneous chemical systems. That recognition was achieved by J. W. Gibbs through an inspired change of perspective. Whereas Clausius and his scientific forebears formulated thermodynamic principles in terms of cycles, engine efficiency, limits on perpetual motion, irreversibility, and other aspects of time-dependent (and time-ordered) nonequilibrium *changes* of state, Gibbs realized that the central conceptual object of theoretical thermodynamics is the thermal *equilibrium state*, divested of all time-like considerations. Gibbs therefore focused attention firmly on the analytical properties of a *single* equilibrium state, extracting from Clausius and pre-Clausius formulations only those inferences that were applicable to such *t*-independent limits.

By removing the excess baggage of presumed t -dependent dynamical, nonequilibrium “events” that give rise to equilibrium states, Gibbs opened the door to quantitative description of a vast array of chemical and phase phenomena. The Gibbs perspective sharply delineated the boundary between equilibrium thermodynamics (the subject of this book) and nonequilibrium thermodynamics (still a scientific frontier land, beleaguered by difficult questions of ergodicity, time-reversibility of dynamical laws, and the deeper physical and physiological significance of “time” variations). In separating thermodynamics from any presumed underlying dynamical theory of heat and frictional phenomena, Gibbs showed how great progress could be achieved despite the pathologically inadequate (pre-quantal) picture of molecular dynamical interactions that was then current. While Clausius himself seemed to show little interest in this development, his arch-rival Maxwell clearly recognized the transcendent power of Gibbs’ methods. Hence, we shall hardly hear any further mention in this book of the pre-Gibbsian pioneers through whom the inductive base of equilibrium thermodynamics was originally achieved. Instead, based on the inductive fact that “equilibration happens” (i.e., that there exist macroscopic equilibrium states satisfying the operational definitions of Section 2.10), we shall now embark in the new direction set by Gibbs.

GIBBSIAN THERMODYNAMICS OF CHEMICAL AND PHASE EQUILIBRIA

Analytical Criteria for Thermodynamic Equilibrium

5.1 THE GIBBS PERSPECTIVE

The Clausius statement of the second law, although logically able to serve as a basis for the general equilibrium theory, was couched in terms of *nonequilibrium* processes that themselves lay outside the scope of such a theory. Attempts to derive the consequences of the Clausius statement were therefore tortuous and indirect, making further progress difficult.

It was the principal genius of J. W. Gibbs (Sidebar 5.1) to recognize how the Clausius statement could be recast in a form that made reference only to the analytical properties of *individual equilibrium states*. The essence of the Clausius statement is that an isolated system, in evolving toward a state of thermodynamic equilibrium, undergoes a steady increase in the value of the entropy function. Gibbs recognized that, as a consequence of this increase, the entropy function in the eventual equilibrium state must have *the character of a mathematical maximum*. As a consequence, this *extremal* character of the entropy function makes possible an analytical characterization of the second law, expressible entirely in terms of state properties of the individual equilibrium state, without reference to cycles, processes, perpetual motion machines, and the like.

Here is how Gibbs expresses this change of perspective in the opening sentences of his abstract:

It is an inference naturally suggested by the general increase of entropy which accompanies the changes occurring in any isolated material system that when the entropy of the system has reached a maximum, the system will be in a state of equilibrium. Although this principle has by no means escaped the attention of physicists, its importance does not appear to have been duly appreciated. Little has been done to develop that principle as a foundation for the general theory of thermodynamic equilibrium.

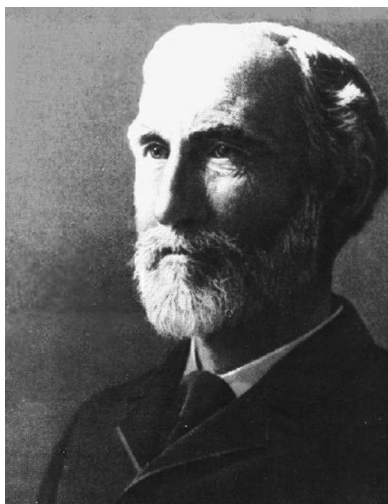
With these few words, the conceptual terrain is profoundly altered, and the way is opened for a broad array of new and profound thermodynamic relationships to emerge for systems of very general physical and chemical constitution.

We can state Gibbs' "criterion of equilibrium" (he did not speak of a "law of thermodynamics") as follows:

Gibbs criterion (I): In an isolated equilibrium system, the entropy function has the mathematical character of a *maximum* with respect to variations that do not alter the energy.

In summary, Clausius states that entropy *strives toward a maximum* in isolated processes tending toward equilibrium, while Gibbs states that entropy *is at a maximum* in isolated equilibrium states.

SIDEBAR 5.1: J. WILLARD GIBBS (1839–1903)



Josiah Willard Gibbs, like his father of the same name, had a lifelong association with Yale University and the scholarly community of New Haven, Connecticut. Gibbs' father, professor of linguistics at Yale, played a notable role in the defense of slave mutineers in the *Amistad* trial of 1839–40. The son Willard, although somewhat retiring and frail of health, exhibited early excellence in both Latin and mathematics, vacillating for a time between the two before finally committing to a career in science. (His meticulous scientific prose style would later be characterized by demanding Latin-style constructions that English syntax was scarcely able to support.) He entered Yale in 1854 and remained there to take his doctorate in 1863, the first doctorate of engineering to be conferred in the United States.

The death of his widower father in 1861 left young Gibbs and his two sisters a comfortable inheritance at the family home near the Yale campus, where the three siblings were to spend out their lives. Gibbs' only extended leave from New Haven was for European post-doctoral studies in 1866–69, when he came under the influence of Kirchhoff, Helmholtz, and other leading thermodynamicists during extended stays (accompanied by his sisters) in Paris, Berlin, and Heidelberg. Soon after his return to New Haven, Gibbs was appointed (unpaid!) professor of mathematical physics at Yale, two years before his first scientific publications (at age 34) were to reveal to the world the scope and depth of his European-style thermodynamic interests.

Gibbs' first published papers appeared in 1873 in the local (and little-known) *Transactions of the Connecticut Academy*, of which Gibbs was a steadfast member. In these two papers, Gibbs introduced highly innovative graphical and geometrical methods for representing abstract thermodynamic relationships. This was followed in 1875–78 by his monumental paper, "On the Equilibrium of Heterogeneous Substances," which stretched (in two major segments) over more than three years and 320 journal pages of *Trans. Conn. Acad.* The heavy publication expenses of the first segment (October 1875–May 1876) forced the Academy to conduct public solicitations among the New Haven

townspeople before the second segment could be completed (May 1877–July 1878). An 18-page abstract of this paper appeared separately in the *American Journal of Science*. With this epochal paper, modern chemical thermodynamics and the discipline of physical chemistry (as later identified by Ostwald, Van't Hoff, Nernst, and others) came into existence.

Gibbs' thermodynamic papers—difficult, abstract, and buried in an obscure journal on the fringes of European scientific awareness—remained for a time virtually unknown except among a small circle of admirers. Fortunately, the latter included physicist James Clark Maxwell, who advocated effectively for Gibbs' insights and methods. Major centers of Gibbsian influence began to appear in Germany, Holland, and elsewhere, as Nobel Prize-winning careers were launched from a passing remark or footnote in Gibbs' monumental masterpiece.

Gibbs later turned his attention increasingly to mathematical matters, particularly to development of now-standard methods of vector algebra (in opposition to then-prevalent quaternion methods) for representing physical phenomena. He devoted his 1887 term as vice-president (mathematics section) of the American Association for the Advancement of Science to strongly advocate methods of “multiple algebra” (today called matrix algebra). He was fond of quoting the observation that “the human mind has never invented a labor-saving machine equal to algebra.”

Gibbs' final years were devoted to his remarkable book, *Elementary Principles in Statistical Mechanics*, published shortly before his death. In this work, Gibbs masterfully reformulated Boltzmann's theory in the more general and rigorous form that was later found to seamlessly accommodate the discovery of quantum mechanics, forming the basis of modern statistical thermodynamics. Einstein, unaware of Gibbs' work, had undertaken a similar generalization of Boltzmann's theory in three papers of 1902–04, but as he later observed:

I only wish to add that the road taken by Gibbs in his book, which consists in one's starting directly from the canonical ensemble, is in my opinion preferable to the road I took. Had I been familiar with Gibbs' book at that time, I would not have published those papers at all, but would have limited myself to the discussion of just a few points.

(Fortunately for science, Einstein's deflection toward other problems ended productively, for the ensuing 1905 *Annus Mirabilis* witnessed publication of Einstein's three papers on the special theory of relativity, Brownian motion, and the quantum theory of the photoelectric effect.)

Although Gibbs briefly considered an invitation to move to an endowed position at Johns Hopkins University (which finally induced Yale to provide him a salary!), he showed little inclination to depart from his unassuming life in New Haven. A biographer remarked that

[he] remained a bachelor, living in his surviving sister's household. In his later years he was a tall, dignified gentleman, with a healthy stride and ruddy complexion, performing his share of household chores, approachable and kind (if unintelligible) to students. Gibbs was highly esteemed by his friends, but U.S. science was too preoccupied with practical questions to make much use of his profound theoretical work during his lifetime. He lived out his quiet life at Yale, deeply admired by a few able students but making no immediate impress on U.S. science commensurate with his genius.

5.2 ANALYTICAL FORMULATION OF THE GIBBS CRITERION FOR A SYSTEM IN EQUILIBRIUM

According to the Gibbs criterion, the entropy function S is a maximum (with respect to certain allowed variations). Recall that for a general function $f(x, y, \dots)$, the conditions that f be a maximum are:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = \dots = 0 \quad (\text{vanishing 1st derivatives}) \quad (5.1)$$

$$\frac{\partial^2 f}{\partial x^2} < 0, \quad \frac{\partial^2 f}{\partial y^2} < 0, \dots \quad (\text{negative 2nd derivatives and determinants of derivatives}) \quad (5.2)$$

In order to obtain the analogous conditions for maximization of S with respect to “allowed variations,” we consider in turn the following questions:

- (i) What are the *variables* on which S depends?
- (ii) What are the “allowed variations” of these variables?
- (iii) What are the conditions for *constrained* maximization of S , subject to the constraints of “allowed variations”?

(i) Variables of S For a closed system (fixed N), the first law

$$dU = T dS - P dV \quad (5.3)$$

may be rewritten as

$$dS = \frac{1}{T} dU + \frac{P}{T} dV \quad (5.4)$$

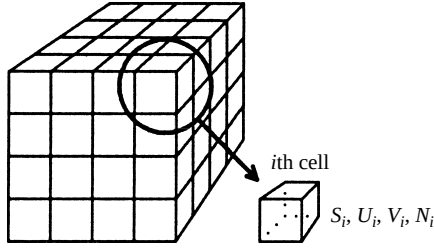
Variations of S can therefore be expressed in terms of its “natural” variables U , V for a closed system. However, for the more general case of an open system (variable N), we should include the total mass variable N to write

$$S = S(U, V, N) \quad (5.5)$$

as the desired functional expression.

(ii) Allowed Variations For a system in isolation, no exchange of heat, work, mass, or any other form of energy with the surroundings is allowed. The allowed variations are therefore those that *rearrange* the distribution of energy U , volume V , and mass N within the system, while leaving the *total* values unchanged.

We can envision these variations by partitioning the system into a large number of *cells*, as follows:



The *i*th cell has extensive properties of entropy S_i , energy U_i , volume V_i , and mass N_i , where

$$S = \sum_i S_i \quad (5.6a)$$

$$U = \sum_i U_i = U_{\text{tot}} \quad (5.6b)$$

$$V = \sum_i V_i = V_{\text{tot}} \quad (5.6c)$$

$$N = \sum_i N_i = N_{\text{tot}} \quad (5.6d)$$

We therefore obtain for S the extended set of variables

$$S = S(U_1, V_1, N_1, U_2, V_2, N_2, \dots) \quad (5.7)$$

which are subject to the constraints (5.6b–d) of constant total U_{tot} , V_{tot} , and N_{tot} .

(iii) Constrained Maximization: Method of Lagrange Undetermined Multipliers The problem of constrained maximization may be posed in its most general form as follows:

Problem: Find the maximum of a function $f(x_1, x_2, \dots, x_n)$ of n variables, subject to the c constraint equations

$$\begin{aligned} g_1(x_1, x_2, \dots, x_n) &= 0 \\ g_2(x_1, x_2, \dots, x_n) &= 0 \\ &\dots \\ g_c(x_1, x_2, \dots, x_n) &= 0 \end{aligned}$$

We might attempt to solve this problem in mundane fashion by first solving each of the c constraint equations for one of the variables $x_1, x_2, \dots, x_c$ in terms of the remaining $x_{c+1}, x_{c+2}, \dots, x_n$, then substituting these into $f(x_1, x_2, \dots, x_n)$ to obtain the function $f_{\text{red}}(x_{c+1}, x_{c+2}, \dots, x_n)$ of reduced variability, and finally solving the remaining $n - c$

derivative equations

$$\frac{\partial f_{\text{red}}}{\partial x_{c+1}} = \frac{\partial f_{\text{red}}}{\partial x_{c+2}} = \dots = \frac{\partial f_{\text{red}}}{\partial x_c} = 0$$

This mundane approach is seldom practical.

A more elegant and useful method was suggested by Lagrange. The fundamental difficulty is that there are fewer variables than the number of derivative conditions. As suggested by Lagrange, we can therefore introduce new constants $\lambda_1, \lambda_2, \dots, \lambda_c$ (“Lagrange undetermined multipliers,” one for each constraint) to define a new *constrained* function $\tilde{f}$ given by

$$\tilde{f}(x_1, x_2, \dots, x_n) \equiv f(x_1, x_2, \dots, x_n) - \sum_{k=1}^c \lambda_k g_k(x_1, x_2, \dots, x_n) \quad (5.8a)$$

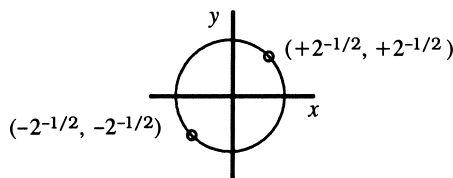
The first derivatives of $\tilde{f}$ may now be *freely* set to zero to obtain the desired solution:

$$\frac{\partial \tilde{f}}{\partial x_1} = \frac{\partial \tilde{f}}{\partial x_2} = \dots = \frac{\partial \tilde{f}}{\partial x_n} = 0 \quad (5.8b)$$

We shall not prove why Lagrange’s method works (see your advanced calculus book). But one can judge that differentiating Lagrange’s $\tilde{f}$ is much easier than finding and differentiating f_{red} . One can also see that, *if* the constraint equations $g_k = 0$ are satisfied (as will be assured by proper choice of the multipliers λ_k), then Lagrange’s solution (5.8) must be equivalent to solution of the original problem. Sidebar 5.2 illustrates the Lagrange method for a simple example.

SIDEBAR 5.2: ILLUSTRATION OF LAGRANGE’S METHOD OF UNDETERMINED MULTIPLIERS

Problem Find the maximum value of the product of x and y on the unit circle, i.e., maximize $f(x, y) = xy$ subject to the constraint equation $x^2 + y^2 = 1$.



(This problem is simple enough for one to see, by inspection, that the desired maximum is located at either of the circled points $x = y = \pm 2^{-1/2}$ in the diagram.)

Mundane Solution We first solve the constraint equation for $y = \pm(1 - x^2)^{1/2}$ and substitute into $f(x, y)$ to obtain a function $f_{\text{red}}(x)$ of a single variable:

$$f(x, y) = f_{\text{red}}(x) = \pm x(1 - x^2)^{1/2}$$

We may then set the ordinary derivative of f_{red} to zero:

$$\frac{df_{\text{red}}}{dx} = 0 = \pm(1 - x^2)^{1/2} \pm \frac{x(-2x)}{2(1 - x^2)^{1/2}}$$

This equation can be solved for x to obtain the final solutions $x = \pm 2^{-1/2} = y$ and $f_{\text{max}} = 2$. (This is ugly, but still doable for this simple case.)

Lagrange Solution We introduce an undetermined multiplier λ to write the constrained function $\tilde{f}$ as

$$\tilde{f} = xy - \lambda(x^2 + y^2 - 1)$$

The first-derivative conditions for maximizing $\tilde{f}$ are

$$\left(\frac{\partial \tilde{f}}{\partial x}\right)_y = y - 2\lambda x = 0$$

$$\left(\frac{\partial \tilde{f}}{\partial y}\right)_x = x - 2\lambda y = 0$$

which can be easily solved to obtain

$$\lambda = \pm \frac{1}{2}, \quad x = y$$

We finally employ the constraint equation to solve for the remaining unknown x ,

$$1 = x^2 + y^2 = 2x^2$$

which gives the same solutions ($x = y = \pm 2^{-1/2}$) as above. Note that Lagrange's $\tilde{f}$ will always be as easy to differentiate as the original f , whereas the mundane f_{red} (if it can be found at all) may present a daunting challenge to differentiate.

Let us now return to the original problem of maximizing the entropy function (5.7) subject to the constraints (5.6b–d). With Lagrange multipliers λ_U , λ_V , and λ_N , the constrained function $\tilde{S}$ is

$$\tilde{S} \equiv S - \lambda_U \left(\sum_i U_i - U_{\text{tot}} \right) - \lambda_V \left(\sum_i V_i - V_{\text{tot}} \right) - \lambda_N \left(\sum_i N_i - N_{\text{tot}} \right) \quad (5.9)$$

We can now freely set all first derivatives of $\tilde{S}$ to zero:

$$\left(\frac{\partial \tilde{S}}{\partial U_i}\right) = \left(\frac{\partial \tilde{S}}{\partial V_i}\right) = \left(\frac{\partial \tilde{S}}{\partial N_i}\right) = 0, \quad \text{all } i \quad (5.10)$$

Note that the constraint terms in (5.9) are linear in each extensive variable U_i , V_i , or N_i , so partial derivatives of $\tilde{S}$ are related simply to those of S by the additive constant multiplier.

For example, for the partial derivative with respect to cell energy U_i (holding all other U_j , V , and N constant), we obtain

$$\left(\frac{\partial \tilde{S}}{\partial U_i}\right) = \left(\frac{\partial S}{\partial U_i}\right) - \lambda_U = 0, \quad \text{all } i \quad (5.11)$$

However,

$$\left(\frac{\partial S}{\partial U_i}\right)_{U_j, V, N} = \frac{1}{T_i} = \text{temperature of } i\text{th cell} \quad (5.12)$$

which, according to (5.11), must have the *same* constant value λ_U in every cell:

$$\lambda_U = \frac{1}{T_i} = \frac{1}{T}, \quad \text{all } i \quad (5.13)$$

This establishes that the temperature T must be a uniform constant in every part of the system, as its identification as an intensive property (Section 2.10) demands. Note that under the conditions of partial differentiation,

$$\left(\frac{\partial \tilde{S}}{\partial U_i}\right)_{U_j, V, N} = \left(\frac{\partial S}{\partial U}\right)_{V, N} = \frac{1}{T} \quad (5.14)$$

because U and U_i differ only by variables being held constant.

For the partial derivative with respect to cell volume V_i (holding all V_j , U , and N constant), we obtain similarly

$$\lambda_V = \frac{P_i}{T_i} = \frac{P}{T}, \quad \text{all } i \quad (5.15)$$

showing that at equilibrium the pressure P must also be a uniform constant in every part of the system. Similar constancy must also hold for the partial derivative of S with respect to cell mass N_i , that is, for the intensive property conjugate to N (later to be identified as “chemical potential”).

In summary, the first-derivative conditions (5.10) imply uniform values of the derivative (intensive) properties of S throughout the system. In this way, the system-wide uniformity of temperature, pressure, and other intensive properties is obtained from the Gibbs criterion of equilibrium as a *deduction*, not an assumption.

The remaining condition for the constrained entropy to be a maximum is that the second derivatives of $\tilde{S}$ must all be negative:

$$\frac{\partial^2 \tilde{S}}{\partial U_i^2} < 0, \quad \frac{\partial^2 \tilde{S}}{\partial V_i^2} < 0, \quad \frac{\partial^2 \tilde{S}}{\partial N_i^2} < 0 \quad (5.16)$$

Because the constraint terms depend only linearly on the differentiation variables, the Lagrange parameters do not contribute to the second derivatives. From (5.14), we can

therefore write

$$\frac{\partial^2 \tilde{S}}{\partial U_i^2} = \frac{\partial^2 S}{\partial U_i^2} = \frac{\partial^2 S}{\partial U^2} < 0, \quad \text{etc.} \quad (5.17)$$

Second-derivative conditions such as (5.17) are known as “stability conditions,” expressing the self-restorative property of thermal equilibrium.

For example, from the stability condition $(\partial^2 S / \partial U^2)_{V,N} < 0$, we obtain

$$\left(\frac{\partial^2 S}{\partial U^2} \right)_{V,N} = \frac{\partial}{\partial U} \left(\frac{\partial S}{\partial U} \right)_{V,N} = \left(\frac{\partial T^{-1}}{\partial U} \right)_{V,N} = -\frac{1}{T^2} \left(\frac{\partial T}{\partial U} \right)_{V,N} = -\frac{1}{T^2 C_V} < 0 \quad (5.18)$$

from which we infer (since $T > 0$)

$$C_V > 0 \quad (5.19)$$

The positivity of C_V , i.e., the fact that temperature must *rise* when heat is added, is a well-known property of thermally stable systems. Like the uniformity of temperature (5.13), the positivity of C_V (5.19) is a deduced (rather than assumed) feature of thermal equilibrium in the Gibbs formulation.

Other stability conditions are obtained from the negativity of second derivatives with respect to V or N . (More generally, determinants of such second derivatives must also be negative in order to guarantee stability with respect to arbitrary combinations of energy, volume, and mass changes.) In summary, we can say that the Gibbs criterion of equilibrium for a closed system is equivalent to conditions of uniform intensive properties T, P ,

$$\left(\frac{\partial S}{\partial U} \right)_{V,N} = \frac{1}{T}, \quad \left(\frac{\partial S}{\partial V} \right)_{U,N} = \frac{P}{T} \quad (\text{constant throughout the system}) \quad (5.20)$$

and the thermal and mechanical stability conditions

$$\left(\frac{\partial^2 S}{\partial U^2} \right)_{V,N} < 0 \quad (5.21a)$$

$$\left(\frac{\partial^2 S}{\partial V^2} \right)_{U,N} < 0 \quad (5.21b)$$

$$\left(\frac{\partial^2 S}{\partial U^2} \right)_{V,N} \left(\frac{\partial^2 S}{\partial V^2} \right)_{U,N} < \left(\frac{\partial^2 S}{\partial U \partial V} \right)_N^2 \quad (5.21c)$$

5.3 ALTERNATIVE EXPRESSIONS OF THE GIBBS CRITERION

Let us now attempt to re-express the Gibbs criterion of equilibrium in alternative analytical and graphical forms that are more closely related to Clausius-like statements of the second law. For this purpose, we write the constrained entropy function $\tilde{S}$ in terms of its

dependence on an arbitrary extensive argument X :

$$\tilde{S}(X) = S(X) - \lambda_X(X + \cdots) \quad (5.22)$$

where the Lagrange multiplier λ_X equals the first derivative (S') of the entropy function [as in (5.13) or (5.15)] for the properly constrained $\tilde{S}$:

$$\lambda_X = S'(X) \quad (5.23)$$

Let us now consider the effect ($\delta\tilde{S}$) on constrained $\tilde{S} = S_{U,V,N}$ of a small variation δX around X :

$$\delta\tilde{S} \equiv \tilde{S}(X + \delta X) - \tilde{S}(X) = (\delta S)_{U,V,N} \quad (5.24)$$

Using the definition (5.22), a Taylor series expansion (Section 1.4) of $S(X + \delta X)$, and the expression (5.23) for the Lagrange multiplier λ_X , we obtain

$$\begin{aligned} \delta\tilde{S} &= S(X + \delta X) - S(X) - \lambda_X \delta X \\ &= [S(X) + S'(X)\delta X + \frac{1}{2}S''(X)(\delta X)^2 + \cdots] - S(X) - S'(X)\delta X \\ &= \frac{1}{2}S''(X)(\delta X)^2 + \cdots \end{aligned} \quad (5.25)$$

where we have neglected terms beyond the second order. For sufficiently small δX , it is evident from (5.25) that the constrained $\delta\tilde{S} = (\delta S)_{U,V,N}$ must have the same (negative) sign as the second derivative $S''(X)$, i.e.,

$$(\delta S)_{U,V,N} < 0 \quad (5.26)$$

The inequality (5.26) merely says that the entropy function was *at* a maximum before the variation, which is the counterpart of the Clausius statement [cf. (4.48)]

$$(\Delta S)_{U,V,N} > 0 \quad (5.27)$$

for the entropy increase *toward* equilibrium under isolation constraints of constant U , V , and N .

The essential feature of isolation constraints is constancy with respect to any form of *energy* exchange, which can be denoted by subscript U . The variations of constrained entropy S_U with respect to variations of X about an equilibrium position X_{eq} are schematically depicted in the S_U – X diagram of Fig. 5.1a. As shown in the diagram, S_U is a maximum at X_{eq} , and nonequilibrium displacements δX lead to a *decrease* in constrained entropy [cf. (5.26)]. Figure 5.1b similarly depicts the unconstrained S – X diagram, with the dashed line showing the tangent S' at equilibrium. Only the constrained entropy S_U achieves a maximum at equilibrium, but in each case the downward curvature of the entropy function (constrained or unconstrained) causes the plotted curve to fall *below* the dashed equilibrium tangent line. This curvature expresses the essence of the second law.

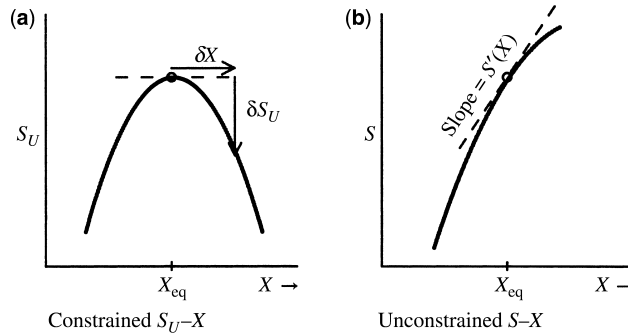


Figure 5.1 Schematic plots of (a) constrained entropy S_U and (b) unconstrained entropy S as functions of a general extensive property X near equilibrium, X_{eq} . In each case, the negative curvature of the entropy function (constrained or unconstrained) carries it *below* its equilibrium tangent (dashed line).

Single-variable plots of $S_U(X)$ or $S(X)$ such as those shown in Fig. 5.1 do not yet convey a geometrical picture of the multivariate entropy function in higher dimensions. Figure 5.2 shows a more complete 3-dimensional SUX view of the $S(U, X)$ surface for a general extensive variable X .

As shown in the figure, the curvature of the entropy function always causes it to fall *below* its tangent planes. A mathematical object having such distinctive global curvature (such as an eggshell or an upside-down bowl) is called “convex.” Accordingly, we may restate the Gibbs criterion in terms of this intrinsic convexity property of the entropy function $S = S(U, V, N)$:

Gibbs criterion (II): The entropy is a convex function of its arguments, having second derivatives of all the same (negative) sign.

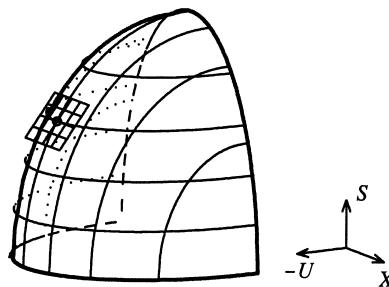


Figure 5.2 Schematic 3-dimensional depiction of the entropy function $S(U, X)$, showing the tangent plane (planar grid) at an equilibrium state (small circle). According to the curvature (stability) condition, the entropy function always falls *below* its equilibrium tangent planes, and thus has the form of a “convex function.”

5.4 DUALITY OF FUNDAMENTAL EQUATIONS: ENTROPY MAXIMIZATION VERSUS ENERGY MINIMIZATION

The preceding sections have made clear that the functional dependence of the entropy on its natural arguments U , V , N ,

$$S = S(U, V, N) \quad (5.28)$$

is indeed (in the terminology employed by Gibbs) a “fundamental equation” from which equations of state and all other thermodynamic relationships can be derived. This seems to give entropy a unique thermodynamic status and importance, as the second-law formulations of Clausius would also suggest.

However, Gibbs demonstrated the equal importance of a second “fundamental equation” that reveals a beautiful duality of the thermodynamic formalism: the deep symmetry between entropy (5.28) and internal energy U (5.29):

$$U = U(S, V, N) \quad (5.29)$$

This duality can be stated as follows:

Claim: If entropy $S = S(U, V, \dots)$ is *maximized* under the constraint of constant U , then the internal energy $U = U(S, V, \dots)$ is correspondingly *minimized* under the constraint of constant S , i.e.,

$$(\delta S)_{U,V,\dots} \leq 0 \quad \text{if, and only if, } (\delta U)_{S,V,\dots} \geq 0 \quad (5.30)$$

and either quantity (5.28) or (5.29) can therefore serve equivalently as the “fundamental equation” for a complete thermodynamic description.

The proof of this important claim is sketched in Sidebar 5.3.

SIDEBAR 5.3: PROOF OF ENTROPY–ENERGY DUALITY

To prove (5.30), let us begin by writing the known fundamental equation for entropy in the more general form $S = S(U, X_1, X_2, \dots)$ to focus on its dependence on U , where $\{X_i\}$ are the remaining extensive arguments (e.g., $V, N_1, N_2, \dots, N_c$ for a system with c independent chemical components). By the chain rule, the differential variations dS can be written as

$$dS = \left(\frac{\partial S}{\partial U} \right)_{X_i} dU + \sum_i \left(\frac{\partial S}{\partial X_i} \right)_{U, X_{j \neq i}} dX_i \quad (\text{S5.3-1})$$

where [cf. (5.14)]

$$\left(\frac{\partial S}{\partial U} \right)_{X_i} = \frac{1}{T} \quad (\text{S5.3-2})$$

or, under the constraints of constant S ($dS = 0$, denoted with a subscript S),

$$0 = \frac{1}{T} dU_S + \sum_i \left(\frac{\partial S}{\partial X_i} \right)_{U, X_{j \neq i}} dX_i \quad (\text{S5.3-3})$$

The partial derivatives in the summation may similarly be written in abbreviated form as

$$\left(\frac{\partial S}{\partial X_i}\right)_{U, X_{j \neq i}} \equiv \frac{\partial S_U}{\partial X_i} \quad (\text{S5.3-4})$$

to denote the isolation constraints of constant U that define the constrained entropy S_U :

$$0 = \frac{1}{T} dU_S + \sum_i \left(\frac{\partial S_U}{\partial X_i}\right) dX_i \quad (\text{S5.3-5})$$

For any chosen variable X_i , we therefore deduce from the above equation

$$\frac{\partial U_S}{\partial X_i} = (-T) \frac{\partial S_U}{\partial X_i}, \quad \text{all variations } X_i \quad (\text{S5.3-6})$$

From this equation, and the fact that $T > 0$, we can see that

- (i) $(\partial U/\partial X_i)_S$ and $(\partial S/\partial X_i)_U$ always have *opposite* signs (for arbitrary variations X_i);
- (ii) if S_U is curving *downward*, then U_S is curving *upward*; and, therefore,
- (iii) S_U and U_S necessarily have extrema at the same (equilibrium) points, but one (S_U) is a *maximum* and the other (U_S) is a *minimum*.

From these mandated relationships between S_U and U_S , the fundamental duality (5.30) between their variations is established.

The entropy–energy duality can also be illustrated graphically from the general representation of the Gibbs USX surface shown in Fig. 5.2. If we cut through this surface in a plane of constant U , as shown in Fig. 5.3a, we obtain the curve of constrained $S_U(X)$ (heavy line), where the equilibrium state (small circle) occurs at a *maximum*.

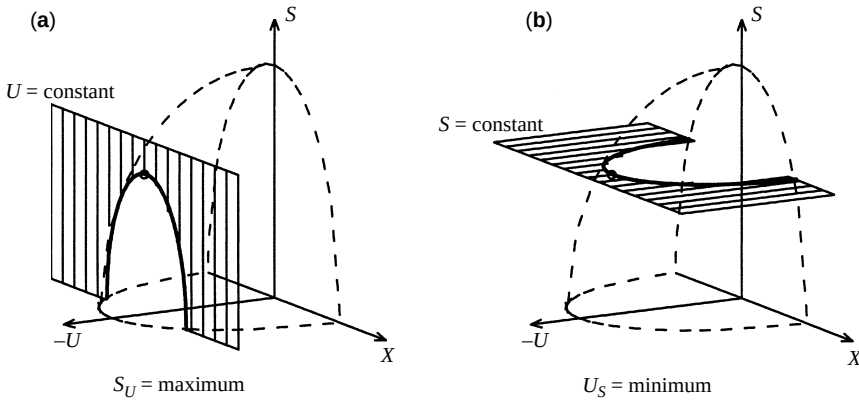


Figure 5.3 Planar slices through the Gibbs USX surface (cf. Fig. 5.2) for (a) $U = \text{constant}$, (b) $S = \text{constant}$, showing that the chosen equilibrium state (small circle) is a *maximum* in the first case (S_U curve, heavy line) but a *minimum* in the second case (U_S curve, heavy line).

However, if the surface is sliced through a plane of constant S , as shown in Fig. 5.3b, the same equilibrium point occurs at a *minimum* of the curve of $U_S(X)$ (or, what is the same, a maximum of $-U_S$). Thus, we can recognize how, for every equilibrium point, the entropy is maximized in the plane of constant U , while the energy is simultaneously minimized in the plane of constant S , as asserted in (5.30).

From this far-reaching duality, we now have two independent (but equivalent) ways of defining an equilibrium state:

- (i) An equilibrium state is a state of *maximum entropy* (if energy is held constant).
- (ii) An equilibrium state is a state of *minimum energy* (if entropy is held constant).

This is reminiscent of the two independent (but equivalent) ways of defining a circle:

- (i) A circle is the shape of *maximum area* (if circumference is held constant).
- (ii) A circle is the shape of *minimum circumference* (if area is held constant).

Just as the circle represents a uniquely simple, symmetric, and optimal limit in the general theory of geometrical shapes, so is the *equilibrium state* to be recognized as a corresponding limit in the general theory of thermodynamic phenomena.

5.5 OTHER THERMODYNAMIC POTENTIALS: GIBBS AND HELMHOLTZ FREE ENERGY

The variations of entropy S and energy U express the content of the second law in a deep way, whether as displacements (δ) from equilibrium (e.g., for a closed system),

$$(\delta S)_{U,V} \leq 0, \quad (\delta U)_{S,V} \geq 0 \quad (\text{displacements from equilibrium}) \quad (5.31)$$

or as the direction (d) of spontaneous change

$$(dS)_{U,V} \geq 0, \quad (dU)_{S,V} \leq 0 \quad (\text{direction of spontaneous change}) \quad (5.32)$$

However, these “potentials” do not yet express the second law in the form most convenient for chemical applications. Open laboratory vessels exposed to the temperature and pressure of the surroundings are subject neither to constraints of isolation (as required for entropy maximization) nor to adiabatic constant-volume conditions (as required for energy minimization). Hence, we seek alternative thermodynamic potentials that express the criteria for equilibrium under more general conditions.

Let us first introduce a useful short-cut to the constrained optimization procedure employed in Section 5.2, based on the general Clausius inequality [cf. (4.43)] for spontaneous changes toward equilibrium:

$$dq \leq T dS \quad (> \text{irreversible}; = \text{reversible}) \quad (5.33)$$

If we write the first law ($dU = dq + dw$) for the case of PV work only, we obtain from (5.33), under conditions of constant S and V ,

$$dU = dq - P dV \leq T dS - P dV = 0, \quad \text{at constant } S, V \quad (5.34)$$

This is just the energy-minimization inequality

$$(dU)_{S,V} \leq 0 \quad (5.35)$$

of (5.32), which may be re-expressed (by the method of Section 5.2) as the corresponding criterion of equilibrium $(\delta U)_{S,V} \geq 0$ of (5.31).

Let us now consider the *Helmholtz free energy* A , defined as

$$A \equiv U - TS \quad (5.36)$$

Under the restriction to PV work only, we obtain

$$dA = dU - d(TS) = dq + dw - T dS - S dT = (dq - T dS) - P dV - S dT \quad (5.37)$$

or, with the help of (5.33),

$$dA \leq -P dV - S dT = 0 \quad \text{at constant } T, V \quad (5.38)$$

Under conditions of constant T and V , we may therefore conclude that

$$(dA)_{T,V} \leq 0 \quad (5.39)$$

i.e., that A strives toward a *minimum* at constant T, V . Thus, Helmholtz free energy A is the governing potential of spontaneous change under isothermal and isochoric conditions.

However, the more important quantity for ordinary laboratory conditions is the *Gibbs free energy* G , defined as

$$G \equiv H - TS \quad (5.40a)$$

or, in alternative forms,

$$G = A + PV = U + PV - TS \quad (5.40b)$$

In the same manner as above, we obtain

$$\begin{aligned} dG &= dU + d(PV) - d(TS) \\ &= (dq - P dV) + (P dV + V dP) - (T dS + S dT) \\ &= (dq - T dS) - S dT + V dP \\ &\leq -S dT + V dP = 0 \quad \text{at constant } T, P \end{aligned} \quad (5.41)$$

We therefore conclude that

$$(dG)_{T,P} \leq 0 \quad (5.42)$$

which establishes that G strives toward minimization at constant T and P . Thus, the Gibbs free energy G is the desired potential that governs spontaneous changes in ordinary laboratory processes under isothermal and isobaric conditions.

For completeness, let us also examine the variations of enthalpy $H = U + PV$,

$$dH = (dq - P dV) + (P dV + V dP) \leq T dS + V dP = 0 \quad \text{at constant } S, P \quad (5.43)$$

which leads to

$$(dH)_{S,P} \leq 0 \quad (5.44)$$

TABLE 5.1 Second-Law Statements for Various Constraint Conditions, Expressed in Terms of the Appropriate Thermodynamic Potential

Constraint Conditions	Thermodynamic Potential	Criterion of Spontaneity	Criterion of Equilibrium
Constant U, V	S	$\Delta S \geq 0$	S maximized
Constant S, V	U	$\Delta U \leq 0$	U minimized
Constant S, P	H	$\Delta H \leq 0$	H minimized
Constant T, V	A	$\Delta A \leq 0$	A minimized
Constant T, P	G	$\Delta G \leq 0$	G minimized

Thus, spontaneous changes under adiabatic, isobaric conditions are always accompanied by net decrease in the enthalpy, which serves as the governing potential under these conditions.

Table 5.1 summarizes the various constraint conditions and the associated thermodynamic potentials and second-law statements for direction of spontaneous change or condition of equilibrium. All of these statements are equivalent to Carnot's theorem ($\oint dq/T \leq 0$) or to Clausius' inequality ($dq \leq T dS$), but each expresses the second law in the form most appropriate for the stated experimental conditions.

Table 5.1 is not exhaustive, for one can envision other constraint conditions (such as those involving constancy of chemical potential, a property to be introduced in Chapter 6) that would require still other modified potentials. The general technique for deriving such potentials involves successive replacements (called "Legendre transformations") of the form

$$L = U - RX \quad (5.45)$$

where the starting potential $U(\dots, X, \dots)$ is replaced by a modified potential $L(\dots, R, \dots)$ having intensive $R = \partial U / \partial X$ in place of extensive X as its natural constraint variable. However, further elaboration of the Legendre transformation technique beyond the elementary examples given above is beyond the scope of the present treatment.

5.6 MAXWELL RELATIONS

As established by the first law, the key feature of energy and other Legendre-transformed thermodynamic potentials is their *state* character (Section 2.10), i.e., their *conservation* under cyclic changes of state. For the leading potentials (U, H, A, G) of chemical interest, the differentials of these conserved quantities are given at equilibrium (under the usual conditions of PV -work only) by the expressions

$$dU = dq + dw = T dS - P dV \quad (5.46a)$$

$$dH = d(U + PV) = T dS + V dP \quad (5.46b)$$

$$dA = d(U - TS) = -S dT - P dV \quad (5.46c)$$

$$dG = d(H + TS) = -S dT + V dP \quad (5.46d)$$

The first law is expressed most succinctly by the mathematical requirement that such differentials are *exact* (Section 1.3).

As described in Section 1.3, the mathematical condition for exactness of a 2-variate differential dZ of the general form

$$dZ = M dX + N dY \quad (5.47)$$

TABLE 5.2 Coefficients M, N for a General Differential $dZ = M dX + N dY$ of Thermodynamic Potentials U, H, A, G (Closed Single-Component System)

$Z(X, Y)$	M	N
$U(S, V)$	$T = (\partial U / \partial S)_V$	$-P = (\partial U / \partial V)_S$
$H(S, P)$	$T = (\partial H / \partial S)_P$	$V = (\partial H / \partial P)_S$
$A(T, V)$	$-S = (\partial A / \partial T)_V$	$-P = (\partial A / \partial V)_T$
$G(T, P)$	$-S = (\partial G / \partial T)_P$	$V = (\partial G / \partial P)_T$

is given by the *Euler condition*,

$$\left(\frac{\partial M}{\partial Y}\right)_X = \left(\frac{\partial N}{\partial X}\right)_Y \quad (5.48)$$

For the differentials dU, dH, dA, dG , the appropriate coefficients M, N can be read directly from the expressions (5.46a–d), as shown in Table 5.2.

Let us now apply the Euler condition (5.48) to each differential (5.46a–d) in turn:

(a) $dU = T dS - P dV$ ($X = S, Y = V; M = T, N = -P$)

$$\boxed{\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V} \quad (5.49a)$$

(b) $dH = T dS + V dP$ ($X = S, Y = P; M = T, N = V$)

$$\boxed{\left(\frac{\partial T}{\partial P}\right)_S = \left(\frac{\partial V}{\partial S}\right)_P} \quad (5.49b)$$

(c) $dA = -S dT - P dV$ ($X = T, Y = V; M = -S, N = -P$)

$$\boxed{\left(\frac{\partial P}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T} \quad (5.49c)$$

(d) $dG = -S dT + V dP$ ($X = T, Y = P; M = -S, N = V$)

$$\boxed{\left(\frac{\partial V}{\partial T}\right)_P = -\left(\frac{\partial S}{\partial P}\right)_T} \quad (5.49d)$$

The mathematical identities (5.49a–d) are known as *Maxwell relations*.

Maxwell relations are a powerful tool for deriving thermodynamic relationships. Their use should be considered whenever it is desirable to replace thermodynamic derivatives involving S with equivalent derivatives involving variables P, V, T only. Sidebars 5.4–5.6 illustrate this derivation techniques for a number of standard thermodynamic identities.

SIDEBAR 5.4: PROOF OF (3.69)

Problem Prove that the Joule–Thomson coefficient μ_{JT} satisfies the identity (3.69): $\mu_{JT} = (T\alpha_P - 1)(V/C_P)$.

Solution From the definition of μ_{JT} and the Jacobi identity (1.14b), we obtain

$$\mu_{JT} \equiv \left(\frac{\partial T}{\partial P} \right)_H = - \frac{(\partial H / \partial P)_T}{(\partial H / \partial T)_P} = - \frac{1}{C_P} \left(\frac{\partial H}{\partial P} \right)_T$$

To evaluate $(\partial H / \partial P)_T$, we start from the expression (5.46b) for dH in terms of its natural variables S, P [or, equivalently, use the identity (1.13) to change the variable held constant from S to T], then use the Maxwell relation (5.49d) to replace the entropy derivative as follows:

$$\left(\frac{\partial H}{\partial P} \right)_T = T \left(\frac{\partial S}{\partial P} \right)_T + V = -T \left(\frac{\partial V}{\partial T} \right)_P + V = -TV\alpha_P + V$$

Substitution of the second equation into the first gives the desired result.

SIDEBAR 5.5: PROOF OF (3.56)

Problem Prove that the heat capacity difference $C_P - C_V$ satisfies the thermodynamic identity (3.56): $C_P - C_V = TV\alpha_P^2/\beta_T$.

Solution It was previously proved that [(3.52)]

$$C_P - C_V = \left[P + \left(\frac{\partial U}{\partial V} \right)_T \right] \left(\frac{\partial V}{\partial T} \right)_P$$

To evaluate the derivative $(\partial U / \partial V)_T$, we start from expression (5.46a) for dU in terms of its natural variables S, V , differentiate with respect to V at constant T , and use the Maxwell relation (5.49c) to replace the entropy derivative,

$$\left(\frac{\partial U}{\partial V} \right)_T = T \left(\frac{\partial S}{\partial V} \right)_T - P = T \left(\frac{\partial P}{\partial T} \right)_V - P$$

giving $C_P - C_V$ entirely in terms of PVT derivatives:

$$C_P - C_V = T \left(\frac{\partial P}{\partial T} \right)_V \left(\frac{\partial V}{\partial T} \right)_P = TV\alpha_P \left(\frac{\partial P}{\partial T} \right)_V$$

The Jacobi identity for $(\partial P / \partial T)_V$ gives finally

$$\left(\frac{\partial P}{\partial T} \right)_V = \frac{(\partial V / \partial T)_P}{(\partial V / \partial P)_T} = \frac{V\alpha_P}{V\beta_T} = \frac{\alpha_P}{\beta_T}$$

to obtain the desired result.

SIDEBAR 5.6: PROOF OF EQ. (3.58)

Problem Prove that for an ideal gas, (3.58) holds, i.e.,

$$\left(\frac{\partial U}{\partial V}\right)_T \stackrel{\text{IG}}{=} 0$$

Solution As shown in Sidebar 5.5, the desired derivative can be generally expressed [with the help of the Maxwell relation (5.49c)] as

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P$$

For an ideal gas, with $P \stackrel{\text{IG}}{=} nRT/V$, the derivative $(\partial P/\partial T)_V$ is evaluated as

$$\left(\frac{\partial P}{\partial T}\right)_V \stackrel{\text{IG}}{=} \frac{nR}{V} = \frac{P}{T}$$

from which

$$\left(\frac{\partial U}{\partial V}\right)_T \stackrel{\text{IG}}{=} T(P/T) - P = 0$$

which proves the desired result. Note that this property of an ideal gas (originally introduced as a plausible assumption based on Joule's experiment) is in fact a rigorous consequence of the ideal gas equation of state, requiring no new assumptions.

The Maxwell relations are powerful tools of thermodynamic derivation. With the help of these relations and derivation techniques analogous to those illustrated in Sidebars 5.3–5.6, the skilled student of thermodynamics can (in principle!) re-express practically *any* partial derivative in terms of a small number of “base properties” involving only *PVT* variables. Consider, for example, the eight most common variables

$$U, H, A, G, P, V, T, S \tag{5.50}$$

of a closed single-component system. From the eight variables in (5.50), we could select any three (X, Y, Z) to obtain a rather large number ($8.7.6 = 336$) of possible thermodynamic derivatives $(\partial X/\partial Y)_Z$ that might superficially appear to be independent properties of the system. Nevertheless, we can in principle reduce any of these to a small number (*three!*) of independent properties, as asserted in the following claim:

Claim: By suitable manipulations, *all* 336 derivatives $(\partial X/\partial Y)_Z$ arising from (5.50) can be reduced to expressions involving only a basic set of *three* properties (which can be taken to be, e.g., C_P , α_P , and β_T).

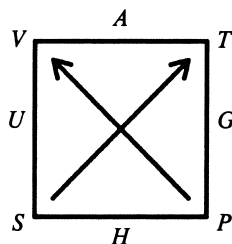
The general proof of this claim will be presented (Chapter 11) after introduction of the metric geometrical formulation of equilibrium thermodynamics, which makes the basis of the claim rather obvious. More general and powerful geometrical methods of

deriving thermodynamic relationships will also be given. Nevertheless, the examples of Sidebars 5.4–5.6 illustrate the general reduction strategy that the claim implies is possible. With sufficient practice, the student of thermodynamics can prove the claim for any specific $(\partial X/\partial Y)_Z$.

The Maxwell relations (5.49a–d) are easy to rederive from the fundamental differential forms (5.46a–d). However, these relations are used so frequently that it is useful to employ a simple mnemonic device to recall their exact forms as needed. Sidebar 5.7 describes the thermodynamic “magic square,” which provides such a mnemonic for Maxwell relations and other fundamental relationships of simple (closed, single-component) systems.

SIDEBAR 5.7: THE THERMODYNAMIC MAGIC SQUARE

The labeled diagram shown below provides a useful “thermodynamic magic square” mnemonic for recalling the fundamental mathematical relationships of a closed, single-component system. [This mnemonic apparently has roots going back to M. Born (1929) or before, but its first published explication is found in F. O. Koenig, *J. Chem. Phys.* **3**, 29 (1935).]



Let us first describe the principles of labeling each geometrical feature of the square:

- **Edges:** Each edge of the square is labeled by a potential A , G , H , or U in clockwise alphabetical order (from the top).
- **Corners:** Corners are labeled by variables T , P , S , or V such that each potential (edge) is flanked by its own two natural variables (corners). (This can only be done in a unique way, with, for example, T at the corner shared between potentials A and G .)
- **Arrows:** Two arrows point diagonally upward toward the opposite corner of the square, from S to T and from P to V .

Let us now describe how various thermodynamic relationships are “read” from geometrical and symmetry properties of the square:

Natural Variables To obtain the natural variables for each potential (edge), read their symbols from the flanking corners.

For example, the right edge shows that T and P are the natural variables of the Gibbs energy $G = G(T, P)$.

Differential Forms To obtain the coefficients of the differential form, follow the arrows from each natural variable to the variable at the opposite end, assigning a positive (+) sign if the coefficient is at the head of the arrow, or a negative (–) sign if it is at the tail of the arrow.

For example, from the right (G) edge, the arrow from natural variable T ends up at $-S$ (taken minus because it lies at the tail of the arrow), whereas that from variable P ends up at $+V$ (taken plus because it lies at the head of the arrow), giving the differential form $dG = (-S)dT + (+V)dP$. Similarly, from the top (A) edge, we can read the differential form for the Helmholtz energy as $dA = (-S)dT + (-P)dV$, because both coefficients fall at arrow tails.

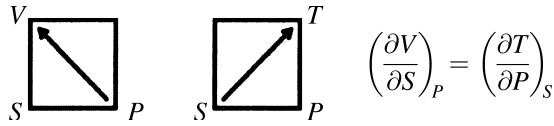
Alternatively, we can obtain the equivalent partial derivative identities of Table 5.2 for the corner variables. For example, S sits at the tail (negative) end of the arrow pointing to T , considered as a variable of either A (at constant V) or of G (at constant P), the two edges that meet at this corner. We can therefore write S as either of the partial derivatives

$$S = -\left(\frac{\partial A}{\partial T}\right)_V = -\left(\frac{\partial G}{\partial T}\right)_P$$

These are equivalent to the coefficients of the differential forms given above.

Maxwell Relations To obtain the Maxwell relations for a partial derivative $(\partial X/\partial Y)_Z$, first reorient the square so that the triangle of variables X , Y , and Z have the same spatial relationship as in the partial derivative (i.e., X to upper left, Y to lower left, and Z to lower right, with a diagonal arrow as hypotenuse). Then read the Maxwell relation $(\partial X/\partial Y)_Z = (\text{sign})(\partial X'/\partial Y')_{Z'}$ from the corner variables of the corresponding $X'Y'Z'$ triangle that is related by *left-right reflection symmetry* of the square, with “sign” taken positive (+) if the arrow direction is preserved by mirror reflection, or negative (−) if the arrow is reversed.

For example, let us consider the Maxwell relation for $(\partial V/\partial S)_P [= (\partial T/\partial P)_S]$, as represented in the following diagram:

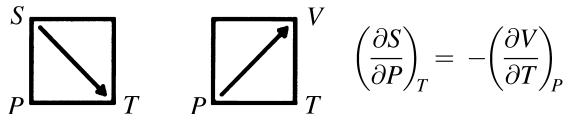


For the partial derivative $(\partial V/\partial S)_P$, focus first on the VSP triangle at the left (with V at upper left, S at lower left, and P at lower right). One can then see that the “corresponding” (mirror-reflectd) triangle on the right has corners T, P, S (with T at upper right mirroring V , P at lower right mirroring S , and S at lower left mirroring P), and with the original arrow (from P to V) having the same (upward) orientation as its mirrored counterpart (from S to T). Hence, the original and mirrored derivatives are related by a plus sign,

$$\left(\frac{\partial V}{\partial S}\right)_P = +\left(\frac{\partial T}{\partial P}\right)_S$$

which gives the proper Maxwell relation (5.49b).

If, instead, we consider the Maxwell relation for $(\partial S/\partial P)_T$, we rotate the square to obtain the symmetry-related corners



where the arrow for the *SPT* triangle at the left is *reversed* relative to that of the mirror *VTP* triangle on the right, leading to the sign change

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P$$

for the Maxwell relation (5.49d).

In the case where the $(\partial X/\partial Y)_Z$ variables do not properly match a corner pattern (i.e., do not have *Y* between *X* and *Z*), it is only necessary to find the Maxwell relation for the “inverted” form $(\partial Y/\partial X)_Z$,

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{1}{(\partial Y/\partial X)_Z}$$

then invert the Maxwell mirror derivative to find the desired relation. For example, to find the Maxwell relation for $(\partial S/\partial P)_V$ (for which there is no *SPV* corner), we can instead write

$$\left(\frac{\partial S}{\partial P}\right)_V = \frac{1}{(\partial P/\partial S)_V} = \frac{1}{-(\partial T/\partial V)_S} = -\left(\frac{\partial V}{\partial T}\right)_S$$

where we have used the Maxwell mirror symmetry between *PSV* and (minus) *TVS* triangles in the middle step.

5.7 GIBBS FREE ENERGY CHANGES IN LABORATORY CONDITIONS

The Gibbs formalism is nominally couched in terms of criteria of equilibrium and the analytical properties of equilibrium states. Nevertheless, the complementary relationship between Gibbs-like and Clausius-like statements of the second law (Table 5.1) establishes that Gibbs free energy *G* is equally useful in characterizing spontaneous *changes* of state under laboratory conditions of constant *T*, *P*. In this section, we wish to summarize more general aspects of the Gibbs free energy concept and its applications to spontaneous and equilibrium phenomena in the laboratory environment.

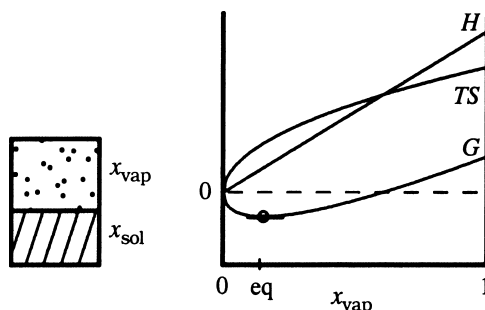
Under conditions of constant *T*, *P*, the criterion for general Gibbs free energy changes ΔG can be stated as

$$\Delta G = \Delta H - T \Delta S \leq 0 \quad (< \text{spontaneous}; = \text{equilibrium}) \quad (5.51)$$

where ΔH , ΔS are the corresponding enthalpy and entropy changes. Sidebars 5.8 and 5.9 illustrate some simple applications of this criterion for familiar physical situations. Many other consequences of Gibbs free energy minimization will be given in ensuing chapters.

SIDEBAR 5.8: VAPORIZATION OF A SOLID

Consider a simple 2-phase system consisting of some number of moles of solid and vapor, expressed in terms of mole fractions x_{sol} , x_{vap} . The following is a schematic graph of the various energetic quantities G, H, TS as functions of x_{vap} (with pure solid as the reference “0”):



For given T, P , will the sample completely vaporize, completely condense, or something intermediate? To answer this question, let us first consider the separate variations of enthalpy and entropy at any particular x_{vap} . As shown in the graph, the overall enthalpy variation is monotonically positive, so $\Delta H = \Delta H(x_{\text{vap}})$ is always positive:

$$H_{\text{vap}} > H_{\text{sol}}, \quad \Delta H > 0$$

which favors the pure solid limit ($x_{\text{vap}} = 0$). The overall entropy change is also monotonically positive:

$$S_{\text{vap}} > S_{\text{sol}}, \quad \Delta S > 0$$

which (in view of the negative sign in $\Delta G = \Delta H - T\Delta S$) favors the pure vapor limit ($x_{\text{vap}} = 1$). However, the actual criterion of equilibrium that G be *minimized* is satisfied at some *intermediate* $x_{\text{vap}}^{\text{eq}}$, as shown by the small circle in the graph. The final system will therefore consist of a non-vanishing quantity of solid ($x_{\text{sol}}^{\text{eq}} = 1 - x_{\text{vap}}^{\text{eq}}$) in equilibrium with its own vapor (vapor pressure P_{eq}). If the external pressure P is decreased below P_{eq} , the sample will begin to vaporize (because $\Delta G < 0$),

$$P < P_{\text{eq}} \longrightarrow \text{sublimation} \quad (\Delta G < 0)$$

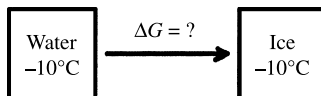
whereas if P is increased above P_{eq} , then vapor will begin to condense to solid (e.g., as hoarfrost),

$$P > P_{\text{eq}} \longrightarrow \text{frost-formation} \quad (\Delta G < 0)$$

(again because $\Delta G < 0$), in accordance with (5.51). Similarly, if the external temperature T is increased (amplifying the entropic term), vaporization is promoted, and so forth. All these changes are in accord with the Gibbs criterion (5.51) and with experience.

SIDEBAR 5.9: SUPERCOOLED WATER

Consider a sample of rainwater from a cold cloud that is initially at temperature $T = -10^\circ\text{C}$ and pressure $P = 1\text{ atm}$ as it comes to earth. Will it freeze? We can pose the question in terms of the initial state (water) and final state (ice) of the system:



For the entropy change $\Delta S = S_{\text{ice}} - S_{\text{water}}$, we find

$$\Delta S = -4.9\text{ cal mol}^{-1}\text{C}$$

a negative change (entropy loss) that favors the liquid water form. For the enthalpy change $\Delta H = H_{\text{ice}} - H_{\text{water}}$, we find

$$\Delta H = -1350\text{ cal mol}^{-1}$$

an exothermic change that favors the solid ice form. Which will win? For the Gibbs free energy change, we find

$$\Delta G = \Delta H - T\Delta S = -1350 - (263)(+4.9) = -60\text{ cal mol}^{-1}$$

which confirms that the conversion of “freezing rain” to ice is spontaneous.

At what temperature T would the water $\rightarrow$ ice conversion no longer be spontaneous? The condition of equilibrium, $\Delta G = \Delta H - T\Delta S = 0$, is satisfied when

$$T = \frac{\Delta H}{\Delta S} = \frac{-1350}{-4.9} = 273\text{K} = 0^\circ\text{C}$$

As expected, this switchover occurs at the normal melting point of ice.

What is the physical nature of the Gibbs free energy, and what is “free” about it? We can consider this question first from the viewpoint of fundamental thermodynamic definitions, with no microscopic molecular connotations. For a reversible change of state carried out under conditions of constant T and P , we can write

$$\begin{aligned}\Delta G &= \Delta H - T\Delta S = \Delta U + \Delta(PV) - T\Delta S \\ &= (q_{\text{rev}} + w_{\text{rev}}) + P\Delta V + V\Delta P - T\Delta S\end{aligned}\tag{5.52}$$

However, we recognize that $V\Delta P = 0$ and $q_{\text{rev}} = T\Delta S$ under the stated conditions. Furthermore, $P\Delta V = -w_{PV}$ is just the (negative of) reversible PV work under isobaric conditions. Hence, (5.52) can be rewritten as

$$\Delta G = w_{\text{rev}} - w_{PV}\tag{5.53}$$

Because w_{rev} is the maximum available work of *any* type, we can say from (5.53) that ΔG is the *maximum available non-PV work*. Here, “available” (or “free”) refers to the idealized reversible limit in which no useful work is dissipated. Practically speaking, the major non-PV work of interest to chemists is the *chemical energy* (as manifested, for example, in electrochemical or osmotic phenomena), associated with the *chemical potential* terms that will be introduced in Chapter 6.

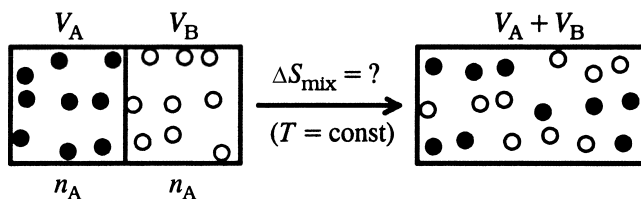
The conceptual mystery of the Gibbs free energy is largely associated with the entropic contribution. Sidebars 5.10–5.13 describe some alternative ways to think about entropy. Sidebar 5.10 evaluates the entropy change ΔS_{mix} for the prototypical mixing of ideal gases A, B under isothermal conditions,

$$\Delta \bar{S}_{\text{mix}} = -R(x_A \ln x_A + x_B \ln x_B) \quad (5.54)$$

which depends only on the final mole fractions x_A , x_B of the gas mixture. ΔS_{mix} can thereby be recognized as a measure of “mixed-up-ness”, akin to the degree of shuffling of a deck of cards from initially separated red and black cards. Alternatively, Sidebar 5.11 describes entropy from Boltzmann’s statistical mechanical viewpoint, showing that the result (5.54) can be derived from purely probabilistic considerations. Boltzmann’s interpretation underlies the common association of S with “randomness,” “chaos,” or “disorder” at the molecular level. Sidebar 5.12 describes Shannon’s information-theoretic entropy, which again associates S with probabilistic notions of “information loss” in the patterns of computer bits. Finally, Sidebar 5.13 presents a more specific molecular picture of entropy as a measure of *vibrational flexibility*, related to the thermal sampling of the near-equilibrium potential energy surface that contributes to “average thermal energy” (Gibbs free energy) at any finite T . Each of these ways of thinking about S suggests that it is a deep measure of “disorganization,” “lack of structural rigidity,” or “mixed-up-ness,” related to the possible T -dependent *distributions* of energy at the molecular level. Gibbs free energy G is correspondingly the *average thermal energy*, which depends not only on the overall depth of the potential energy minimum point (enthalpic contribution) but also the amplitude of low-lying vibrational excursions from the absolute minimum that will be sampled in the T -dependent fluctuations about equilibrium (entropic contribution).

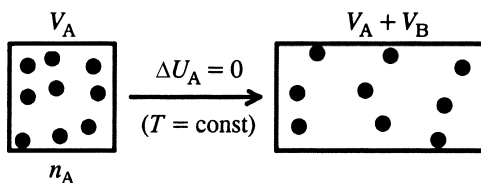
SIDEBAR 5.10: ISOTHERMAL ENTROPY OF MIXING OF IDEAL GASES

Consider the process in which quantities n_A , n_B of ideal gases A, B, initially in separated containers of volume V_A , V_B , are allowed to mix under isothermal conditions, as shown in the diagram below. What is the entropy of mixing ΔS_{mix} ?



To answer this question, we may consider the situation for each ideal gas separately, since each (being ideal) has no interaction with the other. For gas A, the process is merely

isothermal expansion from initial volume V_A to final volume $V_A + V_B$ (with $\Delta U_A = 0$),



which can be carried out reversibly to obtain, according to (3.77),

$$q_A = -w_A = n_A RT \ln \left(\frac{V_A + V_B}{V_A} \right)$$

For gas B, we obtain similarly

$$q_B = -w_B = n_B RT \ln \left(\frac{V_A + V_B}{V_B} \right)$$

The overall entropy change is therefore

$$\begin{aligned} \Delta S_{\text{mix}} &= \Delta S_A + \Delta S_B = \frac{q_A}{T} + \frac{q_B}{T} = n_A R \ln \left(\frac{V_A + V_B}{V_A} \right) + n_B R \ln \left(\frac{V_A + V_B}{V_B} \right) \\ &= -R(n_A \ln x_A + n_B \ln x_B) \end{aligned}$$

where we have used Avogadro's hypothesis ($V \propto n$) to rewrite the volume ratios in terms of mole fractions. We can also divide through by $n_{\text{tot}} = n_A + n_B$ to obtain $\Delta \bar{S}_{\text{mix}}$ on a per-mole basis:

$$\Delta \bar{S}_{\text{mix}} = -R(x_A \ln x_A + x_B \ln x_B)$$

For a more general mixture of c ideal gases, this formula obviously generalizes to

$$\Delta \bar{S}_{\text{mix}} = -R \sum_{i=1}^c x_i \ln x_i$$

SIDEBAR 5.11: BOLTZMANN'S STATISTICAL MECHANICAL ENTROPY

L. Boltzmann (Sidebar 13.7) first presented the famous mathematical equation for S ,

$$S = k \ln \Omega \quad (\text{S5.11-1})$$

that marked a radically new approach to the description of microphysical phenomena. In this equation, k is Boltzmann's constant (molecular gas constant),

$$k = R/N_0 = 1.38 \times 10^{-23} \text{ J mol}^{-1} \text{ K}^{-1}$$

(where we are temporarily denoting Avogadro's number by N_0 to avoid confusion with N_A used below for number of A-type particles) and Ω (written as W , "*Wahrscheinlichkeit*," on Boltzmann's tombstone in Vienna) is a measure of statistical

likelihood:

Ω = relative *probability* of state U, V

= number of ways of distributing energy among molecules (consistent with bulk U, V)

= number of “microstates” consistent with specified macrostate U, V

for the macrostate described by the entropy function $S = S(U, V)$.

With this bold stroke, Boltzmann escaped the futile attempt to describe microscopic molecular phenomena in terms of then-known Newtonian mechanical laws. Instead, he injected an essential probabilistic element that reduces the description of the microscopic domain to a statistical distribution of “microstates,” i.e., alternative microscopic ways of partitioning the total macroscopic energy U and volume V among the unknown degrees of freedom of the molecular domain, all such partitionings having equal a priori probability in the absence of definite information to the contrary.

Boltzmann’s expression for S thereby reduces the description of the molecular micro-world to a statistical *counting* exercise, abandoning the attempt to describe molecular behavior in strict mechanistic terms. This was most fortunate, for it enabled Boltzmann to avoid the untenable assumption that classical mechanics remains valid in the molecular domain. Instead, Boltzmann’s theory successfully incorporates certain quantal-like notions of probability and indeterminacy (nearly a half-century before the correct quantum mechanical laws were discovered) that are necessary for proper molecular-level description of macroscopic thermodynamic phenomena.

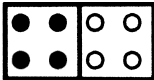
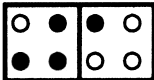
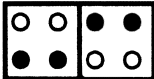
Let us illustrate the application of Boltzmann’s formula for an elementary example: isothermal mixing of ideal gases (Sidebar 5.10). For this purpose, consider a system of $N_A = N_B = 4$ particles. For a particular partition $n:4-n$ of the four A-type particles between the V_A and V_B containers, the number of possible ways Ω of choosing n A-type particles and $4-n$ B-type particles for the first container is given by the product of binomial coefficients

$$\Omega = \binom{4}{n} \binom{4}{4-n}$$

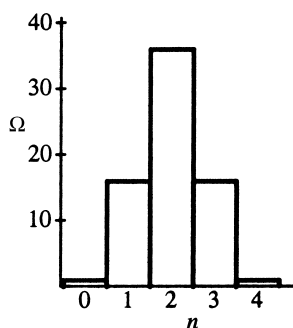
where the binomial coefficient is defined as

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} \text{ (“}n \text{ choose } m\text{”)}$$

In this manner, one can evaluate Ω for each possible partition of particles, as follows:

4 : 0		$\Omega = 1$
3 : 1		$\Omega = 16 = \binom{4}{3} \binom{4}{1}$
2 : 2		$\Omega = 36 = \binom{4}{2} \binom{4}{2}$

The graph of Ω versus n is seen to be steeply peaked at the most mixed (2 : 2) partition:



This dominance of the most-mixed partition becomes more pronounced as the number of particles increases.

The initial unmixed state $\Omega_1 = \Omega_{AB}$ corresponds to the (4 : 0) partition,

$$\Omega_1 = \Omega_A \Omega_B = 1$$

whereas for the final mixed state $\Omega_2 = \Omega_{AB}$ (with the dividing wall removed), all distinct partitions (4 : 0, 3 : 1, 2 : 2, 1 : 3, 0 : 4) contribute with equal likelihood, namely,

$$\Omega_2 = \Omega_{AB} = 1 + 16 + 36 + 16 + 1 = 70$$

This shows that the mixed state is much more probable ($70\times$) even for such small numbers of particles. More generally, it can be shown that if x_A and x_B are the respective fractions of A and B particles, then, to an increasingly accurate approximation for large N_A, N_B ,

$$\frac{\Omega_{AB}}{\Omega_A \Omega_B} \simeq \frac{1}{(x_A)^{N_A} (x_B)^{N_B}}$$

which becomes essentially exact for the large numbers (of order 10^{23}) of a macroscopic system. We therefore conclude that

$$\begin{aligned} \Delta S_{\text{Boltz}} &= S_2 - S_1 = k \ln \Omega_{AB} - k \ln (\Omega_A \Omega_B) = k \ln \left(\frac{\Omega_{AB}}{\Omega_A \Omega_B} \right) \\ &= -k \ln [(x_A)^{N_A} (x_B)^{N_B}] = -N_A k \ln x_A - N_B k \ln x_B \\ &= -n_A R \ln x_A - n_B R \ln x_B = \Delta S_{\text{mix}} \end{aligned}$$

in agreement with the thermodynamic result derived in Sidebar 5.10.

SIDEBAR 5.12: SHANNON'S INFORMATION-THEORETIC ENTROPY

C. E. Shannon (1916–2001) developed an information-theoretic definition of entropy that (although not equivalent to the physical quantity) carries similar associations with microstates and probability theory. Shannon recognized that Boolean bit patterns (sequences of 1's and 0's) can be considered the basis of all methods for encoding “information.”

The information content of such bit strings can be usefully quantified in terms of “entropy” H as defined by

$$H = - \sum_i p_i \log_2 p_i \quad (\text{S5.12-1})$$

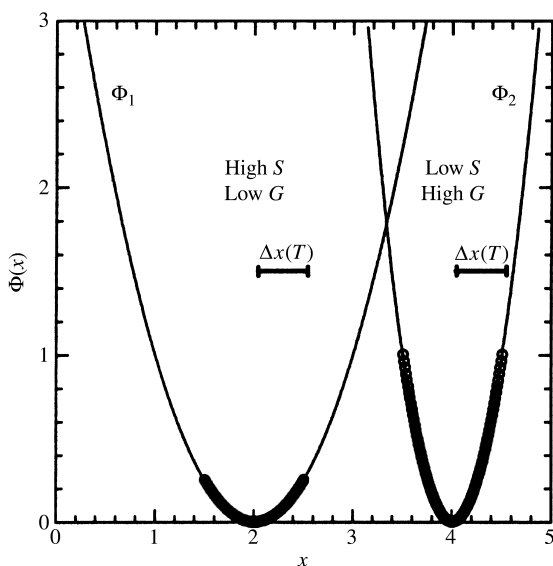
where p_i is the probability of event i . High probability (and high entropy) is thereby associated with high predictability (or low “surprisal”) and loss of information content. Shannon’s formula closely resembles the quantity appearing in Boltzmann’s H -theorem or Gibbs’ statistical mechanics.

The inverse relationship between probability and information content can be illustrated by a simple example. Suppose a policeman asks a witness to describe the runaway suspect. If the answer is “He was 5 feet, 10 inches tall and had brown hair,” the policeman is disappointed, for this description fits so many people that it provides little useful information content to identify a specific suspect. But if the answer is “He was 7 feet tall and had green hair,” the answer is highly informative—perhaps even sufficiently unique to identify a single individual.

SIDEBAR 5.13: MOLECULAR VIBRATIONAL ENTROPY

What is the physical nature of the entropic factor? It is useful to think of thermal effects as an “agitating” factor, leading to characteristic fluctuations Δx from equilibrium position x_{eq} in a potential energy function $\Phi(x)$. The entropic factor describes how well the near-equilibrium region can accommodate these thermal fluctuations without significant energy penalty.

The figure below shows two potential energy functions, Φ_1 and Φ_2 , each having identical well depths and thus identical energies at their respective x_{eq} . However, under the effect of thermal fluctuations, these potentials are obviously inequivalent. As shown in the figure, thermal sampling over the chosen range $\pm \Delta x$ leads to higher thermal-average energy (free energy) in potential Φ_2 than in Φ_1 , because the tighter curvature of the former forces a given displacement to sample to higher portions of the potential energy surface.



We can say that Φ_2 corresponds to lower entropy and higher free energy because of its vibrational *stiffness* (or equivalently, its high vibrational frequency). Conversely, Φ_1 corresponds to higher entropy and lower free energy because of its vibrational softness. Thus, the entropic factor is generally dominated by the character of the low-frequency vibrational spectrum, particularly the softest modes. The number of “floppy” vibrational motions generally constitutes the most important entropy factor for the student of physical chemistry to be aware of in a given molecular or supramolecular species.

How does ΔG depend on the ambient temperature and pressure? Like all thermodynamic properties of a given process, $\Delta G = \Delta G(T, P)$ is expected to depend on the actual T and P values at which the process occurs. Let us examine these dependences separately.

P Dependence of ΔG From the general differential expression $dG = V dP - S dT$ of (5.46d), we obtain for the P dependence (under isothermal conditions)

$$\left(\frac{\partial G}{\partial P}\right)_T = V$$

For the finite change $P_1 \rightarrow P_2$, we therefore obtain

$$\Delta G = G(P_2) - G(P_1) = \int_{P_1}^{P_2} V dP \quad \text{at constant } T \quad (5.55a)$$

For an ideal gas, for example, this gives

$$\Delta G = \int_{P_1}^{P_2} V dP \stackrel{\text{IG}}{=} nRT \int_{P_1}^{P_2} \frac{dP}{P} \stackrel{\text{IG}}{=} nRT \ln \frac{P_2}{P_1} \quad (5.55b)$$

Consistent with experience, this change is spontaneous ($\Delta G < 0$) if $P_2 < P_1$.

T Dependence of ΔG : Gibbs–Helmholtz Equation For the T dependence (under isobaric conditions), we obtain analogously

$$\left(\frac{\partial G}{\partial T}\right)_P = -S = \frac{G - H}{T}$$

Equivalently, we can evaluate the T dependence of G/T as

$$\begin{aligned} \left(\frac{\partial(G/T)}{\partial T}\right)_P &= -\frac{1}{T^2}G + \frac{1}{T}\left(\frac{\partial G}{\partial T}\right)_P = -\frac{G}{T^2} + \frac{G - H}{T^2} = -\frac{H}{T^2} \\ &= H\left(\frac{\partial(1/T)}{\partial T}\right)_P \end{aligned} \quad (5.56)$$

which leads to the more compact form

$$\boxed{\left(\frac{\partial(G/T)}{\partial(1/T)}\right)_p = H} \quad (5.57)$$

called the *Gibbs–Helmholtz equation*. Equations (5.55)–(5.57) are equivalent in physical content. For chemical reactions ($\Delta = \text{products} - \text{reactants}$), the form (5.55) is often preferred:

$$\left(\frac{\partial \Delta G}{\partial T}\right)_p = -\Delta S \quad (5.58)$$

From the base equation (5.46d), the ΔG variations can also be evaluated for a wide variety of other changes. Analogous variations of other thermodynamic potentials (U , H , A) are easily derived from the corresponding differential expressions (5.46a–c). Sidebar 5.14 illustrates such evaluations for reversible volume expansion of an ideal gas, and Sidebar 5.15 shows how to carry out the analogous evaluations for irreversible conditions. Once the power of the differential expressions (5.46a–d) is understood, they will be found sufficient for the solution of nearly any practical problem that the student may encounter.

SIDEBAR 5.14: REVERSIBLE ISOTHERMAL GAS EXPANSION PROBLEMS

Problem One mole of ideal gas expands isothermally and reversibly at 300K from 1.00 L to 10.00 L. Calculate q , w , ΔU , ΔH , ΔA , ΔG , and ΔS for this process.

Solution For isothermal expansion of an ideal gas, both ΔU and $\Delta(PV)$ are zero, so that

$$\begin{aligned} \Delta U &= 0 \\ \Delta H &= \Delta U + \Delta(PV) = 0 \end{aligned}$$

From (3.77), the heat and work are related by

$$q_{\text{rev}} = -w_{\text{rev}} = \int_{V_1}^{V_2} P dV = RT \ln \frac{V_2}{V_1} = (1.987)(300) \ln(10/1) = 1373 \text{ cal mol}^{-1}$$

To find ΔA , we first note from $dA = -S dT - P dV$ that $(\partial A / \partial V)_T = -P$, so that

$$\Delta A = \int_{V_1}^{V_2} \left(\frac{\partial A}{\partial V}\right)_T dV = - \int_{V_1}^{V_2} P dV = w_{\text{rev}} = -1373 \text{ cal mol}^{-1}$$

Given ΔA and the fact that $\Delta(PV) = 0$ for these conditions, we readily infer that

$$\Delta G = \Delta A + \Delta(PV) = \Delta A = -1373 \text{ cal mol}^{-1}$$

but (from $dG = -S dT + V dP$), we might also have evaluated

$$\left(\frac{\partial G}{\partial V}\right)_T = V\left(\frac{\partial P}{\partial V}\right)_T = V\left(-\frac{RT}{V^2}\right) = -P$$

to obtain the same result by integration. Finally, for the entropy change,

$$\Delta S = \frac{q_{\text{rev}}}{T} = \frac{1373 \text{ cal mol}^{-1}}{300\text{K}} = 4.6 \text{ cal mol}^{-1} \text{ K}^{-1}$$

(Note that many alternative methods of solution are possible using various thermodynamic identities.)

SIDEBAR 5.15: IRREVERSIBLE ISOTHERMAL GAS EXPANSION PROBLEMS

Problem Suppose that the same isothermal gas expansion is performed as in Sidebar 5.14, but this time under *irreversible* conditions of abrupt expansion against $P_{\text{ext}} = 0$. Calculate q , w , ΔU , ΔH , ΔA , ΔG , and ΔS for this process.

Solution Since $P_{\text{ext}} = 0$, we can see that no useful work was performed,

$$w_{\text{irrev}} = 0$$

and hence (since ΔU is still zero)

$$q_{\text{irrev}} = 0$$

All other requested quantities are *state* properties. Hence, ΔU , ΔH , ΔA , ΔG , and ΔS must all have the *same* values as calculated in Sidebar 5.14.

(Note that to solve *any* problem involving an irreversible process, you should first evaluate the equivalent change of state for a reversible process, where the differentials can be easily integrated from initial to final state.)

5.8 POST-GIBBSIAN DEVELOPMENTS

Gibbs' epochal paper of 1873–76 represents a watershed in the history of thermodynamics. The following seven decades of physical chemistry research were largely devoted to working out the rich consequences of that work, which served as the master blueprint for the “classic edifice of thermodynamics.” As described by G. N. Lewis and M. Randall in the near-mystical opening words of their Preface to *Thermodynamics and the Free Energy of Chemical Substances* (McGraw-Hill, New York, 1923):

There are ancient cathedrals which, apart from their consecrated purpose, inspire solemnity and awe. Even the curious visitor speaks of serious things, with hushed voice, and as each whisper reverberates through the vaulted nave, the returning echo seems to bear a message of mystery . . . Seeing only the perfection of the completed whole, we are impressed as by some superhuman

agency. But sometimes we enter such an edifice that is still partly under construction . . . to realize that these great structures are but the result of giving to ordinary human effort a direction and a purpose. Science has its cathedrals, built by the efforts of a few architects and of many workers.

Lewis and Randall also evoke the imagery of Gibbs' paper as a thermodynamic gold mine, in a section entitled "The Modern Stage of Thermodynamics":

Except for some addenda of very recent date, the whole foundation of thermodynamics was laid before the middle of the nineteenth century . . . Next came the task of building up from these cardinal principles a great body of thermodynamic theorems . . . especially [by] J. Willard Gibbs, whose great monograph on "The Equilibrium of Heterogeneous Substances" has proved a rich and still unexhausted mine of thermodynamic material.

The following three chapters will be devoted to summarizing some leading theorems and results from Gibbs' great monograph.

Nevertheless, as noted by Lewis and Randall, certain post-Gibbsian "addenda" appeared, which will be discussed in the present section. Some of these innovations, such as "activity" and "fugacity" (Section 5.8.1), were designed primarily to satisfy practical needs of representing experimental thermochemical data, with no deeper claims on the underlying structure of the theory. In contrast, the developments initiated by Nernst's "heat theorem," culminating in what became widely known as the "third law of thermodynamics," appear to call into question the structural completeness of the Gibbsian formalism. These developments will be critically discussed in Section 5.8.2.

5.8.1 The Fugacity Concept

Let us recall the basic expression (5.53) for the P dependence of ΔG under isothermal conditions. For certain purposes, it is convenient to choose a "standard state" pressure, denoted P° , such that the free energy $G = G(P)$ for any other pressure P is given by

$$G = G(P) = G^\circ + \int_{P^\circ}^P V dP \quad (5.59)$$

If the standard state is chosen consistently for all species, the arbitrary choice of P° will cancel out of ΔG_{rxn} , ΔG_f° [compound], and related free energy changes in chemical reactions.

For an ideal gas, (5.59) can be rewritten as

$$\Delta G_{\text{ideal}} = G - G^\circ = \int_{P^\circ}^P \frac{nRT}{P} dP \quad (5.60)$$

or

$$\Delta G_{\text{ideal}} = nRT \ln \frac{P}{P^\circ} \quad (5.61)$$

Note that the argument x of $\ln(x)$ must always be a dimensionless number, and this requirement is properly satisfied for the dimensionless ratio P/P° in (5.61) for any choice of standard state. This also remains true for the special choice " $P^\circ = 1$ " (in the chosen units), where the denominator of the pressure ratio does not appear explicitly in the equation.

For real gases, we can write analogously *by definition*

$$\Delta G_{\text{real}} \equiv nRT \ln \frac{f}{f^\circ} \quad (5.62)$$

which thereby defines the “fugacity” f . By comparison with (5.61), it can be recognized that f is a type of “idealized” or “effective” pressure, and it is assigned the units of pressure. The (dimensionless) ratio γ of fugacity f and the physical pressure P ,

$$\gamma \equiv \frac{f}{P} \quad (5.63)$$

is called the “fugacity coefficient.” We can also define the “activity” a ,

$$a \equiv \frac{f}{f^\circ} \quad (5.64)$$

as the dimensionless ratio with respect to standard-state fugacity f° . From the definition (5.64), it is evident, perforce, that

$$\text{standard state} \equiv \text{state of “unit activity”} \quad (5.65)$$

The chain of definitions (5.62)–(5.65) allow us to rewrite

$$\Delta G_{\text{real}} = nRT \ln a \quad (5.66a)$$

or

$$a = \exp\left(\frac{\Delta G_{\text{real}}}{RT}\right) \quad (5.66b)$$

In effect, activity a is a way of “encoding” Gibbs free energy of real gases in equations that appear to be of ideal gas form!

It must be emphasized that the introduction of fugacity or activity (or standard state) has no physical significance, but is merely a definitional device for representing Gibbs free energies of real substances by ideal-gas-type expressions. This has the unfortunate effect of *masking* some aspects of the physical free energy relationships for real substances in equations of superficial ideal gas form. In effect, introduction of the fugacity is as superficial (and physically uninformative) as the replacement

$$\begin{array}{ccc} PV = nRT & \longrightarrow & fV = nRT \\ \text{ideal} & & \text{real} \end{array} \quad (5.67)$$

where f appears as a kind of “fudge factor” to disguise the fact that the *form* of the ideal gas equation of state should be altered for real gases (as, for example, in the Van der Waals or virial equation of state).

We can obtain a more informative expression for fugacity f or fugacity coefficient γ in terms of the general virial expansion (Section 2.4.3)

$$\frac{PV}{RT} = 1 + \frac{B}{V} + \frac{C}{V^2} + \cdots \quad (5.68)$$

where we take $V = \bar{V}$ ($n = 1$) and where B , C , ... are the empirical 2nd, 3rd, ... virial coefficients for the real gas. This equation can be inverted to obtain the corresponding

power series for $V = V(P)$:

$$V = \frac{RT}{P} + B + \frac{P}{RT}(C - B^2) + \dots \quad (5.69)$$

From the general equation (5.53) or (5.59), we therefore obtain (at constant T)

$$\begin{aligned} \frac{\Delta G_{\text{real}}}{n} &= \int_{P^\circ}^P V dP = \int_{P^\circ}^P \left[\frac{RT}{P} + B + \frac{P}{RT}(C - B^2) + \dots \right] dP \\ &= RT \ln \frac{P}{P^\circ} + B(P - P^\circ) + \frac{C - B^2}{2RT} [P^2 - (P^\circ)^2] + \dots \end{aligned} \quad (5.70)$$

The first term on the right-hand side is the idea gas limit, and the remaining *non*-logarithmic terms express the successive virial corrections for the real gas behavior. It is evidently most convenient for this problem to choose the standard state pressure as $P^\circ = 0$, where all gases are ideal. With this choice, we can write the relationship between fugacity and pressure as

$$f = P \exp \left[\frac{BP}{RT} + \frac{(C - B^2)P^2}{2R^2T^2} + \dots \right] \quad (5.71)$$

or, from the leading terms in the power series expansion of the exponential,

$$f = P \left(1 + \frac{BP}{RT} + \frac{P^2}{2R^2T^2} + \dots \right) \quad (5.72)$$

where the virial expansion for the fugacity coefficient $\gamma = \gamma(P, T)$ appears in the parentheses. Equation (5.72) gives a more incisive expression for the fugacity factor in terms of the physical virial coefficients of the gas.

5.8.2 The “Third Law” of Thermodynamics: A Critical Assessment

The most important post-Gibbsian “addenda” alluded to by Lewis and Randall (1923) are introduced in their section entitled “The Third Law of Thermodynamics” as follows:

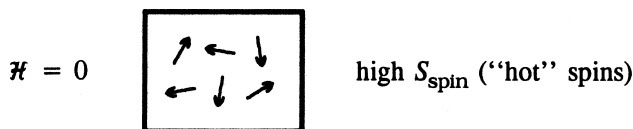
Up to this point it has been our purpose to develop as fully as possible the science which may be called classical thermodynamics, and which consists in a study of the consequences and applications of the first and second laws. It has seemed desirable to build broadly upon these two principles, which are universally accepted, before introducing certain other principles of more recent discovery, the validity of which has not yet been so completely demonstrated.

The “third law,” associated most prominently with the name of W. H. Nernst, remains to this day the most problematic and controversial of the post-Gibbsian developments.

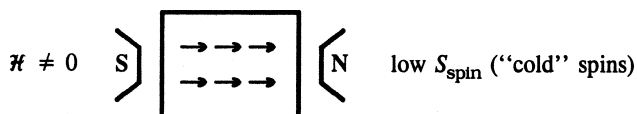
The third law is concerned with the nature of entropy (Sidebars 5.10–5.13) and thermodynamic behavior in the limiting approach toward $T = 0\text{K}$. Although Joule–Thomson expansion (Section 3.6.3) is a useful refrigeration technique down to about 20K (T_1 for H_2), more specialized cryogenic techniques are required to approach the sub-microkelvin (around 10^{-6}K) domain of extreme low temperatures. The most important such technique, *adiabatic demagnetization*, is described in Sidebar 5.16.

SIDEBAR 5.16: ADIABATIC DEMAGNETIZATION

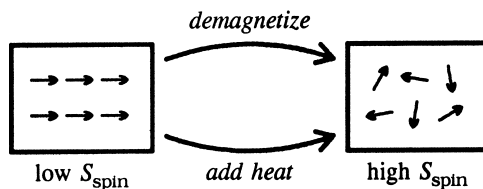
The technique of adiabatic demagnetization was developed by W. F. Giauque (Chemistry Nobel Prize 1949) and P. J. W. Debye (Chemistry Nobel Prize 1936) as a means of achieving extremely low temperatures. The method employs a paramagnetic salt (such as gadolinium sulfate) having unpaired spins at the cationic sites of the crystalline lattice. Each unpaired electron spin behaves as a tiny magnet, with the spin magnetic moment oriented *randomly* in the absence of an external magnetic field ($\mathcal{H} = 0$). The “demagnetized” state therefore corresponds to high spin-entropy S_{spin} ,



whereas the magnetized state ($\mathcal{H} \neq 0$) corresponds to ordered spin alignments of low S_{spin} :



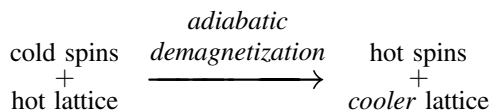
However, the spin-disordered state of high S_{spin} can alternatively be achieved by adding *heat* to the spin system:



The total entropy of the system is the sum of contributions from spins and from the crystal lattice vibrations (the latter being characterized by the temperature T of the crystal):

$$S_{\text{total}} = S_{\text{spin}} + S_{\text{lattice}}$$

Hence, if the crystal is initially in a magnetically ordered state at (lattice) temperature T (hot lattice + cold spins), but is then demagnetized under adiabatic conditions ($q = 0$), the entropy of spin disordering must be drawn from the crystal lattice (because no heat can exchange with the surroundings), and the lattice temperature *drops*:



Thus, repeated cycles of isothermal magnetization and adiabatic demagnetization lead to the desired cooling of the salt.

Entropy Changes Near Absolute Zero and “Third Law” Statements For reversible heat changes dq under constant- P conditions, with

$$dq = T dS = C_P dT \quad (5.73)$$

the entropy change dS for any substance is given by

$$dS = \frac{C_P}{T} dT \quad (5.74)$$

Integration from absolute zero to a finite T gives the absolute entropy $S(T)$:

$$S(T) = S_0 + \int_0^T \frac{C_P}{T} dT \quad (5.75)$$

where $S_0 = S(0)$ is the famous *zero-point entropy*. As cryogenic advances made it possible to measure integrals of C_P/T in the neighborhood of $T = 0$, the value of S_0 naturally became a matter of interest and speculation.

It is easy to recognize from general thermodynamic principles that both C_P and T are intrinsically non-negative. The integral on the right-hand side of (5.75) is therefore a positive number for any finite T . It is also easy to see that the entropy change $\Delta S_{\text{rxn}}(T)$ for any chemical reaction is

$$\Delta S_{\text{rxn}}(T) = S_{\text{products}}(T) - S_{\text{reactants}}(T) = \Delta S_0 + \int_0^T \frac{\Delta C_P}{T} dT \quad (5.76)$$

where ΔS_0 , the net change in zero-point entropies, is again unknown.

What are the observations near $T \rightarrow 0$? The most striking general observation is that the integrand in (5.75) is found to *vanish* as $T \rightarrow 0$,

$$\lim_{T \rightarrow 0} \frac{C_P}{T} = 0 \quad (5.77a)$$

which is to say,

$$C_P \rightarrow 0 \quad \text{faster than} \quad T \rightarrow 0 \quad (5.77b)$$

As a result, entropy changes evaluated from (5.75) become ever smaller near absolute zero, as summarized by Nernst in his “heat theorem” of 1906:

Nernst’s heat theorem (1906): As $T \rightarrow 0$, the entropy change in any reversible process tends to zero.

An important motivation for studying entropy changes at low temperature is to obtain reaction entropies ΔS_{rxn} (5.76) that could be combined with thermochemically measured reaction enthalpies ΔH_{rxn} to give the Gibbs free energy changes for chemical reactions. Starting from the observation that

$$\frac{C_P}{T} \rightarrow 0 \quad \text{as } T \rightarrow 0 \text{ for every reactant or product} \quad (5.78a)$$

one concludes that

$$\frac{\Delta C_P}{T} \rightarrow 0 \quad \text{as } T \rightarrow 0 \text{ for every chemical reaction} \quad (5.78b)$$

and therefore, consistent with Nernst's heat theorem, that

$$\Delta S_{\text{rxn}}(T \rightarrow 0) \rightarrow 0 \text{ for every reaction} \quad (5.78c)$$

including the standard formation reaction of a compound from its elements. It is therefore tempting to *assume* that the integration constant S_0 can be ignored (set to zero for each element), because the same assumption would then become valid for all substances when (5.78c) is applied to the formation reaction. This is the essence of Lewis and Randall's highly tentative formulation of the third law, stating what "may" be true "if" S_0 is assumed zero for the elements:

Lewis and Randall (1923): If the entropy of each element in some crystalline state be taken as zero at $T = 0$, then every substance has a finite positive entropy; but at $T = 0$, the entropy may be zero, and is for the case of perfect crystalline substances.

This statement also introduces the notion of "perfect crystalline substance" that underlies much of the ensuing controversy.

R. H. Fowler and E. A. Guggenheim [*Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1939)] criticized this statement as well as similar statements (to be quoted below) which imply that the entropy of perfect crystalline substances is zero. According to Fowler and Guggenheim, the only valid third-law inference is the "unattainability of absolute zero," as expressed in the following statement:

Fowler and Guggenheim (1939): It is impossible to reduce the temperature of any system to absolute zero in a finite number of steps.

The Fowler–Guggenheim statement follows immediately from (5.77b), because lowering the temperature by heat extraction cannot be effective if the heat capacity C_P vanishes faster than the desired temperature lowering.

Eventually, the rather tentative and confusing language of Lewis and Randall was replaced by the form of the third law that is quoted in nearly all textbooks today, namely

Third law (many textbooks): The entropy of a perfect crystal is zero.
or

$$S_0(\text{perfect crystal}) = 0 \quad (5.79)$$

A slight variation is given in the statement of Buckingham:

Third law (Buckingham): The entropy of a true equilibrium state is zero.

Still more questionable statements (e.g., “ $S = 0$ at $T = 0$ ”) can be found in other textbooks. The multiplicity of statements of the third law suggests its problematic character compared with other laws.

Critical Assessment of Third Law Statements Each statement of the purported third law may be critically tested in terms of three criteria:

- Is the statement *meaningful*?
- If meaningful, is the statement a *valid* general inductive law of experience?
- If meaningful and valid, does the statement convey *new* information that is independent of other inductive laws previously incorporated in the formalism?

Let us begin with the common statement (5.79) of the third law. If we inquire “What is the meaning of a ‘perfect’ crystal?”, the most direct answer appears to be “It is a crystal with $S_0 = 0$.” This circular definition insures that (5.79) is impervious to falsification, but reduces the statement to a meaningless tautology.

Moreover, even if this tautological character is accepted, the statement (5.79) apparently lacks validity for any real substance. Indeed, as shown in Sidebar 5.17, it is probable that *every* real substance has $S_0 \neq 0$, and is therefore “imperfect” in this respect. (The specific case of H_2O is described more completely in Sidebar 5.18.) We conclude that statement (5.79) is meaninglessly tautological as well as inapplicable or invalid for every known physical system. Hence, this statement fails to exhibit the rigorous inductive generality that is inherent in other thermodynamic laws.

[The Buckingham statement fares no better in this regard, for the concept of a “true” equilibrium state is no less tautological than that of a “perfect” crystal. Moreover, the implied restriction to “true” equilibrium states (presumably, those for which *no* kinetic conversion is possible on *any* timescale) is even more strongly at odds with fundamental thermodynamic definitions, as outlined in Sections 2.10 and 2.11. Indeed, such a restriction, if enforced zealously, would preclude application of thermodynamics to *any* chemical system—past, present, or future—except for the final universal *Wärmetod* state.]

Finally, it will be shown (Section 11.8) that the basic observation (5.77a) is already a *consequence* of inductive laws that were previously incorporated in the Gibbsian formalism. Thus, even the Nernst heat theorem and Fowler–Guggenheim unattainability statement (although meaningful and valid) are essentially *superfluous*, bringing no new content to the thermodynamic formalism. We therefore conclude that *all* formulations of the third law fail one or more of the above criteria, and thus play no useful thermodynamic role as “addenda” to the Gibbsian formalism.

Why then should the student of physical chemistry be acquainted with third-law statements such as (5.79)? Two motivations may be cited:

1. Equation (5.79) provides the basis for thermochemical measurements of “third-law entropies” $S_{3\text{rd}}(T)$, as described in Sidebar 5.19. More rigorous values of $S(T)$ are obtainable by T -dependent electrochemical studies, as described in Section 8.7.

The difference

$$S(T) - S_{3rd}(T) = S_0 \quad (5.80)$$

then furnishes a direct experimental measurement of S_0 , which is often of theoretical interest.

- Equation (5.79) is the thermodynamic counterpart of an idealized statistical mechanical limit, as described in Sidebar 5.20. In practice, this limit appears to be unobtainable for any real substance, but it serves as a useful limiting concept for solids that is somewhat analogous to the ideal gas “law” for gases.

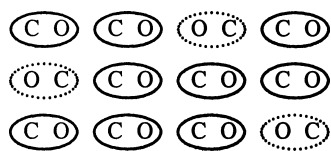
In conclusion, we may say that the “third law” in the form (5.79) is an idealized limit that is made plausible by statistical mechanics, and that underlies thermochemical measurements of “third-law entropies” for comparison with more accurate electrochemical values. However, it seems to play an essentially disposable role in the formal structure of equilibrium thermodynamics, somewhat analogous to the ideal gas concept in this respect. Equation (5.79) should not be considered a “law” in the sense that is used elsewhere in thermodynamic theory.

SIDEBAR 5.17: EXCEPTIONS TO THE THIRD LAW?

Question Are there any known “imperfect” substances for which $S_0 \neq 0$?

Answer Yes. The following is a partial listing:

- Carbon monoxide (CO) is a well-studied species in which the two ends of the diatomic molecule have similar coordinative properties. In the formation of any real crystal of solid CO, a statistical fraction of the molecules are therefore found to enter the lattice in “backward” orientation:



This orientational disorder remains “frozen” in the crystalline lattice down to $T \rightarrow 0$, leading to conspicuously nonzero $S_0 = 5.7 \text{ J mol}^{-1} \text{ K}^{-1}$. Similar well-documented cases of orientational disorder (and $S_0 \neq 0$) are exhibited by N_2O and many other species.

- Water (H_2O) exhibits a well-known zero-point entropy $S_0 = 3.4 \text{ J mol}^{-1} \text{ K}^{-1}$ that arises from the intrinsically random hydrogen-bonding network in the ice I crystal lattice, as described in Sidebar 5.18. Similar residual entropy affects the solid phase of all known aqueous mixtures, representing a large fraction of terrestrial organic or inorganic materials.
- Helium (the simplest possible elemental species) remains a *liquid* at $T = 0$ (unless the pressure is increased above about 20 atm). Bizarre quantum fluid effects and nonzero entropies are exhibited by both ^3He and ^4He in the $T \rightarrow 0$ limit.

4. H_2 (the simplest possible compound) also exhibits a well-known $S_0 \neq 0$ associated with the ortho: para distribution of nuclear spins in the crystalline lattice, arising from the fact that each H nucleus (proton) has intrinsic nuclear spin $I = \frac{1}{2}$. According to the Pauli restriction for identical fermions, the two nuclear spins of diatomic H_2 can couple into singlet (“ortho”) or triplet (“para”) spin states in statistical 3 : 1 proportions. Because the nuclear spin couplings are essentially independent of the electronic interactions that lead to formation of molecules and crystals, the ortho and para nuclear spin states distribute *randomly* throughout the H_2 lattice, leading to conspicuous $S_0 \neq 0$.
5. Related nonzero S_0 of random nuclear spin orientations must occur for *any* substance with identical nuclei of nonzero spin (i.e., any substance giving an NMR spectrum!). Thus, *all* organic and biochemical substances are expected to exhibit significant $S_0 \neq 0$ on this basis.
6. All substances containing isotopes (i.e., elements of mixed isotopic composition) must have $S_0 \neq 0$, because the isotopic variants are again distributed *randomly* through the crystalline lattice. Practically all elements of the periodic table are known to be composed of terrestrial mixtures of two or more isotopes, so the intrinsic isotopic randomness must lead to residual $S_0 \neq 0$ in practically every imaginable compound formed from terrestrial elements.
7. In addition, all glasses, all alloys, and all crystals with impurities or defects must have intrinsic $S_0 \neq 0$. (Even the most precisely controlled vapor deposition techniques are known to give a finite incidence of defect structures or impurity sites that contribute to nonvanishing S_0 .)

Each listed type of “randomization” (orientational, hydrogen-bonding network, nuclear spin statistics, isotopes, impurities, defects, and others that could be cited) makes *independent* contributions to $S_0 \neq 0$. Hence, it seems safe to conclude that *no macroscopic sample of real substance that ever appeared on Earth* satisfies $S_0 = 0$, i.e., that *every* real substance represents an “imperfect” exception to the third law as commonly stated.

SIDEBAR 5.18: ZERO-POINT ENTROPY OF ICE

The zero-point entropy of ice I was theoretically estimated by L. Pauling [*J. Am. Chem. Soc.* **57**, 2680 (1935)] by the following argument. Consider a sample composed of N water molecules in the tetrahedrally hydrogen-bonded structure of ice-I. The sample contains $2N$ hydrogen bonds (four emanating from each water molecule, but each shared between two water molecules to avoid overcounting). Each hydrogen bond, identified by two connected oxygen atoms, offers *two* possible proton positions, ($\text{O}-\text{H} \cdots \text{O}$ or $\text{O} \cdots \text{H}-\text{O}$), providing a total possible number of proton “microstates”

$$\text{number possible} = 2^{2N} = 4^N \quad (\text{S5.18-1})$$

However, many such imaginable microstates would correspond to *ionic* arrangements that are inconsistent with the “ice rule” constraints of two covalent $\text{O}-\text{H}$ bonds and two $\text{O} \cdots \text{H}$ hydrogen bonds to each oxygen atom. To evaluate the fraction of allowed microstates that are consistent with the ice rules, let us consider a chosen O atom and its four tetrahedrally

hydrogen-bonded neighbor O atoms. The total number of microstates (neutral and ionic) for these four hydrogen bonds is

$$\text{number of neutral} + \text{ionic} = 2^4 = 16 \quad (\text{S5.18-2})$$

whereas the number of ways to choose two of these four neighbors for covalent bonding (consistent with a neutral central monomer) is only

$$\text{number of neutral} = \binom{4}{2} = \frac{4!}{2!2!} = 6 \quad (\text{S5.18-3})$$

The allowed fraction of neutral microstates is therefore $6/16 = 3/8$ for this central monomer, or

$$\text{neutral fraction} = \left(\frac{3}{8}\right)^N \quad (\text{S5.18-4})$$

for the entire sample. The total number Ω of allowed neutral microstates of the sample is

$$\Omega = (\text{number possible})(\text{neutral fraction}) = 4^N \left(\frac{3}{8}\right)^N = \left(\frac{3}{2}\right)^N \quad (\text{S5.18-5})$$

From the Boltzmann expression (Sidebar 5.11) for the statistical mechanical entropy in this lowest-possible ($T = 0$) macrostate, we therefore obtain

$$S_{T=0} = k \ln \Omega = k \ln \left[\left(\frac{3}{2}\right)^N \right] = Nk \ln \frac{3}{2} = N \frac{R}{N_A} (0.405) = nR(0.405) \quad (\text{S5.18-6})$$

where n is the number of moles in the sample, and so, on a per-mole basis,

$$\bar{S}_{T=0} = 0.405R = 3.37 \text{ J mol}^{-1} \text{ K}^{-1} \quad (\text{S5.18-7})$$

which is very close to the experimentally measured value [W. F. Giauque and M. F. Ashley. *Phys. Rev.* **43**, 81 (1933); W. F. Giauque and J. W. Stout. *J. Am. Chem. Soc.* **58**, 1144 (1936)]. One can see from this argument that the essential origin of the zero-point entropy of ice is the *random proton ordering* in a tetrahedral hydrogen-bonded network. Although the overall symmetry of the ice I lattice appears tetrahedral, the local symmetry around each water molecule is broken by the ice-rule constraints, resulting in a massively degenerate set of *equivalent* “frustrated” (lowered-symmetry) ground-state proton configurations that contribute equally to zero-point entropy.

SIDEBAR 5.19: THIRD-LAW ENTROPIES

From (5.75), we can write

$$S(T) = S_0 + \int_0^T \frac{C_P}{T} dT \quad \{+ \text{ latent heat contributions} \} \quad (\text{S5.19-1})$$

where “latent heat contributions” (of the form $\Delta S_{\text{latent}} = \Delta H_{\text{latent}}/T_{\text{latent}}$) arise from the isolated points of discontinuous phase changes in the (Lebesgue) integral. The “third-law entropy” $S_{3\text{rd}}(T)$ is the corresponding estimate that results from setting $S_0 = 0$. For a substance with melting point T_m (latent heat $\Delta H_m = \Delta H_{\text{fusion}}$) and boiling point T_b (latent heat $\Delta H_b = \Delta H_{\text{vap}}$) in the temperature interval 0 to T , the third-law entropy is

$$S_{3\text{rd}}(T) = \int_0^{T_m} \frac{C_P^{\text{solid}}}{T} dT + \frac{\Delta H_{\text{fusion}}}{T_m} + \int_{T_m}^{T_b} \frac{C_P^{\text{liquid}}}{T} dT + \frac{\Delta H_{\text{vap}}}{T_b} + \int_{T_b}^T \frac{C_P^{\text{gas}}}{T} dT \quad (\text{S5.19-2})$$

Thermochemical measurements of $C_P(T)$ and latent heats ΔH_{fusion} , ΔH_{vap} therefore suffice to evaluate $S_{3\text{rd}}(T)$ for comparison with the more accurate experimental value

$$S_{\text{exp}}(T) = -\left(\frac{\partial G}{\partial T}\right)_P \quad (\text{S5.19-3})$$

obtained, for example, from T -dependent electrochemical measurements (Section 8.7).

Let us illustrate this procedure for the standard example of $S_{3\text{rd}}(298)$ of HCl. The following table summarizes contributions 1–9 to $S_{3\text{rd}}$ for the temperature ranges of individual phases [solid I (0–98.36K), solid II (98.36–158.91K), liquid (158.91–188.07K), and gas (188.07– T)] and their transitions.

	T Range (K)	Contribution Type	Value (J mol ⁻¹ K ⁻¹)	Cumulative (J mol ⁻¹ K ⁻¹)
1	0	S_0	[0.0]	[0.0]
2	0 → 16.00	(Debye extrapolation)	1.3	1.3
3	16.00 → 98.36	Solid I, $\int_{16.00}^{98.36} (C_P^{\text{I}}/T) dT$	29.5	30.8
4	98.36	Solid I → Solid II ($\Delta H_{\text{I} \rightarrow \text{II}} = 1190 \text{ J mol}^{-1}$)	12.1	42.9
5	98.36 → 158.91	Solid II, $\int_{98.36}^{158.91} (C_P^{\text{II}}/T) dT$	21.1	64.0
6	158.91	Melting ($\Delta H_{\text{fusion}} = 1992 \text{ J mol}^{-1}$)	12.6	76.6
7	158.91 → 188.07	Liquid, $\int_{158.91}^{188.07} (C_P^{\text{liq}}/T) dT$	9.9	86.5
8	188.07	Boiling ($\Delta H_{\text{vap}} = 16\,150 \text{ J mol}^{-1}$)	85.9	172.4
9	188.07 → 298.15	Gas, $\int_{188.07}^{298.15} (C_P^{\text{gas}}/T) dT$	13.5	185.9

As shown in the table, contribution 1 [$S_0 = 0$] is the third-law convention; contribution 2 is a theoretical approximation (based on the Debye heat capacity of long-wavelength vibrational modes) for the low-temperature region 0–16K, contributing only 1.3 J mol⁻¹ K⁻¹; contribution 3 is the C_P/T integral for the low-temperature solid I form of HCl; contribution 4 is the $\Delta H_{\text{I} \rightarrow \text{II}}/T_{\text{I} \rightarrow \text{II}}$ latent heat contribution (12.1 J mol⁻¹ K⁻¹) for the enantiomorphic solid I → solid II phase transition at 98.36K, and so forth.

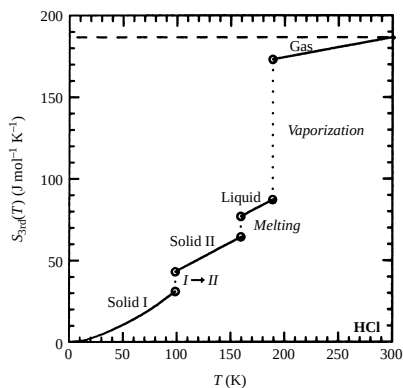
The cumulative third-law estimate, as shown in the final column, is found to be

$$S_{3\text{rd}}(298.15) = 185.9 \text{ J mol}^{-1} \text{ K}^{-1}$$

In this case, the third-law estimate is found to be in good agreement with the experimental electrochemical value, showing that the error of the third-law assumption is rather

negligible (i.e., of the order of other errors in the heat capacity measurements or the theoretical Debye extrapolation).

The following plot displays the increments in $S_{3rd}(T)$ versus T in graphical form, showing the $\int (C_P/T) dT$ contributions over single-phase regions (solid lines) and the latent-heat contributions at isolated phase transitions (dotted lines). One can see from this graph that the latent heat of vaporization at 188K makes by far the largest contribution to S_{3rd} .

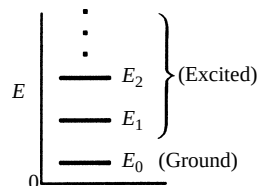


SIDEBAR 5.20: STATISTICAL MECHANICAL MODEL OF ZERO-POINT ENTROPY

According to Boltzmann's statistical mechanical theory of the entropy (Sidebar 5.11),

$$S = k \ln \Omega$$

the zero-point entropy S_0 is associated with the number of possible ways (Ω) of distributing the available thermal energy among available molecular energy levels in the limit ($T \rightarrow 0$) when all molecules are reduced to their lowest possible ("ground") energy state. From the modern perspective, energy E in molecules is described in terms of *quantized* energy levels $E_0, E_1, E_2, \dots$:



Molecular energy can only change by the discrete *quanta* ΔE that take the molecule from level E_{n-1} to the next higher energy level E_n . If the quantal energy-level spacings ΔE are equal (as in the model energy levels of the diagram), the molecular energy

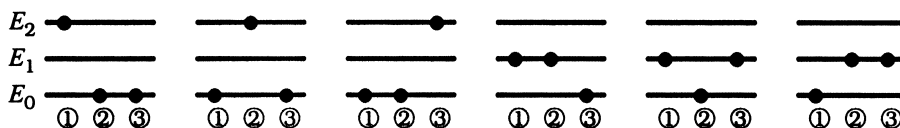
$$E_n = E_0 + n\Delta E$$

is conveniently described by the quantum number n , the "number of energy quanta" stored in the molecule.

The quantal energy “packets” are so small that the total stored molecular energy (the sum of all the molecular excitation quanta) is perceived at the macroscopic level as the continuously variable temperature T rather than a “countable” microscopic quantity. However, this countable aspect of molecular-level energy excitations underlies proper evaluation of Boltzmann’s Ω (number of possible molecular “microstates” consistent with total macrostate energy), and thus the entropy.

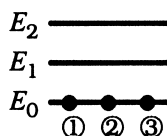
As a very simple example of this evaluation, consider a system of three molecules, each with three nondegenerate quantum energy levels (E_0 , E_1 , E_2) of equal spacing. Consider first a “high-temperature” system with $n = 2$ quanta of energy. As shown in the diagram below, these two quanta can be distributed among the available energy levels of molecules 1, 2, 3 in any of *six* possible ways ($\Omega_{n=2} = 6$), where the filled circles show the excitation level above each molecular symbol:

$$n = 2 \text{ quanta ("high } T\text{")} : \Omega_{n=2} = 6$$



Now consider a “ $T = 0$ ” system with $n = 0$ quanta of energy. As shown in the diagram below, there is only *one* possible way ($\Omega_{n=0} = 1$) to assign each molecule to its non-degenerate ground state E_0 :

$$n = 0 \text{ quanta ("} T = 0\text{")}: \Omega_{n=0} = 1$$



Thus, assuming that the ground level E_0 is truly nondegenerate (i.e., there is no alternative of equal energy), the $T = 0$ microstate is *unique*, and the corresponding Boltzmann entropy is therefore *zero*:

$$S(T = 0) = k \ln \Omega_{n=0} = k \ln 1 = 0$$

which is the statistical mechanical equivalent of the third law. However, quantal degeneracies arise in many ways, and the relevance of the idealized nondegenerate limit to realistic macroscopic systems is problematic.

Thermodynamics of Homogeneous Chemical Mixtures

The previous chapters have generally restricted attention to closed systems of a single chemical component. This restriction is quite unrealistic for chemically interesting systems, where the ebb and flow of chemical reactants and products is an essential feature of the thermodynamic behavior. However, the method of Gibbs allows this restriction to be overcome in a quite general and elegant manner, extending the domain of thermodynamics to *chemical equilibria* of arbitrary chemical complexity. In this chapter, we shall deal with such complex chemical mixtures of homogeneous (single-phase) form, whereas the following Chapter 7 extends the treatment to the *phase equilibria* characteristic of more general heterogeneous (multiphase) chemical systems.

6.1 CHEMICAL POTENTIAL IN MULTICOMPONENT SYSTEMS

The Gibbsian perspective leads naturally to the supposition that the fundamental Gibbs-space equation $U = U(S, V)$ of a single-component system can be generalized to incorporate the expected dependences on the *chemical composition* of a multicomponent chemical system. If the c independent chemical species $A_1, A_2, \dots, A_c$ are described with composition variables (mole numbers) $n_1, n_2, \dots, n_c$, we expect that the corresponding Gibbsian fundamental equation (in the energy representation) should be formulated with additional arguments $\{n_i\}$,

$$U = U(S, V, n_1, n_2, \dots, n_c) \quad (6.1)$$

to represent the dependence of thermodynamic behavior on chemical changes. Note that the energy function (6.1) is of the expected general form (cf. Section 3.3)

$$U = U(X_1, X_2, \dots, X_i, \dots, X_{c+2}) \quad (6.2)$$

in terms of the enlarged list of extensive properties

$$\{X_i\} = \{S, V, n_1, n_2, \dots, n_c\} \quad (6.3)$$

that now includes chemical composition variables.

The differential variations dU of (6.1) are given as usual by the chain rule

$$dU = \left(\frac{\partial U}{\partial S} \right)_{V, \underline{n}} dS + \left(\frac{\partial U}{\partial V} \right)_{S, \underline{n}} dV + \sum_{i=1}^c \left(\frac{\partial U}{\partial n_i} \right)_{S, V, \underline{n}} dn_i \quad (6.4)$$

where $\underline{n}$ denotes the vector list of composition variables held constant (i.e., all but an “active” dn_i) during partial differentiation. Analogous to the single-component case, we obtain the intensive variables R_i conjugate to each extensive argument X_i of U in the form

$$R_i = \left(\frac{\partial U}{\partial X_i} \right)_{\underline{X}} \quad (6.5)$$

Temperature T and pressure P are obtained as before [cf. (4.31a,b) or Table 5.2],

$$T = \left(\frac{\partial U}{\partial S} \right)_{V, \underline{n}} \quad (6.6a)$$

$$P = - \left(\frac{\partial U}{\partial V} \right)_{S, \underline{n}} \quad (6.6b)$$

to give the generalized form of the first and second laws [cf. (4.28)],

$$dU = TdS - PdV + \sum_{i=1}^c \left(\frac{\partial U}{\partial n_i} \right)_{S, V, \underline{n}} dn_i \quad (6.7)$$

The coefficients $(\partial U / \partial n_i)_{S, V, \underline{n}}$ of the dn_i in (6.7) are evidently an important new set of intensive properties that control the *chemical flows* (analogous to the manner in which T controls the entropy flow and P the volume flow). Following Gibbs, we identify each coefficient $(\partial U / \partial n_i)_{S, V, \underline{n}}$ as the *chemical potential* (μ_i) of the corresponding species A_i :

$$\mu_i = \left(\frac{\partial U}{\partial n_i} \right)_{S, V, \underline{n}}, \quad i = 1, 2, \dots, c \quad (6.8)$$

With this notation, the combined first/second law is written as

$$\boxed{dU = TdS - PdV + \sum_{i=1}^c \mu_i dn_i} \quad (6.9a)$$

or, more generally, in terms of conjugate extensive (X_i) and intensive (R_i) properties,

$$dU = \sum_{i=1}^{c+2} R_i dX_i \quad (6.9b)$$

With the inclusion of chemical potential terms, (6.9a,b) now incorporate the important variations of internal energy with chemical changes.

The modification of U for chemical composition variations leads also to corresponding modifications of other thermodynamic potentials, such as enthalpy $H = U + PV$,

$$dH = TdS + VdP + \sum_{i=1}^c \mu_i dn_i \quad (6.10a)$$

Helmholtz free energy $A = U - TS$,

$$dA = -SdT - PdV + \sum_{i=1}^c \mu_i dn_i \quad (6.10b)$$

and Gibbs free energy $G = H - TS$,

$$dG = -SdT + VdP + \sum_{i=1}^c \mu_i dn_i \quad (6.10c)$$

Chemical potential μ_i can therefore be expressed by any of the alternative definitions

$$\mu_i = \left(\frac{\partial H}{\partial n_i} \right)_{S,P,\underline{n}} \quad (6.11a)$$

$$\mu_i = \left(\frac{\partial A}{\partial n_i} \right)_{T,V,\underline{n}} \quad (6.11b)$$

$$\boxed{\mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{T,P,\underline{n}}} \quad (6.11c)$$

Although each definition (6.11a–c) is fully equivalent to (6.8), the G -based equation (6.11c) is most convenient for defining μ_i under the usual laboratory conditions of constant T and P .

6.2 PARTIAL MOLAR QUANTITIES

The variations $(\partial/\partial n_i)$ with respect to chemical species (usually at constant T, P) are referred to as “partial molar” quantities. For any molar extensive property $\bar{X}$, the partial molar derivative $\bar{X}_i$ with respect to chemical component i is given by

$$\bar{X}_i \equiv \left(\frac{\partial \bar{X}}{\partial n_i} \right)_{T,P,\underline{n}} \quad (6.12)$$

Thus, (6.11c) shows that chemical potential μ_i could also be described as “partial molar Gibbs energy”:

$$\mu_i = \bar{G}_i \quad (6.13)$$

However, the implications and usefulness of partial molar quantities go far beyond this example.

Let us consider the special case of partial molar volume $\bar{V}_i$:

$$\bar{V}_i \equiv \left(\frac{\partial \bar{V}}{\partial n_i} \right)_{T,P,\underline{n}} \quad (6.14)$$

We can also describe $\bar{V}_i$ in non-calculus terms as follows:

Partial molar volume $\bar{V}_i$: The change in molar volume $\bar{V}$ when one mole of A_i is added to an “infinitely large” sample at constant T and P .

From the definition (6.14) and the general chain rule under conditions of constant T and P , we can write the general differential of volume variation dV as

$$dV = \bar{V}_1 dn_1 + \bar{V}_2 dn_2 + \cdots + \bar{V}_c dn_c \quad (6.15)$$

Let us now integrate dV “drop by drop” at *constant concentration* from initial droplet size $n_i = 0$ to final solution size n_i for each component:

$$\begin{aligned} V &= \int_{\text{droplet}}^{\text{full size}} dV = \int_0^{n_1} \bar{V}_1 dn_1 + \int_0^{n_2} \bar{V}_2 dn_2 + \cdots + \int_0^{n_c} \bar{V}_c dn_c \\ &= n_1 \bar{V}_1 + n_2 \bar{V}_2 + \cdots + n_c \bar{V}_c \end{aligned} \quad (6.16)$$

where we have used the fact that each $\bar{V}_i$ = constant throughout the integration range, which allows $\bar{V}_i$ to be taken outside the integral sign. (The foregoing is *not* the way the solution is usually prepared chemically, but it is an imaginable process for the present mathematical purposes.) From the integral (6.16), we conclude that

$$V = \sum_{i=1}^c n_i \bar{V}_i \quad (6.17)$$

showing that total volume V is strictly *additive* in the partial molar volumes $\bar{V}_i$.

It may seem dilettantish to decompose volume V in terms of partial molar volumes, as in (6.17). Would it not just be simpler to write the total volume as the sum of the volumes of the pure components? Although it may seem “obvious” that the total volume of, for example, a binary (two-component) solution is just the sum of solute and solvent volumes,

$$V_{\text{solution}} \stackrel{?}{=} V_{\text{solute}} + V_{\text{solvent}} \quad (6.18)$$

experiment shows (see Sidebar 6.1) that (6.18) is generally *not* true! More specifically, if $V_A^\bullet$ denotes the molar volume of pure A and $V_B^\bullet$ that of pure B, then the volume of a solution composed of n_A moles of A and n_B moles of B will generally *not* equal the sum of volumes of initial components:

$$V \neq n_A V_A^\bullet + n_B V_B^\bullet \quad (6.19)$$

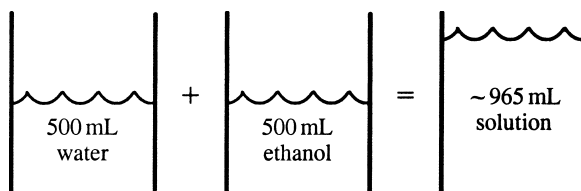
or, for the analogous solution of c components,

$$V \neq \sum_{i=1}^c n_i V_i^\bullet \quad (6.20)$$

However, the analogous additivity relation (6.17) for partial molar volumes is strictly satisfied for all solutions.

SIDEBAR 6.1: NONADDITIVE VOLUME OF SOLUTION

Problem Suppose that 500 mL of pure H₂O are added to 500 mL of pure C₂H₅OH. What is the volume of the resulting solution? How does the result change if the components are ethyl acetate (C₂H₅OOCCCH₃) and ethyl iodide (C₂H₅I)?



Answer As shown in the diagram, the ethanol/water solution volume is found to be conspicuously *less* (by 3–4%) than 1000 mL, showing that $V_A^\bullet + V_B^\bullet > V_{\text{solution}}$ in this case. However, for components A = ethyl acetate and B = C₂H₅I or CS₂, one finds instead that $V_A^\bullet + V_B^\bullet < V_{\text{solution}}$ (by 1–2%). The deviations from additivity may therefore be of either sign, and are attributable in each case to changes in coordination and packing patterns at the molecular level, particularly dramatic for hydrogen-bonding liquids.

Similar partial molar additivity relationships can be derived for *any* extensive property X :

$$X = n_A \bar{X}_A + n_B \bar{X}_B + \cdots \quad (\text{constant } T, P) \quad (6.21)$$

Thus, for example, for entropy, enthalpy, or Gibbs free energy (all at constant T, P),

$$S = n_A \bar{S}_A + n_B \bar{S}_B + \cdots \quad (6.22a)$$

$$H = n_A \bar{H}_A + n_B \bar{H}_B + \cdots \quad (6.22b)$$

$$G = n_A \bar{G}_A + n_B \bar{G}_B + \cdots = n_A \mu_A + n_B \mu_B + \cdots \quad (6.22c)$$

Analogous additivity relationships also apply to any *change* of state ΔX

$$\text{initial} \xrightarrow{\Delta X} \text{final} \quad (6.23a)$$

namely,

$$\Delta X = n_A \Delta \bar{X}_A + n_B \Delta \bar{X}_B + \cdots \quad (6.23b)$$

(for $\Delta = \Delta_{\text{soln}}, \Delta_{\text{phase change}}, \Delta_{\text{rxn}}, \text{etc.}$). Sidebar 6.2 illustrates such relationships for heats of solution ($\Delta X = \Delta H_{\text{soln}}$), showing the relationship between integral and differential heats of solution (previously introduced in Section 3.6.7 in somewhat less general notation) as a special case of (6.23). The analogous Gibbs free energy relationships are expressed most directly in terms of shifts $\Delta \mu_A, \Delta \mu_B$ in chemical potential (Sidebar 6.2).

How are partial molar quantities determined experimentally? Sidebar 6.3 illustrates the general procedure for the special case of the partial molar volumes $\bar{V}_A, \bar{V}_B$ of a binary solution (analogous to the graphical procedure previously employed in Section 3.6.7 for finding differential heats of solution). As indicated in Sidebar 6.3, each partial molar

quantity $\bar{X}_A = \bar{X}_A(T, P, \underline{n})$, $\bar{X}_B = \bar{X}_B(T, P, \underline{n})$, ... is generally dependent on the concentration variables $\underline{n}$ as well as the chosen T, P values held constant during differentiation. Despite their apparent simplicity, partial molar equations such as (6.21)–(6.23) accurately describe all details of actual solution composition and external T, P conditions.

SIDEBAR 6.2: INTEGRAL AND DIFFERENTIAL HEATS AND FREE ENERGIES OF SOLUTION

The “integral” heat (enthalpy) of solution ΔH_{soln} is defined as the total heat liberated (under constant- P conditions) when a solution is formed from its pure components A, B:

$$\Delta H_{\text{soln}} = H_{\text{solution}} - H_{\text{pure components}} \quad (\text{S6.2-1})$$

Using (6.22b) for the solution and pure components ($H_A^\bullet, H_B^\bullet$), we obtain

$$\begin{aligned} \Delta H_{\text{soln}} &= (n_A \bar{H}_A + n_B \bar{H}_B) - (n_A H_A^\bullet + n_B H_B^\bullet) \\ &= n_A (\bar{H}_A - H_A^\bullet) + n_B (\bar{H}_B - H_B^\bullet) \\ &= n_A \Delta \bar{H}_A + n_B \Delta \bar{H}_B \end{aligned} \quad (\text{S6.2-2})$$

where the “differential heats of solution” $\Delta \bar{H}_A, \Delta \bar{H}_B$ of the two components are thereby defined as

$$\begin{aligned} \Delta \bar{H}_A &\equiv \bar{H}_A - H_A^\bullet \\ \Delta \bar{H}_B &\equiv \bar{H}_B - H_B^\bullet \end{aligned} \quad (\text{S6.2-3})$$

Similar relationships can be derived for the Gibbs free energy changes ($\Delta G = \Delta G_{\text{soln}}$) of solution formation. Starting from (6.22c), we obtain in a similar manner

$$\Delta G = n_A \Delta \mu_A + n_B \Delta \mu_B \quad (\text{S6.2-4})$$

where the shifts in chemical potential

$$\begin{aligned} \Delta \mu_A &\equiv \mu_A - \mu_A^\bullet \\ \Delta \mu_B &\equiv \mu_B - \mu_B^\bullet \end{aligned} \quad (\text{S6.2-5})$$

correspond to the differential Gibbs free energy changes $\Delta \bar{G}_A, \Delta \bar{G}_B$ in solution formation.

SIDEBAR 6.3: EXPERIMENTAL DETERMINATION OF PARTIAL MOLAR VOLUME

For a two-component solution composed of A (“solvent”) and B (“solute”), (6.17) becomes

$$V = n_A \bar{V}_A + n_B \bar{V}_B \quad (\text{S6.3-1})$$

It is convenient to express this on a per-mole basis, dividing through by the total number of moles ($n_A + n_B$) to obtain the equivalent equation in terms of mole fraction x_B

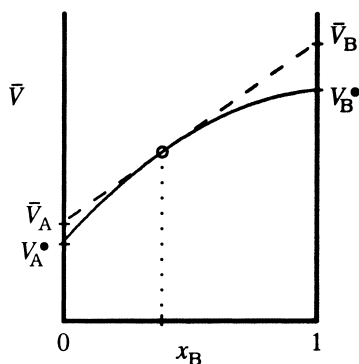
(with $x_A = 1 - x_B$):

$$\bar{V} = x_A \bar{V}_A + x_B \bar{V}_B = (1 - x_B) \bar{V}_A + x_B \bar{V}_B = \bar{V}_A + x_B (\bar{V}_B - \bar{V}_A) \quad (\text{S6.3-2})$$

As shown by the final equation on the right, the equation for $\bar{V} = \bar{V}(x_B)$ is that of a *straight line* with slope

$$\frac{d\bar{V}}{dx_B} = \bar{V}_B - \bar{V}_A \quad (\text{S6.3-3})$$

and intercept $\bar{V}_A$. Thus, this straight line corresponds to the *tangent* of $\bar{V} = \bar{V}(x_B)$ at the concentration of interest, with slope and intercept expressed in terms of the desired partial molar volumes $\bar{V}_A$, $\bar{V}_B$. The graph below shows the plotted behavior of $\bar{V}(x_B)$ (solid line) from $\bar{V}_A^\bullet$ (at $x_B = 0$) to $\bar{V}_B^\bullet$ (at $x_B = 1$), with its corresponding tangent line (dashed) at the concentration of interest (circle, dotted line). As shown in the graph, extrapolation of the tangent line to the axes yields $\bar{V}_A$ (on the left) and $\bar{V}_B$ (on the right). The extrapolated-tangent construction therefore makes it easy to read off the partial molar volumes $\bar{V}_A$, $\bar{V}_B$ at any desired concentration x_B .



6.3 THE GIBBS–DUHEM EQUATION

In this section, we wish to derive the Gibbs–Duhem equation, the fundamental relationship between the allowed variations dR_i of the intensive properties of a homogeneous (single-phase) system. Paradoxically, this relationship (which underlies the entire theory of phase equilibria to be developed in Chapter 7) is discovered by considering the fundamental nature of *extensive* properties X_i , as well as the intrinsic “scaling” property of the fundamental equation $U = U(S, V, n_1, n_2, \dots, n_c)$ that derives from the extensive nature of U and its Gibbs-space arguments.

Recall from Section 2.10 that the characteristic feature of extensive properties X_i is their uniform *scaling* with respect to the “size” of the system, expressible in terms of a multiplicative positive scale factor λ . “Re-sizing” the macroscopic system merely means that all extensive properties are multiplied by the common scale factor λ ,

$$X_i \longrightarrow \lambda X_i \quad (6.24)$$

or, more specifically,

$$U \longrightarrow \lambda U \quad (6.25a)$$

$$S \longrightarrow \lambda S \quad (6.25b)$$

$$V \longrightarrow \lambda V \quad (6.25c)$$

$$n_i \longrightarrow \lambda n_i \quad (i = 1, 2, \dots, c) \quad (6.25d)$$

We begin by rewriting the Gibbs fundamental energy equation $U = U(S, V, n_1, n_2, \dots, n_c)$ in symbolic form, with $t = c + 2$ extensive arguments X_i :

$$U = U(X_1, X_2, \dots, X_t) \quad (6.26)$$

The general scaling property of a macroscopic system can then be expressed by the mathematical relationship

$$U(\lambda X_1, \lambda X_2, \dots, \lambda X_t) = \lambda U(X_1, X_2, \dots, X_t), \text{ all } \lambda \quad (6.27)$$

Equation (6.27) merely says that if the independent extensive arguments of U are multiplied by λ [cf. (6.25b–d)], then U itself must be multiplied by the same factor [cf. (6.25a)]. [Mathematically, the property (6.27) identifies the internal energy function (6.26) as a “homogeneous function of first order,” and the consequence to be derived is merely a special case of what is called “Euler’s theorem for homogeneous functions” in your college algebra textbook.]

Let us implicitly differentiate both sides of (6.27) with respect to λ , using the general chain rule to obtain

$$\sum_{i=1}^t \frac{\partial U(\dots \lambda X_i \dots)}{\partial(\lambda X_i)} \frac{\partial(\lambda X_i)}{\partial \lambda} = \sum_{i=1}^t \frac{\partial U(\dots \lambda X_i \dots)}{\partial(\lambda X_i)} X_i = U(X_1, X_2, \dots, X_t) \quad (6.28)$$

Equation (6.28), being true for all positive λ , must also be true for the special case $\lambda = 1$, namely,

$$\sum_{i=1}^t \frac{\partial U}{\partial X_i} X_i = \sum_{i=1}^t R_i X_i = U \quad (6.29)$$

where we have used the general conjugacy relationship

$$R_i = \left(\frac{\partial U}{\partial X_i} \right)_{\underline{X}}, \quad i = 1, 2, \dots, t \quad (6.30)$$

to rewrite (6.29) (Euler’s theorem) in terms of intensive properties R_i .

Let us rewrite the symbolic equation (6.29) in terms of explicit thermodynamic properties R_i , X_i :

$$U = \sum_{i=1}^t R_i X_i = TS - PV + \sum_{i=1}^c \mu_i n_i \quad (6.31)$$

This is a rather surprising result, as we can see by differentiating (6.31) to obtain

$$dU = TdS + SdT - PdV - VdP + \sum_{i=1}^c (\mu_i dn_i + n_i d\mu_i) \quad (6.32)$$

However, we know [from the general first/second law, (6.7)]

$$dU = TdS - PdV + \sum_{i=1}^c \mu_i dn_i \quad (6.33)$$

Comparing the two expressions (6.32), (6.33) for dU , we can see that the “extra” terms in (6.32) must *vanish*, i.e.,

$$\boxed{SdT - VdP + \sum_{i=1}^c n_i d\mu_i = 0} \quad (6.34)$$

Equation (6.34) is the *Gibbs–Duhem equation*, showing that the intensive variations dT , dP , $d\mu_i$ are *not all independent*. Equation (6.34) therefore expresses a deep interconnectedness of the intensive properties that underlies the Gibbsian formalism.

Let us now derive some simple consequences of the Gibbs–Duhem equation for special cases.

For the simple single-component ($c = 1$) systems previously considered in Chapters 1–5, the Gibbs–Duhem equation (6.34) reduces to

$$SdT - VdP + nd\mu = 0 \quad (6.35a)$$

for n moles of pure substance. Dividing through by n to express quantities in molar terms, we obtain the equation for chemical potential in this simple system:

$$d\mu = -\bar{S}dT + \bar{V}dP \quad (c = 1) \quad (6.35b)$$

Equation (6.35b) shows that the “new” intensive variable, chemical potential μ , as introduced in this chapter, is actually *superfluous* for the case $c = 1$, because its variations can always be expressed in terms of the “old” variations dT , dP . Thus, as stated in Inductive Law 1 (Table 2.1), only *two* degrees of freedom (independently variable intensive properties) suffice to describe the thermodynamic variability of a simple $c = 1$ system. This confirms (as expected) that chemical potential μ only becomes an informative thermodynamic variable when chemical *change* is possible, that is, for $c \geq 2$ chemical components.

We may also note the special form of the Gibbs–Duhem equation (6.34) under laboratory conditions of constant T and P , namely,

$$\sum_{i=1}^c n_i d\mu_i = 0 \quad (\text{at constant } T, P) \quad (6.36a)$$

or, if we divide by the total number of moles to express the relationship in mole fraction terms,

$$\sum_{i=1}^c x_i d\mu_i = 0 \quad (\text{at constant } T, P) \quad (6.36b)$$

For a binary (two-component) solution of solute B and solvent A, for example, this establishes that the solute variations $d\mu_B$ are always calculable from the solvent variations $d\mu_A$ by the equation

$$d\mu_B = -\frac{1-x_B}{x_B}d\mu_A \quad (c=2; \text{constant } T, P) \quad (6.37)$$

Because variations in *solvent* chemical potential are generally much easier to determine experimentally (e.g., by osmotic pressure measurements, as described in Section 7.3.6), (6.37) gives the recipe for determining the more difficult $d\mu_{\text{solute}}$ from its Gibbs–Duhem dependence on other easily measured thermodynamic intensities. Equations such as (6.35)–(6.37) are sometimes referred to as “Gibbs–Duhem equation(s),” but they are really only special cases of (and thus less general than) “the” Gibbs–Duhem equation (6.34).

Finally, let us note some interesting identities for other thermodynamic potentials that follow from Equation (6.31). From the energy identity $U = TS - PV + \sum_i \mu_i n_i$ and basic thermodynamic definitions, we can readily infer that

$$H = U + PV = TS + \sum_{i=1}^c \mu_i n_i \quad (6.38)$$

$$A = U - TS = -PV + \sum_{i=1}^c \mu_i n_i \quad (6.39)$$

$$G = U + PV - TS = \sum_{i=1}^c \mu_i n_i \quad (6.40)$$

Equation (6.40) for G was previously obtained as (6.22c), and will serve as the starting point for discussion of chemical equilibria (Chapter 8).

6.4 PHYSICAL NATURE OF CHEMICAL POTENTIAL IN IDEAL AND REAL GAS MIXTURES

In order to better understand the physical nature of the chemical potential μ_i of a chemical substance, let us first review the major mathematical features of the Gibbsian thermodynamics formalism. The starting point is the Gibbs “fundamental equation” for the internal energy function

$$U = U(X_i), \quad X_i = \{S, V, n_1, n_2, \dots, n_c\}, \quad c+2 \text{ arguments} \quad (6.41)$$

From the fundamental equation, we derive the “equations of state” for the intensive fields R_i as the successive partial derivatives

$$R_i = \left(\frac{\partial U}{\partial X_i} \right)_X, \quad i = 1, 2, \dots, c+2 \quad (6.42)$$

each extensive variable X_i being thereby linked to its “conjugate” intensive field variable R_i . In terms of these conjugate pairs (R_i, X_i) , the combined first/second law is given by

$$dU = \sum_{i=1}^{c+2} R_i dX_i \quad (6.43)$$

while the Gibbs–Duhem equation is given by

$$0 = \sum_{i=1}^{c+2} X_i dR_i \quad (6.44)$$

Equations (6.41)–(6.44) summarize the basic mathematical structure of the theory.

Let us now consider a system that is initially partitioned into two regions α , β of differing R_i values,

$$\begin{array}{c|c} \alpha & \beta \\ \hline \text{high } R_i & \text{low } R_i \end{array} \quad (6.45)$$

If the partition is removed, we know by the Gibbs criterion that quantity X_i will spontaneously migrate from the high- R_i α region to the low- R_i β region, allowing extraction of useful work

$$dw_i = R_i dX_i \quad (6.46)$$

until the values of R_i in the two regions are *equalized* at equilibrium. Generic equilibration processes corresponding to (6.46) are depicted schematically in Fig. 6.1 for three specific types of “driving force”: (a) thermal, (b) electrical, and (c) chemical. Each panel of the figure may suggest the image of water falling over a waterwheel under the driving force of the gravitational field. From the analogies presented in Fig. 6.1, one can verbally characterize chemical potential μ_i as

$$\mu_i = \text{the driving force (field) for } \textit{chemical} \text{ migration} \quad (6.47)$$

Chemical potential is analogous to the temperature gradient that drives heat flow or the cell emf potential that drives electrical current flow, in that it provides the driving force for *diffusive migration of chemical species* from one region of the system to another.

Other characterizations of chemical potential follow from the single-component limit of (6.22c), where

$$\mu = G/n = \overline{G} \quad (6.48)$$

	(a) Thermal	(b) Electrical	(c) Chemical
α (high R_i)	High T	High emf	High μ_i
Transfer quantity X_i	Entropy S	Charge Q	Chemical n_i
			
β (low R_i)	Low T	Low emf	Low μ_i
$R_i dX_i$	$T dS$	$\text{emf } dQ$	$\mu_i dn_i$

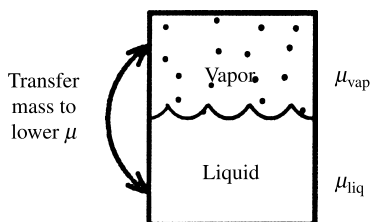
Figure 6.1 Schematic depiction of Gibbs equilibration for three driving forces: (a) thermal (T difference), (b) electrical ($\mathcal{E}_{\text{emf}}$ difference), (c) chemical (μ_i difference), showing the transported quantity X_i and available work $R_i dX_i$ for each driving field R_i .

We can therefore describe μ in terms similar to those used previously for Gibbs free energy (Section 5.7), namely,

$$\mu = \text{capacity for additional non-}PV \text{ work} \quad (6.49)$$

(by migration of mass to a phase of lower μ)

One can also consider the liquid/vapor equilibria in a simple ($c = 1$) two-phase system with $\mu_{\text{vap}} \neq \mu_{\text{liq}}$:



In this case, molecules of the chemical species will “escape” from the phase of higher μ to that of lower μ until the chemical potentials are equalized (with altered liquid/vapor proportions):

$$\mu = \text{“escaping tendency” of the chemical species} \quad (6.50)$$

Each verbal characterization (6.47), (6.49), (6.50) complements the strict mathematical definition (6.11c) of chemical potential.

Both (6.47) and (6.50) suggest that μ_i is a kind of “chemical pressure.” In single-component gaseous systems, there is indeed a close connection between μ and P , which can be seen as follows. For $c = 1$, where (6.48) is valid, let us consider the effect of a pressure variation $P_1 \rightarrow P_2$ on μ (or $\bar{G}$) under isothermal conditions:

$$\Delta\mu = \Delta\bar{G} = \int_{P_1}^{P_2} \left(\frac{\partial \bar{G}}{\partial P} \right)_T dP = \int_{P_1}^{P_2} \bar{V} dP \quad (6.51)$$

If we choose (for convenience) a “standard state” pressure P° , this can also be rewritten as

$$\mu(P) = \mu^\circ + \int_{P^\circ}^P \bar{V} dP \quad (6.52)$$

for any pure gas. For the special case of an ideal gas, we can also write

$$\mu(P) \stackrel{\text{IG}}{=} \mu^\circ + \int_{P^\circ}^P \frac{RT}{P} dP \stackrel{\text{IG}}{=} \mu^\circ + RT \ln \frac{P}{P^\circ} \quad (6.53)$$

which shows the logarithmic dependence of μ on P in this limit.

Let us now consider a c -component mixture of ideal gases. In this case, Dalton’s law of partial pressures tells us that each gas exerts a *partial* pressure P_i proportional to its

mole fraction,

$$P_i^{\text{IG}} \equiv x_i P \quad (6.54)$$

as though it occupies the entire volume by itself, oblivious to the presence of other components. In this ideal noninteracting limit, each gas $i = 1, 2, \dots, c$ of the mixture has chemical potential μ_i equal to its value (6.53) in the single-component limit,

$$\mu_i = \mu_i(P)^{\text{IG}} \equiv \mu_i^\circ + RT \ln \frac{P_i}{P^\circ} \quad (6.55)$$

where a common standard-state P° is chosen for all components. We can substitute (6.54) into (6.55) to obtain

$$\mu_i^{\text{IG}} \equiv \left[\mu_i^\circ + RT \ln \frac{P_i}{P^\circ} \right] + RT \ln x_i \quad (6.56)$$

which can be rewritten as

$$\mu_i^{\text{IG}} \equiv \mu_i^* + RT \ln x_i \quad (6.57)$$

where

$$\mu_i^* \equiv \mu_i^\circ + RT \ln \frac{P}{P^\circ} \quad (6.58)$$

is the (ideal) chemical potential of *pure* component i at the chosen pressure P . On the right-hand side of (6.57), the first term μ_i^* refers to pure component i (at the chosen P), whereas the second term $RT \ln x_i$ is the *dilution effect* on chemical potential.

The defining characteristic of ideal gas mixtures is the absence of any interactions. Thus, all thermodynamic properties *separate* into their “partial” contributions; for example,

$$P \equiv \sum_{i=1}^c P_i \quad (\text{Dalton's law}) \quad (6.59a)$$

$$U \equiv \sum_{i=1}^c U_i \quad (6.59b)$$

$$S \equiv \sum_{i=1}^c S_i = -R \sum_{i=1}^c n_i \ln x_i \quad (6.59c)$$

$$H \equiv \sum_{i=1}^c H_i \quad (6.59d)$$

and so forth. Equation (6.59c) corresponds to the isothermal mixing of ideal gases that was previously derived in Sidebar 5.10. Thus, it is rather trivial to calculate thermodynamic properties of ideal gas mixtures from the expressions derived previously for pure components (see, e.g., Section 3.6.4).

Following the philosophy of Section 5.8.1, it is also possible (if uninformative) to express μ_i for *real* gas mixtures in a form that mimics the ideal gas expression (6.57), namely,

$$\mu_i = \mu_i^\circ + RT \ln a_i \quad (6.60)$$

Here, a_i is called the “activity” (“effective” concentration) of species i . Activity a_i is essentially *defined* by (6.60), or the equivalent

$$a_i \equiv \frac{\exp(\mu_i - \mu_i^\circ)}{RT} \quad (6.61)$$

i.e., it allows one to re-express (or “encode”) μ_i in formulas of ideal gas form. The related “fugacity” concept was discussed in Section 5.8.1, and further discussion of activity and activity coefficients will be given in Section 7.3.7. Expressions such as (6.60) have become time-honored in chemical thermodynamics textbooks, and their usage is rather standard in tabulations of experimental thermochemical data, even if the conceptual gain of replacing chemical potential by activity is debatable.

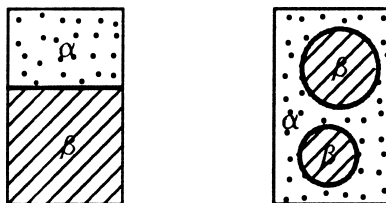
Thermodynamics of Phase Equilibria

The concept of “phase” (*φασις*: “appearance”) underlies some of the most remarkable phenomena of thermodynamics, and the complete elucidation of phase equilibrium phenomena represents the most famous achievement of Gibbsian thermodynamics. This chapter describes how the Gibbs principles extend almost effortlessly to such complex multiphase systems.

Let us begin with a formal definition of thermodynamic phase:

Phase: a *uniform* region of the system, with distinct boundaries (7.1)

Phase uniformity extends to all aspects of visual appearance (“look”), physical texture (“feel”), electrical and magnetic properties, and chemical composition. The phase boundary (interface) may be a planar surface (as on the left in the following diagram), the normal separation geometry between phases α , β of different density in a gravitational field, or a *domain* structure (as on the right), with regular or irregular islands of β phase embedded in a surrounding matrix of α phase:



Each system may be characterized as “homogeneous” or “heterogeneous” according to the number of distinct phases, p :

Homogeneous: single phase ($p = 1$) (7.2a)

Heterogeneous: two or more phases ($p \geq 2$) (7.2b)

The Gibbsian equilibration principles as developed previously (Chapter 5) apply straightforwardly to heterogeneous as well as homogeneous systems. Consider, for example, the heterogeneous two-phase system of phases α , β . Extensive properties such

as S , V , and mole numbers n_i must be simply *additive* (Section 2.10) in the subsystems:

$$S = S_\alpha + S_\beta \quad (7.3a)$$

$$V = V_\alpha + V_\beta \quad (7.3b)$$

$$n_i = n_{i\alpha} + n_{i\beta} \quad (i = 1, 2, \dots, c) \quad (7.3c)$$

The Gibbs conditions of total entropy maximization under isolation conditions are (following the example of Section 5.2)

$$\tilde{S}' = (S)'_{U,V,\underline{n}} = (S_\alpha + S_\beta)'_{U,V,\underline{n}} = 0 \quad (7.4a)$$

$$S'' = S''_\alpha + S''_\beta < 0 \quad (7.4b)$$

These lead, as before, to equalization of the respective conjugate intensive properties [from the first-derivative requirement (7.4a)],

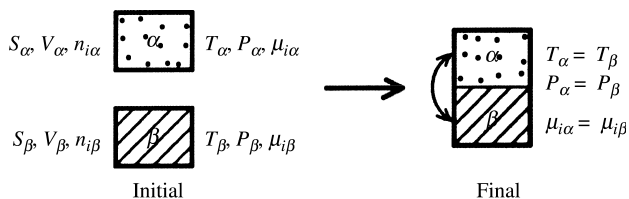
$$T_\alpha = T_\beta, \quad \text{uniform temperature throughout} \quad (7.5a)$$

$$P_\alpha = P_\beta, \quad \text{uniform pressure throughout} \quad (7.5b)$$

$$\mu_{i\alpha} = \mu_{i\beta}, \quad \text{uniform chemical potential } \mu_i \text{ throughout} \quad (7.5c)$$

as well as to various “stability conditions” [from the second-derivative requirement (7.4b)]. Alternatively, the optimizations might have been performed equivalently in the energy representation (Section 5.4). Exchange of extensities (7.3a–c) and equalization of intensities (7.5a–c) throughout the multiphase system therefore follow directly from Gibbs’ criterion of equilibrium for the heterogeneous case.

Consider, for example, the following change of state from an initial state consisting of separated α and β phases (left) to the final state consisting of the two phases in full equilibrium contact (right):



In accordance with the equilibration conditions, if initial $T_\alpha > T_\beta$, then entropy S will be transferred from α to β until the temperatures are equalized; similarly, if initial $P_\alpha > P_\beta$, then volume V will be transferred $\alpha \rightarrow \beta$; and if initial $\mu_{i\alpha} > \mu_{i\beta}$, then species n_i will be transferred $\alpha \rightarrow \beta$ until all intensive variables are equalized between the coexisting phases, as shown at the right. (Note that the densities of coexisting phases are generally distinct, so density certainly lacks “intensive” status in the sense of the equilibration criterion and the degrees of freedom enumerated by the Gibbs phase rule, as discussed in the following section.)

7.1 THE GIBBS PHASE RULE

The Gibbs phase rule

$$f = c - p + 2 \quad (7.6)$$

expresses the simple theoretical relationship between

$$\begin{aligned} f &= \text{“degrees of freedom”} \\ &= \text{number of independent intensities} \\ &\quad \text{that specify the thermodynamic state} \end{aligned} \quad (7.7a)$$

$$\begin{aligned} c &= \text{minimum number of independent chemical components} \\ &\quad \text{to specify composition (of any phase)} \end{aligned} \quad (7.7b)$$

$$p = \text{number of distinct phases} \quad (7.7c)$$

that underlies the thermodynamics of phase equilibria.

Proper evaluation of the number of “degrees of freedom” constitutes the important first step in characterizing the equilibrium states of a given thermodynamic system. Operationally, the number and type of macroscopic properties needed to uniquely identify a thermodynamic state might be determined as follows. Suppose that numerical values of a selected set of macroscopic properties are known, and that many copies of the system are now prepared having all the same values of these properties. If an independent observer can detect a macroscopic difference in any of these copies (other than spatial location, shape, etc.), the overall procedure can be repeated with the value of this new property being specified in each member of the ensemble. After a sufficient number of properties have thereby been added, it will eventually be found that no further macroscopic differences can be detected, and the copies are then said to share the same thermodynamic state. Empirically, it is found that only a small, fixed number f of distinct macroscopic properties is required to uniquely specify any such state, as predicted by the Gibbs phase rule (7.6). As remarked in Section 2.11, a characteristic feature of thermodynamic systems is that this number is remarkably *small*, far less than might have been anticipated from the complexity (of the order of 10^{23} variables) of the microscopic description.

Before proving the phase rule (7.6), we list some elementary guidelines and *caveats* regarding the definitions (7.7a–c):

- In the definition of c , one must generally assume that *any* chemical component is free to penetrate into *any* phase, even if the partitioning between phases varies strongly from component to component. Hence, c is the number of independent chemical components found in *any* phase, to account for the limiting case in which different chemical components partition “completely” (within limits of experimental detection) into different phases.
- The number of *independent* components, c , in a given system of interest can generally be evaluated as the *total* number of chemical species minus the number of *relationships* between concentrations. The latter may consist of *initial conditions* (defined by conditions of preparation of the system) or by *chemical equilibrium conditions* (for chemical reactions that are “active” in the actual system). Sidebar 7.1 provides illustrative examples of how c is determined in representative cases.

- In counting the number of phases, p , it is only important that *some* amount of each phase be present, even if only a droplet or granule. This merely reflects the general definition of thermodynamic “state” (Section 2.10) as being *independent of size*, whether of the total system or its phase subsystems. Sidebar 7.2 further illustrates this size independence for a two-phase system, showing why systems containing *any* quantity of either phase (but sharing common values of intensive state properties) are merely different “samples” of the *same* thermodynamic state, and are thus *equivalent* for thermodynamic purposes.

Attention to these points will avoid many common misconceptions and misuses of phase rule concepts.

Let us now proceed to prove the Gibbs phase rule (7.6). For each phase α , the total possible intensities are $c + 2$ in number (namely, $T, P, \mu_1, \dots, \mu_c$),

$$f_{\alpha}^{\text{tot}} = c + 2 \quad (7.8a)$$

but, in view of the Gibbs–Duhem equation (6.34), only $c + 1$ are *independent* (in accordance with Inductive Law IL-1 for $c = 1$):

$$f_{\alpha} = c + 1 \quad (7.8b)$$

Therefore, the maximum possible (f^{max}) for p phases ($\alpha, \beta, \dots, \pi$) is

$$f^{\text{max}} = f_{\alpha} + f_{\beta} + \dots + f_{\pi} = p(c + 1) \quad (7.9)$$

However, this total number f^{max} is subject to the *equilibrium conditions* between phases, namely,

$$T_{\alpha} = T_{\beta} = \dots = T_{\pi} \quad [p - 1 \text{ conditions}] \quad (7.10a)$$

$$P_{\alpha} = P_{\beta} = \dots = P_{\pi} \quad [p - 1 \text{ conditions}] \quad (7.10b)$$

$$\mu_{i\alpha} = \mu_{i\beta} = \dots = \mu_{i\pi} \quad \text{for } i = 1, 2, \dots, c \text{ } [c(p - 1) \text{ conditions}] \quad (7.10c)$$

[Note that only $p - 1$ equations (the number of equal signs) are needed to express the equality of the p phase temperatures in (7.10a), and similarly for the number of conditions given for (7.10b, c).] From the equilibrium conditions (7.10a–c), we therefore obtain the total number of equilibrium conditions as

$$\text{number of equilibrium conditions} = (c + 2)(p - 1) \quad (7.11)$$

Subtraction of (7.11) from (7.9) therefore leads to the *independent* number of intensities (f^{indpt}), namely,

$$f^{\text{indpt}} \equiv f = p(c + 1) - (c + 2)(p - 1) = c - p + 2 \quad (7.12)$$

which is the Gibbs phase rule (7.6)—*QED*.

SIDEBAR 7.1: COUNTING INDEPENDENT CHEMICAL COMPONENTS

The number of independent chemical components c can generally be determined from the equation

$$c = (\text{total number of species}) - (\text{number of constraint relationships between species})$$

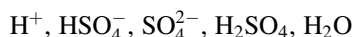
where “constraint relationships” may refer to (a) *initial conditions* pertaining to the mode of preparation of the system, or (b) *equilibrium conditions*, pertaining to chemical reactions that are active (not merely possible) in the system.

Problem Determine the number of independent chemical components, c , for each of the following systems:

- (a) pure helium gas
- (b) a sealed bottle of water in equilibrium with its vapor
- (c) an aqueous solution of H_2SO_4
- (d) an aqueous solution of NaCl and KBr
- (e) an aqueous solution of NaCl , NaBr , KCl , and KBr
- (f) a gaseous mixture of H_2 , O_2 , and H_2O

Solutions

- (a) $c = 1$. [The system consists of a single phase ($p = 1$) and a single chemical component, pure He ($c = 1$); hence $f = 2$.]
- (b) $c = 1$. [The system consists of two phases, liquid and vapor ($p = 2$), but only a single component, pure H_2O ($c = 1$); hence $f = 1$.]
- (c) $c = 2$. This answer may seem surprising to the chemistry student who would list the following as actual chemical species present in sulfuric acid solution (among others):



But no more than *two* of these five species are actually under independent control of the experimenter, because the dissociated ionic species concentrations are fixed by *equilibrium conditions* and are therefore not “independent variables” from the viewpoint of the experimenter. For example, if H_2O and H_2SO_4 are chosen as the two independently variable components, then the concentrations $[\text{H}^+]$, $[\text{HSO}_4^-]$, $[\text{SO}_4^{2-}]$ of the remaining species could be determined by one ionic balance condition

$$[\text{H}^+] + [\text{HSO}_4^-] = [\text{H}_2\text{SO}_4]$$

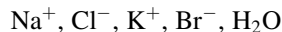
and two equilibrium conditions

$$K_1 = \frac{[\text{H}^+][\text{HSO}_4^-]}{[\text{H}_2\text{SO}_4]}, \quad K_2 = \frac{[\text{H}^+][\text{SO}_4^{2-}]}{[\text{HSO}_4^-]}$$

Owing to these three constraint conditions, only two of the five total species can be considered independent ($c = 2$). (The same answer would be obtained by realizing

that the solution could be prepared from *two* pure reagent bottles from the stockroom: H₂O and H₂SO₄.)

- (d) $c = 3$. This answer may again seem surprising if we think of the actual species present in the aqueous ionic solution, which would include the five species



However, the concentrations of two of these five species are constrained by two initial conditions of ionic balance,

$$[\text{Na}^+] = [\text{Cl}^-], \quad [\text{K}^+] = [\text{Br}^-]$$

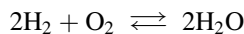
so that only *three* species remain independently variable ($c = 3$). (The same answer would be obtained by realizing that the solution could be prepared from three pure reagent bottles: NaCl, KBr, and H₂O.)

- (e) $c = 4$. It is clear from chemical considerations that the total list of species would be identical to that in (d), but the initial conditions have changed (so that neither $[\text{Na}^+] = [\text{Cl}^-]$ nor $[\text{K}^+] = [\text{Br}^-]$ is now a required condition). In this case, the only remaining constraint on the ion concentrations is that of *overall* ionic balance, i.e.,

$$[\text{Na}^+] + [\text{K}^+] = [\text{Cl}^-] + [\text{Br}^-]$$

so that the five total species are reduced to four independent species ($c = 4$) in this case. (The same answer would be obtained by realizing that one could choose any three of the four pure salts, together with water, to prepare the desired ionic mixture.)

- (f) Here the correct answer is “It depends.” Under ordinary low-temperature conditions, the correct answer is $c = 3$. However, at higher temperatures, or in the presence of a catalyst, where the chemical reaction



is “active,” the correct answer would be $c = 2$. [Note that the reaction must be occurring at an appreciable rate (not just as a theoretical possibility) on the actual timescale of the thermodynamic measurements.] Thus, as the temperature of a gas-phase mixture is raised (or a catalyst introduced), one expects to see crossover behavior from $c = 3$ ($f = 4$ degrees of freedom) to $c = 2$ ($f = 3$ degrees of freedom) in the equilibrium thermodynamic properties.

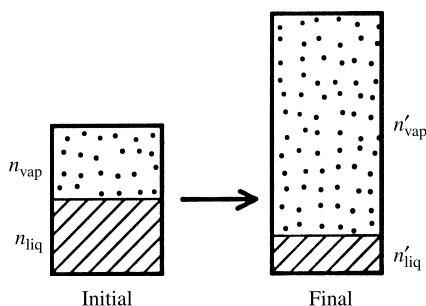
Note that the above considerations establish only the *number* of independent chemical variations c , not *which* species should be chosen as variables in the phase diagram, the latter choice being purely a matter of convenience. Note also from the final problem (f) that the question of effective chemical variability c can be addressed *experimentally*, based on the “active” chemistry of the actual system (rather than, for example, abstract theoretical

possibilities). A change of c (for fixed p) leads to a change in f , and thus to expected changes in the form of the experimental phase diagram (namely, coexistence regions having the geometrical character of points vs. lines vs. areas vs. volumes ...) that can readily distinguish whether a “correct” choice of c has been made for the given experimental conditions. (Whether a given theoretical reaction is active on the relevant experimental timescale can usually be determined by independent measurements.)

SIDEBAR 7.2: “SIZE” VERSUS “INTENSITY” AS “STATE” DESCRIPTORS OF A HETEROGENEOUS SYSTEM

We have previously emphasized (Section 2.10) the importance of considering only *intensive* properties R_i (rather than size-dependent *extensive* properties X_i) as the proper state descriptors of a thermodynamic system. In the present discussion of heterogeneous systems, this issue reappears in terms of the size dependence (if any) of individual phases on the overall state description. As stated in the caveat regarding the definition (7.7c), the formal thermodynamic “state” of the heterogeneous system is wholly *independent* of the quantity or “size” of each phase (so long as at least *some* nonvanishing quantity of each phase is present), so that the formal state descriptors of the multiphase system again consist of intensive properties only. We wish to see why this is so.

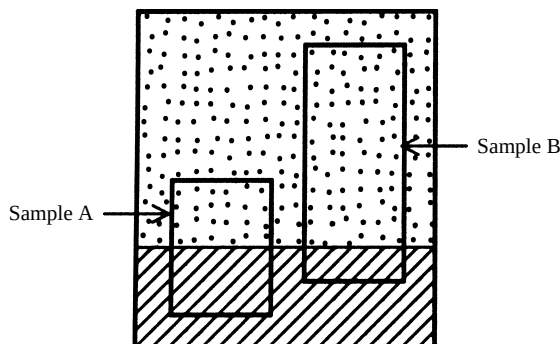
It is indeed somewhat surprising that the quantity of each phase is in some sense *irrelevant* to thermodynamic description of the phase-transition phenomenon. Consider, for example, a 1 kg sample of pure water in equilibrium with its own vapor at, say, the normal boiling point ($T = 100^\circ\text{C}$, $P = 1\text{ atm}$), initially with n_{vap} moles of vapor and n_{liq} moles of liquid, as shown at the left:



Under constant T , P conditions, a portion of the liquid is converted to vapor to give the final system shown at the right, with n'_{vap} moles of vapor and n'_{liq} moles of liquid ($n_{\text{liq}} + n_{\text{vap}} = n'_{\text{liq}} + n'_{\text{vap}}$). Because some liquid was converted to vapor, hasn't some “change of state” occurred in this process?

The answer can be judged by considering a larger sample of the *same* two-phase state (say, a 5 kg liquid–vapor sample at the *same* temperature and pressure). As shown in the following diagram, we can arbitrarily select for study two “aliquots” of the system,

sample A and sample B, that correspond precisely to the “initial” and “final” states above:



Thus, the apparent “change of state” is equivalent to merely selecting two “samples” of the *same* state. In this way, one can see that the *intensities* (i.e., the shared T , P of the coexisting phases) are the appropriate descriptors of the special “state” (condition) of the two-phase system, without regard to the arbitrarily chosen amounts of each phase (which are described rather trivially by extensive factors). It is essentially *irrelevant* for thermodynamic purposes whether we choose a thimble-full, a liter container, or an entire geyser pool of water to characterize this special state of the two-phase system.

7.2 SINGLE-COMPONENT SYSTEMS

To explore the ramifications of the phase rule (7.6), we shall first consider the phase equilibria of pure chemical substances ($c = 1$). Subsequent sections will examine the more complex behavior of binary ($c = 2$) and ternary ($c = 3$) multiphase systems.

The primary tool for representing the phase behavior of a chemical system is the *phase diagram*, a graphical “roadmap” of phase stability domains. For a pure substance, with

$$f = c - p + 2 = 3 - p = \begin{cases} 2 & (p = 1) \\ 1 & (p = 2) \\ 0 & (p = 3) \end{cases} \quad (7.13)$$

degrees of freedom, the phase diagram is simply a plot of P versus T , the two independent intensive variables of a homogeneous single-component system, showing the phase(s) of greatest stability (lowest Gibbs free energy) at each point (T , P).

Some phenomenological features of a representative phase diagram (for CO_2) were previously described in Section 2.5. In the present section, we shall first review key topological features of the phase diagram for H_2O from the perspective of the phase rule (Section 7.2.1). The general theory of phase boundaries will then be developed (Section 7.2.2) and illustrated (Section 7.2.3) for some simple elemental and molecular substances. These representative examples will serve to illustrate the bewildering multiplicity of phase forms and properties that are possible even in the “simple” $c = 1$ limit.

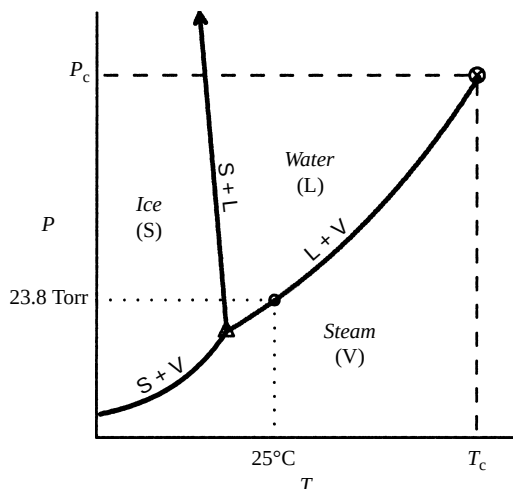


Figure 7.1 Schematic phase diagram of water (not to scale), showing phase boundaries (heavy solid lines), triple point (triangle), critical point (circle- $\times$), and a representative point (circle, dotted lines) at 25°C on the liquid–vapor coexistence curve.

7.2.1 The Phase Diagram of Water

A schematic representation of the phase diagram for pure H_2O (not to scale) is shown in Fig. 7.1. Let us examine the topological features of this diagram in terms of increasing complexity from single- to multiphase character.

Homogeneous Areas For $p = 1$, the phase rule establishes that

$$f = c - p + 2 = 1 - 1 + 2 = 2 \quad (7.14)$$

corresponding to a two-dimensional *area* for each pure phase, such as liquid (L, water), vapor (V, steam), or solid (S, ice). The phase diagram is accordingly divided by phase boundaries (solid lines) into the separate domains of stability of ice (low T , high P), steam (high T , low P), and liquid water (intermediate T and P), each occupying a two-dimensional area.

Heterogeneous Lines and Points For $p = 2$, the phase rule requires that

$$f = 1 - 2 + 2 = 1 \quad (p = 2) \quad (7.15)$$

corresponding to one-dimensional *lines* (curves) of coexisting phases, such as

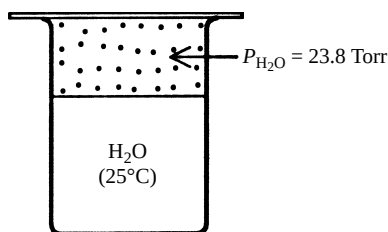
$$\text{L} + \text{V} \quad (\text{liquid/vapor}): \text{vapor-pressure curve} \quad (7.16a)$$

$$\text{L} + \text{S} \quad (\text{liquid/solid}): \text{melting curve} \quad (7.16b)$$

$$\text{S} + \text{V} \quad (\text{solid/vapor}): \text{sublimation curve} \quad (7.16c)$$

On such a two-phase coexistence curve, the system has only a single degree of freedom, so that, for a given T , the pressure P is fixed, and vice versa. For example, if the temperature of a liquid–vapor system is chosen as $T = 25^\circ\text{C}$, the corresponding P (read from the vapor-pressure curve) must be 23.8 Torr, as shown by the dotted line in Fig. 7.1.

What does it mean that $(25^\circ\text{C}, 23.8 \text{ Torr})$ is a point on the liquid–vapor coexistence line? Consider a beaker of liquid water at 25°C , covered with a lid and allowed to come into equilibrium with its own vapor:



When equilibrium is achieved, the “saturation vapor pressure” (partial pressure $P_{\text{H}_2\text{O}}$ of water vapor above the liquid) will be found to be

$$P_{\text{H}_2\text{O}} = 23.8 \text{ Torr (at } 25^\circ\text{C)} \quad (7.17)$$

If the lid is removed, and the external surroundings have partial pressure $P_{\text{H}_2\text{O}}$ *less* than 23.8 Torr (“relative humidity < 100%”), then water will evaporate from the beaker into the surroundings until the beaker is empty, because only vapor is stable under these conditions. However, if the external surroundings have partial pressure $P_{\text{H}_2\text{O}} > 23.8 \text{ Torr}$, water will condense from the surroundings to fill the beaker, because only liquid is stable under these conditions. Thus, the saturation vapor pressure (“100% relative humidity”) corresponds to the unique concentration (partial pressure) of water vapor that can coexist at equilibrium in the atmosphere above liquid water at 25°C . Other (T, P) points on the vapor-pressure curve can be interpreted analogously.

For $p = 3$, one also finds from the phase rule

$$f = 1 - 3 + 2 = 0 \quad (p = 3) \quad (7.18)$$

corresponding to a zero-dimensional (invariant) *point* of three-phase coexistence, a “triple point.” The triangle in Fig. 7.1 marks the triple point of solid–liquid–vapor coexistence in water, which occurs at $T = 0.01^\circ\text{C}$, $P = 4.56 \text{ Torr}$. As described in Section 2.3, the formal invariance ($f = 0$) of the triple point, as guaranteed by the phase rule, makes this special coexistence junction an excellent reference state for thermometric purposes.

Figure 7.1 also shows the “critical point” (circle- $\times$, dashed lines) of water, the terminus of the liquid–vapor coexistence line. Beyond this point (which occurs at $T_c = 374^\circ\text{C}$, $P_c = 217.7 \text{ atm}$), there is no longer a sensible distinction between “liquid” and “vapor,” so one should only speak of a “supercritical fluid” (or simply “fluid”) beyond the dashed lines. A sample of water above P_c can never exhibit a “boiling point,” no matter how far the temperature is increased, nor can a sample above T_c exhibit “condensation,” no matter how far the pressure is increased.

The critical state is evidently an invariant “point” (terminus of a line) in this case, because it lies at a dimensional *boundary* between states of $f = 2$ ($p = 1$) and $f = 1$ ($p = 2$). The critical point is therefore a uniquely specified state for a pure substance, and it plays an important role (Section 2.5) as a type of “origin” or “reference state” for description of all thermodynamic properties. Note that a limiting critical terminus appears to be a universal feature of liquid–vapor coexistence lines, whereas (as shown in Fig. 7.1) solid–liquid and solid–vapor lines extend indefinitely or form closed networks with other coexistence lines.

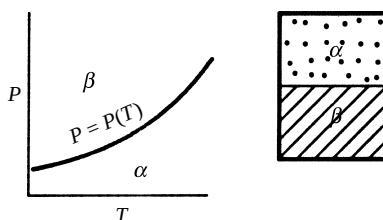
One might also ask whether states of simultaneous coexistence of *four* phases are possible. However, the phase rule for this case gives “ -1 ” degrees of freedom,

$$f = 1 - 4 + 2 = -1 \quad (p = 4) \quad (7.19)$$

which is a geometrical absurdity. The problem is not lack of possible phases, for (as will be seen in Section 7.2.3) water exhibits at least *nine* distinct phase manifestations that offer abundant opportunities for $p > 3$ phase coexistence outside the range depicted in Fig. 7.1. Nevertheless, the phase rule assures us that no “quadruple point” will *ever* be found for water or any other $c = 1$ system (whereas such four-phase states become routinely possible for $c \geq 2$; cf. Section 7.3). Thus, the Gibbs phase rule provides an invaluable guide as to what types of phase boundary geometry, coexistence junctions, or critical features may be expected as the phase diagram “roadmap” is extended to unexplored T, P regions.

7.2.2 Clapeyron and Clausius–Clapeyron Equations for Phase Boundaries

The defining features of phase diagrams are the phase boundaries that delineate phase domains and mark the conditions of coexistence with adjacent phases. Theoretical description of a phase diagram is therefore tantamount to finding the “equations of coexistence” that describe these phase boundaries. For a simple phase equilibrium between phases α and β , as shown below, the $\alpha + \beta$ coexistence curve is described by an equation of the form $P = P(T)$, whose form we now wish to determine:



We know that if $\mu_\alpha < \mu_\beta$, then only phase α is stable, whereas if $\mu_\alpha > \mu_\beta$, then only phase β is stable. The desired coexistence condition is thereby determined from

$$\mu_\alpha(T, P) = \mu_\beta(T, P) \quad (7.20)$$

which implicitly gives the desired relationship between T and P .

How can we compare the values of μ_α and μ_β for given T or P ? Since $\mu = \bar{G}$ for each phase, we expect that crossing points (7.20) will be determined by the *slope* of $\bar{G}$ versus T or P for each phase. At constant P , for example, the temperature slope $(\partial\bar{G}/\partial T)_P$ is determined [cf. (5.55)] by the negative *entropy* of the phase:

$$\left(\frac{\partial\bar{G}}{\partial T}\right)_P = -\bar{S} = \left(\frac{\partial\mu}{\partial T}\right)_P \quad (7.21)$$

We expect on general physical grounds (Sidebars 5.11 and 5.13) that the entropy of the gaseous or liquid state far exceeds that of the solid state,

$$\bar{S}_{\text{gas}} > \bar{S}_{\text{liquid}} > \bar{S}_{\text{solid}} \quad (7.22)$$

so that μ_{gas} and μ_{liquid} are expected to plunge more steeply downward than μ_{solid} as T increases. Accordingly, a plot of μ versus T is expected to have the form shown in Fig. 7.2a. The most strongly sloping μ_{gas} curve is expected to become most stable in the high- T limit (above T_b), the least-sloping μ_{solid} should similarly be most stable at the low- T limit (below T_m), and the intermediate-sloping μ_{liquid} should be most stable at intermediate temperatures (between T_m and T_b), as shown in Fig. 7.2a.

Similar reasoning applies to the P dependence of μ_{phase} (at constant T). In this case, the slopes $(\partial\mu/\partial P)_T$ are dictated by the phase *volumes* [cf. (5.52)]:

$$\left(\frac{\partial\bar{G}}{\partial P}\right)_T = \bar{V} = \left(\frac{\partial\mu}{\partial P}\right)_T \quad (7.23)$$

The *usual* order of phase volumes is

$$\bar{V}_{\text{gas}} \gg \bar{V}_{\text{liquid}} > \bar{V}_{\text{solid}} \quad (7.24)$$

leading to the expectation that the gas phase will become most stable at sufficiently low P , the solid will become favored in the high- P limit, and the liquid will be favored in an intermediate region, as shown in Fig. 7.2b. The special values of P where the coexistence

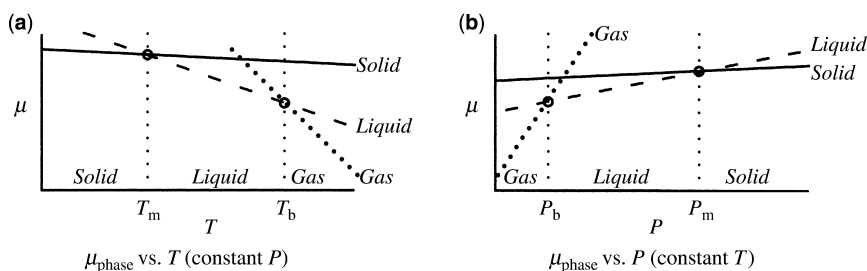


Figure 7.2 Phase coexistence conditions (circles), showing phase of lowest chemical potential as a function of (a) T , (b) P . The μ_{gas} (heavy dots), μ_{liquid} (dashes), and μ_{solid} (solid line) curves are plotted with slopes (7.21) [or (7.23)] consistent with the expected order (7.22) [or (7.24)] for the derivative slope $(\partial\mu/\partial T)_P$ [or $(\partial\mu/\partial P)_T$].

conditions (7.20) are satisfied are shown by the circles and vertical dotted lines in Fig. 7.2b, and these crossing points must shift as the phase volumes shift with temperature.

It is evident from these considerations that the coexistence condition (7.20) can only *remain* satisfied under variations dT or dP if these variations are chosen to insure that

$$d\mu_\alpha = d\mu_\beta \quad (7.25)$$

However, from the general Gibbs–Duhem equation (6.34) for each phase, we can write

$$d\mu_\alpha(T, P) = -\bar{S}_\alpha dT + \bar{V}_\alpha dP \quad (7.26a)$$

$$d\mu_\beta(T, P) = -\bar{S}_\beta dT + \bar{V}_\beta dP \quad (7.26b)$$

From (7.25) and (7.26), we infer that, on the coexistence line,

$$-(\bar{S}_\beta - \bar{S}_\alpha) dT + (\bar{V}_\beta - \bar{V}_\alpha) dP = 0 \quad (7.27a)$$

or

$$\boxed{\left(\frac{dP}{dT}\right)_{\text{coexistence}} = \frac{\bar{S}_\beta - \bar{S}_\alpha}{\bar{V}_\beta - \bar{V}_\alpha} = \frac{\Delta\bar{S}}{\Delta\bar{V}}} \quad (7.27b)$$

which is *Clapeyron's equation* for the phase boundary.

We can write Clapeyron's equation in a more compact notation by letting “ σ ” symbolize the $\alpha \rightarrow \beta$ phase transition (i.e., the constraint of remaining *on* the coexistence curve of α , β phases)

$$\alpha \xrightarrow{\sigma} \beta \quad (7.28a)$$

We can also write the transition entropy $\Delta\bar{S}_\sigma$ for this phase transition in terms of the associated *latent heat* $\Delta\bar{H}_\sigma$ and transition temperature T_σ as

$$\Delta\bar{S}_\sigma = \frac{\bar{H}_\sigma}{T_\sigma} \quad (7.28b)$$

This allows us to rewrite Clapeyron's equation

$$\boxed{\left(\frac{dP}{dT}\right)_\sigma = \frac{\Delta\bar{H}_\sigma}{T_\sigma \Delta\bar{V}_\sigma}} \quad (7.29)$$

as a differential equation for the phase boundary σ in compact and general form. [Note that the left-hand side of (7.29) might also be written in more conventional partial differential notation as $(\partial P/\partial T)_\sigma = (\partial P/\partial T)_{\alpha \rightarrow \beta}$, but we have chosen the “ordinary” dP/dT notation as a reminder that the derivative refers to the *single-variable* function $P = P(T)$ for the coexistence curve σ of $f = 1$ degree of freedom.]

Melting Transition Let us first apply (7.29) to the melting transition (“m”: $\sigma = \text{S} \rightarrow \text{L}$):

$$\left(\frac{dP}{dT}\right)_m = \frac{\Delta\bar{H}_m}{T_m(\bar{V}_{\text{liquid}} - \bar{V}_{\text{solid}})} \quad (7.30)$$

Noting that melting temperature $T_m > 0$ and latent heat $\Delta\bar{H}_m > 0$ (endothermic) are both positive, we can infer from (7.30) that

$$\left(\frac{dP}{dT}\right)_m > 0 \quad \text{if } \bar{V}_{\text{liquid}} > \bar{V}_{\text{solid}} \quad (\text{usual case: } \rho_{\text{liquid}} < \rho_{\text{solid}}) \quad (7.31a)$$

but

$$\left(\frac{dP}{dT}\right)_m < 0 \quad \text{if } \bar{V}_{\text{solid}} > \bar{V}_{\text{liquid}} \quad (\text{e.g., H}_2\text{O: } \rho_{\text{solid}} < \rho_{\text{liquid}}) \quad (7.31b)$$

Equations (7.31a, b) imply that the melting curve slopes *forward* in the *usual* case (e.g., Fig. 2.9 for CO₂) where the density of the solid exceeds that of the liquid ($\rho_{\text{solid}} > \rho_{\text{liquid}}$), but this curve tilts *backward* in the anomalous case of water (Fig. 7.1), where $\rho_{\text{solid}} < \rho_{\text{liquid}}$ (ice cubes float!). The relationships (7.31a, b) can also be paraphrased and generalized to any phase transition σ in the statement

σ tilts to cover the *less dense* phase

(7.32)

As will be illustrated in Section 7.2.3, this provides a powerful mnemonic for judging the relative densities of adjacent phases from the qualitative features (slopes) of coexistence lines in the phase diagram (or vice versa).

Vaporization Transition: Clausius–Clapeyron Equation For the liquid–vapor coexistence line (“vapor-pressure curve”), the Clapeyron equation (7.29) becomes

$$\left(\frac{dP}{dT}\right)_{\text{vap}} = \frac{\Delta\bar{H}_{\text{vap}}}{T_b(\bar{V}_{\text{gas}} - \bar{V}_{\text{liquid}})} \quad (7.33)$$

where $\Delta\bar{H}_{\text{vap}}$ is the latent heat of vaporization and T_b the boiling temperature.

We introduce three approximations to the exact equation (7.33):

1. Because $\bar{V}_{\text{gas}} \gg \bar{V}_{\text{liquid}}$, we *neglect* the liquid volume in the denominator of (7.33):

$$\Delta\bar{V}_{\text{vap}} \simeq \bar{V}_{\text{gas}} \quad (7.34a)$$

2. For the remaining term $\bar{V}_{\text{gas}}$, we employ the *ideal gas* approximation in the form

$$\bar{V}_{\text{gas}} \simeq \frac{RT}{P} \quad (7.34b)$$

3. Because the latent heat $\Delta\bar{H}_{\text{vap}}(T, P)$ is nearly constant over the usual effective range of T and P variations, we assume that it can be replaced by the *fixed* molar enthalpy of vaporization $\Delta\bar{H}_{\text{vap}}^\circ$ for standard state conditions (e.g., 25°C, 1 atm):

$$\Delta\bar{H}_{\text{vap}}(T, P) \simeq \Delta\bar{H}_{\text{vap}}^\circ \quad (7.34c)$$

With these approximations, (7.33) becomes

$$\frac{dP}{dT} \simeq \frac{\Delta H_{\text{vap}}^{\circ}}{T(RT/P)} = \frac{P}{T^2} \frac{\Delta H_{\text{vap}}^{\circ}}{R} \quad (7.35a)$$

which can be rewritten as

$$\frac{dP}{P} \simeq \frac{dT}{T^2} (\Delta H_{\text{vap}}^{\circ}/R) = -d(1/T)(\Delta H_{\text{vap}}^{\circ}/R) \quad (7.35b)$$

or alternatively as

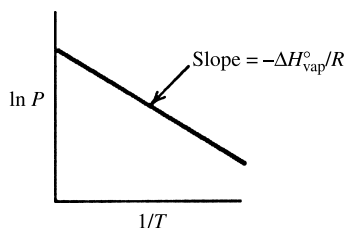
$$\boxed{\frac{d \ln P}{d(1/T)} \simeq -\frac{\Delta H_{\text{vap}}^{\circ}}{R}} \quad (7.35c)$$

Equation (7.35c) is known as the *Clausius–Clapeyron equation*, which may also be expressed in integrated form

$$\ln \frac{P_2}{P_1} \simeq -\frac{\Delta H_{\text{vap}}^{\circ}}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (7.35d)$$

for the vapor-pressure curve segment from (T_1, P_1) to (T_2, P_2) . The Clausius–Clapeyron equation (7.35c), although only approximate, provides a useful tool for estimating vapor-pressure changes with temperature in typical near-ambient laboratory conditions.

According to (7.35c), a plot of $\ln P$ versus $1/T$ is expected to give a *straight line* whose slope (in multiples of R) is the latent heat of vaporization:



Sidebar 7.3 illustrates a simple type of problem for which the Clausius–Clapeyron approximation is useful.

SIDEBAR 7.3: HUMIDITY PROBLEM

Problem How much is the absolute humidity (saturation vapor pressure of H_2O) *lowered* from a hot summer day (100°F) to a cool evening (50°F)? (Assume $\Delta H = 10 \text{ kcal mol}^{-1}$)?

Solution The tricky part is to convert Fahrenheit to Kelvin temperature:

$$T_1 = 100^\circ\text{F} = 311\text{K}, \quad T_2 = 50^\circ\text{F} = 283\text{K}$$

From the Clausius–Clapeyron equation (7.35d), we obtain (with $R = 1.987 \text{ cal mol}^{-1} \text{ K}^{-1}$)

$$\ln \frac{P_2}{P_1} = -\frac{10000}{1,987} \left(\frac{1}{283} - \frac{1}{311} \right) = -1.59$$

which gives

$$\frac{P_2}{P_1} = e^{-1.59} = 0.203$$

Accordingly, the evening humidity is only about 20% that of the hot daytime humidity, a great relief! (Note that “relative humidity” expresses vapor pressure $P_{\text{H}_2\text{O}}$ as a *percentage* of the saturation vapor pressure or “absolute humidity” for given T .)

7.2.3 Illustrative Phase Diagrams for Pure Substances

Water Qualitative aspects of the low-pressure H_2O phase diagram were previously examined in Section 7.2.1, focusing on the restricted region near the gas–liquid coexistence curve and its terminating triple point with ice I, the familiar solid form of H_2O . However, the existence of numerous additional solid phases is revealed as one extends investigation of the H_2O phase diagram to the more extreme conditions accessible to current high-pressure technology.

Figure 7.3 provides a more complete picture of the H_2O phase diagram in the range of pressures up to 10 kbar. (For comparison, the pressure at the bottom of Earth’s deepest oceans is about 0.5 kbar and that in planetary interiors ranges above 1 Mbar.) The region covered in Fig. 7.1 is the narrow sliver under the dotted line in the lower right corner, with the entire liquid–vapor coexistence curve visually indistinguishable from

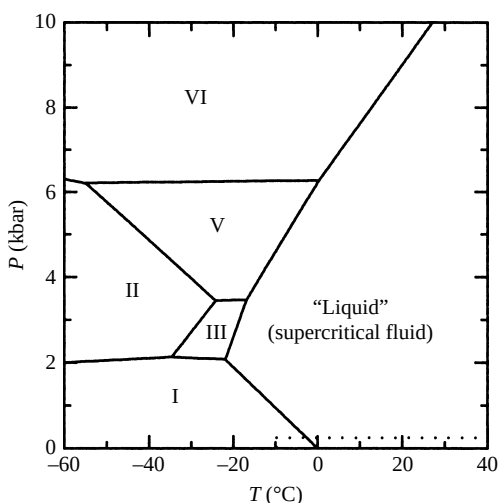


Figure 7.3 Qualitative phase diagram of H_2O in the 0–10 kbar range, showing additional equilibrium solid phases (ice II, ice III, ice V, and ice VI) beyond “normal” ice I (hexagonal ice Ih) of the low-pressure limit. (The dotted line marks the approximate upper limit of pressures in Fig. 7.1.) Note that the phase boundaries are shown here as straight lines, although slight curvature is present in the actual coexistence curves.

the horizontal axis. Although we shall refer to the high- T region as the “liquid” phase, it should be noted that virtually all of this region lies above the supercritical limit $P_c \simeq 0.22$ kbar, where the distinction between “liquid” and “vapor” is no longer meaningful.

Figure 7.3 exhibits the new ice phases, ice II, ice III, and ice V, that appear in the 2–3 kbar range, as well as the additional ice VI phase that dominates above about 6 kbar. At still higher pressures (up to about 500 kbar), there is evidence for at least four other stable ice phases—conventionally labeled ice VII, ice VIII, ice IX, and ice X. [Fortunately, ice IX is not the mythical “ice-nine” described in Kurt Vonnegut’s novel, *Cat’s Cradle*: “Ice-nine was the last gift Felix Hoenicker created for mankind before going to his just reward . . . He had made a chip of ice-nine. It was blue–white. It had a melting point of 114.4°F.”] Additional *metastable* phases are also found, such as ice IV and ice XII (which overlap the stable ice V region). Each distinct ice phase corresponds to subtle differences in microscopic hydrogen-bonding coordination pattern or proton ordering, and each exhibits characteristic macroscopic differences in density and other mechanical, optical, and dielectric properties. Understanding the microscopic basis of such macroscopic diversity is a principal objective of statistical thermodynamic models (Chapter 13).

In accordance with the Clapeyron equation and Le Chatelier’s principle, the more highly ordered (low-entropy) phases tend to lie further to the left (at lower T), whereas the higher-density phases tend to lie further upward (at higher P). The mnemonic (7.32) allows us to anticipate the relative densities of adjacent phases. From the slope, for example, of the ice II–ice III coexistence line (which “tilts” forward to cover ice III), we can expect that ice II is denser than ice III ($\rho_{\text{II}} > \rho_{\text{III}}$). Similarly, from the forward slopes of the liquid coexistence lines with the high-pressure ices II, V, and VI, we can expect that cubes of ice II, ice V, and ice VI would all *sink* in a glass of water, whereas ice I floats (in accord with the backward tilt of its phase boundary). Many such inferences can be drawn from the slopes of the various phase boundaries in Fig. 7.3, all consistent with the measured phase densities ρ_{phase} (in g L^{-1}), namely,

$$\rho_{\text{liq}} = 1.00, \quad \rho_{\text{I}} = 0.92, \quad \rho_{\text{II}} = 1.17, \quad \rho_{\text{III}} = 1.14, \quad \rho_{\text{V}} = 1.23, \quad \rho_{\text{VI}} = 1.31 \quad (7.36)$$

Six new triple points are seen to appear in Fig. 7.3, and still others are found at higher P and lower T . However, *no* quadruple points (or coexistence features of higher order) are found, as predicted by the Gibbs phase rule.

Surprising behavior of liquid and ice phases is found if we follow various T, P paths in this extended phase diagram. Sidebar 7.4 illustrates how to determine the expected phase transitions and properties of H_2O for various temperatures and pressures far outside the realm of ordinary experience. It is remarkable that such multiplicity of forms and properties can result from a pure substance composed of only a single type of molecule.

SIDEBAR 7.4: PHASE-DIAGRAM PROBLEMS

Problem At what temperature will water freeze to ice if the pressure is (a) 2 kbar, (b) 10 kbar? What phase of ice is formed in each case?

Solution

- (a) -20°C (-4°F); ice I
- (b) $+25^\circ\text{C}$ ($+77^\circ\text{F}$); ice VI

Problem An ice cube, initially of ordinary ice I, is subjected to increasing pressures at various fixed temperatures:

- (a) -10°C
- (b) $+30^{\circ}\text{C}$
- (c) $+60^{\circ}\text{C}$

What successive phases will be encountered (and at what pressures) in each case?

Solutions

- (a) -10°C : ice I $\rightarrow$ liquid (~ 1 kbar) $\rightarrow$ ice V (~ 4.5 kbar) $\rightarrow$ ice VI (~ 6.2 kbar)
- (b) -30°C : ice I $\rightarrow$ ice III (~ 2.1 kbar) $\rightarrow$ ice II (~ 2.7 kbar) $\rightarrow$ ice V (~ 4 kbar) $\rightarrow$ ice VI (~ 6.2 kbar)
- (c) -60°C : ice I $\rightarrow$ ice II (~ 2 kbar) $\rightarrow$ ice VI (~ 6.3 kbar)

Helium One could hardly imagine a simpler elemental substance than helium, which is viewed at the microscopic level as consisting of spherically symmetric closed-shell 2-electron atoms that interact most weakly of all known atoms. It is therefore surprising that helium exhibits thermodynamic phase behavior that is spectacularly unusual and perplexing.

Figure 7.4 illustrates the phase diagram of the ^4He isotope in the low-temperature condensation region. The thermodynamic properties of ^4He are fundamentally distinct from those of the trace isotope ^3He , and the two isotopes spontaneously phase-separate near 1K. Both isotopes exhibit the spectacular phenomenon of *superfluidity*, the near-vanishing of viscosity and frictional resistance to flow. The strong dependence on fermionic (^3He) or bosonic (^4He) character and bizarre hydrodynamic properties are manifestations of the “quantum fluid” nature of both species. ^3He is not discussed further here.

As shown in Fig. 7.4, helium exhibits a conventional liquid–gas coexistence curve that terminates in a critical point (circle- $\times$) at very low T , P (about 5K, 2 atm); hence, ordinary

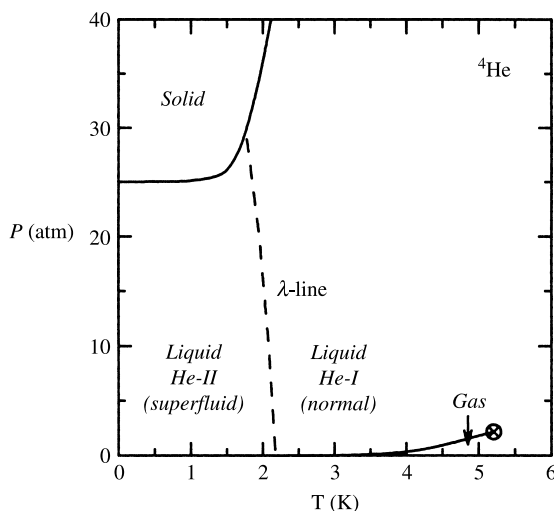


Figure 7.4 Phase diagram of ^4He , showing the solid, gas, and two liquid phases (He-I, superfluid He-II), the λ -line (dashed) of liquid–liquid transitions (upper terminus: 1.76K, 29.8 atm; lower terminus: 2.17K, 0.0497 atm), and the gas–liquid critical point (circle- $\times$: 5.20K, 2.264 atm).

helium “gas” is a supercritical fluid in all near-ambient conditions. Surprisingly, ^4He exhibits not just one, but *two* distinct liquid phases, labelled He-I and He-II, separated by an unusual “ λ -line” (dashed) that will be described below. Still more surprisingly, the low-temperature He-II liquid does *not* freeze to a solid (unless the pressure is increased to about 25 atm), remaining a liquid down to absolute zero! This behavior is unprecedented, as every other known substance freezes at 10 K or higher (far above even the *critical* temperature for ^4He). Moreover, liquid He-II itself behaves as a complex “two-component” mixture of normal fluid and *superfluid*, with extraordinary hydrodynamic and electrical (i.e., superconducting) properties.

From the slopes of the phase boundaries, one can judge [using the Clapeyron mnemonic (7.32)] that $\rho_{\text{solid}} > \rho_{\text{I}}, \rho_{\text{II}}$ (i.e., a high-pressure “ice cube” of frozen helium will sink in either He-I or He-II) and that $\rho_{\text{II}} > \rho_{\text{I}}$ (i.e., the low- T He-II superfluid floats on the high- T He-I normal fluid). One can also judge from its placement at lower T that He-II is more highly “ordered” than He-I ($\bar{S}_{\text{II}} < \bar{S}_{\text{I}}$), despite its superfluid proclivities.

The mysteries of the helium phase diagram further deepen at the strange “ λ -line” that divides the two liquid phases. In certain respects, this “coexistence curve” (dashed line) exhibits characteristics of a line of critical points, with *divergences* of heat capacity and other properties that are normally associated with critical-point limits (so-called “second-order transitions,” in Ehrenfest’s classification). Sidebar 7.5 explains some aspects of the Ehrenfest classification of phase transitions and the distinctive features of λ -transitions (such as the characteristic lambda-shaped heat-capacity curve that gives the transition its name) that defy classification as either “first-order” or “second-order.” Such anomalies suggest that microscopic understanding of phase behavior remains woefully incomplete, even for the simplest imaginable atomic components.

SIDEBAR 7.5: EHRENFEST CLASSIFICATION AND THE HELIUM λ -LINE

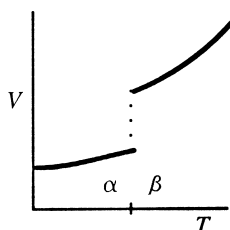
Paul Ehrenfest suggested a widely used classification of thermodynamic transition phenomena according to the lowest derivative of Gibbs free energy that exhibits a mathematical discontinuity at the phase transition.

In the case of a “first-order transition,” a discontinuity at the phase boundary is seen in G' , a *first* derivative of free energy:

$$\left(\frac{\partial G}{\partial T}\right)_p = -S$$

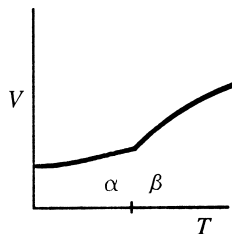
$$\left(\frac{\partial G}{\partial P}\right)_T = V$$

For example, the first pressure-derivative $V = (\partial G / \partial P)_T$ varies with temperature as follows:



with finite jump-discontinuity $V_\beta - V_\alpha$ (and corresponding density difference) at the transition temperature $T_{\alpha \rightarrow \beta}$. The second derivative $G'' = (\partial^2 G / \partial P^2)_T$ is therefore undefined (“infinite”) at this temperature. Ordinary phase transitions have such “first-order” character.

In the case of a “second-order transition,” each first-derivative G' is continuous at $T_{\alpha \rightarrow \beta}$,

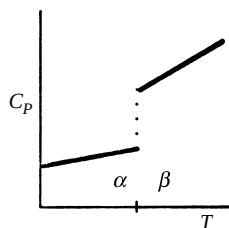


but the slope is envisioned as discontinuous, corresponding to a discontinuous jump in the *second* derivative G'' :

$$\left(\frac{\partial^2 G}{\partial T^2}\right)_P = -\left(\frac{\partial S}{\partial T}\right)_P = -\frac{C_P}{T}$$

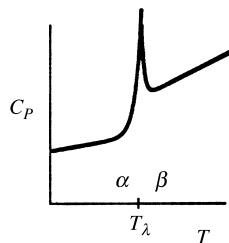
$$\left(\frac{\partial^2 G}{\partial P^2}\right)_T = \left(\frac{\partial V}{\partial P}\right)_T = -V\beta_T$$

For example, the heat capacity C_P in a second-order Ehrenfest transition might be expected to behave as follows:



(Such an Ehrenfest second-order transition, although formally imaginable, is not believed to describe any real phase transition.)

However, the heat capacity near the λ -transition has a quite different form (resembling the Greek letter lambda):



As shown, C_P varies in a continuous manner, but appears to *diverge* at T_λ ($= 2.17\text{K}$ at 1 atm), resembling in this respect the behavior at the gas–liquid *critical point* (discussed

extensively in Section 11.9). Such critical-like divergence of C_P and other thermodynamic properties is observed along the entire λ -line in Fig. 7.4. Thus, the helium λ -transition conforms to *neither* a first-order nor a second-order transition in the Ehrenfest classification. The microscopic nature of the λ -line, and indeed of many other features in the ^3He – ^4He phase diagram, remains deeply mysterious.

Sulfur Figure 7.5 displays a portion of the phase diagram of elemental sulfur. Pressure is shown on a logarithmic scale, spanning about ten orders of magnitude from the millitorr to kilobar ranges. The sulfur diagram serves to illustrate the general phenomenon of *metastability* and the sluggishness of transitions between solid phases, which often adds considerable complexity (as well as interest!) to the solid-phase region.

The principal features of elemental sulfur in the displayed T, P range are the usual liquid and vapor phases and two solid forms, α -sulfur (“red sulfur,” of orthorhombic crystalline form) and β -sulfur (“yellow sulfur,” monoclinic needle-like crystals), both of which are available as common stockroom species. The stable phase ranges for each elemental form are shown by the solid lines in Fig. 7.5. The liquid–vapor coexistence line terminates in a critical point at 1041°C , and will not be discussed further.

Under ordinary laboratory conditions (at lower temperatures than shown in Fig. 7.5), it is seen that only the red α -sulfur form is stable. The yellow β -sulfur needles are the most stable phase only in a narrow temperature range around 96 – 120°C , but they persist as a “supercooled” *metastable* phase well below this range (surviving, for example, on a stockroom shelf for indefinite periods). Given a sample of yellow β -sulfur, it is easy to detect the melting point near 120°C , but is far from easy to detect the enantiomorphic α/β solid–solid conversion near 96°C , unless a nucleating crystallite or catalyst is introduced. Once produced, the β -phase tends to persist as a supercooled (undercooled) metastable extension

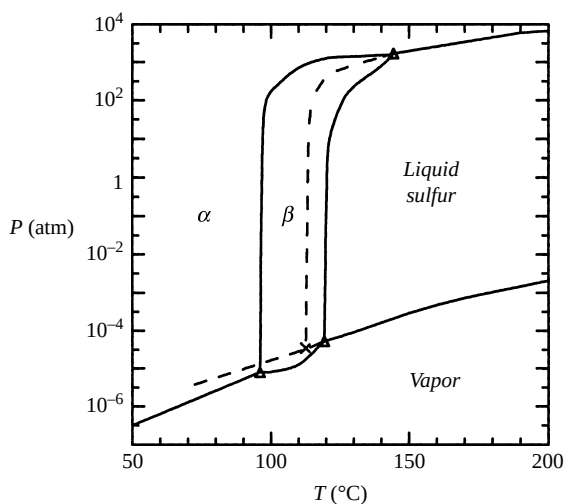


Figure 7.5 Phase diagram of elemental sulfur, showing the stable solid phases α -sulfur (orthorhombic “red sulfur”) and β -sulfur (monoclinic “yellow sulfur”) and equilibrium phase boundaries (solid lines) as well as the *metastable* phase boundary (dashed line) that connects α -sulfur to liquid and vapor phases.

well below its island of stability in the phase diagram. Indeed, kinetic conversion of β -sulfur to α -sulfur is so sluggish that the former shows no perceptible inclination to “convert” as it cools below 96°C.

Similarly, when rhombic red α -sulfur is heated above 100°C, it usually fails to exhibit the expected thermodynamic conversion to yellow β -sulfur at 96°C. Instead, it persists as a “superheated” metastable phase up to 114°C (dashed line), where it exhibits an apparently normal melting point to the liquid form (unless extreme patience or a nucleating seed crystal of β -sulfur is employed). The dashed lines in Fig. 7.5 therefore mark out *metastable* phase transition boundaries between forms of sulfur that are not true Gibbs free energy minima at the cited temperature and pressure (e.g., superheated α -sulfur and supercooled liquid sulfur at 114°C, 1 atm). The metastable phase domains can overlap stable phase domains in a quite complex and confusing manner. A kinetically facile metastable phase boundary will often appear more “real” and relevant to actual chemical phenomena than will the idealized boundary between (kinetically inaccessible) phases of lowest Gibbs free energy.

Consistent with the stable/metastable distinction, *four* distinct triple points are seen in Fig. 7.5: three (triangles) of stable type (α - β -vapor point at 96°C, 0.01 Torr; β -liquid-vapor point at 120°C, 0.025 Torr; and α - β -liquid point at 155°C, 1290 atm) and one (cross) of metastable type (β -liquid-vapor point at 120°C, 0.02 Torr).

Still another type of bewildering complexity (not shown in Fig. 7.5) is found in the liquid region, where *two* distinct liquid allotropic forms are known, called “ λ -sulfur” and “ μ -sulfur.” The former is the molten form commonly encountered in the laboratory: pale yellow, diamagnetic, easily soluble in nonpolar solvents such as carbon disulfide, and of rather ordinary viscosity. The μ -sulfur variant, by contrast, is *extremely* viscous, deep orange-red, paramagnetic (indicative of radical unpaired-electron character), and essentially insoluble in *any* common solvent. These surprising differences point to essentially different coordination and bonding patterns at the microscopic level, as will be discussed in Chapter 13.

Despite the complexities introduced by metastable solid and liquid phases, topological features of the phase diagram can be thermodynamically interpreted in a standard manner from Clapeyron-type equations. Thus, from the forward slopes of α - β , β -liquid (stable), and α -liquid (metastable) phase boundaries, we can infer from (7.32) that

$$\rho_{\alpha} > \rho_{\beta} > \rho_{\text{liquid}} \quad (7.37)$$

i.e., that crystals of α -sulfur or β -sulfur will both sink into (form at the bottom of) molten sulfur. Simple thermodynamic principles therefore continue to guide the use of the phase diagram in both stable and metastable free-energy “layers.” However, a fundamental understanding of the atomic-level origins of the observed phase complexity certainly presents a daunting theoretical challenge, even for such a “simple” elemental system.

Carbon One of the most industrially and economically important phase transformations is that involving graphite and diamond, both allotropes of pure carbon. Graphite is a soft, slippery material, drab black and with poor insulating properties, familiar in the form of soot, lampblack, and pencil lead. Diamond is among the hardest materials known, with excellent heat/electrical insulating properties, known for its dazzling optical properties as a gemstone as well as other industrial applications. It is therefore remarkable that these vastly different materials are merely different phases of the same atomic-level constituents, separated by an enantiotropic solid–solid phase transition in the carbon phase diagram.

[In this case, the chemical origin of the distinctive mechanical, electrical, and optical phase differences is rather clearly understood. Graphite consists of sp^2 -hybridized C atoms

π -bonded in extended planar benzenoid sheets but only weakly interacting in the third dimension, whereas diamond consists of sp^3 -hybridized C atoms in a saturated three-dimensional tetrahedral network of sigma bonds. The phase differences therefore correspond to distinctive *molecular* character of the two solids, and at low temperatures (where covalent bonds remain intact) each of these molecular solids underlies an *independent* phase diagram of “single-component” character. However, at the high temperatures where covalent bond breaking is active, the graphitic and diamond bond networks are no longer independently variable chemical components, and the carbon phase diagram corresponds to only *one* independently variable species ($c = 1$, chosen, for example, as the carbon *atom* rather than a molecular species). Thus, the carbon phase diagram is not strictly comparable to more conventional low- T “single-component” chemical systems based on a single molecular species (such as H_2O or CO_2), but it serves as an illustrative example of high- T phase behavior where chemical bond-breaking is facile.]

Figure 7.6 depicts known portions of the carbon phase diagram pertaining to the graphite $\rightleftharpoons$ diamond transition region. The diagram spans temperatures up to 6000K (comparable to that of the Sun’s surface) and pressures up to 200 kbar (roughly 5–10% that of the Earth’s core). At still higher pressures, there is evidence of still another “metallic” phase of carbon, beginning around 400 kbar and 2000K. Sidebar 7.6 describes specialized techniques that are required for laboratory studies of the high pressures characteristic of diamond formation and other important geological phenomena.

As shown in Fig. 7.6, graphite is the stable form of carbon under ordinary terrestrial surface conditions (which occupy an imperceptibly small sliver in the lower left corner of the phase diagram). Under our near-ambient conditions, diamond is only metastable (not “forever”), becoming the phase of absolute minimum Gibbs free energy only at pressures above about 15 kbar. Both graphite and diamond coexist with liquid carbon in a triple point near 4000K, 140 kbar. As shown by the slopes of the phase boundaries in this triple-point region,

$$\rho_{\text{diamond}} < \rho_{\text{graphite}} < \rho_{\text{liquid}} \quad (7.38)$$

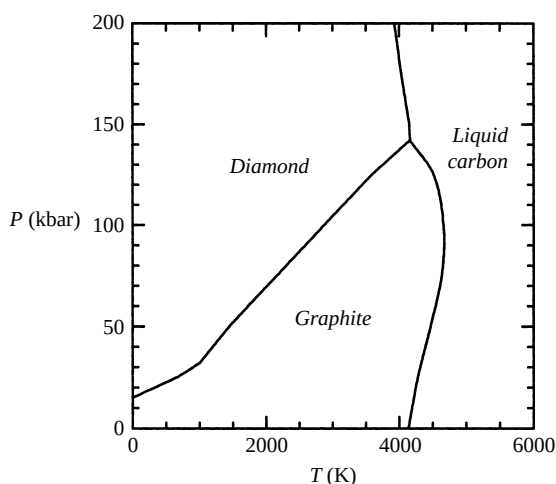


Figure 7.6 Phase diagram of elemental carbon [after F. P. Bundy. *J. Geophys. Res.* **85**, 6930 (1980)].

that is, both diamond and graphite crystals will *float* on liquid carbon under these conditions. However, at lower pressures, the slope of the graphite–liquid coexistence curve slopes forward, so that graphite sinks in its coexisting melt below about 90 kbar.

The economic valuations of the two carbon allotropes also differ significantly, by more than five orders of magnitude at current prices for gem-quality natural diamonds. Hence, there is considerable economic incentive to employ the natural “alchemy” of the phase diagram to convert graphite to diamond, by subjecting the former to the high pressures at which it should spontaneously transform to the more stable diamond form. However, the same kinetic sluggishness that preserves diamonds at low T , P tends to make the graphite $\rightarrow$ diamond conversion impractically slow unless suitable catalysts and/or more extreme temperatures and pressures are employed. The successful production of synthetic diamonds was carried out by R. H. Wentorf and co-workers at the General Electric Company (Sidebar 7.6), employing a molten nickel catalyst at about 1500°C and 60 kbar. Synthetic production of diamonds and other gemstones is currently a multi-billion dollar business. Accurate knowledge of phase diagram boundaries and kinetic barriers is also important in many other areas of metallurgy and materials technology.

SIDEBAR 7.6: P. W. BRIDGMAN AND THE THERMODYNAMICS OF HIGH PRESSURES



Pressure, the force per unit area, varies widely in everyday experience. We continually experience the 1 atm ($\simeq$ 1 bar) pressure of the Earth’s atmosphere, we periodically inflate tires to three times this value, and we may occasionally feel still stronger pressures in a stubbed toe or other mishap. The transient pressure of a strong hammer blow can be 5 kbar and that of an impacting rifle bullet on armor plate as much as 100 kbar. But these “extreme” pressures are still small compared with the relevant pressures of many geophysical phenomena involving the Earth’s interior, which rise into the megabar range. How are such high pressures achieved and measured in the laboratory?

Early studies of pressure effects on gases were carried out in apparatus with glass components, which could not withstand pressures exceeding about 0.4 kbar. In order to extend T. Andrews’

studies of critical point behavior, French physicist E. H. Amagat (1841–1915) developed a glass-free hydraulic manometer that was able to achieve static pressures up to 3 kbar, the practical limit until about 1905.

Harvard professor Percy W. Bridgman (1882–1961) achieved the next great breakthrough with development of the “Bridgman cell.” The principal novelty of Bridgman’s hydraulic anvil device was a sealing ring made of pipestone (Catlinite, the distinctive red claystone quarried from ancient times at Pipestone, Minnesota, and traded throughout pre-European America for making peace-pipes). Pipestone was found to possess unique

self-sealing deformation properties that actually *tighten* the seal as pressure increases, completely eliminating gasket failure as a limiting factor. With this device, Bridgman was able to extend systematic studies of pressure effects on chemical substances up to about 100 kbar. For his pioneering studies of high-pressure thermodynamics, Bridgman was awarded the Nobel Prize in Physics in 1948. Bridgman also contributed to thermodynamic teaching through his influential books, *The Logic of Modern Physics* (1927) and *The Nature of Thermodynamics* (1941), in which he strongly advocated the importance of *operational* definitions. Bridgman is properly recognized as the founding father of high-pressure thermodynamics.

Bridgman's anvil was finally supplanted in the 1950s by the tetrahedral anvil (H. T. Hall, General Electric, 1954) and the diamond anvil cell [C. E. Weir, R. R. Lippincott, A. Van Valkenburg, and E. N. Bunting. *J. Res. Natl Bur. Stand.* **63A**, 55–62 (1959)], which were able to attain another order-of-magnitude pressure increase, surpassing the 1 Mbar milestone. However, the ability to generate high pressures became less critical than the ability to reliably *measure* the pressures thus attained. A breakthrough was achieved in 1971 with the development of ruby optical fluorescence as a reliable pressure measurement standard [R. A. Forman, G. J. Piermarine, J. D. Barnett, and S. Block. *Science* **176**, 284–5 (1972)]. These technological advances have been combined with improved laser-spectroscopic and heating techniques to achieve sustained multi-Mbar pressures at about 1000K temperature in modern high-pressure facilities.

For more than a century, a leading impetus to high-pressure studies has been the desire to convert graphite to diamonds. This was repeatedly attempted, unsuccessfully, by Bridgman himself. Successful synthesis of diamond grit (useful for industrial grinding purposes) was first announced in 1955 by scientists at the General Electric Company [F. P. Bundy, H. T. Hall, H. M. Strong, and R. H. Wentorf. *Nature* **176**, 51–4 (1955)]. However, carat-size diamonds of gemstone quality were not synthesized until about 1970 by R. H. Wentorf and co-workers at General Electric, under conditions described in the text. Synthetic diamonds also became an important source of supply for improved diamond anvil cells, which in turn allowed further extension of the pressure range. (The earliest prototypes of the diamond anvil cell at the US National Bureau of Standards were fabricated from confiscated natural diamonds seized by federal agents from diamond smugglers.)

A quite different set of *dynamic* high-pressure techniques are based on the use of chemical or nuclear explosions to produce transient shock waves of high peak pressure but short duration. With such methods, one can often penetrate the high- T , P regions where kinetic barriers become unimportant and a catalyst is unnecessary. However, the same kinetics that allows facile conversion of graphite to diamonds as the shock front arrives also allows the facile back-conversion as the shock wave passes. As a pioneer of shock-wave diamond synthesis remarked ruefully, “We were millionaires for one microsecond” [B. J. Alder and C. S. Christian. *Phys. Rev. Lett.* **7**, 367 (1961); B. J. Alder, in W. Paul and D. M. Warschauer (eds). *Solids under Pressure* (McGraw-Hill, New York, 1963), p. 385].

7.3 BINARY FLUID SYSTEMS

Let us return to the simple case of a two-component ($c = 2$) A/B chemical solution (A = “solvent,” B = “solute”), as considered in Chapter 6. We focus primarily on the fluid (liquid–vapor) region of the phase diagram, neglecting for the moment the melting/freezing phenomena that lie at lower T and higher P .

From the Gibbs phase rule (7.6), we deduce for this case that

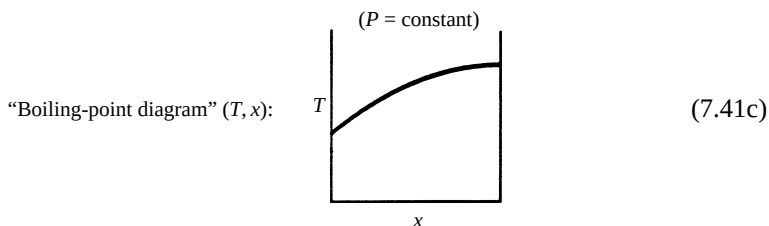
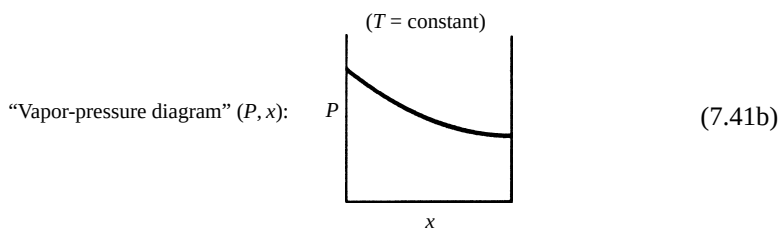
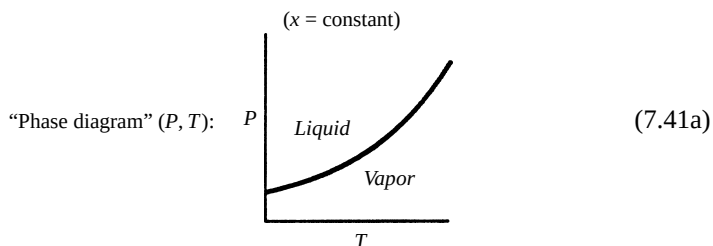
$$f = c - p + 2 = 4 - p = \begin{cases} 3, & \text{if } p = 1 \text{ (phase volume)} \\ 2, & \text{if } p = 2 \text{ (coexistence area)} \\ 1, & \text{if } p = 3 \text{ (triple line)} \\ 0, & \text{if } p = 4 \text{ (quadruple point)} \end{cases} \quad (7.39)$$

Compared with corresponding f values for $c = 1$ (Section 7.2), we can see that each phase multiplicity p “enters a new dimension,” and in addition *four*-phase coexistence becomes possible for binary solutions. Thus, each degree of phase coexistence presents new challenges to geometric visualization for $c = 2$.

For the base $p = 1$ case with $f = 3$, the three independent variables of the phase diagram might be chosen as intensives T , P , and μ_A (or μ_B). However, it is instead convenient to choose a *composition* variable, such as mole fraction

$$x = x_B = x_{\text{solute}} \quad (7.40)$$

as the third independent variable. With (T, P, x) as chosen variables, we can envision describing the system with *three* planar (2D) diagrams, namely



However, the description of the system composition variable $x = x_B$ is actually somewhat more complicated than implied in (7.41b, c). The problem is that solute generally partitions *unequally* between phases, so that the concentration “ x_B ” is *different* in different

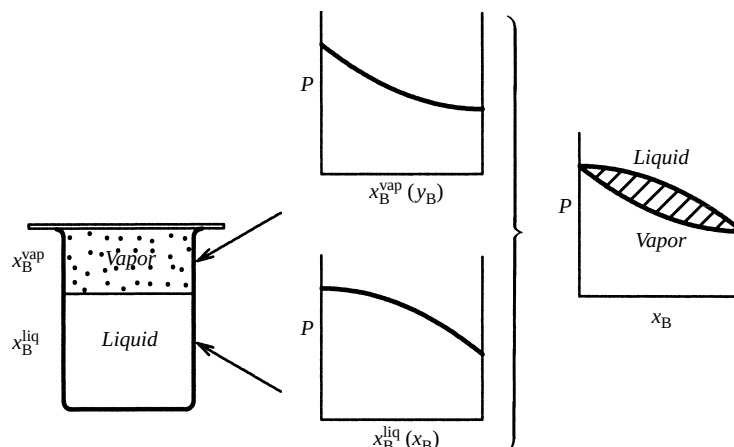
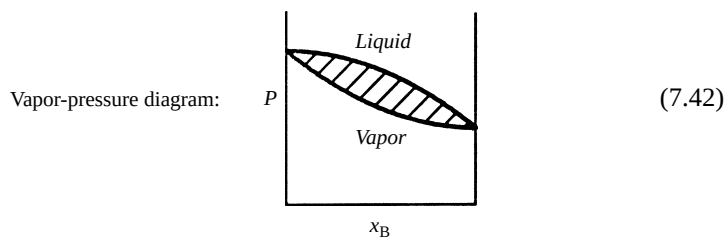
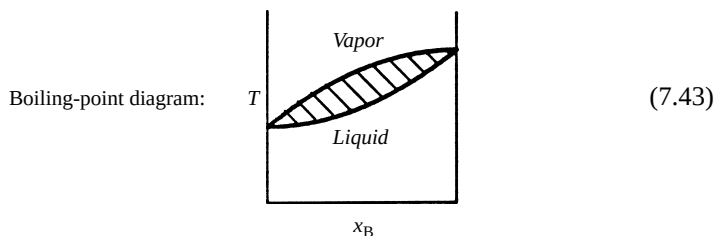


Figure 7.7 Vapor-pressure diagram of a binary fluid. The two-phase sample (left, with vapor pressure P above solution) has different solute concentration in the vapor phase ($x_B^{\text{vap}} \equiv y_B$) and liquid phase (x_B^{liq}). This leads to two distinct P - x_B plots (center): the P - x_B^{vap} vapor plot (above) and P - x_B^{liq} liquid plot (below). However, the separate vapor and liquid plots are conveniently combined in the composite P - x_B diagram (right), with a “hole” (hatch marks) between the vapor and liquid curves.

phases. Accordingly, the vapor pressure P above the solution might be expressed either as a function of x_B^{liq} (as a P - x_B^{liq} diagram) or of x_B^{vap} (as a P - x_B^{vap} diagram), as shown in Fig. 7.7. (The notation $x_B \equiv x_B^{\text{liq}}$, $y_B \equiv x_B^{\text{vap}}$ is sometimes employed to distinguish the two composition variables.) As shown at the far right of Fig. 7.7, these two diagrams can be deftly combined into a *single* diagram. Because the vapor and liquid curves “connect” only at the endpoints (pure A or B, where x_B^{liq} and x_B^{vap} must be identical), a “hole” is left in the diagram between the two curves (as shown with hatch marks). Thus, (7.41b) is replaced by the more complete vapor-pressure diagram



In an analogous manner, the oversimplified boiling-point diagram (7.41c) is replaced by



For a concentration value x_B , the ordinate value (P or T) is to be read in each case from the vapor curve if $x_B = x_B^{\text{vap}}$ is the vapor concentration, or from the liquid curve if $x_B = x_B^{\text{liq}}$ is the liquid composition. In either case (7.42) or (7.43), a “hole in the phase diagram” results from use of a *nonintensive* composition variable x_B , whereas no such hole results if, as in (7.41a), only intensive plotting variables are chosen.

From the three distinct 2D cross-sectional views (7.41a), (7.42), (7.43) of the P – T – x surface, we can now visualize the full 3D form of the surface as shown in Fig. 7.8. The surface is seen to resemble a curved envelope, clipped at each end to reveal the “inside” of the envelope through the hatched “holes.” Viewed toward the P – T plane, only the curved edge of the envelope is seen, as in (7.41a). However, viewed toward the P – x_B plane or the T – x_B plane, the inside of the envelope is seen as the hatch marks in (7.42) or (7.43), respectively. The upper P – T – x_B^{liq} surface of the envelope is called the “bubble-point” surface, in reference to the first vapor bubbles that are seen as the liquid is heated to its boiling point. The P – T – x_B^{vap} underside of the envelope is correspondingly called the “dew-point” surface, in reference to the first dewy droplets of liquid as the vapor is cooled to its condensation temperature. Although we normally see only the flat P – T , P – x_B , or T – x_B projections on the blackboard or book page, it is useful to keep in mind the full 3D form of the P – T – x_B surface that underlies these 2D projections of the $f = 3$ system.

The generic phase-diagram relationships illustrated in Figs. 7.7 and 7.8 underlie all binary fluid thermodynamics. We first focus on the P – x aspects of real and ideal solutions (Sections 7.3.1–7.3.3), then on the complementary T – x aspects (Sections 7.3.4 and 7.3.5),

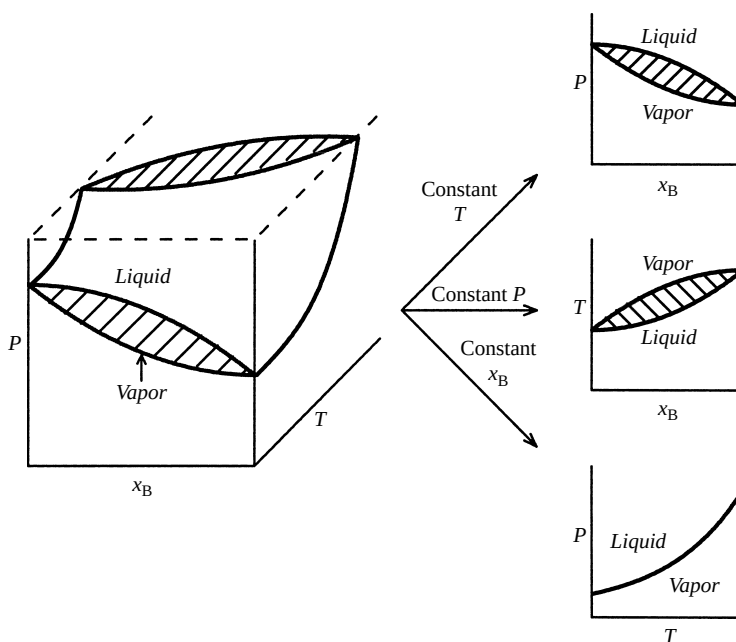


Figure 7.8 Three-dimensional “curved envelope” of the binary fluid P – T – x_B surface (left), showing the upper bubble-point (liquid) surface, the lower dew-point (vapor) surface, and the hatched “inside” of the envelope, together with the three 2D projections (right) that result from slicing the envelope through the plane of constant temperature (upper), pressure (middle), or composition (lower).

and finally on the combined P – T – x behavior of binary fluids as manifested in colligative and osmotic properties (Sections 7.3.6 and 7.3.7). Further aspects of the P – T – x diagram in the low-temperature domain of solid–liquid equilibria will be considered in Section 7.4.

7.3.1 Vapor–Pressure (P – x) Diagrams: Raoult and Henry Limits

Let us now consider the general P – x dependence (vapor-pressure curve) of the binary A/B solution under fixed isothermal conditions (e.g., $T = 25^\circ\text{C}$), with

$$x = x_B = x_{\text{solute}} \quad (7.44a)$$

$$x_A = x_{\text{solvent}} = 1 - x \quad (7.44b)$$

As indicated, we often drop the subscript “B” on the concentration variable x , with the understanding that x ranges from 0 (pure solvent A) to 1 (pure solute B) in the P – x diagram. Initially, we focus on $x = x^{\text{liq}} = x_B^{\text{liq}}$, the concentration of the *liquid* phase, in accord with the usual convention for labeling solution concentration.

We first define the concept of an *ideal solution*. This is the ultrasimplified limiting case in which the partial pressure of each component is simply proportional to its concentration in the liquid,

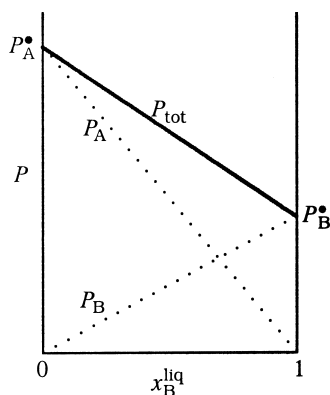
$$P_A = x_A^{\text{liq}} P_A^\bullet \quad (7.45a)$$

$$P_B = x_B^{\text{liq}} P_B^\bullet \quad (7.45b)$$

so that total vapor pressure $P = P_A + P_B$ is

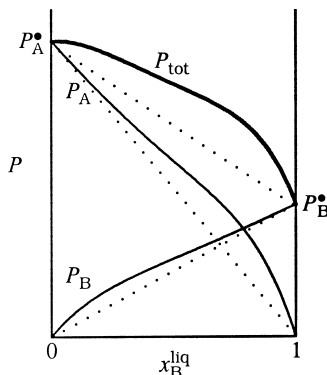
$$P = x_A^{\text{liq}} P_A^\bullet + x_B^{\text{liq}} P_B^\bullet \quad (\text{ideal}) \quad (7.45c)$$

where $P_A^\bullet$ and $P_B^\bullet$ are the vapor pressures of pure components A and B, respectively. Equations (7.45a, b) are reminiscent of Dalton’s law of partial pressures (Section 2.4), but the proportionality is to liquid composition (rather than vapor, as in Dalton’s law). As a result, the total vapor pressure (7.45c) above the solution is simply a straight-line interpolation between $P_A^\bullet$ and $P_B^\bullet$:



The dotted lines represent the component partial pressures (7.45a, b) and the heavy solid line the total vapor pressure (7.45c).

For real solutions, the partial pressures P_A , P_B and total vapor pressure P deviate from the idealized limit (7.45a–c), with deviations of either sign. The following is an illustrative diagram for “positive deviations” ($P > P_{\text{ideal}}$), showing the partial pressures (solid lines) and total pressure (heavy solid line) *exceeding* ideal values (dotted lines):



Two useful limits can be recognized in the partial-pressure curves. At high concentration ($x_B^{\text{liq}} \rightarrow 1$), the P_B curve approaches the limiting ideal limit (7.45b), and is said to satisfy Raoult's (rhymes with “grows”) law:

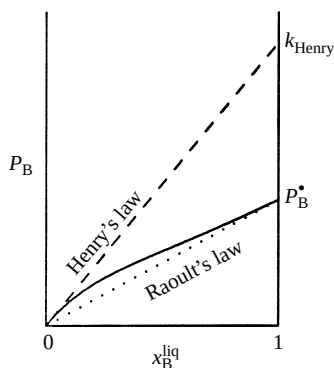
Raoult's law:

$$P_B = x_B^{\text{liq}} P_B^* \quad (\text{accurate for } x_B^{\text{liq}} \rightarrow 1) \quad (7.46a)$$

At low concentration ($x_B^{\text{liq}} \rightarrow 0$), the P_B curve approaches limiting *linear* behavior, characterized by slope $k = k_{\text{Henry}}$; in this limit, the vapor pressure is said to satisfy Henry's law:

Henry's law:

$$P_B = x_B^{\text{liq}} k_{\text{Henry}} \quad (\text{accurate for } x_B^{\text{liq}} \rightarrow 0) \quad (7.46b)$$



An ideal solution therefore satisfies both Raoult's *and* Henry's laws at *all* concentrations, i.e.,

$$k_{\text{Henry}} = P_B^* \quad (\text{ideal}) \quad (7.47)$$

Although Raoult's law (7.46a) seems intuitively obvious for the high-concentration limit $x_B^{\text{liq}} \rightarrow 1$, it may not seem obvious why Henry's law (7.46b) is the expected behavior for the extreme-dilution limit $x_B^{\text{liq}} \rightarrow 0$. However, on general physical grounds we expect that, as $x_B^{\text{liq}} \rightarrow 0$,

- (a) solute–solute interactions become negligible
- (b) solute–solvent interactions “saturate,” with no incremental effect of further dilution

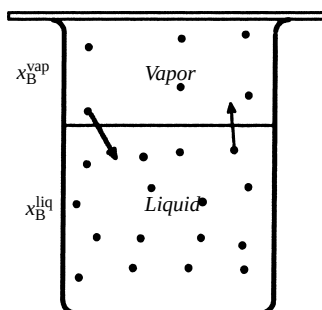
Sidebar 7.7 describes how the “independent solute particle” assumption (a) leads to expected Henry's law behavior in the dilute-solution limit. The “saturation” assumption (b) leads to the expectation (in agreement with observations) that

$$\Delta H_{\text{dilution}} \rightarrow 0, \Delta V_{\text{dilution}} \rightarrow 0, \dots \text{ as } x_B^{\text{liq}} \rightarrow 0, x_A^{\text{liq}} \rightarrow 1 \quad (7.48)$$

consistent with Raoult's law behavior for the solvent.

SIDEBAR 7.7: HENRY'S LAW IN THE INFINITE-DILUTION LIMIT

Consider the following liquid–vapor system, with a small concentration of solute molecules B (small circles) in each phase:



The random thermal motions of each B-particle will occasionally cause a vapor-phase molecule to be “captured” in the liquid phase (heavy arrow), or a liquid-phase molecule to “escape” into the vapor phase (light arrow). In the limit that solute molecules are so dilute that solute–solute interactions can be neglected, the probability of each such capture/escape event is simply proportional to the number density (or mole fraction) of solute particles in the originating phase. We therefore expect that

$$\begin{aligned} \text{capture rate} &\propto x_B^{\text{vap}} \propto P_B \quad (\text{by Dalton's law}) \\ \text{escape rate} &\propto x_B^{\text{liq}} \end{aligned} \quad (S7.7-1)$$

At equilibrium, where the capture and escape rates must be equal, this leads to

$$P_B \propto x_B^{\text{liq}} \quad (S7.7-2)$$

which is Henry's law.

From the above considerations, we can therefore conclude:

- All real solutions obey Henry's law (for the *solute*) when sufficiently dilute.
- All real solutions obey Raoult's law (for the *solvent*) when $x_A^{\text{liq}} \rightarrow 1$.

It therefore follows that:

- For a *sufficiently dilute* solution, it is always safe to assume “Henry's law solution” behavior, namely,

$$\text{solute: } P_B = k_{\text{Henry}} x_B^{\text{liq}} \quad (\text{Henry's law}) \quad (7.49a)$$

$$\text{solvent: } P_A = P_A^{\bullet} x_B^{\text{liq}} \quad (\text{Raoult's law}) \quad (7.49b)$$

As defined by (7.49a, b), a “Henry's law solution” is a more general and useful approximation than an “ideal solution” as defined by (7.45) or (7.47), but each of these approximations is often inadequate for real solutions at concentrations of chemical interest.

The attentive student may be concerned that the concept of “ideal solution” as introduced in Chapter 6 is possibly inconsistent with the usage of that term as defined in (7.45a–c). However, Sidebar 7.8 demonstrates that these definitions of solution “ideality” are in fact consistent. We may therefore regard chemical potential-based definitions of binary solution ideality [cf. (6.57)],

$$\mu_A = \mu_A^{\circ} + RT \ln x_A \quad (7.50a)$$

$$\mu_B = \mu_B^{\circ} + RT \ln x_B \quad (7.50b)$$

as equivalent to (7.45a–c).

SIDEBAR 7.8: EQUIVALENCE OF ALTERNATIVE DEFINITIONS OF SOLUTION IDEALITY

Solution “ideality” was previously defined [cf. (6.57)] in terms of the relationship between chemical potential and mole fraction:

$$\mu_A = \mu_A^{\circ} + RT \ln x_A \quad (\text{S7.8-1})$$

where μ_A° is the chemical potential for component A in the chosen standard state. Alternatively, ideality was also defined [cf. (7.45a)] in terms of the Raoult's-law relationship between partial pressure and mole fraction:

$$P_A = x_A^{\text{liq}} P_A^{\bullet} \quad (\text{S7.8-2})$$

How are these definitions related?

To see the equivalence of these definitions, let us first use the more general (activity-based) equation (6.60) for the liquid phase,

$$\mu_A(\text{liq}) = \mu_A^{\circ}(\text{liq}) + RT \ln a_A^{\text{liq}}$$

together with the standard ideal gas equation (6.55) for the vapor phase,

$$\mu_A(\text{gas}) = \mu_A^\circ(\text{gas}) + RT \ln P_A$$

to express the equality of chemical potentials in the two phases as

$$\mu_A(\text{liq}) = \mu_A(\text{gas}) = \mu_A^\circ(\text{gas}) + RT \ln (P_A^\bullet x_A^{\text{liq}})$$

where we have used the definition (S7.8-2) to replace the partial pressure in the logarithm. The logarithmic product can be broken up to rewrite this equation as

$$\mu_A(\text{liq}) = \{ \mu_A^\circ(\text{gas}) + RT \ln P_A^\bullet \} + RT \ln x_A^{\text{liq}}$$

However, the term in braces (a numerical constant) must be merely the standard state chemical potential for the *liquid* phase (where we choose *pure A* as “standard state” for liquid at the chosen T and P),

$$\mu_A^\circ(\text{liq}) = \mu_A^\circ(\text{gas}) + RT \ln P_A^\bullet$$

which leads finally to

$$\mu_A(\text{liq}) = \mu_A^\circ(\text{liq}) + RT \ln x_A^{\text{liq}}$$

i.e., to the definition (S7.8-1). Thus, use of the ideal gas approximation for the vapor and Raoult’s law (S7.8-2) for the liquid phase is indeed equivalent to the earlier definition (6.57) (or equivalently, $a_A^{\text{liq}} = x_A^{\text{liq}}$ or $\gamma_A^{\text{liq}} = 1$) for “ideality.”

7.3.2 The Lever Rule

Let us now combine the P – x_B^{liq} diagram with the P – x_B^{vap} diagram (cf. Fig. 7.7) to obtain the composite P – x_B diagram shown in Fig. 7.9. As shown in Fig. 7.9a, there are now *three* distinct “ x_B ” composition variables: x_B^{vap} (vapor phase), x_B^{liq} (liquid phase), and x_B^{tot} (total system). How can we read all three composition values from the P – x_B diagram in Fig 7.9b?

As shown in Fig. 7.9, for a given vapor pressure P (dotted line), the compositions x_B^{liq} , x_B^{vap} of the coexisting phases are found from the intersections (small circles) with the liquid and vapor boundaries of the hatched two-phase region. These intersections are connected by a horizontal *tie-line* (heavy solid line) that spans the two-phase “hole” in the diagram. All points along this tie-line represent the *same* thermodynamic *state* (i.e., same temperature, pressure, chemical potentials, and compositions of each phase), but each differs only in the *relative amounts* of each phase (cf. Sidebar 7.2), whether nearly all vapor (at the extreme left of the tie-line), nearly all liquid (at the extreme right), or roughly equimolar amounts of liquid and vapor (near the middle).

More precisely, we can determine the relative molar amounts n^{liq} , n^{vap} of the two phases (and therefore the remaining composition variable x_B^{tot} of the total system) by means of a simple “lever rule” that expresses the overall mass balance of the system. Intuitively, we can see from Fig. 7.9 that x_B^{tot} must be *intermediate* between x_B^{vap} and x_B^{liq} , as expressed

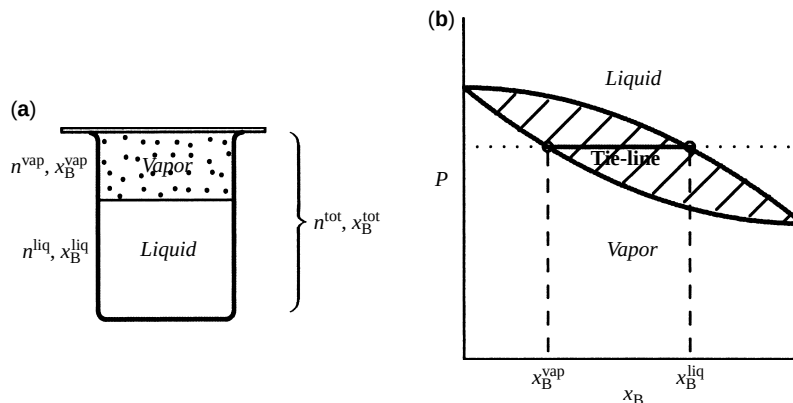
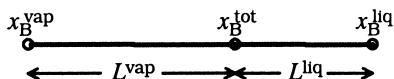


Figure 7.9 P - x_B diagram (b) for a binary two-phase system (a), showing the compositions of coexisting vapor (x_B^{vap}) and liquid (x_B^{liq}) phases for a particular vapor-pressure value (dotted line), and the connecting “tie-line” (heavy solid line) that connects vapor and liquid compositions at this pressure. Varying amounts ($n^{\text{vap}}, n^{\text{liq}}$) of the two phases correspond to different positions along the tie-line, as determined by the lever rule (see text).

by the equation

$$x_B^{\text{tot}} = \frac{n^{\text{vap}} x_B^{\text{vap}} + n^{\text{liq}} x_B^{\text{liq}}}{n^{\text{vap}} + n^{\text{liq}}} \quad (7.51)$$

We can therefore locate x_B^{tot} along the tie-line between x_B^{vap} and x_B^{liq} , dividing the tie-line into segments of length L^{vap} and L^{liq} , respectively, as follows:



In terms of these tie-line segments, the lever rule can now be stated as follows:

$$n^{\text{liq}} L^{\text{liq}} = n^{\text{vap}} L^{\text{vap}} \quad (7.52)$$

Equation (7.52) is equivalent to the simple rule for balancing a schoolyard seesaw: if the fulcrum divides the board into lengths L_1 , L_2 , then the masses m_1 , m_2 at the two ends should satisfy $m_1 L_1 = m_2 L_2$ (i.e., the heavier weight should be at the shorter end) to balance the seesaw. The proof of (7.52) is presented in Sidebar 7.9. The lever rule makes it easy to determine the relative amounts of each phase present from the tie-line ratio, and thus to determine the final unknown composition variable x_B^{tot} from (7.51).

SIDEBAR 7.9: PROOF OF THE LEVER RULE

To prove the lever rule (7.52), we first note that we may express the total number of moles of solution, n^{tot} , either as the sum of chemical components n_A , n_B ,

$$n^{\text{tot}} = n_A + n_B$$

or as the sum of phases $n^{\text{liq}}, n^{\text{vap}}$,

$$n^{\text{tot}} = n^{\text{liq}} + n^{\text{vap}}$$

Equivalently, the quantity of solute n_B can be expressed either as

$$n_B = n^{\text{tot}} x_B^{\text{tot}} = (n^{\text{liq}} + n^{\text{vap}}) x_B^{\text{tot}}$$

or as

$$n_B = n_B^{\text{liq}} + n_B^{\text{vap}} = n^{\text{liq}} x_B^{\text{liq}} + n^{\text{vap}} x_B^{\text{vap}}$$

Comparing the right-hand side of each equation, we obtain

$$0 = n^{\text{vap}} (x_B^{\text{vap}} - x_B^{\text{tot}}) + n^{\text{liq}} (x_B^{\text{liq}} - x_B^{\text{tot}})$$

which, with the substitutions

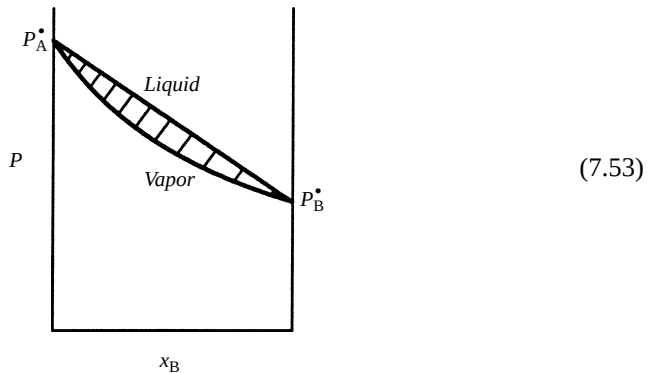
$$\begin{aligned} L^{\text{liq}} &= x_B^{\text{liq}} - x_B^{\text{tot}} \\ L^{\text{vap}} &= x_B^{\text{tot}} - x_B^{\text{vap}} \end{aligned}$$

becomes the lever rule (7.52).

7.3.3 Positive and Negative Deviations

Let us now briefly describe some broader phenomenological aspects of the P - x diagrams for binary solutions, ranging from the ideal solution limit to extreme nonideal deviations of either positive or negative sign.

The limiting linear behavior of the ideal P - x_B^{liq} liquid diagram was previously described in Section 7.3.1. The corresponding P - x_B^{vap} vapor curve can be added to give the full P - x_B diagram for an ideal solution as follows:



Ideal solution behavior is observed only when the solute and solvent molecules have similar sizes and intermolecular interactions, as in benzene/toluene or hexane/octane solutions.

As shown in (7.53), the P – x_B^{vap} boundary always curves *below* the linear P – x_B^{liq} boundary, in such a manner that the vapor phase is always *enriched in the more volatile component*, as physical intuition would suggest:

$$x_B^{\text{vap}} > x_B^{\text{liq}} \quad \text{if } P_B^\bullet > P_A^\bullet \quad (7.54a)$$

$$x_A^{\text{vap}} > x_A^{\text{liq}} \quad \text{if } P_A^\bullet > P_B^\bullet \quad (7.54b)$$

Sidebar 7.10 describes the mathematical relationship between x_B^{liq} and x_B^{vap} for an ideal solution, showing how (7.54a, b) are achieved in this simple case. However, the physically reasonable relationships (7.54a, b) between coexisting liquid and vapor compositions are also satisfied in more general nonideal solutions described below.

SIDEBAR 7.10: LIQUID AND VAPOR COMPOSITIONS FOR IDEAL SOLUTIONS

To derive the relationship between x_B^{vap} and x_B^{liq} for an ideal solution, we first use Dalton's law of partial pressures and Raoult's law (7.45a, b) to write x_B^{vap} as

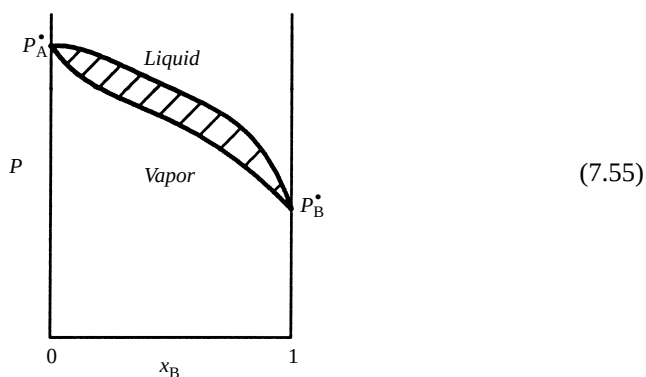
$$x_B^{\text{vap}} = \frac{P_B}{P} = \frac{P_B}{P_A + P_B} = \frac{x_B^{\text{liq}} P_B^\bullet}{x_A^{\text{liq}} P_A^\bullet + x_B^{\text{liq}} P_B^\bullet}$$

If we divide numerator and denominator by $P_B^\bullet$ and substitute $x_B^{\text{liq}} = 1 - x_A^{\text{liq}}$ in the denominator, we obtain

$$x_B^{\text{vap}} = x_B^{\text{liq}} \left[\frac{1}{1 - x_A^{\text{liq}}(1 - P_A^\bullet/P_B^\bullet)} \right]$$

From the right-hand side, one can see that if $P_B^\bullet > P_A^\bullet$, then the denominator is less than unity and the term in brackets is greater than unity, so that $x_B^{\text{liq}} < x_B^{\text{vap}}$, which is (7.54a). The contrary condition $P_A^\bullet > P_B^\bullet$ leads similarly to (7.54b).

Positive deviations from ideality were also depicted in Section 7.3.1 in the form of P – x_B^{liq} curves. The full P – x_B diagram in this case may have the following qualitative appearance:



Qualitatively, such positive deviations can be associated with “phobicity” between components (i.e., solute and solvent that are “unhappy to mix”), corresponding to a tendency toward immiscibility. At the molecular level, this tendency is associated with like-molecule solute–solute or solvent–solvent interactions that are *stronger* than the corresponding solute–solvent interactions in solution, i.e.,

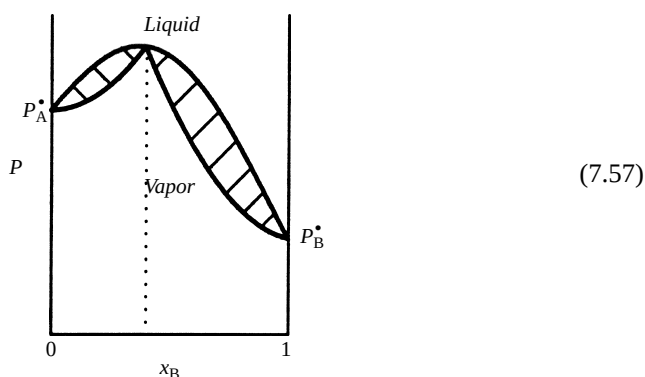
$$A-A, A-B > A-B \text{ interactions} \quad (7.56a)$$

This is also paraphrased in the verbal mnemonic

$$\text{“A is trying to squeeze B out of solution”} \quad (7.56b)$$

to express the increased volatility compared with an ideal solution.

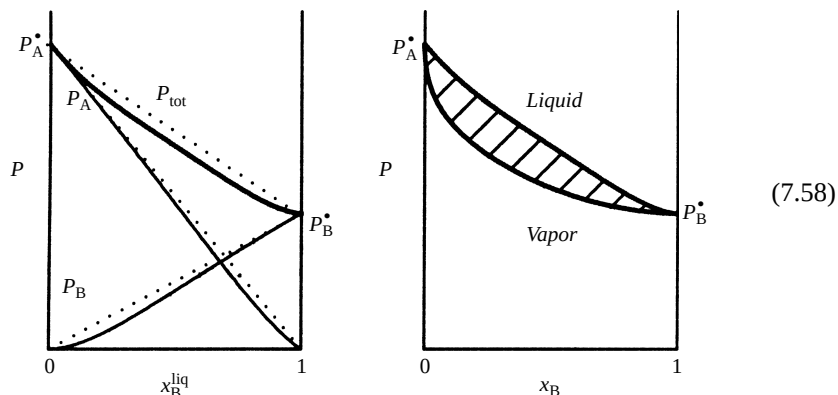
Still more extreme positive deviations can finally lead to a *maximum* in the vapor-pressure curve:



At such a maximum (dotted line) the compositions of coexisting liquid and vapor phases necessarily *coincide*, and the two-phase “hole” closes as liquid and vapor curves meet at this point. Examples of such maxima occur frequently, for example, in alcohol–water mixtures, as in the C_2H_5OH/H_2O system at $x_{H_2O} = 0.20$.

The necessity for liquid and vapor compositions to become identical at a point of vapor-pressure maximum can be readily understood from the Gibbs phase rule. At such a stationary point, the pressure variations perforce *vanish*, corresponding to a formal loss of intensive variability, i.e., to formal reduction of the number of degrees of freedom, f . Since the number of phases remains unchanged, reduction of f must correspond to reduction of c , the number of independent chemical components. The “independence” of components A, B is thereby reduced at the special stoichiometry of a stable A:B complex (e.g., of formula A_nB_m) that is, in principle, an isolable “chemical reaction product” of an active reaction. (Note however, that the actual equilibrium mixture is that of a “parent” A_nB_m complex and its equilibrium dissociation products, so the special “ $x_{A:B}$ ” value may well vary with pressure.) In effect, the A–B phase diagram can be separated at the special A:B-complex stoichiometry (dotted line) into two *independent* binary diagrams: one for A/A:B, the other for B/A:B. Visualizing the separability of the phase diagram at the A:B-complex composition makes it easy to see why liquid and vapor curves must coincide at this composition, as expected for a single component.

Deviations from ideality may also occur in the opposite *negative* sense, as illustrated here for the P - x_B^{liq} curve (left) and the full P - x_B diagram (right):



Such negative deviations can be generally associated with “philicity” between components (i.e., solute and solvent that are “happy to mix”), corresponding to a tendency toward compound formation. At the molecular level, this tendency is associated with unlike-molecule solute–solvent interactions in solution that are *stronger* than the corresponding solute–solute or solvent–solvent interactions in the pure components, i.e.,

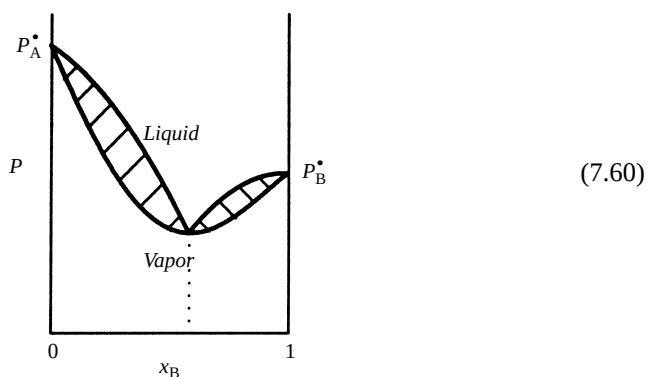
$$A-B > A-A, B-B \text{ interactions} \quad (7.59a)$$

as paraphrased in the verbal mnemonic

$$\text{“A is trying to absorb B into solution”} \quad (7.59b)$$

which expresses the reduced volatility compared with an ideal solution.

As these negative deviations become more extreme, the vapor-pressure curve may exhibit a *minimum*:



At such a minimum (dotted line), the liquid and vapor must again have matching compositions ($x_B^{\text{liq}} = x_B^{\text{vap}}$), corresponding to formal reduction of f (and c) at the point where the pressure variations are vanishing. As in (7.57), this permits separating the phase diagram (7.60) into independent A/A:B and B/A:B diagrams for the specific A:B complex as

a separable component. Examples of systems exhibiting extreme negative deviations include aqueous strong-acid solutions such as $\text{HNO}_3/\text{H}_2\text{O}$ and $\text{HCl}/\text{H}_2\text{O}$.

7.3.4 Boiling-Point Diagrams: Theory of Distillation

We turn now to the T - x_B “boiling-point diagram” (7.43) at fixed pressure (e.g., $P = 1$ atm). Under these conditions, the T - x_B^{liq} curve displays the quantitative variation of the normal boiling point of solution as a function of liquid concentration. The boiling-point diagram therefore provides a basis for exact theoretical analysis of the ancient art of *distillation*, a staple of chemical laboratory practice.

To describe the formal theory of fractional distillation, let us consider the boiling-point diagram of a near-ideal A/B binary solution, as shown in Fig. 7.10. The solution is initially at high concentration x_1 of the high-boiling component B. Consider the following four steps, as illustrated in the figure:

1. The solution at composition x_1 is heated to its boiling point at T_1 , producing vapor of composition x_2 (enriched in low-boiling component A).
2. The vapor of composition x_2 is collected and condensed to liquid at temperature T_2 .
3. The condensate from Step 2 is boiled at T_2 , producing vapor of composition x_3 .
4. The vapor from Step 3 is condensed to a liquid at T_3 ; and so forth.

Steps 1–4 constitute two “theoretical plates” of a fractional distillation sequence that successively reduces the concentration of B (enriches the concentration of A) in the distillate.

It is clear that we can purify the low-boiling component A to any desired degree by employing a sufficient number of theoretical plates, if the boiling point diagram has the qualitative form shown in Fig. 7.10. More generally, for any specified initial concentration x_i and final concentration x_f , we can see how to determine the exact number of fractional distillation steps required to achieve the desired degree of separation, and the exact temperature needed at each distillation step. Although laboratory distillations are often carried out in continuous reflux mode, rather than in the theoretically optimal sequence of discrete

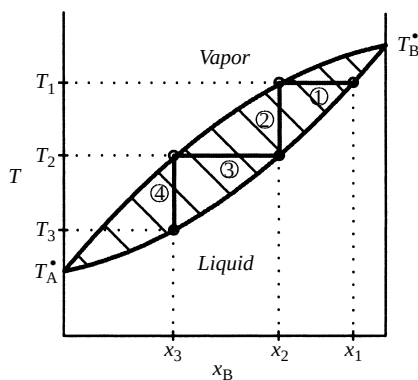
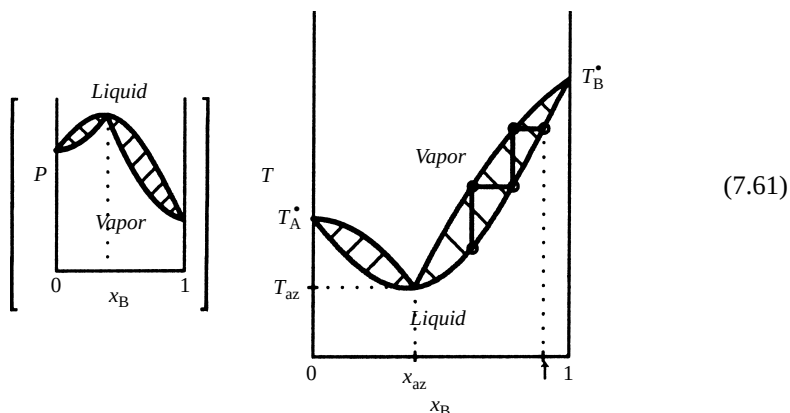


Figure 7.10 Boiling-point diagram for a binary solution, illustrating two “theoretical plates” of an idealized fractional distillation process that progressively enriches the distillate in the low-boiling component A (see text).

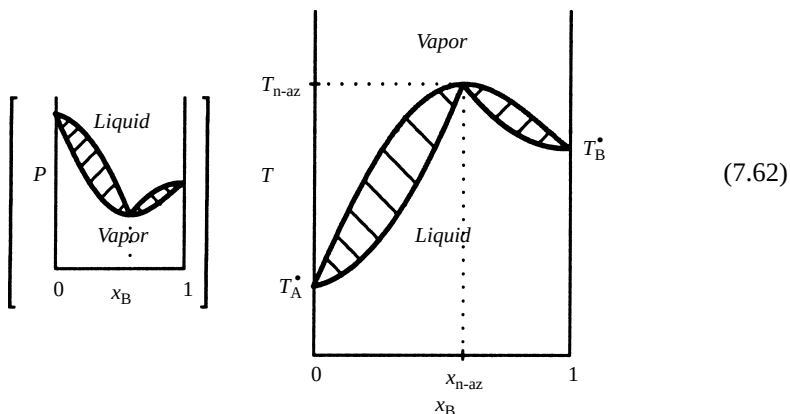
steps, accurate knowledge of the boiling point diagram underlies efficient design of modern industrial distillation processes.

The distillation process becomes more complex as deviations from ideality become more extreme. For the case (7.57) of extreme positive deviations (shown in brackets below), the T - x_B diagram has the form shown at the right:



As shown in the diagram, the boiling-point curve now exhibits a *minimum* (at the same point where the P - x_B curve exhibits a *maximum*, for reasons that will be discussed below). The special minimum-boiling mixture is known as an *azeotrope*, with composition x_{az} and boiling point T_{az} . If a distillation process is initiated at high concentration $x_B > x_{az}$ (arrow), the distillation steps lead, as shown, down to the azeotrope. Similarly, a distillation initiated at low $x_B < x_{az}$ must again lead to the azeotrope. In short, *all distillations end at the azeotrope* (a special low-boiling mixture that boils to a vapor of the same composition), and all further attempts to purify the solution by distillation are futile. A well-known azeotropic mixture is 4 wt% H_2O in C_2H_5OH , which limits the best efforts of moonshiners. Other azeotropes occur for various ketone/water, alcohol/water, alcohol/chloroform, and related mixtures.

For the contrary case (7.59) of extreme negative deviations (in brackets below), the T - x_B diagram takes the form shown at the right:



In this case, the boiling-point curve passes through a *maximum* (again, at the same composition where P - x_B passes through a minimum) where $x_B^{liq} = x_B^{vap}$. This maximum-boiling

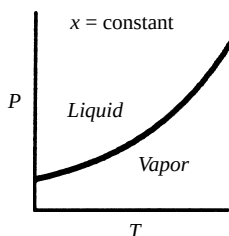
mixture is called a *negative azeotrope*, labeled $(x_{n\text{-az}}, T_{n\text{-az}})$ in the diagram. It is clear from the diagram that any distillation process must move *away from* the negative azeotrope, i.e., toward pure A if $x_{\text{initial}} < x_{n\text{-az}}$ or toward pure B if $x_{\text{initial}} > x_{n\text{-az}}$. Well-known cases of such negative azeotropic behavior include 20.2 wt% HCl in H_2O , 68 wt% HNO_3 in H_2O , 79 wt% chloroform in acetone, and many other hydrogen-bonded mixtures. Except for the rather vague statements summarized in (7.56a, b), (7.59a, b), the molecular-level details of positive and negative azeotropic behavior are at present rather obscure.

Why have we made such sweeping assumptions about the close relationships between maximum/minimum features in the P - x_B and T - x_B diagrams? Sidebar 7.11 sketches the thermodynamic proof that P - x_B and T - x_B diagrams are generally related by a kind of “upside-down” symmetry, such that the slopes and extrema of one type of diagram are precisely mirrored (in an upside-down sense) by the other type of diagram, as illustrated in (7.61) or (7.62). This thermodynamic symmetry makes it simple to sketch the expected topological features of either type of diagram from the known features of the other.

SIDEBAR 7.11: THERMODYNAMIC SYMMETRY OF P - x AND T - x DIAGRAMS

What is the relationship between the P - x and T - x diagrams? Thermodynamics dictates a deep symmetry, based on general properties of the P - T - x surface for a binary solution (Fig. 7.8).

Consider first the P - T behavior at constant composition x :



On general physical grounds, we expect that the curve $P = P(T)$ must curve *upward* at any fixed composition x :

$$\left(\frac{\partial P}{\partial T}\right)_x > 0$$

For a binary fluid with $c = 2$, $p = 2$, the phase rule gives $f = 2$ degrees of freedom. Under these conditions, we know from the Jacobi cyclic identity (1.14b) that

$$\left(\frac{\partial P}{\partial T}\right)_x = \frac{(\partial x / \partial T)_P}{(\partial x / \partial T)_T} > 0$$

From this inequality, we deduce that the derivatives $(\partial P / \partial x)_T$ and $(\partial T / \partial x)_P$ must have *opposite* signs, i.e., if the P - x curve is *increasing*, then the T - x curve is simultaneously

decreasing, and vice versa. In particular, if the P - x curve is passing through a *maximum*, then the T - x curve is simultaneously passing through a *minimum*, as in (7.61), or if P - x is at a *minimum*, then T - x is at a *maximum*, as in (7.62). Thus, the P - x and T - x curves are locked into a complementary symmetry of *opposed* slopes (and extrema) at every point of the phase diagram, consistent with their depictions in (7.61), (7.62).

7.3.5 Immiscibility and Consolute Behavior

As positive deviations (7.56) increase (as in oil–water mixtures), the tendency toward immiscibility and phase separation increases. In this case, the simple “liquid” region of (7.61) may separate into two distinct liquid phases: one “A-rich” and one “B-rich” in composition. Such liquid–liquid *partial* miscibility is a ubiquitous feature of aqueous solutions of hydrophobic solutes, particularly at lower temperatures.

Figure 7.11 illustrates the T - x phase diagram for a binary A/B solution with partially miscible liquid phases. As shown in the figure, the T - x diagram now has *two* two-phase “hole” (hatched) regions: the upper for liquid–vapor coexistence and the lower for A-rich/B-rich liquid–liquid coexistence. The two hatched regions (between T_{az} and T_c) are separated by a “miscibility gap” in which the two components are miscible in all proportions, leading to a single homogeneous liquid phase. However, below the *consolute temperature* T_c (at the top of the lower coexistence dome), the components are only miscible for compositions lying at the left (“A-rich liquid” region) or right (“B-rich liquid” region) of the coexistence dome. Within the dome region, separation into two immiscible

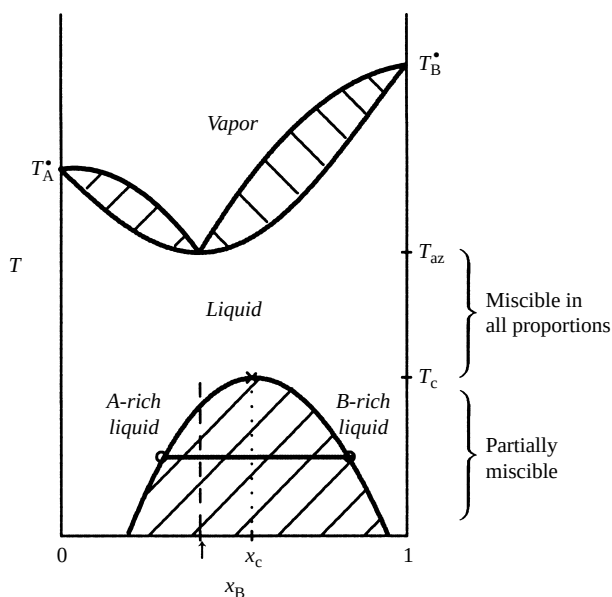


Figure 7.11 Schematic T - x phase diagram for a binary A/B solution exhibiting partial immiscibility and liquid–liquid phase separation below the consolute temperature T_c . The horizontal tie-line (heavy solid line) connects the compositions of coexisting A-rich and B-rich liquid phases (small circles) in the lower liquid–liquid coexistence dome. (See text for description of behavior along vertical dashed and dotted lines.)

liquids is observed, with the horizontal tie-line (heavy solid line) giving the compositions of coexisting A-rich and B-rich liquids in the usual way.

The special *upper consolute point* ($\times$) atop the coexistence dome at (x_c, T_c) is in fact a *critical* point that terminates the two-phase liquid–liquid coexistence line in the P – T diagram. At this point, the distinction between “A-rich” and “B-rich” liquids disappears, corresponding to *vanishing* of the meniscus that separates these phases at lower temperatures. The critical discontinuity from $p = 2$ phases (or $f = 1$ degree of freedom) below T_c to $p = 1$ phase (or $f = 2$ degrees of freedom) above T_c is accompanied by the usual spectacular critical-point phenomena, including divergences in optical scattering (opalescence), heat capacity, compressibility, and other properties.

In contrast, a heterogeneous solution of *noncritical* composition (e.g., $x < x_c$, as shown by the arrow and dashed line in Fig. 7.11) shows a qualitatively different behavior as it rises through the coexistence boundary and into the homogeneous region near and above T_c . For each increase in temperature along the dashed line in Fig. 7.11, a horizontal tie-line yields both the compositions of the A-rich and B-rich liquids (from the two ends of the tie-line), as well as the relative amounts of each phase (from the lever rule). Clearly, the critical composition x_c remains near the middle of the tie-line as T increases toward T_c , whereas a noncritical composition $x \neq x_c$ moves toward one or other terminus of the tie-line as the temperature is raised.

The upper portion of Fig. 7.12 depicts a succession of “snapshots” of the solution along the dashed line of Fig. 7.11, showing, for example, the greater ratio of A-rich (lighter color,

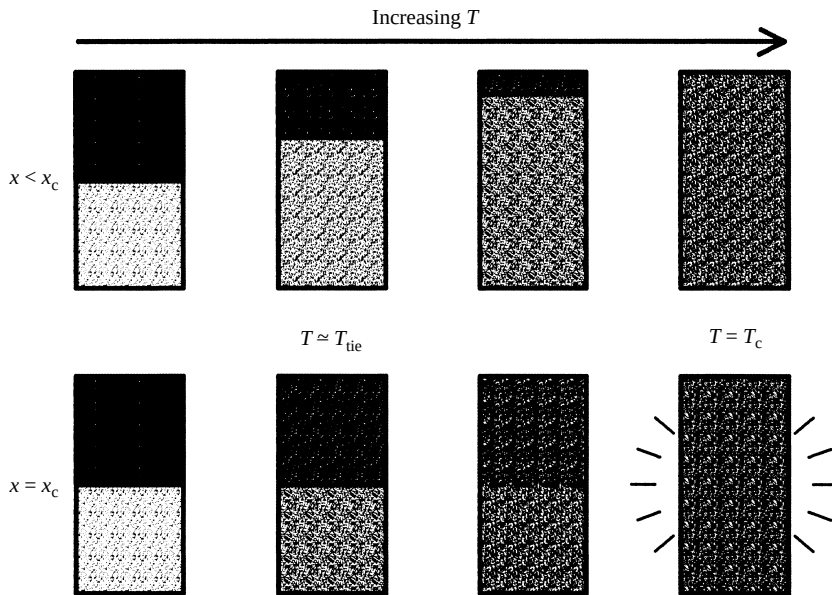
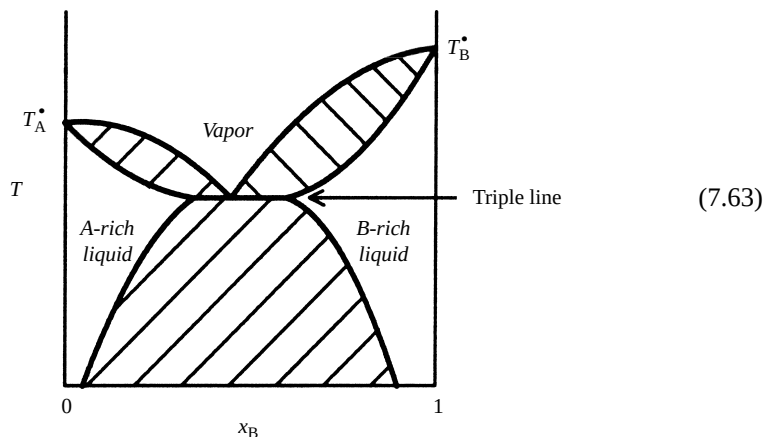


Figure 7.12 Successive “snapshots” of a binary A/B system in the immiscible region of coexisting A-rich (lighter) and B-rich (darker) liquid phases (Fig. 7.11), comparing the T -dependent behavior for noncritical (upper sequence; cf. dashed line in Fig. 7.11) versus critical composition (lower sequence; cf. dotted line in Fig. 7.11), and showing how the meniscus “rises out of the container” in the first case but “vanishes” at the critical point T_c in the second case. [Both solutions are taken to be fairly comparable in the starting low- T snapshot, but their deviations are readily apparent in the second snapshot at the temperature T_{tie} of the horizontal tie-line in Fig. 7.11.]

higher density) to B-rich (darker color, lower density) liquid needed to satisfy the lever rule at the tie-line temperature T_{tie} . As the temperature continues to rise, the meniscus is seen to “rise through the top of the container,” but with no other particular noteworthy change as the system enters the single-phase region, the meniscus still being plainly visible as it escapes through the top of the container. (By the same reasoning, the meniscus would instead “fall out of the bottom of the container” if $x > x_c$.) However, the behavior at the critical composition $x = x_c$ (dotted line in Fig. 7.11) is fundamentally different, as shown in the lower snapshots of Fig. 7.12. Now, the meniscus hardly moves, but it becomes less and less distinct as T increases toward the consolute point T_c . At that point, the meniscus vanishes, the system erupts in the characteristic display of critical-point phenomena, and the distinction between “A-rich” and “B-rich” liquids is no longer meaningful. The consolute point ($\times$) in Fig. 7.11 therefore represents a critical singular feature that is strongly distinguished from other points along the liquid–liquid coexistence boundary.

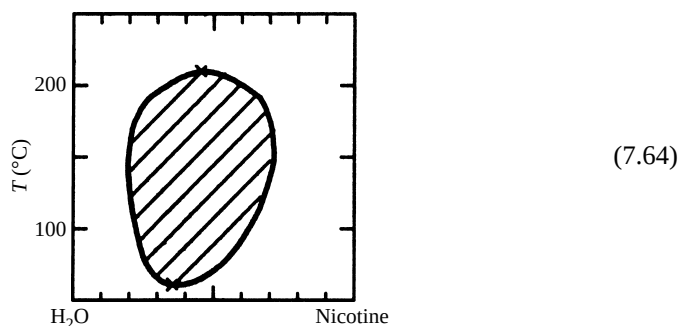
We can also envision the limit in which positive deviations are still more extreme (i.e., A–B interactions are still less favorable compared with A–A, B–B interactions), leading finally to *closing* of the miscibility gap, as follows:



In this limit, the two coexistence regions of Fig. 7.11 have merged at a *triple line* ($p = 3$), and the two homogeneous liquids (A-rich and B-rich) now occupy separated regions of the phase diagram, with no connecting region of continuous A/B variability. One can picture the diagram (7.63) as arising from a two-phase liquid solution that *vaporizes* before reaching its consolute point of complete liquid/liquid miscibility. Further positive deviations from ideality will generally cause the A-rich and B-rich homogeneous liquid regions to further narrow and separate toward opposite edges of the diagram, corresponding to ever-stronger aversion to mixing and more complete partitioning of the two components into separate phases (as in oil/water mixtures).

Lower temperature tends to reduce miscibility, so the liquid–liquid coexistence region tends to spread to a wider composition range at lower T , as shown in Fig. 7.11 or (7.63). However, exceptional cases are known in which the liquid–liquid coexistence region terminates in a *lower* consolute point, so that complete miscibility can be achieved by *cooling* below this point. Perhaps most remarkable in this respect is the famous nicotine/water

system, which exhibits *both* upper and lower consolute points [C. S. Hudson. *Z. Phys. Chem.* **47**, 113 (1904)], marked by $\times$'s in the following schematic phase diagram:



In the hatched two-phase region of limited miscibility, the system separates heterogeneously into water-rich and nicotine-rich layers. However, at temperatures below the lower consolute point (about 61°C) or above the upper consolute point (about 210°C), the components become miscible in all proportions, resulting in a uniform homogeneous phase. The molecular-level origins of this extraordinary behavior, as well as more general aspects of consolute behavior in other (typically, hydrogen-bonded) systems, remain deeply obscure.

7.3.6 Colligative Properties and Van't Hoff Osmotic Equation

For the limiting case of a dilute Henry's law solution of a *nonvolatile* solute (e.g., salt or other soluble solid) with $P_A^{\bullet} \simeq 0$, the expression for the vapor pressure of the solution reduces to the simplified form

$$P = x_A P_A^{\bullet} = (1 - x_B) P_A^{\bullet} \quad (7.65)$$

with a (Raoult-type) contribution from the solvent only. In this limit, the solution properties depend on the solute B only through the mole fraction x_B , i.e., through the *number* of solute particles (not their chemical identity). The characteristic solution properties of this limit are known as “colligative properties” (referring to the dependence on solute particle number only).

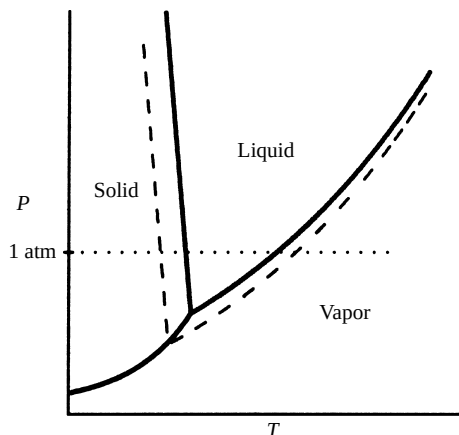
For aqueous solutions of nonvolatile solutes, the important colligative properties include

- vapor-pressure lowering
- boiling-point elevation
- freezing-point depression
- osmotic pressure

The first three colligative effects follow rather straightforwardly from (7.65), as described in Sidebar 7.12. The remainder of this section will be devoted to describing osmotic pressure and related dialysis phenomena.

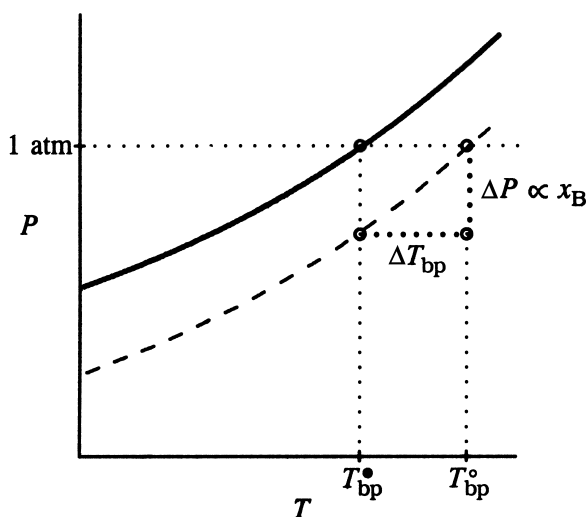
SIDEBAR 7.12: VAPOR-PRESSURE LOWERING AND ASSOCIATED BOILING AND FREEZING POINT DISPLACEMENTS

The following diagram depicts the general dependence of P versus T for pure solvent A (solid lines) and for the binary A/B solution (dashed lines) with $P_A^\bullet \approx 0$:



According to (7.65), the vapor-pressure lowering $\Delta P = P_A - P$ is directly proportional to x_B . As shown in the diagram, the dashed vapor-pressure curve for solution therefore falls *below* the solid pure-solvent curve, corresponding to the well-known effect of a nonvolatile solute in reducing the vapor pressure of solution.

Some additional details of the vapor-pressure curve in the neighborhood of the normal boiling point ($P = 1$ atm) are shown in the following expanded diagram:



Consistent with the vapor-pressure lowering, the boiling point ($T_{bp}^\bullet$) of pure solvent is perforce shifted to a *higher* value $T_{bp}^\circ(x_B)$ for any finite solute concentration x_B . As shown by the diagram, the vapor-pressure lowering ΔP (proportional to x_B) and boiling-point

elevation $\Delta T_{\text{bp}} \equiv T_{\text{bp}}^{\circ} - T_{\text{bp}}^{\bullet}$ form the edges of a parallelogram (heavy dotted lines). For sufficiently small x_{B} , these edges become proportional (with proportionality constant that can be determined from the Clapeyron equation for the slope of the coexistence boundary),

$$\Delta T_{\text{bp}} \propto \Delta P \propto x_{\text{B}} \quad (\text{S7.12-1})$$

In the limit of small x_{B} , the solution *molality* m_{B} (= moles of solute per kilogram of solution) is also proportional to x_{B} :

$$m_{\text{B}} = \frac{n_{\text{B}}}{\text{kg solution}} \propto n_{\text{B}} \propto x_{\text{B}} \quad (\text{S7.12-2})$$

The overall proportionality between ΔT_{bp} and solute molality m_{B} can therefore be expressed as

$$\Delta T_{\text{bp}} = k_{\text{pb}} m_{\text{B}} \quad (\text{S7.12-3})$$

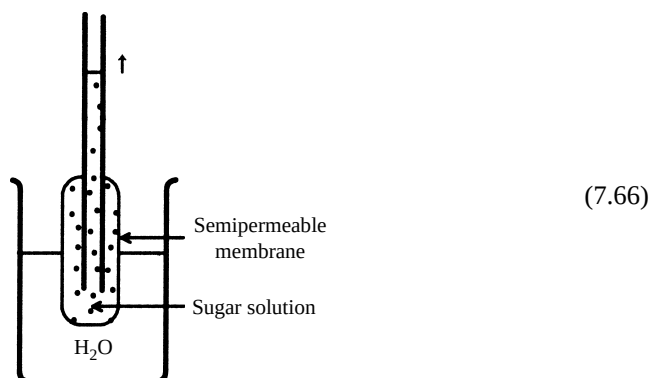
where k_{bp} is the “molal boiling-point constant,” dependent on properties of the solvent. For water, for example, $k_{\text{bp}}(\text{H}_2\text{O}) = 0.512^{\circ}\text{C m}^{-1}$. This is the effect that acts to prevent radiator boil-over when a component of low volatility is added to the cooling water.

As shown in the first diagram in this sidebar (for water), the melting point of solution is also displaced toward *lower* values by addition of a nonvolatile solute, the “freezing-point depression” effect. By arguments similar to those given above, the freezing-point depression $\Delta T_{\text{fp}} \equiv T_{\text{fp}}^{\bullet} - T_{\text{fp}}^{\circ}(x_{\text{B}})$ will also be found to be proportional to solute molality:

$$\Delta T_{\text{fp}} = k_{\text{fp}} m_{\text{B}} \quad (\text{S7.12-4})$$

where k_{fp} is the “molal freezing-point constant” for the solvent. For water, for example, $k_{\text{fp}}(\text{H}_2\text{O}) = 1.86^{\circ}\text{C m}^{-1}$. This is the effect that acts to keep streets ice-free when road-salt is applied under wintry conditions.

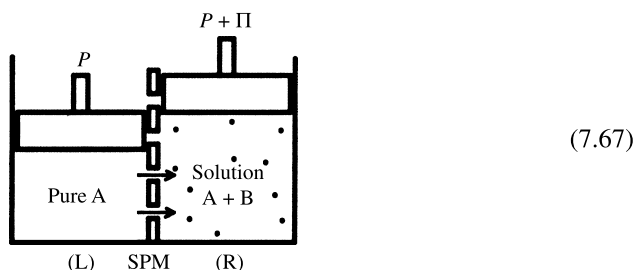
The phenomenon of osmotic pressure is often demonstrated with apparatus such as the following:



An open beaker of pure water, under ordinary atmospheric pressure, contains a glass tube to which is attached a semipermeable membrane filled with an aqueous solution of, for example, sugar (or salt or other solute of low volatility). As its name implies, the membrane

is permeable (porous) to pure water but not to solute, such semipermeability being an innate property of many biological membranes. Under these conditions, pure water will be seen to spontaneously diffuse (“osmose”) through the membrane into the sugar solution, causing the latter to rise in the glass tube until it is far above the level in the surrounding beaker, seemingly defying gravity! How *high* the sugar solution will ultimately be lifted by osmosis is dependent on the precise solute concentration, in a manner to be described below. However, it is remarkable that fluids can be elevated to considerable heights by this technique, far above the 32 feet or so achievable by a vacuum pump. Indeed, osmotic pressure is thought to be the principal mechanism by which tall trees convey sap up to the leafy crown, often 100 feet or more above their roots.

The osmotic pressure phenomenon can be visualized in terms of the left (L) and right (R) pressure chambers shown below, connected through a semipermeable membrane (SPM) that permits passage of A but not of B:



The left chamber contains pure A at pressure P , whereas the right chamber contains an A/B solution at pressure $P + \Pi$. Under these conditions, solvent A will spontaneously flow from left to right, increasing the pressure head on the solution chamber. The incremental pressure Π needed to *prevent* solvent flow (osmosis) is called the “osmotic pressure.”

Let us now analyze osmotic flow from the viewpoint of the Gibbsian equilibration conditions (Chapter 5). Because volume V cannot be freely exchanged between chambers, the pressure does not equalize: $P^{(L)} \neq P^{(R)}$. Similarly, solute n_B cannot freely exchange, so $\mu_B^{(L)} \neq \mu_B^{(R)}$. However, solvent n_A is free to exchange through the semipermeable membrane, so its chemical potential must *equalize* between left and right chambers:

$$\mu_A^{(L)} = \mu_A^{(R)} \quad (7.68a)$$

If we express $\mu_A = \mu_A(T, P, x_A)$ as a mathematical function of temperature, pressure, and concentration, we can rewrite the equilibration condition (7.68a) as

$$\mu_A(T, P, 1) = \mu_A(T, P + \Pi, 1 - x_B) \quad (7.68b)$$

The physical picture underlying (7.68b) is that $\mu_A^{(R)}$ is *lowered* by the effect of dilution with B, but *raised* by the added pressure Π , with the two effects exactly balanced to maintain equality with $\mu_A^{(L)}$.

To evaluate Π from (7.68b), let us expand $\mu_A(T, P + \Pi, 1 - x_B)$ as a *double* Taylor series in the pressure and concentration dependence, assuming that both $\Delta P (= \Pi)$ and

$\Delta x_A (= -x_B)$ are sufficiently small that corrections beyond the first order can be neglected (cf. Section 1.4):

$$\begin{aligned} \mu_A(T, P + \Delta P, 1 + \Delta x_A) &= \mu_A(T, P, 1) + \left(\frac{\partial \mu_A}{\partial P} \right)_{T, x_A} \bigg|_{(L)} \Delta P \\ &\quad + \left(\frac{\partial \mu_A}{\partial x_A} \right)_{T, P} \bigg|_{(L)} \Delta x_A + \cdots \end{aligned} \quad (7.69)$$

[The vertical bar and subscript “(L)” denote that the partial derivatives are to be evaluated in the infinite-dilution limit of the left-hand chamber.] In order that (7.68b) be satisfied for all sufficiently small ΔP and Δx_A , the two first-order correction terms in (7.69) must *cancel*, i.e. (omitting the evaluation limit for simplicity),

$$\left(\frac{\partial \mu_A}{\partial P} \right)_{T, x_A} \Delta P = - \left(\frac{\partial \mu_A}{\partial x_A} \right)_{T, P} \Delta x_A \quad (7.70a)$$

or

$$\left(\frac{\partial \mu_A}{\partial P} \right)_{T, x_A} (\Pi) = \left(\frac{\partial \mu_A}{\partial x_A} \right)_{T, P} (x_B) \quad (7.70b)$$

For the pressure dependence of μ_A in the pure-A limit, we can use the Gibbs–Duhem equation ($d\mu_A = \bar{V}_A dP - \bar{S}_A dT$) to obtain

$$\left(\frac{\partial \mu_A}{\partial P} \right)_{T, x_A} = \bar{V}_A = \frac{V_A^\bullet}{n_A} \quad (7.71a)$$

For the concentration dependence of μ_A , we can employ the dilute-solution Henry’s law limit ($\mu_A = \mu_A^\bullet + RT \ln x_A$) to obtain

$$\left(\frac{\partial \mu_A}{\partial x_A} \right)_{T, P} = \frac{RT}{x_A} \quad (7.71b)$$

From (7.70b), we obtain therefore

$$\Pi \left(\frac{V_A^\bullet}{n_A} \right) = x_B \left(\frac{RT}{x_A} \right) = RT \left(\frac{n_B}{n_A} \right) \quad (7.72)$$

In the assumed dilute-solution limit, we can also assume that $V_A^\bullet \simeq V$, the total volume of solution. Under these conditions, (7.72) reduces to the *Van’t Hoff osmotic equation*

$$\boxed{\Pi V = n_B RT} \quad (7.73)$$

As the derivation makes clear, the Van’t Hoff equation (7.73) is valid for dilute ideal solutions only. The leading correction for nonideality can be evaluated from the virial

expansion (Section 5.8.1) in terms of the second virial coefficient $B(T)$, namely,

$$\Pi = RT \left[\frac{n_B}{V} + B(T) \left(\frac{n_B}{V} \right)^2 + \cdots \right] \quad (7.74)$$

As in the analogous case of gases (Section 2.4), corrections for nonideality can be obtained by measurements of osmotic pressure at different solute concentrations, with extrapolation toward the infinite-dilution limit. For electrolytes, the correction for ionic dissociation is important.

Because osmotic pressures can be experimentally measured down to rather low values, the Van't Hoff equation proves to be valuable for determining the molecular weights of proteins and other high polymers, as illustrated in Sidebar 7.13. Other practical aspects of osmosis, dialysis, and reverse osmosis phenomena in the physiological and industrial domain are described briefly in Sidebar 7.14.

SIDEBAR 7.13: MOLECULAR WEIGHT DETERMINATION BY OSMOTIC-PRESSURE MEASUREMENTS

The evident formal similarity between the Van't Hoff equation (7.73) and the ideal gas equation (2.2) suggests many parallels in applications, such as molecular-weight (MW) determinations. Just as conventional pressure measurements are used to determine the MW of gaseous species by the familiar Dumas flask method, so can osmotic pressure measurements determine the MW of solutes in dilute aqueous solutions. However, the latter method can be extended to a considerably higher MW range, applicable to nonvolatile macromolecular species. Although now largely supplanted by other techniques, osmotic-pressure measurements were historically important in obtaining the first MW estimates for large biomolecules.

To see how this is done, consider g_B grams of solute B of MW M_B (with $n_B = g_B/M_B$) dissolved in volume V at known temperature T . We can rewrite the Van't Hoff equation (7.73) in the form

$$M_B = \frac{g_B RT}{\Pi V}$$

Because osmotic pressures can be routinely measured down to the sub-torr (about 10^{-3} atm) level, the above equation permits MW determinations ranging up to *kilodaltons* (kDa, 10^3 atomic mass units) and beyond, as shown in the following problem.

Problem 50 mg of an unknown protein is dissolved in 10.0 mL of water and found to give an osmotic pressure of 1.4 Torr at standard state conditions. What is the MW of the protein?

Solution From the Van't Hoff equation as rewritten above, we find

$$M_{\text{protein}} = \frac{(50 \times 10^{-3} \text{ g})(0.082 \text{ L atm mol}^{-1} \text{ K}^{-1})(298 \text{ K})}{(1.4 \text{ Torr})(1 \text{ atm}/760 \text{ Torr})(10 \text{ mL})} \simeq 67 \times 10^3 \text{ g/mol}$$

or 67 kDa. This is the approximate MW of hemoglobin [as first determined by osmotic pressure measurements of G. S. Adair. *Proc. R. Soc. Lond. A* **120**, 595–603 (1928)].

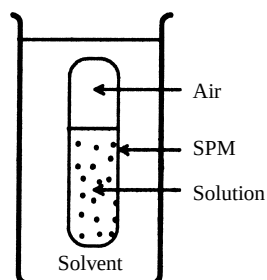
SIDEBAR 7.14: OSMOSIS, REVERSE OSMOSIS, AND DIALYSIS PHENOMENA

Osmotic phenomena of considerable importance are found throughout the natural world. Osmosis also underlies many important medical and industrial applications, as will now be briefly summarized.

Central to the osmosis phenomenon is the semipermeable membrane (SPM), whose physical properties and species-selectivity directly govern the kinetics and thermodynamics of osmotic flow. Naturally occurring biomembranes of high selectivity, permeable to water but not to other solutes, are ubiquitous, for example, in macroscopic stomach linings and blood vessels, as well as in the microscopic cell membranes that encapsulate all known cell types. Some common synthetic membranes, such as Gore-Tex and cellophane, also exhibit selective permeability and osmotic activity.

Although standard “cartoons” of the SPM [as in (7.67)] represent the selectivity as being due to physical hole size (with smaller water molecules able to pass through the pores, while larger solute molecules are blocked), this is surely an unrealistic molecular-level picture for most biomembranes. Indeed, at the cellular level, elaborate membrane-embedded proteins are known to control the transport of components in and out of the cell (with maintenance of crucial concentration gradients across cell boundaries) using highly specific chemical interactions that are able to reject unwanted components of both smaller and larger size. Fortunately, the thermodynamics of osmotic phenomena is independent of our (lack of) present knowledge concerning the molecular-level details of membrane selectivity.

A more general type of “dialysis” exchange between fluids of different solute concentration is illustrated in the following diagram:



In this configuration, the SPM is fashioned into a closed “sausage” containing a solution of known concentration and an air pocket. The dialysis sausage is then immersed in pure solvent (as shown) or in a solution of unknown concentration. The resulting concentration gradient across the SPM induces osmotic flow, which alters the pressure within the dialysis sausage, thereby compressing the air pocket, which acts as the internal “barometer” of osmotic pressure. From the measured shift of the air–solution interface, one can evaluate

the volume decrease (and pressure increase) due to concentration differences inside and outside the sausage. If the component to which the SPM is permeable corresponds to an undesired toxic contaminant, the dialysis sausage acts as a “purifier” to remove the contaminant from the immersion fluid.

In medical applications, the dialysis SPM may be the patient’s own stomach lining. A prepared solution is infused into the abdomen, stimulating osmotic flow of toxins across the stomach lining into the ingested solution, which is subsequently drained from the stomach. Alternatively, the dialyzer for blood dialysis (hemodialysis) may be a prepared membrane with special solution over which the blood flows to osmotically remove impurities.

Osmotic disequilibrium may also become the *source* of serious medical malfunctions. For example, an injection of pure water (or other fluid of improper pH or ionic strength) into the bloodstream can induce osmotic pressure differences that rupture blood cells with fatal consequences. Similarly, drinking salty ocean water creates an osmotic imbalance across the stomach lining that draws water into the stomach from surrounding organs, eventually resulting in death (ironically, from dehydration). In general, the fluid systems of plants and animals participate in a complex web of osmotic equilibria and disequilibria in which pH, ionic strength, and other concentration variables must be carefully controlled for proper biological function.

One can also recognize that application of sufficient pressure (above the equilibrium osmotic pressure Π) to the right-hand chamber in (7.67) must cause the solvent flow to *reverse*, resulting in extrusion of pure solvent from solution. This is the phenomenon of “reverse osmosis,” an important industrial process for water desalination. Reverse osmosis is also used for other purification processes, such as removal of H_2O from ethanol beyond the azeotropic limit of distillation (Section 7.3.4). Reverse osmosis also finds numerous applications in wastewater treatment, solvent recovery, and pollution control processes.

7.3.7 Activity and Activity Coefficients

“Activity” and related post-Gibbsian concepts were previously introduced in Sections 5.8.1 and 6.4. We now wish to describe how these concepts are employed in the general framework of applications to nonideal solutions.

In general, “activity” is merely an alternative way to express “chemical potential.” The general objective is to express μ_i in a form that emulates the ideal gas expression (6.55), but with the *actual* vapor pressure P_i of component i (rather than that assumed from Dalton’s law). The “trick” will be to choose a standard-state divisor in (6.55) that makes this expression valid for the components of a real solution.

One can begin by defining an “absolute activity” $a_{i,\text{abs}}$ of component i through the equation

$$\mu_i \equiv RT \ln a_{i,\text{abs}} \quad (7.75a)$$

where, like $\mu_i = \mu_i(T, P, x)$ itself,

$$a_{i,\text{abs}} = a_{i,\text{abs}}(T, P, x) \quad (7.75b)$$

is a function of variables T, P, x only. One can then define the “relative activity” $a_{i,\text{rel}}$ as the *ratio*

$$a_{i,\text{rel}} = \frac{a_{i,\text{abs}}}{a_{i,\text{abs}}^\circ} \quad (7.76a)$$

where $a_{i,\text{abs}}^\circ$ is the absolute activity of a chosen “standard state” (to be discussed below), as denoted by a small circle/degree symbol (or sometimes by a filled circle, or star, or other symbol according to context). Thus, the numerical value of $a_{i,\text{rel}}$ also becomes dependent on the choice of standard state:

$$a_{i,\text{rel}} = a_{i,\text{rel}}(T, P, x, \text{ standard state}) \quad (7.76b)$$

If $\mu_i^\circ \equiv RT \ln a_{i,\text{abs}}^\circ$ denotes the standard-state chemical potential, then by construction,

$$\mu_i = \mu_i^\circ + RT \ln a_{i,\text{rel}} \quad (7.77)$$

This quantity $a_{i,\text{rel}}$ is what is usually called “the” activity a_i , and we henceforth proceed with this simplified notation. However, the implicit dependence (7.76b) of a_i on the particular choice of standard state should be kept in mind.

More generally, the relationship (7.77) between activity a_i and chemical potential μ_i of chemical component i will be written variously as

$$\mu_i = \begin{pmatrix} \mu_i^\circ \\ \text{or} \\ \mu_i^* \\ \text{or} \\ \mu_i^\bullet \\ \text{etc.} \end{pmatrix} + RT \ln a_i \quad (7.78)$$

The first term (no matter how symbolized) is a *constant*, which can be identified as the “chemical potential at unit activity.” This is the usual starting point for discussing the relationship between chemical potential and activity.

To proceed, we assume that the activity a_i can be expressed as the product of an “activity coefficient” γ_i and the chosen unit of concentration, namely,

$$a_i = \gamma_i \quad (\text{concentration unit}) \quad (7.79)$$

Thus, (7.79) might be written in any of the explicit forms

$$a_i = \gamma_i^x \cdot x_i \quad (\text{mole fraction scale}) \quad (7.80a)$$

$$a_i = \gamma_i^c \cdot c_i \quad (\text{molarity scale}) \quad (7.80b)$$

$$a_i = \gamma_i^m \cdot m_i \quad (\text{molality scale}) \quad (7.80c)$$

depending on the concentration unit of choice. Each activity coefficient γ_i can therefore be expressed in terms of the functional dependence

$$\gamma_i = \gamma_i(T, P, x, \text{ standard state, concentration scale}) \quad (7.81)$$

In this section, we shall adopt the mole-fraction scale (7.80a), so that

$$\gamma_i = \gamma_i^x \quad (7.82)$$

and (7.79) can be written in simplified notation as

$$a_i = \gamma_i x_i \quad (7.83)$$

Note that the successive activity-related quantities involve an ever-deeper chain of dependences on arbitrary conventions [cf. (7.75b), (7.76b), (7.81)].

We must now address the choice of standard state ($^\circ$). Tautologically, this can always be defined as the state of unit (relative) activity, i.e.,

$$a_i^\circ = a_{i,\text{rel}}^\circ \equiv 1 \quad (7.84)$$

From (7.83), we can therefore see that the standard-state activity coefficient γ_i is just the reciprocal of the corresponding standard-state concentration (in our case, mole fraction):

$$\gamma_i^\circ = \frac{a_i^\circ}{x_i^\circ} = \frac{1}{x_i^\circ} \quad (7.85)$$

We now consider two distinct conventions for choosing the standard state.

Convention I (Raoult's Law, "Solvent Convention," or "Rational Convention") In this case, the standard state is taken as the *pure substance* ($x_i^\circ = x_i^\bullet = 1$), so that, by (7.85), we are choosing $\gamma_i^\circ \equiv 1$ [so that $\mu_i^\circ \equiv \mu_i^\bullet$, suggesting the filled-circle notation in (7.78)]. Thus, we can state this convention as

$$\lim_{x_i \rightarrow 1} \gamma_i \rightarrow 1 \quad (\text{convention I}) \quad (7.86)$$

This is a consistent ("rational") physical choice, well adapted to the Raoult's law limit for the solvent.

Convention II (Henry's Law, "Solute Convention," or "Practical Convention") In this case, we make the somewhat counterintuitive ("practical") choice

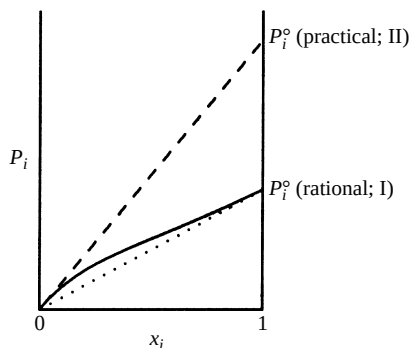
$$\lim_{x_i \rightarrow 0} \gamma_i \rightarrow 1 \quad (\text{convention II}) \quad (7.87)$$

This convention requires that $\gamma_i \rightarrow 1$ as $x_i \rightarrow 0$ [whereas, by definition, $\gamma_i^\circ \rightarrow 1$ as $x_i^\circ \rightarrow 1$; cf. (7.85)]. In words, this paradoxical choice of standard state is often expressed as follows:

Standard state II: The *hypothetical* standard state in which the solute at unit concentration ($x_i = 1$) has the properties it *would* have at infinite dilution ($x_i = 0$).

Such a statement reminds us that the "standard state" is a wholly arbitrary convention that need have *no* physical significance (except as a practical expedient), and may be chosen *differently* for different components of the solution. As indicated, the hypothetical standard state II is specifically adapted to the solute, i.e., to limiting Henry's law behavior as $\gamma_i \rightarrow 1$.

What does all this mean? In the simplest terms, we can express these choices graphically as follows:



As can be seen, the two standard-state choices correspond to the limiting Raoult's law [P_i° (I)] or Henry's law [P_i° (II)] dashed lines at $x_i = 1$. Accordingly, for solving problems involving activities, we can always use the equation

$$a_i = \gamma_i x_i = \frac{P_i}{P_i^\circ} \quad (7.88)$$

where we choose

$$P_i^\circ = P_i^\bullet \quad \text{for solvent } (i = A) \quad (7.89a)$$

$$P_i^\circ = k_{\text{Henry}} \quad \text{for solute } (i = B) \quad (7.89b)$$

For ideal solutions, these two conventions become equivalent ($P_i^\circ = P_i^\bullet = k_{\text{Henry}}$), but for real solutions the distinction between solvent (P_A°) and solute (P_B°) standard states must be kept in mind.

7.4 BINARY SOLID-LIQUID EQUILIBRIA

Binary solid-liquid equilibria encompass the phenomena of melting and zone refinement, as well as ordinary solid-liquid solubility equilibria. The former are of particular interest to the metallurgist, geologist, and materials engineer, while the latter are of particular interest to solution chemists and crystallographers. Many features of liquid-vapor and liquid-liquid phase equilibria (Section 7.3) are echoed in liquid-solid and solid-solid phase diagrams, but only selected aspects of the latter are discussed in the present section. Because solid-liquid pressure effects are generally much less important than those in liquid-vapor equilibria (owing to the similarity in condensed-phase densities; cf. the Clapeyron equation), our primary focus in this section is on the temperature-dependent (T - x) aspect of the phase diagram, e.g., for fixed $P = 1$ atm. Also, because solid phases tend to exhibit complete immiscibility, analogs of partial immiscibility or liquid-liquid consolute behavior are rather uncommon below the melting point, and will be mentioned only in passing.

The T - x diagrams for binary solid–liquid systems can be categorized into four primary types:

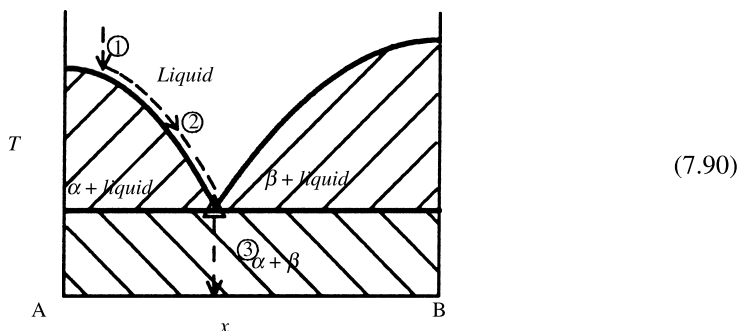
1. eutectic diagrams
2. congruent melting diagrams
3. incongruent melting diagrams
4. solid solution (alloy) diagrams

We shall first describe representative behavior for each type (Sections 7.4.1–7.4.4), then sketch how continuous changes in intermolecular interactions are expected to lead continuously from one type of T - x behavior to another (Section 7.4.5), including rather uncommon features such as solid–solid consolute points.

7.4.1 Eutectic Behavior

Eutectic diagrams (from Greek *ευτηκτοσ*: “easily melted”) represent the T - x melting behavior for binary systems with completely immiscible solid phases α , β . The solid α , β phases often correspond to (virtually) pure components A, B, respectively, so we may treat phase and component labels (rather loosely) as interchangeable in this limit.

The qualitative form of a eutectic diagram is as follows:



The diagram shows the homogeneous liquid (melt) region (with $f = 2$) as well as the three distinct two-phase regions (each with $f = 1$). The junction of these regions is the *eutectic point* (triangle), a triple point (with $f = 0$).

In each hatched two-phase region, the lever rule (Section 7.3.2) can be used as usual to determine the relative amounts of the two phases at opposite ends of the tie-line. However, the quantity of precipitated solid α and/or β is usually of less interest than the composition of the melt, so the principal focus is on the two liquidus lines that meet at the eutectic point. These liquidus lines are also called “solubility curves” or “freezing-point depression curves,” in that they map both the saturation-solubility limits (horizontal variations) as well as the freezing-point depression of the liquid (vertical variations).

The dashed lines in (7.90) illustrate an expected cooling “path” for a liquid chosen to be of initial concentration $x = x_B \simeq 0.1$. As cooling proceeds (1), the first precipitate of solid α appears when the cooling path meets the α -liquidus curve. Thereafter, as solid α proceeds to freeze out of solution, the melt is successively depleted of component A and moves down the α -liquidus curve (2) toward the eutectic point. At this point, the remaining liquid freezes to a mix of solid α and β (in proportion to eutectic composition), and the mix of solids

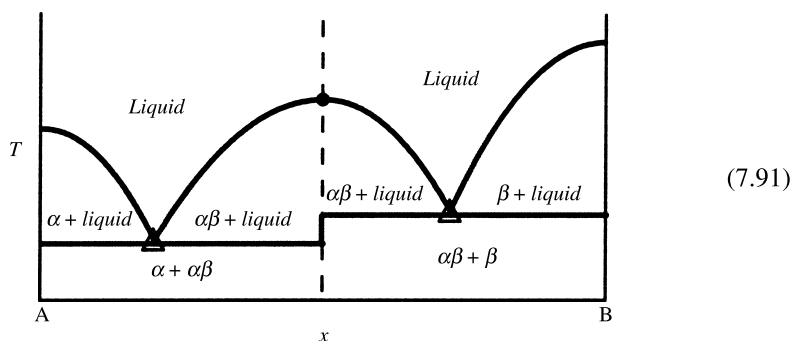
continues to cool (3) uneventfully. (An initial melt of composition $x \simeq 0.9$ would similarly track down the β -liquidus curve until meeting the same eutectic point.)

Binary systems exhibiting simple eutectic behavior are rather common, including metallic pairs such as Bi/Cd or Pb/Sb, inorganic saline solutions such as NaCl/H₂O, and organic systems such as naphthalene/benzene.

7.4.2 Congruent Melting

A congruent melting diagram corresponds to formation of a solid compound $\alpha\beta$. As usual, compound formation corresponds to formal reduction of c and f , leading to a *maximum* (with vanishing temperature variation) at the “congruent melting point” for the specific $\alpha\beta$ stoichiometry. A congruent melting-point maximum in the solid-liquid diagram is formally analogous to a negative azeotrope (Section 7.3.4) in the liquid-vapor diagram. At the special composition of congruent melting, the solid melts to a liquid of identical composition.

The qualitative form of a congruent melting-point diagram is as follows (with hatch-marks in two-phase regions omitted for clarity):



As can be seen, the diagram now contains *two* eutectic points (triangles) connected by the congruent-melting maximum (circle) at the special stoichiometry (dashed line) of the $\alpha\beta$ compound. In effect, the congruent-melting diagram (7.91) can be *divided* at the dashed line into two distinct eutectic diagrams: one for the binary $\alpha/\alpha\beta$ system, the other for the binary $\beta/\alpha\beta$ system.

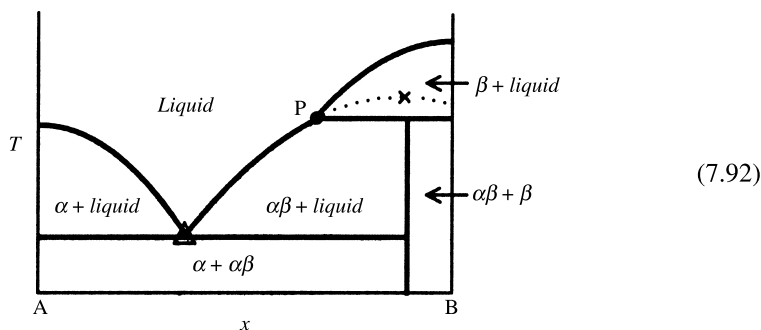
As in Section 7.4.1, one can visualize the path that would be followed if a liquid at or near the special $\alpha\beta$ phase composition ($x_{\alpha\beta}$) is gradually cooled. As shown in (7.91), the first solid to appear is the congruent $\alpha\beta$ compound, which has a maximum freezing point. If the initial $x = x_B$ is slightly less than the congruent composition ($x < x_{\alpha\beta}$), the melt follows the limb of the $\alpha\beta$ -liquidus curve to the left of the dashed line, ending at the α (left) eutectic point, where remaining liquid freezes to solid α . Correspondingly, if $x > x_{\alpha\beta}$, then the melt follows the $\alpha\beta$ -liquidus limb to the right, ending at the (higher) β eutectic point, where remaining liquid freezes to solid β . The student should similarly try to envision the cooling path if the initial x lies close to pure A or pure B. Altogether, there are now *six* distinct two-phase regions (as well as the two distinct eutectic “bays” of the homogeneous liquid region) in which the system might be found, but one can make consistent use of the tie-line and lever rule in each two-phase region to judge the appropriate solid precipitates (α , β , or $\alpha\beta$) and liquid compositions in each case.

Congruent melting points are commonly encountered both in metallic and organic systems. The former is exemplified by the binary Zn/Mg system (with MgZn_2 compound), whereas the latter is exemplified by phenol/aniline mixtures (with 1 : 1 hydrogen-bonded complex).

7.4.3 Incongruent Melting and Peritectics

An “incongruent melting point” or “peritectic point” can be considered to result from a congruent compound that *decomposes* below its own melting point. This produces a characteristic break in slope that (unlike a eutectic) occurs on the *shoulder* of the compound $\alpha\beta$ -liquidus curve, *below* the expected maximum point for the congruent melting point, which no longer appears. While the precipitating solids below the peritectic point continue to reflect $\alpha\beta$ compound formation, no special features occur on the liquidus curve at this composition, which seems instead to represent an “interrupted” form of the liquidus curve for the “missing” β eutectic. It is as though the two halves of the congruent diagram (7.91) are somehow “jammed together,” causing the congruent maximum and adjacent β eutectic to *disappear*, and leaving an awkward peritectic “weld” in their place.

The diagram in (7.92) below illustrates a schematic form of an incongruent melting point (circle), showing the discontinuity of slope that occurs at the peritectic point P. The solid precipitates below this point all continue to be of pure α , β , or compound $\alpha\beta$ type, with no special precipitate of peritectic composition x_p . Also shown in the diagram is a dotted line representing the continuation of the $\alpha\beta$ -liquidus line through the “ghost” congruent melting point ($\times$), which of course is never reached owing to premature decomposition. For clarity, hatch-marks have again been omitted in the five distinct two-phase regions of the diagram.



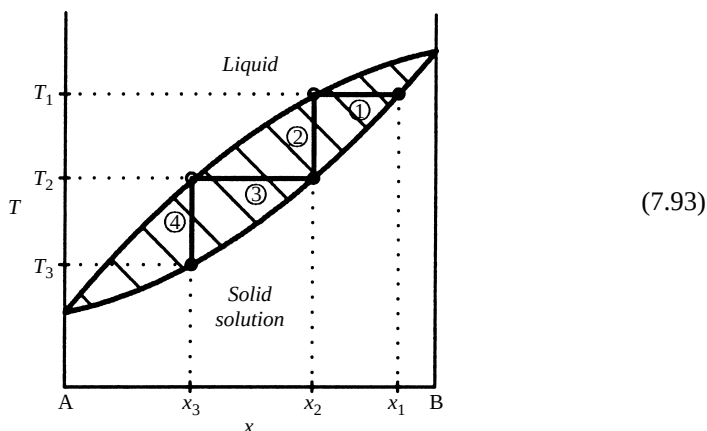
As shown in the diagram, a melt with composition just above that of the peritectic point ($x > x_p$) will initially freeze out solid β (only). However, below the peritectic temperature, only solid compound $\alpha\beta$ freezes out. The peritectic point is therefore a triple point for β , $\alpha\beta$, and liquid phases, representing the highest temperature at which solid $\alpha\beta$ compound survives. Unlike the case of congruent melting, the coexisting liquid and solid(s) at an incongruent melting point are of differing composition.

Well-known examples of peritectic behavior occur in the binary Na/K system (from decomposition of Na_2K) and the $\text{Na}_2\text{SO}_4/\text{H}_2\text{O}$ system (from decomposition of the decahydrate).

7.4.4 Alloys and Partial Miscibility

In contrast to the highly immiscible cases considered above, we may also consider the limit of A/B solids that are completely miscible in all proportions, leading to “solid solution” or

“alloy” behavior. In this limit, the solid–liquid T – x diagram has the following qualitative form, analogous to the corresponding liquid–vapor diagram (cf. Fig. 7.10):



The analogy extends also to practical “zone melting” methods (analogous to distillation) for purifying the alloy by successive melting and recrystallization. The diagram illustrates four steps (two theoretical plates) of an idealized zone melting sequence that purifies the alloy from x_1 to x_3 , successively enriched in the lower-melting component.

Alloy formation is of great technological importance. Indeed, the “Bronze Age,” “Iron Age,” and other epochal historical transformations are characteristically identified with metallurgical advances in alloying. Well-known binary or ternary alloys that underlie aspects of modern life include bronze (Cu/Sn), brass (Cu/Zn), stainless steel (Fe/Ni/Cr), coinage and jewelry metals (Au/Ni, Ag/Cu), amalgams (Hg/Ag/Cu), and aluminum can material (Al/Mn), among many others.

7.4.5 Phase Boundaries and Gibbs Free Energy of Mixing

The near-ideal limit depicted in (7.93) corresponds to only a single solid phase, with melting point and other properties that are intermediate between those of its pure components. However, numerous alloys of technological importance are engineered for nonideal deviations, with mechanical and thermal properties *beyond* those of either component. In such a case, multiple phase domains may also occur in the low-temperature solid region, leading to solid–solid *partial* miscibility, miscibility gaps, consolute points, and other features analogous to those well known in liquid–liquid and liquid–vapor systems. In this section, we seek to gain a more general overview of nonideal solid–liquid behavior and the underlying Gibbs free energy of mixing and microscopic A/B interactions in the competing phases.

How can one understand the connection between the phase diagram and the underlying intermolecular forces? Each physical A/B system comes with fixed intermolecular interactions and fixed T – x diagram. However, we can attempt to visualize how a family of continuously varying A/B potential energy surfaces leads to *continuous* changes of phase diagram from one extreme “type” to another. The variations in solid–liquid T – x diagrams can be related to specific mathematical forms of $G_{\text{liq}}(x)$, $G_{\text{sol}}(x)$, the Gibbs free energy of mixing of competing phases. The $G_{\text{liq}}(x)$, $G_{\text{sol}}(x)$ mixing functions, in turn, can be derived from phenomenological solution models (Sidebar 7.15) or more complete statistical

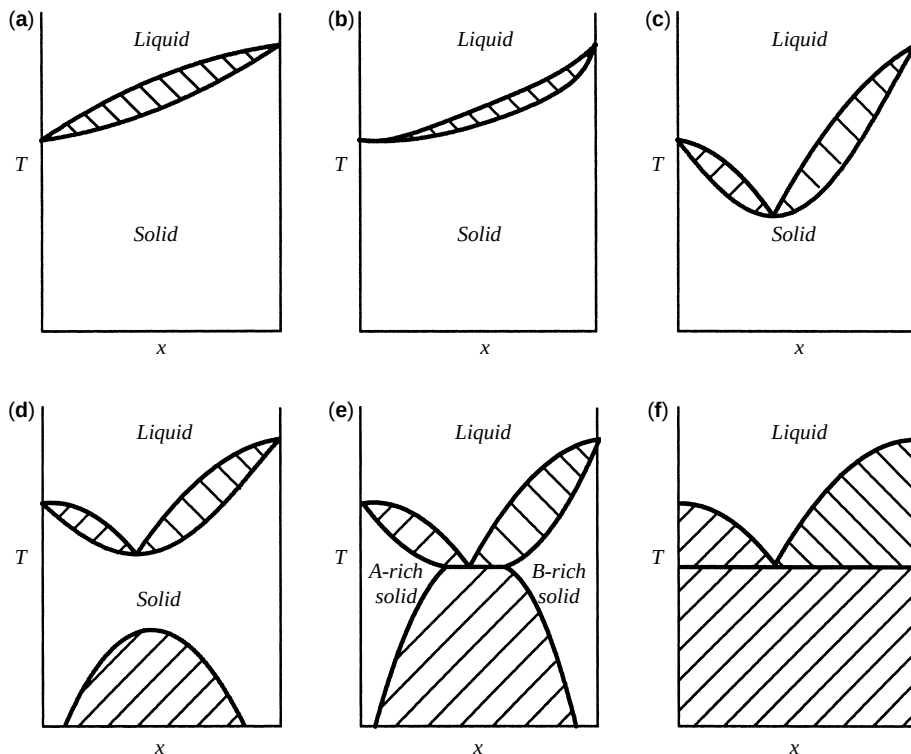


Figure 7.13 Continuous changes of solid–liquid phase-diagram “type” from (a) ideal to (f) eutectic, showing (b) positive nonideality that progressively results in (c) a low-melting minimum, (d) a region of solid–solid immiscibility, or (e) complete closing of the solubility gap, culminating in (f) complete eutectic immiscibility.

thermodynamic treatments (Chapter 13) that incorporate the intermolecular A–A, B–B, and A–B interactions at a microscopic level.

Consider the sequence of phase diagrams shown in Fig. 7.13. This depicts the morphing of the T – x diagram from ideal type (Fig. 7.13a) to eutectic type (Fig. 7.13f) in a continuous movie-like sequence of frames a–f. Starting from the simplest ideal-type dependence (corresponding to complete A/B miscibility) as shown in Fig. 7.13a, we can envision the progressive increase of positive deviations (“A trying to squeeze B out of solution”) as shown in Fig. 7.13b. Increasing nonideality eventually leads to an azeotropic-like *minimum* in the melting curve (Fig. 7.13c). As A–B incompatibility further increases, a region of solid–solid *immiscibility* appears at lower temperature (Fig. 7.13d). The solid–solid miscibility gap eventually closes (Fig. 7.13e), with alloy-type miscibility only at the extremes of A-rich and B-rich concentration. Finally, at the most extreme degree of immiscibility, limiting eutectic behavior is achieved (Fig. 7.13f).

How can the details of Fig. 7.13 be related to the underlying Gibbs free energy changes? As previously seen (Section 7.2.2), the phase boundaries are ultimately traceable to the mathematical form of the Gibbs free energy function for competing phases. For the T – x diagrams shown above, we must focus on the forms of $G_{\text{liq}}(x)$ and $G_{\text{sol}}(x)$ at chosen T and P , and more specifically on the derivatives of these functions that determine the chemical potential of solute B in each phase. (Note that we can restrict attention to solute concentration $x = x_B$, because solvent concentration $x_A = 1 - x_B$ is not an independent variable.)

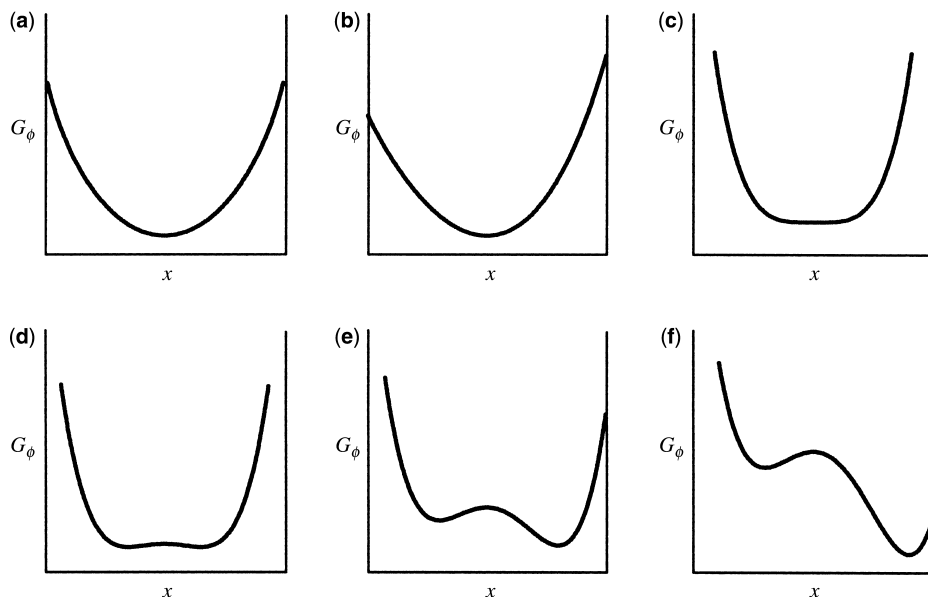


Figure 7.14 Representative forms of Gibbs free energy of mixing $G_\phi(x)$: (a) ideal single well, dominated by entropy of mixing; (b) asymmetric single well; (c) critical (consolute) limit ($G_\phi'' = 0$); (d) shallow symmetric double well; (e) asymmetric double well; (f) highly asymmetric double well.

The general requirement for matching chemical potentials thereby determines how the derivatives (slopes) of $G_{\text{liq}}(x)$ and $G_{\text{sol}}(x)$ must be related in order for these phases to coexist, giving rise to a hatched coexistence region.

Qualitatively, the form of $G_\phi(x)$ ($\phi = \text{liquid or solid}$) might be expected to vary from simple single-well behavior to more complex double-well or multiwell form, as depicted schematically in Fig. 7.14. Figure 7.14a depicts the ideal limiting form that is dominated by the entropy of mixing [cf. (6.59c)], leading to a simple minimum in $G_\phi(x)$ at $x = 0.5$. More complex nonideal A/B interactions progressively alter the form of $G_\phi(x)$ to distorted single-well (Fig. 7.14b) or critical (consolute)-limit behavior (Fig. 7.14c), or to symmetric or asymmetric double-well character (Fig. 7.14d–f), and so forth. In principle, the specific mathematical form of $G_\phi(x)$ (including its temperature and pressure dependence) is obtainable, for example, from phenomenological solution models (Sidebar 7.15). However, for the present illustrative purposes, it is sufficient to consider qualitative features of $G_\phi(x)$ as depicted in the frames of Fig. 7.14.

Given the forms of $G_{\text{liq}}(x)$, $G_{\text{sol}}(x)$ for a binary A/B system, the Gibbs equilibration condition requires minimization of total Gibbs free energy at the chosen T, P . If $G_{\text{liq}}(x) < G_{\text{sol}}(x)$ for all x , then only the liquid phase is stable, and a horizontal cross-section through the T – x diagram lies wholly in the liquid region. Conversely, if $G_{\text{sol}}(x) < G_{\text{liq}}(x)$ for all x , then only the solid phase is stable. However, if $G_{\text{liq}}(x)$ and $G_{\text{sol}}(x)$ curves *cross* at one or more values of x , more complex phase *coexistence* becomes possible that leads to appearance of hatched two-phase regions in the T – x diagram.

As described in Section 7.2.2, the phase-coexistence conditions are expressed in terms of the required matching variations of chemical potential in the two phases. Geometrically, this means that the $G_{\text{liq}}(x)$ and $G_{\text{sol}}(x)$ curves must have a *common tangent* line, which contacts the G_{liq} curve at x_{liq} and the G_{sol} curve at x_{sol} . Figure 7.15 illustrates this

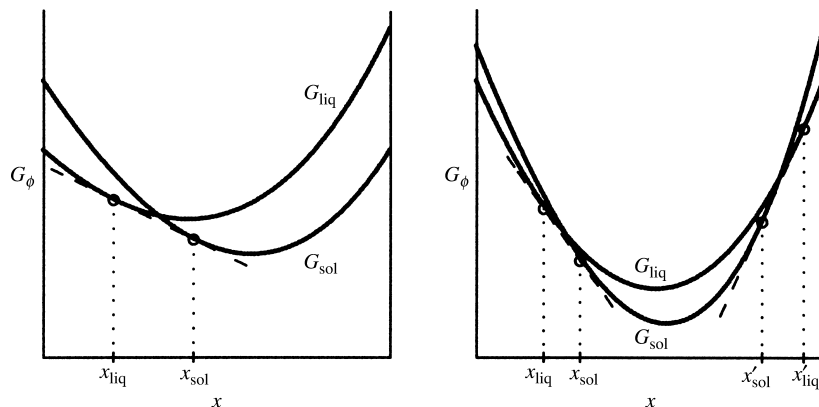


Figure 7.15 Illustrations of the Gibbs double-tangent construction for finding coexistence regions of $G_{\text{liq}}(x)$, $G_{\text{sol}}(x)$ in the case of (a) a single crossing or (b) double crossing.

double-tangent construction for two representative cases. In Fig. 7.15a, the G_{liq} and G_{sol} curves undergo only a single crossing, leading to a single two-phase coexistence region (between x_{liq} and x_{sol}); this corresponds, for example, to a horizontal slice through Fig. 7.13a, b that passes through a single hatched region. In Fig. 7.15b, two crossings occur, leading to two distinct coexistence regions (from x_{liq} to x_{sol} and from x'_{liq} to x'_{sol}), corresponding, for example, to a horizontal slice through the “rabbit ears” in the upper part of Fig. 7.13c–e that crosses two hatched regions. If $G_{\text{sol}}(x)$ develops double-well character (corresponding to two distinct possible solid phases, as in Fig. 7.14a–f), then one can similarly envision their common tangent leading to a solid–solid coexistence region, of the type shown in the lower portion of Fig. 7.13d. Rather simple and continuous changes in the form of $G_{\phi}(x)$ (Fig. 7.14) can thereby be seen to lead to continuous variations in the double-tangent construction (Fig. 7.15) that give rise to the richly varied features of T – x diagrams (Fig. 7.13).

Although examples of solid–solid consolute behavior such as depicted in Fig. 7.13d are known experimentally (e.g., in the binary CH_4/Ar system), such behavior is rather unusual and seems to differ from that of analogous liquid–liquid consolute points in important respects. As previously mentioned (Section 7.2.3), solid-phase equilibria are often complicated by *metastability* and the sluggishness of achieving the significant microstructural rearrangements associated with solid–solid phase transitions. The practical description of solid alloys therefore involves issues of metastability, fatigue, or other kinetic phenomena, as well as domain structure, grain size, surface and defect structures, and so forth, that lie outside the idealized equilibrium thermodynamics framework. A particularly interesting and puzzling type of solid–solid phase transition occurs in “shape memory alloys,” as described briefly in Sidebar 7.16. Further aspects of liquid–solid and solid–solid equilibria, including many rich geological and metallurgical applications, are beyond the scope of this book.

SIDEBAR 7.15: PHENOMENOLOGICAL SOLUTION MODELS

The Gibbs free energy of mixing $G^{\phi}(x)$ contains the basic information needed to evaluate the contribution of phase ϕ to A/B solution properties (for notational convenience here,

we have changed the index “ ϕ ” from a subscript to a superscript). Accordingly, a specific functional form for $G^\phi(x)$ can be considered equivalent to a physical “model” of solution behavior. In principle, $G^\phi(x)$ can be determined from the fundamental statistical thermodynamic partition function (Chapter 13), based on molecular-level description of A/B interactions, but such a detailed molecular model is often impractical. An alternative approach is to postulate the mathematical form of $G^\phi(x)$, with adjustable empirical parameters fitted to experimental solution data. We shall briefly describe some phenomenological solution models that have proven useful in representing experimental data.

The usual starting point for empirical solution models is separation of enthalpic and entropic components of G^ϕ :

$$G^\phi(x) = H^\phi(x) - TS^\phi(x) \quad (\text{S7.15-1})$$

into ideal (id) and “excess” (xs) contributions, namely,

$$\begin{aligned} H^\phi(x) &= H_{\text{id}}^\phi(x) + H_{\text{xs}}^\phi(x) \\ S^\phi(x) &= S_{\text{id}}^\phi(x) + S_{\text{xs}}^\phi(x) \end{aligned} \quad (\text{S7.15-2})$$

where the ideal molar contributions are [cf. (6.59c, d)]

$$\begin{aligned} H_{\text{id}}^\phi(x) &= 0 \\ S_{\text{id}}^\phi(x) &= -R[x \ln(x) + (1-x) \ln(1-x)] \end{aligned} \quad (\text{S7.15-3})$$

Specific solution models therefore involve specific mathematical assumptions concerning the excess functions $H_{\text{xs}}^\phi(x)$, $S_{\text{xs}}^\phi(x)$. (If phase ϕ is not the standard-state form of component A or B, an additional contribution is needed for the free energy phase change of each pure component, but this involves only pure-component properties and can be ignored for the present purposes.)

A particularly simple approximation known as “regular-solution theory” was developed by Hildebrand and co-workers [J. H. Hildebrand. *J. Am. Chem. Soc.* **51**, 66–80 (1929)]. The regular-solution model assumes that the excess enthalpy of mixing can be represented as a simple one-parameter correction

$$H_{\text{xs}}^\phi(x) = \Omega^\phi x(1-x) \quad (\text{S7.15-4})$$

while the entropy of mixing is of ideal form

$$S_{\text{xs}}^\phi(x) = 0 \quad (\text{S7.15-5})$$

(In terms of activity coefficients, the regular-solution approximation can also be expressed as $RT \ln \gamma_A^\phi = \Omega^\phi x_B^2$, $RT \ln \gamma_B^\phi = \Omega^\phi x_A^2$.) The mixing parameter Ω^ϕ expresses nonideality corrections in the usual way: positive deviations ($\Omega^\phi > 0$) correspond to repulsive A–B interactions (favoring phase separation), whereas negative deviations ($\Omega^\phi < 0$) correspond to attractive A–B interactions (favoring solubility). Regular-solution theory is useful for nonpolar components, but tends to fail for more specific solute–solvent interactions. For polymer solutions, a significant improvement is obtained in the related “Flory–Huggins theory” [P. J. Flory. *J. Chem. Phys.* **10**, 51–61 (1942); M. L. Huggins. *Ann. NY Acad.*

Sci. **43**, 1–32 (1942)] by taking account of leading chain-conformational contributions to $S_{xs}^\phi(x)$.

A more accurate type of “quasi-chemical solution model” was introduced by Guggenheim [E. A. Guggenheim, *Mixtures* (Oxford University Press, New York, 1952)] to account for specific $A + B \leftrightarrow AB$ association corrections in both $H_{xs}^\phi(x)$ and $S_{xs}^\phi(x)$. The quasi-chemical approach employs an explicit pair partition function for the equilibrium population of $A:B$ complexes in solution. More general “associated solution models” were also developed to incorporate AB complexes of other than 1 : 1 stoichiometry [A. D. Pelton and M. Blander, *Metall. Trans.* **17B**, 805–15 (1986)]. Although such quasi-chemical and solution models can be more accurate, they are also more difficult to implement.

A still more flexible approach is to expand $H_{xs}^\phi(x)$ and $S_{xs}^\phi(x)$ as general power series “Margules expansions,” with empirical coefficients for each term. Numerous models of this form have been suggested [see, e.g., J. B. J. Thompson, in P. H. Abelson (ed.), *Researches in Geochemistry*, Vol. 2 (Wiley, New York, 1967), pp. 340–61], but the difficulty of fitting the coefficients, as well as assessing their uniqueness and physical significance, becomes increasingly problematic as expansion length increases.

SIDEBAR 7.16: SHAPE MEMORY ALLOYS

Shape memory alloys (SMAs) constitute a novel class of metal alloys (e.g., Ni/Ti, Cu/Zn/Al, Cu/Al/Ni) with the remarkable ability to undergo spontaneous shape change to an earlier form upon being warmed through a solid–solid transition temperature. For example, a Ni/Ti (“nitinol”) wire of distinctive shape might be bent, twisted and crumpled into unrecognizable form at a lower temperature, but will spontaneously “remember” and spring back to its original shape if warmed to the original temperature (see the demonstration video at <http://mrsec.wisc.edu/Edetc/cineplex/NiTi/index.html>).

Near the transition temperature, SMAs also exhibit the curious effect of “pseudo-elasticity,” in which the metal recovers (apparently in the usual manner) from an isothermal bending deformation when the stress is removed. However, the elasticity is not due to the usual elastic modulus of a fixed crystalline form, but instead results from strain-induced solid–solid phase transition to a more deformable crystalline structure, which yields to the stress, then spontaneously returns to the original equilibrium crystal structure (restoring the original macroscopic shape) when the stress is removed.

The SMA effect can be traced to properties of two crystalline phases, called “martensite” and “austenite,” that undergo facile solid–solid phase transition at temperature T_m (dependent on P and x). The low-temperature martensite form is of body-centered cubic crystalline symmetry, soft and easily deformable, whereas the high-temperature austenite form is of face-centered cubic symmetry, hard and immalleable. Despite their dissimilar mechanical properties, the two crystalline forms are of nearly equal density, so that passage from austenite to a “twinned” form of martensite occurs without perceptible change of shape or size in the macroscopic object.

If the SMA is sufficiently close to T_m , an imposed stress is sufficient to cause pressure-induced austenite $\rightarrow$ martensite phase transitions in selected grains of the alloy, relieving the stress through pseudo-elastic deformation of the softer martensite grains. Similarly, if the original austenite-shaped alloy is brought below T_m to convert it to malleable martensite form, many deformations of macroscopic shape leave the martensitic atoms “close” to their

original austenitic positions, ready to “snap back” into original structure and shape when heated above T_m . However, under sufficient distortions, some fraction of the martensitic atoms may fall into new positions in the restored austenitic lattice, leading to “fatigue” and eventual failure to restore the original shape after repeated cycles.

The occurrence of irreversible fatigue phenomena and grain-size effects indicates that important features of the SMA phenomenon lie outside the domain of equilibrium thermodynamics. Nevertheless, details of the SMA T - x (and T - P - x) phase diagram are clearly important for the understanding and engineering of this curious thermal effect.

7.5 TERNARY AND HIGHER SYSTEMS

Ternary A/B/C systems ($c = 3$) present further challenges to thermodynamic description. According to the phase rule, the number of degrees of freedom

$$f = c - p + 2 = 5 - p$$

ranges from $f = 4$ (for $p = 1$) down to $f = 0$ (for $p = 5$). The homogeneous limit requires up to *four*-dimensional graphical representations, and the heterogeneous limit requires consideration of up to *five* coexisting phases (“quintuple point”).

Because phase diagrams on the printed page or blackboard are generally limited to two-dimensional plots, we may initially fix T and P , leaving only two of the three composition variables x_A , x_B , x_C as independent variables. This might suggest plotting phase diagrams as 2D cartesian plots of, for example, x_A versus x_B (or any other pair of concentration variables). However, such a choice is both unsymmetrical and inconvenient for chemical purposes.

Gibbs (1876) and Roozeboom [H. W. B. Roozeboom. *Z. Phys. Chem.* **15**, 145 (1894)] introduced a clever graphical device for representing ternary concentration variables by an equilateral triangle. Figure 7.16 depicts the form of the Gibbs–Roozeboom triangle, with each vertex labeled by chemical component A, B, or C. As shown in Fig. 7.16a, the scale of x_A can be pictured as running from the vertex labeled by A (where $x_A = 1$) down the perpendicular to the opposite edge (where $x_A = 0$), and similar x_B , x_C scales can be pictured as shown in Fig. 7.16b, c. In each case, the grid of lines running parallel to an edge marks the successive increments of concentration toward the labeled vertex.

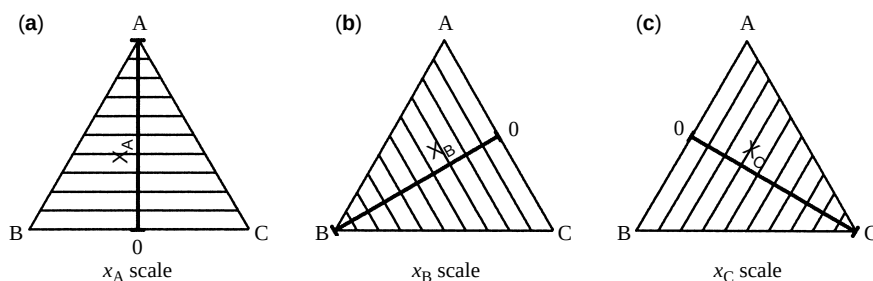
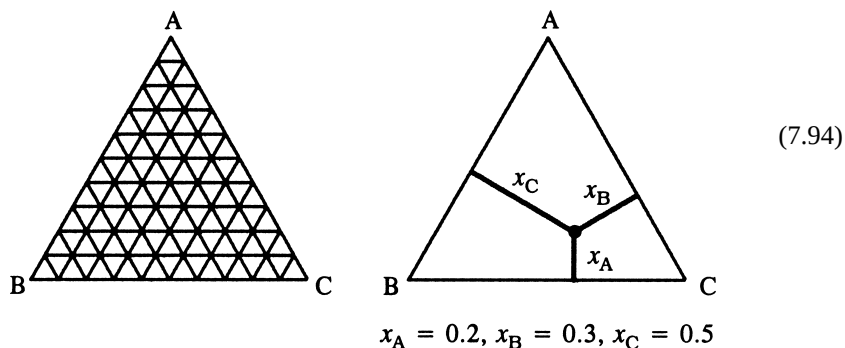


Figure 7.16 Gibbs–Roozeboom grid-lines (at 0.1 increments) for mole fractions of (a) component A, (b) component B, and (c) component C. In each case, the heavy line runs from $x = 0$ (“0”) to $x = 1$ (at the labeled vertex).

A composite triangle that includes all three grid-scales would be rather confusing, as shown at the left in (7.94) below, but one can envision the presence of these grid-lines if they are not drawn explicitly. For any chosen point in the interior, we can draw the perpendiculars to the three edges to obtain the three line segments proportional to the x_A , x_B , x_C values, as shown in the diagram at the right:

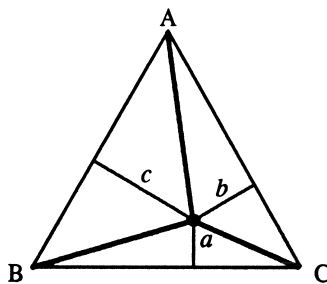


The marked values in (7.94) illustrate how every interior point of the Gibbs–Roozeboom triangle can be uniquely associated with the composition variables x_A , x_B , x_C of a given ternary A/B/C system (in this case, for $x_A = 0.2$, $x_B = 0.3$, $x_C = 0.5$).

Why does the Gibbs–Roozeboom triangle “work”? The answer becomes apparent from the following elementary geometrical theorem:

Theorem: For any interior point of an equilateral triangle, the sum of perpendiculars to the three edges is a constant. (7.95)

The theorem can be proved as follows. Consider an equilateral triangle of side L (and area $3^{1/2}L^2/4$):



For a given point in the interior, draw the perpendicular line segments a , b , c to the three edges. Also, draw the (heavy) lines to the apices that divide the triangle into three smaller triangles t_a , t_b , t_c , as shown. The formula for the area of a triangle (one-half the base times the height) allows us to write

$$\text{area of } t_a = \frac{1}{2}aL \quad (7.96a)$$

$$\text{area of } t_b = \frac{1}{2}bL \quad (7.96b)$$

$$\text{area of } t_c = \frac{1}{2}cL \quad (7.96c)$$

and the total area is therefore

$$\text{total area} = \frac{1}{2}L(a + b + c) = 3^{1/2}L^{2/4} \quad (7.97a)$$

from which we conclude

$$a + b + c = \frac{3^{1/2}L}{2} = \text{constant} \quad (7.97b)$$

which is the desired result—*QED*.

The Gibbs–Roozeboom triangle may be divided as usual into single-phase homogeneous regions, two-phase coexistence regions, and so forth. The two-phase coexistence “holes” are spanned as usual with tie-lines, but in general these are no longer horizontal (as in binary diagrams), but instead must be drawn in explicitly to show the connecting phases. Other features may also appear “skewed” in the ternary diagram, and terms such as “maximum” or “minimum” no longer have their usual significance in the triangular axis system. Despite these slightly disorienting characteristics, the ternary phase diagram is interpreted in a manner that parallels previous experience with binary phase diagrams.

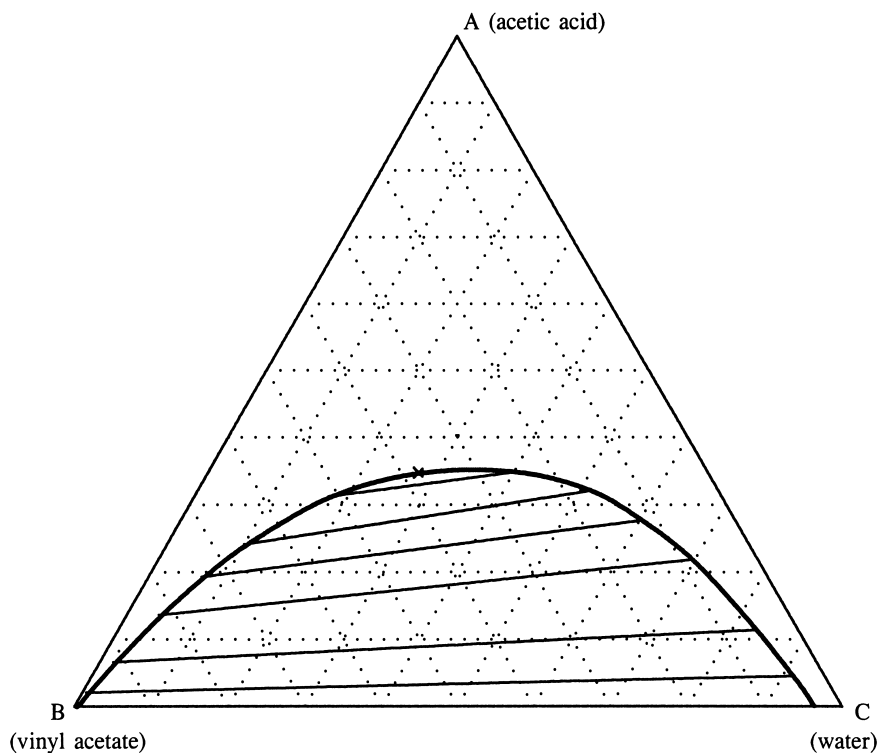


Figure 7.17 Ternary A/B/C phase diagram (at 25°C, 1 atm) for A = acetic acid, B = vinyl acetate, C = water, showing nonhorizontal tie-lines in the immiscible two-phase region (organic liquid + aqueous liquid), culminating at a plait point (×). Concentration grid values (dotted lines) are in wt% at 10% intervals.

In particular, the lever rule is used in the usual way to determine the relative phase quantities along the marked tie-lines.

A simple example of a ternary phase diagram is shown in Fig. 7.17 for the acetic acid/vinyl acetate/water system. The phase diagram is divided into the totally miscible (upper) region and the immiscible two-phase region, the latter being spanned by the nonparallel tie-lines that are drawn to connect coexisting concentrations. The coexistence tie-lines constrict to zero length at the limiting “plait point” ($\times$). (A seam of plait points may connect to a consolute point when a new axis is added to display the temperature dependence of the composition triangle.)

As shown in Fig. 7.17, the limiting plait point appears at a concentration of approximately 35 wt% acetic acid, 38 wt% vinyl acetate, and 27 wt% water, which can be read from the dotted grid of concentration values shown in the background [using the “trick” shown in (7.94)]. Sidebar 7.17 presents some simple problems that provide experience in reading and interpreting the ternary phase diagram.

SIDEBAR 7.17: TERNARY DIAGRAM PROBLEMS FOR IMMISCIBLE LIQUIDS

The following problems all refer to the acetic acid (AA)/vinyl acetate (VA)/water (W) system of Fig. 7.17. Answer each question to “eyeball accuracy.”

Problem 10 g AA, 20 g VA, and 30 g W are mixed together at 25°C, 1 atm. How many phases form, and what is their composition?

Solution The given weights correspond (in wt%) to about 17% AA, 33% VA, and 50% W. This point lies within the hatched two-phase region, slightly below the third tie-line from the bottom. If we interpolate a new tie-line that passes through this point, we find that the organic phase is approximately 85% VA, 12% AA, and 3% W, and the aqueous phase is approximately 73% W, 19% AA, and 8% VA.

Problem A 50:50 (w/w) mix of vinyl acetate and acetic acid is progressively diluted with water at 25°C, 1 atm.

- (a) At what wt% water will the solution spontaneously phase-separate into two distinct phases?
- (b) When the total wt% of water reaches 60%, what is the composition of the two immiscible phases? What percentage of the total solution is present in each phase?
- (c) If still more water is added, at what wt% does the solution again become a single homogeneous phase?

Solution

- (a) Since the solution always contains equal weights of AA, VA, the dilution path follows a line from the bisector of the AB edge to the C (water) vertex. Along this dilution line, the system enters the two-phase region (just beyond the plait point) when the solution is 30% water (and 35% each AA and VA).

- (b) At 60% water, the dilution line is intersecting the third tie-line from the bottom. From the ends of this tie-line, we can see that the organic phase contains about 82% VA, 14% AA, and 4% W, and the aqueous phase contains about 69% W, 22% AA, and 9% VA. From the lever rule and tie-line segments, we can also determine that about 80% of the solution is in the aqueous phase.
- (c) The dilution line re-emerges into the single-phase region near the end of the lowest tie-line, when the solution is about 92% water (and 4% each AA and VA).

Let us briefly examine some elementary aspects of solubility in ternary systems as an illustration of characteristic similarities and differences with respect to corresponding binary systems. Figure 7.18 displays a ternary solubility diagram for two solids (α and β) and a liquid (C), corresponding to the case of an aqueous solution of two partially soluble salts that share a common ion. The diagram shows the homogeneous liquid region ($Cacb$) of complete solubility, the two two-phase regions (αac , βbc) of supernatant liquid plus solid precipitate α or β , and the triangular three-phase region ($c\alpha\beta$) of doubly saturated liquid c and α and β precipitates.

One might suppose that the solubility properties of the ternary system are simply related to those of the two associated binary systems, but Fig. 7.18 shows that this is not so. Point a

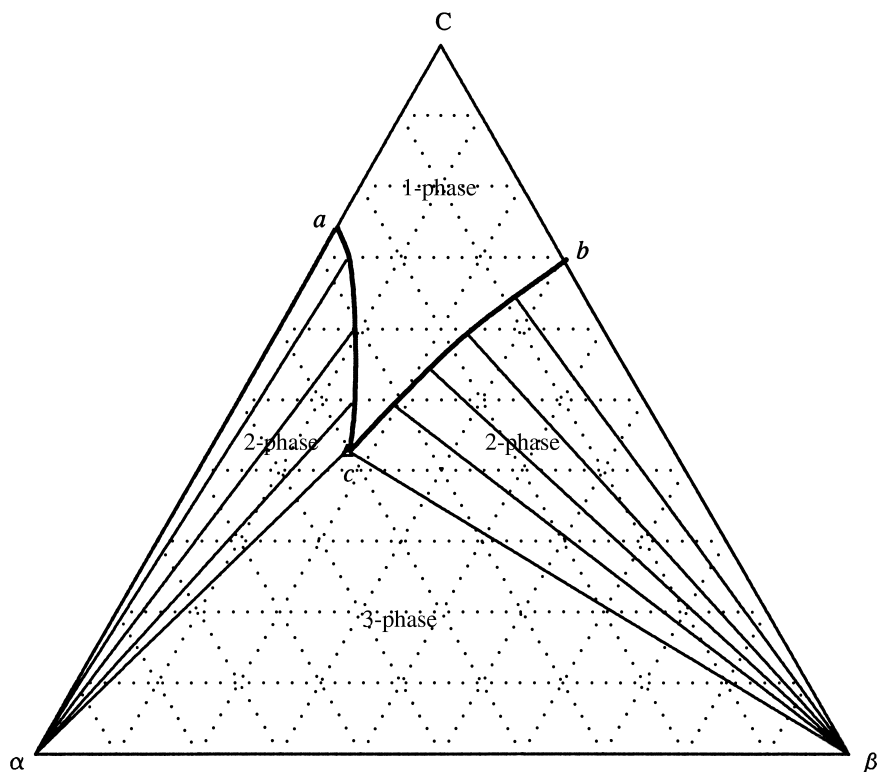


Figure 7.18 Ternary diagram for two partially soluble salts (α , β , sharing a common ion) and solvent (C), illustrating the “common-ion” effect of one salt influencing the solubility of the other. (The plotted behavior corresponds to the $\text{NH}_4\text{Cl}/(\text{NH}_4)_2\text{SO}_4/\text{H}_2\text{O}$ system.)

marks the solubility limit of the binary α/C system and b that of the α/C system, and the two solubility curves (heavy lines) show how these solubility limits vary in the ternary system. Whereas the binary solubility limit of solid α is about 26% (point a), the corresponding solubility in the ternary system rises to 40% (point c). For the β solute, conversely, the solubility in pure C (30%, point b) is reduced (down to 17% at point c) by the presence of dissolved α . Sidebar 7.18 presents additional problems to illustrate the interpretation of the solubility diagram.

SIDEBAR 7.18: TERNARY DIAGRAM PROBLEMS FOR TWO PARTIALLY SOLUBLE SOLIDS

The following problems refer to the $\alpha/\beta/C$ diagram in Fig. 7.18.

Problem An initial unsaturated solution of 10% α , 10% β , and 80% C is allowed to slowly evaporate under isothermal conditions. Describe the sequence of precipitation events that will be observed as %C is reduced to zero.

Solution The evaporation path follows the vertical bisector of the diagram, the line of equal α and β . From the initial solution, the evaporation line enters the two-phase β -C region at 56% C, where solid β begins to precipitate out of solution. As further C evaporates, pure β continues to precipitate (and the liquid composition moves down the solubility curve toward point c) as the evaporation line traverses the β -C coexistence region. When the overall concentration reaches 35% C (the boundary of the three-phase region), solid α begins to precipitate out of solution, and the remaining saturated liquid is of composition c (42% C). Thereafter, both α and β precipitate out of solution until the evaporation path reaches the base of the triangle, where the last drop of liquid disappears. For any point within the three-phase triangle, the relative amounts of α , β , and C must “balance the triangle”; that is, if z is a chosen point in the triangle, with vector line segments $\vec{L}_{z\alpha}$, $\vec{L}_{z\beta}$, $\vec{L}_{zc}$ to the three apices, then the relative amounts (%- α , %- β , %-C) of the three phases must satisfy the vector balance condition

$$(\%\alpha)\vec{L}_{z\alpha} + (\%\beta)\vec{L}_{z\beta} + (\%c)\vec{L}_{zc} = 0$$

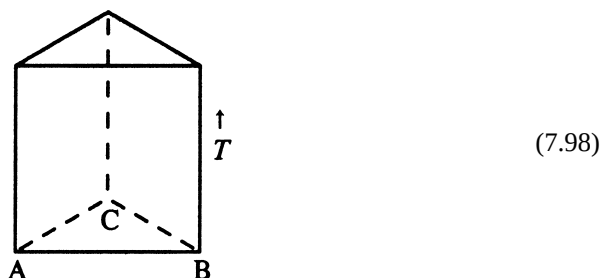
This is just the three-phase (triangle) analog of the two-phase (line) “lever rule” balance condition of Section 7.3.2.

Problem Consider a system originally prepared as 60% C and 40% α , which results in saturated liquid (a) and excess precipitate α . Describe what happens as pure β is added to this system.

Solution The β -addition line passes from the initial point (at 60% C on the left edge of the triangle) toward the β vertex (roughly along the extension of the second tie-line of the β -C region). Initially, the added β naturally dissolves, but, surprisingly, excess α *also* begins to dissolve, until, at about 13% β , *both* salts have completely dissolved and the system is a homogeneous solution. That is to say, by adding salt β , we enabled the liquid to dissolve a greater quantity of salt α ! The homogeneous liquid continues to dissolve added β up to about 19% β (the intersection of the β -addition line with the bc boundary

line), whereupon the liquid is saturated in β , and any added β merely precipitates out of solution. The technique of using a common-ion solute β to modulate solubility of α [e.g., adding $(\text{NH}_4)_2\text{SO}_4$ to alter NH_4Cl solubility, with NH_4^+ as the common ion] is called the “common-ion effect.”

The temperature dependence of ternary phase behavior can be described by affixing a new axis perpendicular to the composition triangle to obtain a three-dimensional triangular prism:



This can be viewed as a triptych of the three possible binary T - x diagrams on the facets, with ternary interactions represented by the prism interior. Many aspects of modern alloying technology can be described in terms of such 3D graphical representations, but these and other aspects of ternary behavior lie beyond the scope of the present treatment.

For quaternary ($c = 4$) and higher systems, the difficulties of graphical representation are further compounded. For $c = 4$ (with $f = 6 - p$), depiction of the composition dependence for fixed T , P necessitates 3D graphics. For a four-component A/B/C/D system, this dependence is most conveniently and symmetrically represented by a regular *tetrahedron* labeled by a pure component at each apex,



which generalizes the three-component Gibbs–Roozeboom triangle. For any interior point of the tetrahedron, the perpendicular distance to the face opposite vertex A gives x_A , and similarly for x_B , x_C , x_D . Each facet of the quaternary tetrahedron displays the Gibbs–Roozeboom composition triangle for one of the four possible ternary subsystems, with the full quaternary dependence contained in the interior. The student should consult advanced treatises for further discussion of complex multicomponent phase diagrams.

While classical phase diagrams provide a powerful methodology for grasping the thermodynamic behavior of few-component systems, it is evident that the restricted 2D or 3D realm of human graphical intuition cannot adequately cope with the complexities of many-component systems. Hence, it is important to find generalized analytical techniques that can accurately represent many-component phase behavior for arbitrary values of c . Such techniques will be considered in the metric geometric representation of multicomponent phenomena (Chapter 12).

Thermodynamics of Chemical Reaction Equilibria

The thermodynamics of chemical reaction equilibria underlies much of the freshman chemistry curriculum, and indeed much of early alchemical and chemical practice. The beginning student gains facility in applying such concepts as “mass action,” “equilibrium constant,” and “Le Chatelier’s principle” long before their thermodynamic foundations (and inherent limitations) are fully explained. As we show in this chapter, the Gibbsian formulation of equilibrium thermodynamics permits this general foundation of chemical and electrochemical practice to be established in a rigorous and elegant manner.

8.1 ANALYTICAL FORMULATION OF CHEMICAL REACTIONS IN TERMS OF THE ADVANCEMENT COORDINATE

Let us begin as usual by expressing the symbolic chemical reaction



in the more general form [cf. (3.92)–(3.94)]

$$0 = \sum_i \nu_i A_i \quad (8.2)$$

where A_i denotes the i th chemical species and ν_i is the associated stoichiometric coefficient (taken negative for a reactant species or positive for a product species) in the balanced chemical equation. If we express the extensive property G in terms of its partial molar contributions [cf. (6.22c)], we can write the Gibbs free energy at any “point” of the reaction as

$$G = \sum_i n_i \mu_i \quad (8.3)$$

where n_i denotes the number of moles and μ_i the chemical potential of species A_i at the given point of advancement of the reaction.

To characterize the “point of advancement of the reaction” in more precise fashion, we introduce an “advancement” or “progress” coordinate ξ (also called the “extent of

reaction”) that varies from $\xi = 0$ for initial reactants to $\xi = 1$ for final products. If we designate the initial amount of A_i as n_i° at $\xi = 0$, we can write the amount $n_i = n_i(\xi)$ present at each point ξ as

$$n_i = n_i^\circ + \nu_i \xi \quad (8.4)$$

From this equation, we see that the advancement rate of each n_i is determined by the stoichiometric coefficient ν_i :

$$\frac{dn_i}{d\xi} = \nu_i \quad (8.5)$$

Note that (8.4) leads properly to the initial and final mole numbers at both limits of reaction:

$$n_i \rightarrow n_i^\circ \quad \text{as } \xi \rightarrow 0 \quad (8.6a)$$

$$n_i \rightarrow n_i^\circ + \nu_i \quad \text{as } \xi \rightarrow 1 \quad (8.6b)$$

From (8.4), the explicit expression for ξ can be written as

$$\xi = \frac{n_i - n_i^\circ}{\nu_i}, \quad \text{all } i \quad (8.7)$$

or, in differential form,

$$d\xi = \frac{dn_1}{\nu_1} = \frac{dn_2}{\nu_2} = \dots \quad (8.8)$$

showing (as expected) that the progress of reaction is the same for any chosen species, whether reactant or product.

8.2 CRITERION OF CHEMICAL EQUILIBRIUM: THE EQUILIBRIUM CONSTANT

With the notation of (8.4), (8.3) becomes

$$G(\xi) = \sum_i (n_i^\circ + \nu_i \xi) \mu_i \quad (8.9)$$

As usual, our goal is to find the *minimum* of $G(\xi)$ in order to determine the equilibrium position ($\xi = \xi_{\text{cq}}$) of the chemical reaction at constant T and P . From (6.10c) [cf. the Gibbs–Duhem equation (6.36a)], the differential dG under these conditions is simply

$$dG = \sum_i \mu_i dn_i \quad (8.10)$$

To find the minimum of $G(\xi)$, we must locate the stationary point $(\partial G/\partial \xi)_{T,P} = 0$. For this purpose we employ (8.10), with help from (8.5):

$$0 = \left(\frac{\partial G}{\partial \xi} \right)_{T,P} = \sum_i \mu_i \left(\frac{\partial n_i}{\partial \xi} \right)_{T,P} = \sum_i \mu_i \nu_i \quad (8.11)$$

The equilibrium criterion can therefore be stated as

$$\boxed{\sum_i \mu_i \nu_i = 0 \quad \text{at equilibrium}} \quad (8.12)$$

As usual, the first-derivative stationary criterion (8.12) must be augmented by a second-derivative criterion (*stability condition*)

$$\boxed{\left(\frac{\partial^2 G}{\partial \xi^2} \right)_{T,P} \geq 0} \quad (8.13)$$

to insure that $G(\xi_{\text{eq}})$ is indeed a *minimum*. The inequality (8.13) forms the basis of *Le Chatelier's principle*, to be discussed in Section 8.6. For the present purposes, the criterion (8.12) is the desired analytical characterization of chemical equilibrium.

To express (8.12) in more recognizable form, we first rewrite each chemical potential μ_i in terms of the associated activity a_i [cf. (7.78)],

$$\mu_i = \mu_i^\circ + RT \ln a_i \quad (8.14)$$

to obtain

$$0 = \sum_i \mu_i^\circ \nu_i + RT \sum_i \nu_i \ln a_i \quad (8.15)$$

The first term is a collection of *constants* that we can identify as the overall standard free energy change ΔG° ,

$$\sum_i \mu_i^\circ \nu_i = G_{\text{products}}^\circ - G_{\text{reactants}}^\circ \equiv \Delta G^\circ = \Delta G_{\text{rxn}}^\circ \quad (8.16)$$

which depends only on the standard-state Gibbs free energies of *unmixed* species. The second term can be rewritten with familiar logarithmic arithmetic as

$$RT \sum_i \nu_i \ln a_i = RT \sum_i \ln a_i^{\nu_i} = RT \ln \left[\prod_i a_i^{\nu_i} \right] \quad (8.17)$$

where “ $\prod$ ” denotes “product of.” Equation (8.15) then becomes

$$\Delta G^\circ + RT \ln \left[\prod_i a_i^{\nu_i} \right] = 0 \quad \text{at } \xi_{\text{eq}} \quad (8.18)$$

The quantity in brackets can be identified as the *reactant quotient* Q , the ratio of product and reactant activities, each raised to its stoichiometric power (as illustrated, for example,

for the symbolic reaction $aA + bB \rightarrow cC + dD$):

$$\mathcal{Q} \equiv \prod_i a_i^{v_i} = \frac{\text{products}}{\text{reactants}} = \frac{a_C^c a_D^d}{a_A^a a_B^b} \quad (8.19)$$

Under the equilibrium condition (8.18), where the remaining terms ΔG° , R , and T are all constants, the reaction quotient $\mathcal{Q}$ must itself become a constant, which we identify as the “equilibrium constant” K_{eq} :

$$\mathcal{Q} \rightarrow K_{\text{eq}} \equiv \prod_i a_i^{v_i} \quad \text{at } a_i = a_i(\xi_{\text{eq}}) \quad (8.20)$$

This allows us to finally rewrite (8.18) as

$$\boxed{\Delta G^\circ = -RT \ln K_{\text{eq}}} \quad (8.21)$$

the central equation of chemical equilibrium theory.

It should be emphasized that K_{eq} must be expressed in terms of *activities* (8.20) for full thermodynamic rigor. Nevertheless, for qualitative purposes, it is often possible to assume dilute near-ideal solution behavior in which activity a_i becomes approximately proportional to *molarity* [near-constant activity coefficient; cf. (7.80b)],

$$a_i \propto c_i = [A_i] = n_i/V \quad (8.22a)$$

for dilute near-ideal solutions. In this case, the proportionality constants can be combined with K_{eq} to give a modified equilibrium constant K_c ; for the $aA + bB \rightarrow cC + dD$ reaction, for example, this gives the familiar Guldberg–Waage “mass action law” of freshman chemistry:

$$K_c = \frac{[C]^c [D]^d}{[A]^a [B]^b} \quad (8.22b)$$

A more accurate approximation of this type also applies to gaseous mixtures, where activity a_i is proportional to partial pressure P_i [cf. (7.88)]:

$$a_i \propto P_i, \quad \text{for gas mixtures} \quad (8.23a)$$

leading to the modified equilibrium constant K_P :

$$K_P = \frac{P_C^c P_D^d}{P_A^a P_B^b} \quad (8.23b)$$

Still other conventional equilibrium constants, such as the “solubility product” K_{sp} for electrolytes and the “ion product” $K_w = [\text{H}^+(aq)][\text{OH}^-(aq)]$ for water, can be defined by absorbing near-constant activities (such as those of pure solids or liquids) into the constant K . Although such modified K_{eq} -type expressions may have practical utility, their

approximate nature should be clearly recognized (particularly for K_{sp}). Here, we shall continue to understand K_{eq} in the sense of (8.20), conferring full thermodynamic authority on its validity.

8.3 GENERAL FREE ENERGY CHANGES: DE DONDER'S AFFINITY

A more general picture of the free energy “driving forces” accompanying chemical reaction can be given in terms of de Donder’s “affinity” $\mathcal{A}$, defined as

$$\mathcal{A} \equiv \left(\frac{\partial G}{\partial \xi} \right)_{T,P} = \mathcal{A}(\xi) \quad (8.24)$$

Although not a literal “force” in the mechanical sense (cf. Sidebar 2.8), affinity plays an analogous role in “driving” the reaction toward equilibrium, where $\mathcal{A} = 0$ [cf. (8.11)]. For general nonequilibrium $\xi \neq \xi_{eq}$, we can recognize that

$$\mathcal{A}(\xi) = \sum_i \mu_i \nu_i = \Delta G = G_{\text{products}}(\xi) - G_{\text{reactants}}(\xi) \quad (8.25)$$

From (8.15)–(8.21), we therefore obtain, for general ξ ,

$$\Delta G = \Delta G^\circ + RT \ln \mathcal{Q} \quad (8.26)$$

The approach to equilibrium is therefore characterized by the following (equivalent) conditions:

$$\xi \longrightarrow \xi_{eq} \quad (8.27a)$$

$$\mathcal{A} \longrightarrow 0 \quad (8.27b)$$

$$\Delta G \longrightarrow 0 \quad (8.27c)$$

$$\mathcal{Q} \longrightarrow K_{eq} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) \quad (8.27d)$$

The expression for K_{eq} in (8.27d) shows that if $\Delta G^\circ < 0$ (“exoergic”), then $K_{eq} > 1$, and the reaction lies toward the *product* side at equilibrium. Conversely, if $\Delta G^\circ > 0$ (“endoergic”), then $K_{eq} < 1$, and equilibrium lies toward the *reactant* side.

Figure 8.1a illustrates the free energy relationships (8.27a–d) in a graphical plot of $G(\xi)$. As shown in the diagram, a “rolling ball” analogy aptly suggests the thermodynamic tendency toward minimum Gibbs free energy. The thermodynamic driving force is toward the right (product side) if $\mathcal{A} < 0$ or the left (reactant side) if $\mathcal{A} > 0$, with all processes proceeding spontaneously toward the equilibrium “resting state” with $\mathcal{A} = 0$. Figure 8.1b shows the corresponding graph of de Donder’s affinity $\mathcal{A}(\xi)$, the derivative of $G(\xi)$. As shown, the sign of the chemical driving force dictates that the direction of spontaneous reaction is toward larger ξ for $\mathcal{A} < 0$ or smaller ξ for $\mathcal{A} > 0$, i.e., always toward the point ξ_{eq} of zero affinity.

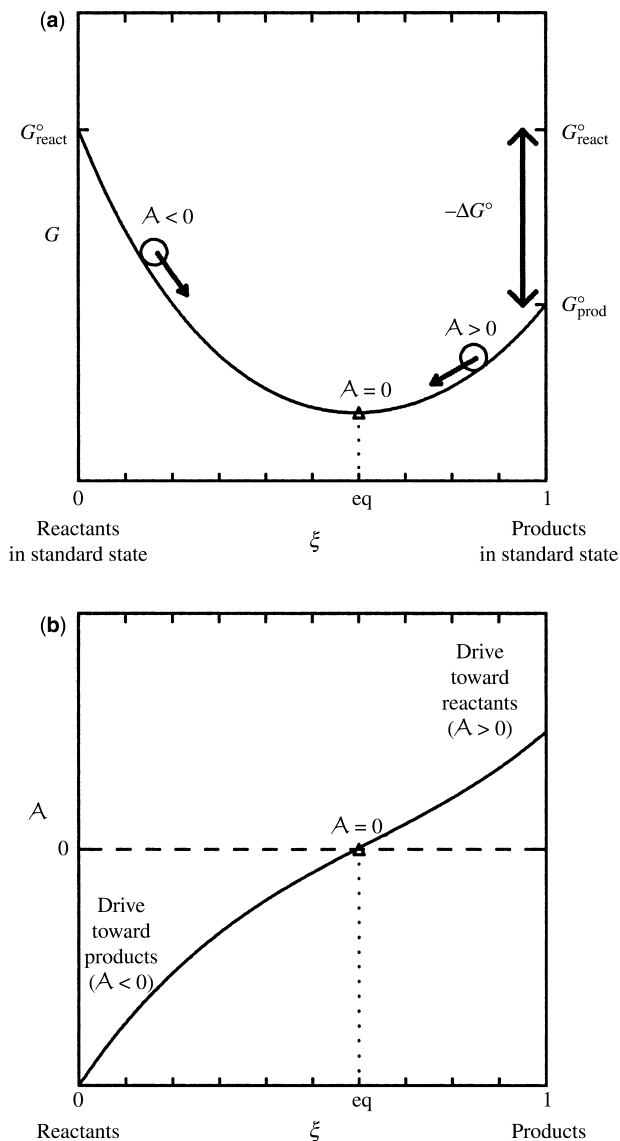


Figure 8.1 Graph of (a) $G(\xi)$, (b) $\mathcal{A}(\xi)$ for an exoergic reaction ($\Delta G^\circ < 0$), showing the equilibrium point of minimum Gibbs free energy (triangle) at ξ_{eq} (dotted line). In (a), the direction of spontaneous reaction (“rolling ball”) is shown both for initial $\xi \simeq 0.15$ (where $\mathcal{A} < 0$) and $\xi \simeq 0.85$ (where $\mathcal{A} > 0$), both tending toward equilibrium at $\xi_{eq} \simeq 0.60$ (where $\mathcal{A} = 0$).

8.4 STANDARD FREE ENERGY OF FORMATION

Equation (8.21) indicates the central importance of the standard Gibbs free energy difference ΔG° of unmixed product and reactant species at standard state conditions. As usual, this quantity can be separated into enthalpic and entropic components:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (8.28)$$

The equilibrium constant K_{eq} , (8.27d), thereby factors into enthalpic and entropic terms:

$$K_{\text{eq}} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) = \exp\left(-\frac{\Delta H^\circ}{RT}\right) \exp\left(\frac{\Delta S^\circ}{RT}\right) \quad (8.29)$$

The enthalpic factor, $\exp(-\Delta H^\circ/RT)$, clearly dictates the T dependence of K_{eq} , whereas the entropic factor, $\exp(\Delta S^\circ/RT)$, is T -independent. The exponential dependences in (8.29) remind us (cf. Sidebar 8.1) that even chemically modest changes in ΔH° , ΔS° , or T can have profound effects on K_{eq} .

SIDEBAR 8.1: EQUILIBRIUM SHIFTS DUE TO HYDROGEN BONDING

Problem Hydrogen bonds are ubiquitous features of aqueous chemistry as well as the recognition and reactive sites of biological enzymes. Estimate the shift of binding equilibrium if an enzymatic binding site is re-engineered to gain one additional specific hydrogen bond (with binding enthalpy about 5 kcal mol^{-1}) to the binding substrate, all else being equal. How much is this shift affected by a 1K change in T ?

Solution An enthalpy increment of 5 kcal mol^{-1} , although modest by chemical standards, is highly significant compared with the ambient RT value of about $0.6 \text{ kcal mol}^{-1}$ near $T \simeq 300\text{K}$. If we denote the new binding constant with a prime, and consider only the net enthalpy change $\Delta\Delta H^\circ = \Delta H'^\circ - \Delta H^\circ \simeq -5 \text{ kcal mol}^{-1}$, we can see from (8.29) that

$$\frac{K'_{\text{eq}}}{K_{\text{eq}}} = \exp\left(-\frac{\Delta\Delta H^\circ}{RT}\right) \simeq \exp\left(\frac{5 \text{ kcal mol}^{-1}}{0.6 \text{ kcal mol}^{-1}}\right) \simeq e^{8.3} \simeq 4000(!)$$

With similar arithmetic, one can find that each 1K temperature change alters ΔK_{eq} by about 3%. The change of binding selectivity is profound and quite T -sensitive, suggesting why careful temperature control is essential to properly functioning biological machinery.

In a similar manner to that employed for thermochemical ΔH° of chemical reactions [cf. (3.106)], the reaction free energy ΔG° can be expressed in terms of the standard *Gibbs free energy of formation* $\Delta G_f^\circ[\text{A}_i]$ for each species A_i , namely,

$$\Delta G^\circ = \sum_i \nu_i \Delta G_f^\circ[\text{A}_i] \quad (8.30)$$

Each ΔG_f° is in turn composed of standard enthalpies ΔH_f° and entropies ΔS_f° of formation,

$$\Delta G_f^\circ[\text{A}_i] = \Delta H_f^\circ[\text{A}_i] - T^\circ \Delta S_f^\circ[\text{A}_i] \quad (8.31)$$

which are tabulated in standard thermochemical databases [e.g., *NIST–JANAF Thermochemical Tables*, 4th edn, Monograph 9 (American Institute of Physics, Melville NY, 1998)]. ΔH_f° values are obtained by standard calorimetric measurements (Section 3.6.5). ΔS_f° values can be *estimated* from heat-capacity measurements combined with “third-law” assumptions (Sidebar 5.19), but rigorous values are obtained from the measured T dependence of K_{eq} or electrochemical potentials, as described in Sections 8.5 and 8.7.

8.5 TEMPERATURE AND PRESSURE DEPENDENCE OF THE EQUILIBRIUM CONSTANT

8.5.1 Temperature Dependence: Van’t Hoff Equation

Let us first examine the temperature dependence of K_{eq} (at constant P), taking as the starting point a rewritten form of (8.29):

$$\ln K_{\text{eq}} = -\Delta G^\circ/RT \quad (8.32)$$

Differentiating with respect to T (at constant P), we obtain

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial T}\right)_P = -\left(\frac{\partial(\Delta G^\circ/RT)}{\partial T}\right)_P \quad (8.33)$$

However, from the general Gibbs–Helmholtz equation (5.56), we can write

$$\left(\frac{\partial(G/T)}{\partial T}\right)_P = -\frac{H}{T^2} \quad (8.34a)$$

or, for the present application,

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial T}\right)_P = \left(\frac{\partial(\Delta G^\circ/T)}{\partial T}\right)_P = -\frac{\Delta H^\circ}{T^2} \quad (8.34b)$$

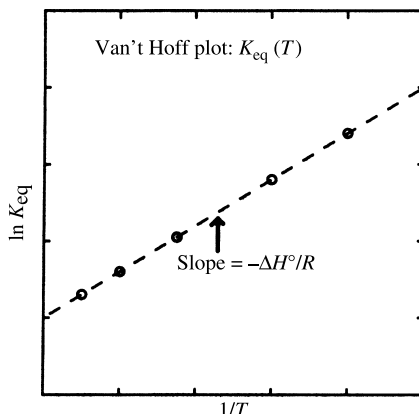
which is one form of the *Van’t Hoff equation*. Analogous to (5.57), we can also rewrite this as

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial(1/T)}\right)_P = -\frac{\Delta H^\circ}{R} \quad (8.34c)$$

As shown in (8.34b, c), the T dependence of K_{eq} is governed solely by the heat of reaction ΔH° under constant- P conditions.

The Van’t Hoff equation (8.34c) can also be expressed in graphical form. If we plot values of $\ln K_{\text{eq}}$ versus $1/T$, (8.34c) establishes that the *slope* of the curve is $-\Delta H^\circ/R$. In many cases, the variation of ΔH° with temperature is slight, so that

the plot of $\ln K_{\text{eq}}$ versus $1/T$ becomes a *straight line* (at least for sufficiently small temperature interval), as depicted below:



In this case, one can also integrate (8.34c) to obtain the form of the Van't Hoff equation most familiar to beginning chemistry students, namely,

$$\ln \left[\frac{K_{\text{eq}}(T_2)}{K_{\text{eq}}(T_1)} \right] = -\frac{\Delta H^\circ}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (8.35)$$

One can also see that the Van't Hoff equation is consistent with the prior conclusion [cf. (8.29), Sidebar 8.1] that ΔH° controls the variations of K_{eq} with temperature.

From (8.34c) or (8.35), it is easy to see that if the chosen reaction is *endothermic* ($\Delta H^\circ > 0$), then a T increase tends to promote *product* formation (the reaction “shifts right”). Conversely, if the reaction is *exothermic* ($\Delta H^\circ < 0$), a temperature increase promotes formation of *reactants* (the equilibrium “shifts left”). Such conclusions appear “intuitive” from the perspective of Le Chatelier’s principle, and indeed we shall show in Section 8.6 that such Le Chatelier-like conclusions arise from deep theoretical roots that permeate the Van't Hoff equation and many other thermodynamic relationships.

8.5.2 Pressure Dependence

For the pressure dependence of K_{eq} (at constant T), we again differentiate (8.32) to obtain

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial P} \right)_T = -\frac{1}{RT} \left(\frac{\partial \Delta G^\circ}{\partial P} \right)_T = -\frac{1}{RT} \Delta V^\circ \quad (8.36)$$

where we employed the standard identity (5.52) to rewrite $(\partial \Delta G^\circ / \partial P)_T = \Delta V^\circ$ in terms of

$$\Delta V^\circ = V_{\text{products}}^\circ - V_{\text{reactants}}^\circ \quad (8.37)$$

the overall volume change of reaction at standard state. This volume change is usually negligible for reactions involving only liquids or solids,

$$\Delta V^\circ \simeq 0 \quad (\text{liquids, solids}) \quad (8.38a)$$

whereas for reactions involving changing number of moles of gaseous species (Δn_{gas}), we can estimate ΔV° from the ideal gas approximation as

$$\Delta V^\circ \simeq \frac{RT}{P} \Delta n_{\text{gas}} \quad (8.38b)$$

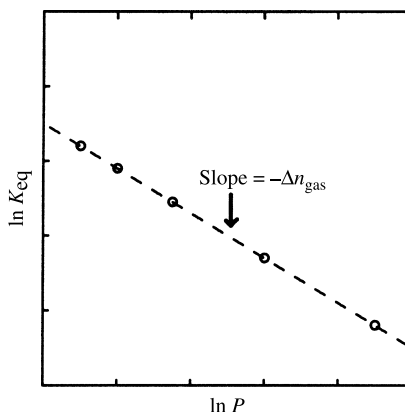
Under these approximations, (8.36) becomes

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial P} \right)_T \simeq -\frac{\Delta n_{\text{gas}}}{P} \quad (8.39a)$$

or, equivalently,

$$\left(\frac{\partial \ln K_{\text{eq}}}{\partial \ln P} \right)_T \simeq -\Delta n_{\text{gas}} \quad (8.39b)$$

Graphically, (8.39b) tells us that a log–log plot of $\ln K_{\text{eq}}$ versus $\ln P$ should lead to straight-line behavior with slope given by (the negative of) the reaction volume change ΔV° :



The shift of K_{eq} with P change can therefore be predicted from the sign of ΔV° : if reaction volume *increases* ($\Delta n_{\text{gas}} > 0$), then a P increase shifts K_{eq} toward the *reactant* side, whereas if reaction volume *decreases* ($\Delta n_{\text{gas}} < 0$), a P increase will promote *product* formation. These inferences are consistent with the expectations of Le Chatelier's principle.

8.6 LE CHATELIER'S PRINCIPLE

In 1884, French chemist Henri Louis Le Chatelier enunciated a qualitative principle of great insight and generality pertaining to the “responses” of chemical equilibrium when subjected

to the “stresses” of changed conditions. As remarked by Pauling [L. Pauling. *College Chemistry*, 3rd edn (W. H. Freeman, San Francisco, 1964), pp. 437–8]:

It is fortunate that there is a general qualitative principle, called Le Chatelier's principle, that relates to all the applications of the principles of chemical equilibrium . . . Some years after you have finished your college work, you may (unless you become a chemist or work in some closely related field) have forgotten all the mathematical equations relating to chemical equilibrium. I hope, however, that you will not have forgotten Le Chatelier's principle.

Le Chatelier's principle has been stated in many forms, some excessively vague or anthropomorphic and subject to misuse. For our present purposes, we adopt the following statement:

Le Chatelier's principle: When a variable affecting the position of equilibrium is changed, the system re-equilibrates in the direction tending to *counteract* the change.

The underlying idea is the “restorative tendency” of equilibrium, tending to counteract the effects of attempted changes on an original equilibrium system. This restorative tendency is associated with the *stability* of chemical equilibrium, and we therefore use the rigorous stability condition (8.13) to prove the above statement of Le Chatelier's principle in a general form.

To express Le Chatelier's principle mathematically, we consider a general “variable” Y and “position of equilibrium” ξ . The direction of equilibrium shift $d\xi$ in response to an attempted change of the variable dY is given by the sign of $d\xi/dY$. To determine the sign of this derivative, let us first recall that affinity $\mathcal{A} = \mathcal{A}(Y, \xi)$ is *invariant* ($d\mathcal{A} = 0$) under changes of Y and ξ for which the system *remains* at equilibrium, namely,

$$0 = d\mathcal{A} = \left(\frac{\partial \mathcal{A}}{\partial Y}\right)_{\xi} dY + \left(\frac{\partial \mathcal{A}}{\partial \xi}\right)_Y d\xi \quad (8.40)$$

Under this constraint, the ratio $d\xi/dY$ is evaluated as

$$\frac{d\xi}{dY} = - \frac{(\partial \mathcal{A} / \partial Y)_{\xi}}{(\partial \mathcal{A} / \partial \xi)_Y} \quad (8.41)$$

[as might also be directly inferred from the Jacobi identity (1.14b) for $x = \xi$, $y = Y$, $z = \mathcal{A}$]. However, from the definition (8.24) of affinity and the general stability condition (8.13) we know that

$$\left(\frac{\partial \mathcal{A}}{\partial \xi}\right)_Y = \left(\frac{\partial^2 G}{\partial \xi^2}\right)_Y > 0 \quad (8.42)$$

We therefore see from (8.41), (8.42) that $d\xi/dY$ and $(\partial \mathcal{A} / \partial Y)_{\xi}$ must have *opposite* signs,

$$\frac{d\xi}{dY} \left(\frac{\partial \mathcal{A}}{\partial Y}\right)_{\xi} < 0 \quad (8.43)$$

which is the essential content of Le Chatelier's principle.

To see how (8.43) corresponds to the usual statement of Le Chatelier's principle, let us consider the specific case of a *pressure* change ($Y = P$) under isothermal conditions.

For this choice, $(\partial G/\partial Y)_\xi$ becomes

$$\left(\frac{\partial G}{\partial Y}\right)_\xi = \left(\frac{\partial G}{\partial P}\right)_{T,\xi} = \left(\frac{\partial G}{\partial P}\right)_{T,\underline{n}} = V \quad (8.44)$$

the system volume. Therefore, we can use the definition of $\mathcal{A}$ and Maxwell-type identities for mixed second derivatives to rewrite $(\partial \mathcal{A}/\partial Y)_{\xi,T}$ as

$$\left(\frac{\partial \mathcal{A}}{\partial Y}\right)_{\xi,T} = \left(\frac{\partial \mathcal{A}}{\partial P}\right)_{\xi,T} = \frac{\partial}{\partial P} \left(\frac{\partial G}{\partial \xi}\right)_T = \frac{\partial}{\partial \xi} \left(\frac{\partial G}{\partial P}\right)_T = \left(\frac{\partial V}{\partial \xi}\right)_{T,P} \quad (8.45)$$

From (8.43), we see that the derivatives $dV/d\xi$ and $d\xi/dP$ must have opposite signs under the stated conditions. Therefore, if $dV/d\xi$ is *positive* (i.e., $V_{\text{products}} > V_{\text{reactants}}$), then $d\xi/dP$ must be *negative* (i.e., the reaction shifts *left* with respect to pressure increases), thereby reducing volume to “oppose” the increased- P stress, as Le Chatelier’s principle leads us to expect. Other special cases can be treated similarly.

Note that the variable Y must be able to *affect* the position of equilibrium ξ ; change of the partial pressure of an *unreactive* species (e.g., inert He gas) cannot affect ξ , and therefore cannot result in any “re-equilibration” change in the system. In all cases, the rigorous thermodynamic formulation (8.43) may be used to guide qualitative applications of the Le Chatelier concept.

8.7 THERMODYNAMICS OF ELECTROCHEMICAL CELLS

In Section 3.3.4, we discussed the inclusion of electrical (emf) contributions to the general expression for work,

$$dw_{\text{emf}} = \mathcal{E} dQ \quad (8.46)$$

as appropriate to an electrochemical cell, where $\mathcal{E}$ is the cell potential and Q is the charge transferred in the cell reaction. Figure 8.2 depicts the schematics of such a cell, showing the directions of current I and electron e^- flow in the external circuit, as well as ion flow in the internal cell reaction. Electrochemical cells provide important practical devices for extracting useful electrical work from a chemical reaction, as well as important laboratory devices for measuring thermodynamic properties of cell reactions, as we now wish to describe.

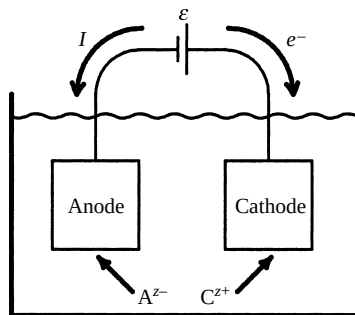


Figure 8.2 Schematic electrochemical cell of potential $\mathcal{E}$, showing directions of current I , electron e^- , cation C^{z+} , and anion A^{z-} flow as the cell redox reaction proceeds, with z electrons transferred to/from the ions at each electrode.

Compared with the cell-free conditions discussed previously, the internal energy changes dU must be modified by inclusion of the electrical work term:

$$dU_{\text{cell}} = dU + \mathcal{E} dQ \quad (8.47)$$

Similarly, the equilibrium condition for the cell reaction ($dG_{\text{cell}} = 0$) must be modified from the earlier cell-free form ($dG = 0$) by inclusion of the electrical work term:

$$dU_{\text{cell}} = dG + \mathcal{E} dQ = 0 \quad \text{at equilibrium (constant } T, P) \quad (8.48)$$

where dQ is the infinitesimal charge increment drawn from a cell at potential $\mathcal{E}$.

The charge increment dQ for current I flowing through time increment dt can be measured as

$$dQ = I dt \quad (8.49)$$

However, from a chemical viewpoint, dQ can also be expressed in terms of the electrons transferred at the electrodes for each increment $d\xi$ of cell reaction. For this purpose, it is convenient to write the overall redox cell reaction as separate oxidation/reduction “half-reactions,” expressing the loss or gain of z electrons at each electrode in the balanced cell reaction (i.e., involving z “equivalents” of charge transferred in oxidation and reduction steps). It is also convenient to quantify total charge in “molar” units (i.e., Avogadro’s number N_A of electrons) as expressed by the Faraday constant $\mathcal{F}$,

$$\mathcal{F} = \text{charge on } N_A \text{ electrons} \simeq 96,490 \text{ C (SI units)} \quad (8.50)$$

With these definitions, we obtain

$$dQ = z\mathcal{F} d\xi \quad (8.51)$$

as the expression for the charge transferred in each advancement increment $d\xi$ of the cell reaction. Accordingly, (8.46) becomes

$$dw_{\text{cell}} = z\mathcal{F}\mathcal{E} d\xi \quad (8.52)$$

and the equilibrium cell condition (8.48) becomes

$$dG = -z\mathcal{F}\mathcal{E} d\xi \quad (8.53)$$

in terms of the earlier (cell-free) $dG = dG_{\text{rxn}}$ of chemical reaction.

To obtain the overall ΔG of chemical reaction, we formally integrate (8.53) from $\xi = 0$ (reactants) to 1 (products) under constant- $\mathcal{E}$ conditions,

$$\int_{\text{reactants}}^{\text{products}} dG = \Delta G = - \int_0^1 z\mathcal{F}\mathcal{E} d\xi = -z\mathcal{F}\mathcal{E} \quad (8.54)$$

recognizing that z , $\mathcal{F}$, and $\mathcal{E}$ are all constants. From this, we obtain the central electrochemical relationship

$$\boxed{\Delta G = -z\mathcal{F}\mathcal{E}} \quad (8.55)$$

for the given chemical reaction of interest. Measuring ΔG is as simple as measuring the potential $\mathcal{E}$ of an electrochemical cell at the chosen temperature and pressure conditions!

To proceed, we can introduce the more general equation (8.26) for ΔG in terms of reaction quotient Q ,

$$\Delta G = \Delta G^\circ + RT \ln Q \quad (8.56a)$$

or, in terms of the *standard electrode potential* $\mathcal{E}^\circ$ satisfying $\Delta G^\circ = -z\mathcal{F}\mathcal{E}^\circ$,

$$-z\mathcal{F}\mathcal{E} = -z\mathcal{F}\mathcal{E}^\circ + RT \ln Q \quad (8.56b)$$

Dividing through by the constants $z\mathcal{F}$, we obtain the celebrated *Nernst equation*

$$\boxed{\mathcal{E} = \mathcal{E}^\circ - \frac{RT}{z\mathcal{F}} \ln Q} \quad (8.57a)$$

which describes the concentration (Q) dependence of electrochemical cell potential for the given cell reaction. For standard $T = 298.15\text{K}$, the physical constants can be combined with the conversion factor from natural to common logarithms to give $2.303RT/\mathcal{F} \simeq 0.05915\text{ V}$, leading to

$$\mathcal{E} = \mathcal{E}^\circ - \frac{0.05915}{z} \log_{10} Q \quad (8.57b)$$

the more familiar textbook form of the Nernst equation.

Equations (8.55), (8.57) make it easy to see the connection between cell potential and reaction spontaneity. When $\mathcal{E}$ is *positive* ($\Delta G < 0$), the cell reaction is *spontaneous*, allowing useful electrical work to be withdrawn (i.e., “the battery is draining”). Conversely, when $\mathcal{E}$ is *negative* ($\Delta G > 0$), the cell reaction is *nonspontaneous*, requiring input work from the surroundings (i.e., “the battery is charging”). When $\mathcal{E} = 0$ ($\Delta G = 0$), the cell reaction is at *equilibrium*, and, in accordance with (8.27d), the Nernst equation reduces to

$$\mathcal{E}^\circ = \frac{RT}{z\mathcal{F}} \ln K_{\text{eq}} \quad (8.58a)$$

or, equivalently,

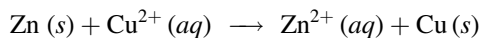
$$K_{\text{eq}} = \exp\left(\frac{z\mathcal{F}\mathcal{E}^\circ}{RT}\right) \quad (8.58b)$$

showing that measurement of standard electrode potential $\mathcal{E}^\circ$ is tantamount to measurement of K_{eq} .

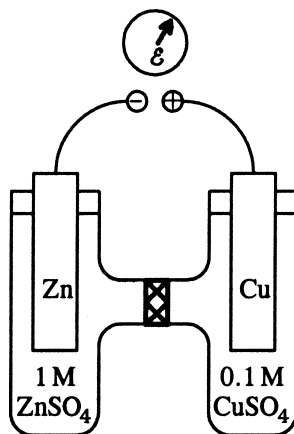
Because $\mathcal{E}$ (like ΔG) refers to a difference in a *state* property, it can be evaluated in additive fashion along many alternative pathways. For this purpose, it is convenient to assign conventional $\mathcal{E}^\circ$ values to each *half-cell* reaction [e.g., “standard oxidation potentials” as compiled in W. M. Latimer, *Oxidation Potentials*, 2nd edn (Prentice-Hall, New York, 1952)], such that the algebraic sum of the two half-reaction potentials equals the overall cell $\mathcal{E}^\circ$. Such half-reaction $\mathcal{E}^\circ$ values can in turn be obtained by choosing some standard electrode reaction as the conventional “zero” of the scale [such as the “standard hydrogen electrode” (SHE) for the $1/2\text{H}_2(\text{g}) \rightarrow \text{H}^+(\text{aq}) + e^-$ oxidation reaction, with $\mathcal{E}_{\text{SHE}}^\circ = 0$]. Sidebar 8.2 illustrates a simple example of this procedure.

SIDEBAR 8.2: NERNST ELECTRODE POTENTIAL PROBLEM

Problem A strip of Zn metal, when dipped into an aqueous CuSO_4 solution, is observed to dissolve vigorously, with production of flakes of pure metallic Cu and aqueous ZnSO_4 , indicating strong spontaneity of the net ionic redox equation



The following diagram depicts an electrochemical cell based on this reaction [denoted $\text{Zn}|\text{Zn}^{2+}(1\text{ M})||\text{Cu}^{2+}(0.1\text{ M})|\text{Cu}$]:



The oxidation $[\text{Zn} (s) \rightarrow \text{Zn}^{2+} (aq) + 2e^-]$ and reduction $[\text{Cu}^{2+} (aq) + 2e^- \rightarrow \text{Cu} (s)]$ half-reactions are confined to the left (anode) and right (cathode) chambers, respectively, separated by a “salt bridge” (allowing passage of SO_4^{2-} counterions to preserve electro-neutrality). Use the Nernst equation and tabulated standard oxidation potentials

$$\mathcal{E}^\circ[\text{Zn} \rightarrow \text{Zn}^{2+} + 2e^-] = 0.763\text{ V}$$

$$\mathcal{E}^\circ[\text{Cu} \rightarrow \text{Cu}^{2+} + 2e^-] = -0.337\text{ V}$$

to estimate the voltmeter reading at the terminals when, as shown, the $[\text{SO}_4^{2-} (aq)]$ concentrations are 1 M in the anode chamber and 0.1 M in the cathode chamber.

Solution From the given half-reaction $\mathcal{E}^\circ$ values, we can algebraically sum the two half-reactions to get the standard-state $\mathcal{E}^\circ$ for the overall reaction as follows:

	Reaction	$\mathcal{E}^\circ$ (V)
Anode:	$\text{Zn} (s) \rightarrow \text{Zn}^{2+} (aq) + 2e^-$	+0.763
Cathode:	$\text{Cu}^{2+} (aq) + 2e^- \rightarrow \text{Cu} (s)$	+0.337
Overall:	$\text{Zn} (s) + \text{Cu}^{2+} (aq) \rightarrow \text{Zn}^{2+} (aq) + \text{Cu} (s)$	+1.100

If we neglect the (unit) activities of solid metals and approximate the ion activities by molarities [cf. (8.22a)],

$$[\text{Zn}^{2+}] = 1.0 \text{ M}, \quad [\text{Cu}^{2+}] = 0.1 \text{ M}$$

we can use the Nernst equation (8.57b) to evaluate the cell potential at the given concentration conditions as

$$\mathcal{E} = \mathcal{E}^\circ - \frac{0.059}{z} \log_{10} \left(\frac{[\text{Zn}^{2+}]}{[\text{Cu}^{2+}]} \right) = +1.100 - \frac{0.059}{2} \log_{10} (10) = +1.070 \text{ V}$$

showing that the cell reaction is strongly spontaneous. (Measured deviations from the predicted $\mathcal{E} = +1.070 \text{ V}$ are indicative of nonideal ion activities, as discussed in the following section.)

8.8 ION ACTIVITIES IN ELECTROLYTE SOLUTIONS

The early studies of Grotthuss, Arrhenius, and others led to general recognition that certain molecular species, termed “electrolytes,” are largely dissociated into ionic subunits in aqueous solutions. Such electrolyte behavior reflects a general tendency for ionic or polar bonds (including those of water) to dissociate heterolytically under the influence of specific solvent coordination (e.g., hydrogen bonds) that can sufficiently stabilize and delocalize ionic charge. The important role of ionic subunits in aqueous chemistry suggests that aqueous ions may also be considered as formal “components” or “reagents” of chemical reactions, with assigned enthalpies of formation (cf. Section 3.6.8), activities, and other thermodynamic descriptors. Because aqueous ionic interactions typically lie far from the “ideal” thermodynamic limit, the discussion of ion activity requires special consideration of strong nonideality effects. Deeper molecular-level understanding of electrolyte nonideality is an important area of modern physical chemistry research.

Rather than mole fraction or molarity, as employed in previous discussions of near-ideal systems, it is convenient to discuss ion activities on the *molality* scale [cf. Section 7.3.7, (7.80c)]. Compared with molarity c_i , the molality m_i , defined as

$$m_i \equiv \frac{\text{moles solute } n_i}{\text{kg solvent}} \quad (8.59)$$

shares the advantages of close association with mole numbers n_i , but avoids implicit dependence on other (non-mass-related) thermodynamic variables, a severe disadvantage of molarity units (Sidebar 8.3). We therefore switch to the molality scale to write chemical potential [cf. (7.78)] as

$$\mu_i = \mu_i^\circ + RT \ln a_i = \mu_i^\circ + RT \ln (\gamma_i m_i) \quad (8.60)$$

where the activity coefficient $\gamma_i = \gamma_i^m$ [cf. Eq. (7.80c)] is now in inverse molality units.

SIDEBAR 8.3: MOLARITY VERSUS MOLALITY UNITS

Problem A careful laboratory technician prepares an ethanol/water solution of precisely known concentration by measuring the number of moles of pure ethanol (n_{et}), mass of pure water (M_{wa} , in kilograms) and volume of the resulting solution (V_{soln}). From these, he calculates the precise molarity and molality of the solution,

$$\text{molarity } c_{\text{et}} \equiv \frac{n_{\text{et}}}{V_{\text{soln}}}, \quad \text{molarity } m_{\text{et}} \equiv \frac{n_{\text{et}}}{M_{\text{wa}}}$$

then carefully seals the solution in a tightly stoppered container, labeled with the accurate molarity and molality, and sends the container to a deep underground storage site for security. However, an equally careful technician at the storage site checks the container contents and finds that one of the two concentration values is now (slightly) incorrect! Which value (c_{et} or m_{et}) is erroneous, and why?

Solution Although n_{et} and M_{wa} (and thus m_{et}) are invariant and remain precisely determined, slight differences in temperature or pressure at the underground facility will cause slight variations of V_{soln} , thus altering the molarity c_{et} of a sealed container. Such extraneous dependence makes molarity an unsatisfactory concentration unit for quantitative physical chemistry purposes.

For a simple 1 : 1 electrolyte undergoing dissociation of the form



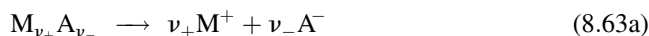
we seek to factor the activity a of MA into constituent ion activities a_+ , a_- for the component ions M^+ , A^- . Although such ion activities cannot be determined individually, we can instead define the “mean ionic activity” $a_{\pm}$ as

$$a = a_+ a_- = a_{\pm}^2 \quad (1 : 1 \text{ electrolyte}) \quad (8.62a)$$

that is, as the geometric mean of cation and anion activities,

$$a_{\pm} \equiv (a_+ a_-)^{1/2} \quad (8.62b)$$

For a more general $\nu_+ - \nu_-$ electrolyte, dissociating as



we can similarly define $a_{\pm}$ through the equation

$$a = a_{\pm}^{\nu} = a_+^{\nu_+} a_-^{\nu_-} \quad (8.63b)$$

or

$$a_{\pm} \equiv a^{1/\nu} \quad (8.63c)$$

with total number of ions ν ,

$$\nu = \nu_+ + \nu_- \quad (8.63d)$$

Although ion activities a_+ , a_- are not separately measurable, the mean ionic activity $a_{\pm}$ can be readily determined through (8.63c) from the known activity a of the neutral electrolyte.

If we now envision the separate activity coefficients (γ_+ , γ_-) and molalities (a_+ , a_-) for the ions, satisfying

$$a_+ = \gamma_+ m_+ \quad (8.64a)$$

$$a_- = \gamma_- m_- \quad (8.64b)$$

we can formally extend the geometric-mean concept to define a “mean ionic activity coefficient” $\gamma_{\pm}$ and a “mean ionic molality” $m_{\pm}$ in a manner similar to (8.63d):

$$\gamma_{\pm} \equiv (\gamma_+^{\nu_+} \gamma_-^{\nu_-})^{1/\nu} \quad (8.65a)$$

$$m_{\pm} \equiv (m_+^{\nu_+} m_-^{\nu_-})^{1/\nu} \quad (8.65b)$$

With these definitions, we obtain

$$a_{\pm} = \gamma_{\pm} m_{\pm} \quad (8.66)$$

where total activity a is expressed as

$$a = (\gamma_+ m_+)^{\nu_+} (\gamma_- m_-)^{\nu_-} = a_{\pm}^{\nu} = (\gamma_{\pm} m_{\pm})^{\nu} \quad (8.67)$$

The use of such equations to determine $\gamma_{\pm}$, $m_{\pm}$ from the known activity a and molality m of a given $\nu_+ - \nu_-$ electrolyte is illustrated in Sidebar 8.4.

SIDEBAR 8.4: MEAN IONIC ACTIVITY PROBLEM

Problem Determine $\gamma_{\pm}$ for a 0.125 m $\text{La}_2(\text{SO}_4)_3$ solution of measured activity a .

Solution For the 2 : 3 electrolyte $\text{La}_2(\text{SO}_4)_3$, we have

$$\nu_+ = 2, \quad \nu_- = 3, \quad \nu = 5$$

so that

$$m_+ = 2m, \quad m_- = 3m$$

and

$$m_{\pm} = [(2m)^2 (3m)^3]^{1/5} = m[(4)(27)]^{1/5} = m(108)^{1/5} = 0.125(2.551) = 0.319$$

For the measured a , one can write

$$a = a_{\pm}^5 = a_{\text{La}}^2 a_{\text{SO}_4}^3 = m_{\pm}^5 \gamma_{\pm}^5$$

so that

$$\gamma_{\pm} = \frac{a^{1/5}}{m_{\pm}} = 3.136a^{1/5}$$

An important aspect of electrolyte solutions can be recognized from the *power-law* dependences on ionic charge in equations such as (8.67). As suggested by Coulomb's law of classical electrostatics, the energetic strength of interactions between ions of charge z_i increases *quadratically* with charge. As a result, ions of larger z_i are expected to make disproportionately large contributions to ionic nonideality effects, far stronger than their molality m_i might suggest.

A useful descriptor of the overall strength of ionic forces in electrolyte solutions is given by the “ionic strength” I of a solution, defined as

$$I \equiv 1/2 \sum_i^{\text{ions}} m_i z_i^2 \quad (8.68)$$

G. N. Lewis first recognized empirically that ion activity seems to depend only on I (not the specific ion identities) in the dilute limit, thus suggesting the importance of I in a general theory of electrolyte solutions. As shown by (8.68), the contribution of each ion i to ionic strength I is linearly proportional to molality m_i , but quadratically proportional to charge z_i . As suggested in the previous paragraph and illustrated in Sidebar 8.5, ionic-strength effects are expected to grow dramatically with increasing charge on the ions.

SIDEBAR 8.5: IONIC STRENGTH COMPARISON

Problem Two aqueous solutions are prepared, each of 0.1 m salt concentration, but with NaCl as the salt in the first case and $\text{La}_2(\text{SO}_4)_3$ in the second. Which salt solution is expected to exhibit larger ionic strength effects? By what factor?

Solution For 0.1 m NaCl, a 1 : 1 electrolyte, we evaluate the ionic strength from (8.68) as

$$I_{1:1} = \frac{1}{2} [0.1(+1)^2 + 0.1(-1)^2] = 0.1$$

whereas for the same molality of the 2 : 3 electrolyte $\text{La}_2(\text{SO}_4)_3$, we obtain

$$I_{2:3} = \frac{1}{2} [0.2(+3)^2 + 0.3(-2)^2] = 1.5$$

that is, 15 *times* as strong in the latter case!

With these preliminaries, we now briefly summarize the nonideality effects in representative electrolyte solutions. Figure 8.3 displays the characteristic behavior of mean ionic activity coefficients $\gamma_{\pm}$ for typical 1 : 1 and 2 : 1 electrolytes as a function of concentration. As shown in the figure, the ideal limit $\gamma_{\pm} \rightarrow 1$ is achieved at $m = 0$ in each case (as expected), but the deviations from ideality are generally *much* larger than those typical for nonelectrolytes at nonzero m values. As shown, the electrolyte $\gamma_{\pm}$ values decay *steeply* from ideality even for small values of m , quite different from the gently sloping deviations exhibited by typical nonelectrolytes in the dilute limit. The nonideality is seen to be much stronger for 2 : 1 electrolytes (e.g., H_2SO_4) than for 1 : 1 electrolytes (e.g., HCl), consistent with strong dependence on ionic charges as expressed by ionic strength I .

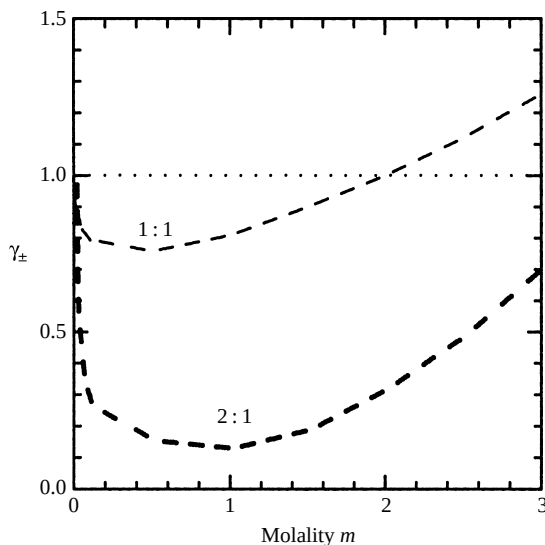


Figure 8.3 Schematic plot of mean ionic activity coefficient $\gamma_{\pm}$ versus molality (m) for typical strong electrolytes [1:1 (e.g., HCl), light dashed line; 2:1 (e.g., H_2SO_4), heavy dashed line], showing the extreme deviations from ideality (dotted line) even in dilute solutions.

The qualitative behavior shown in Fig. 8.3 is representative of many strong electrolytes. In the dilute limit of small m , where solute–solute interactions are negligible, solute–solvent interactions *reduce* the chemical potential of the solute, μ_{solute} , corresponding to strong *negative* deviations from ideality. At higher m , the increasingly strong solute–solute interactions act to *increase* μ_{solute} , resulting in *positive* deviations from ideality. These competing tendencies lead to a *minimum* in $\gamma_{\pm}$ at intermediate m , as well as near-ideality ($\gamma_{\pm} \simeq 1$) in a crossover region of higher m , where competitive solute–solvent and solute–solute effects are effectively cancelling.

A qualitative molecular-level understanding of the behavior of $\gamma_{\pm}$ in the dilute-electrolyte limit is provided by Debye–Hückel theory (Sidebar 8.6). This theory predicts logarithmic proportionality of $\gamma_{\pm}$ to the square root of ionic strength,

$$\log_{10} \gamma_{\pm} = -0.509 |z_+ z_-| I^{1/2} \quad (8.69)$$

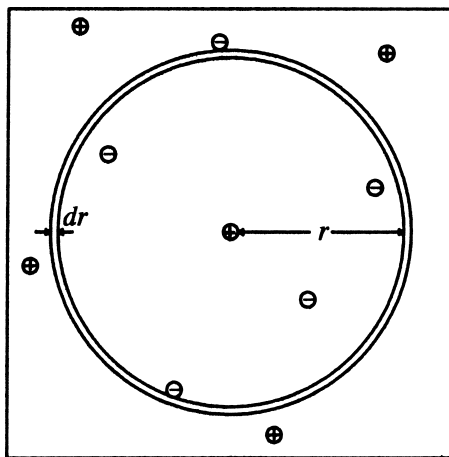
as observed for a dilute electrolyte with ionic charges z_+ , z_- . Although Debye–Hückel theory seems to account satisfactorily for the sharply negative values of $\gamma_{\pm}$ in the extreme-dilution limit, other aspects of nonideality in strong and weak electrolytes remain deeply challenging to molecular-level description.

The extreme nonidealities characteristic of electrolyte solutions warn of the dangers inherent in approximations commonly employed in general chemistry. Except in the crossover region of intermediate m where $\gamma_{\pm} \simeq 1$, blithe replacement of activity by molality is seldom justified for strong electrolytes. Elementary treatments of acid dissociation, solubility products, and the like may therefore be subject to considerable error unless the realistic variations of chemical potential with concentration are properly considered.

SIDEBAR 8.6: DEBYE–HÜCKEL THEORY OF DILUTE ELECTROLYTES

In 1923, Peter Debye and Erich Hückel developed a classical electrostatic theory of ionic distributions in dilute electrolyte solutions [P. Debye and E. Hückel. *Phys. Z.* **24**, 185 (1923)] that seems to account satisfactorily for the qualitative low- m nonideality shown in Fig. 8.3. Although this theory involves some background in statistical mechanics and electrostatics that is not assumed elsewhere in this book, we briefly sketch the physical assumptions and mathematical techniques leading to the Debye–Hückel equation (8.69) to illustrate such molecular-level description of thermodynamic relationships.

For a charged solute species, such as the cationic “test charge” shown in the center of the diagram below, the dominant influence on μ_{solute} arises from the integrated effect of the *electrostatic potential* $\Phi(r)$ due to its Coulombic interactions with the charge density $\rho(r)$ of other ions at distance r . The central assumption of Debye–Hückel theory is that each ion tends to be surrounded by an “atmosphere” of ions of opposite charge (see diagram), resulting in net *attractions* of each solute ion to its surroundings and consequent *lowering* of chemical potential (and mean ionic activity coefficient) compared with the ideal (zero-concentration) limit. We now wish to determine the charge density ρ and associated potential Φ of a distribution of solute ions i with charge $z_i e$ and overall molality m_i .



The mathematical relationship between $\Phi(r)$ and $\rho(r)$ is given by *Poisson's equation* of classical electrostatics:

$$\nabla^2 \Phi = -(\kappa \epsilon_0)^{-1} \rho \quad (\text{S8.6-1})$$

In this equation, $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ denotes the “Laplacian” operator of cartesian second derivatives, $\rho(r)$ is the charge density in a spherical shell of radius r and infinitesimal thickness dr centered at the particle of interest (see diagram), κ is the effective dielectric constant, and ϵ_0 is the “permittivity of free space” (8.854×10^{-12} in SI units). The energy of interaction E_i of ions of charge $z_i e$ with their surroundings,

$$E_i = z_i e \Phi \quad (\text{S8.6-2})$$

leads to expected variations of charge density around the center of interest. We assume these variations can be described by a Boltzmann weighting factor $\exp(-E_i/kT)$ that modulates the average charge density contribution (charge per unit volume: $z_i n_i \mathcal{F}/V$, where $\mathcal{F}$ is the Faraday constant):

$$\rho(r) = \sum_i \frac{z_i n_i \mathcal{F}}{V} \exp\left[-\frac{z_i e \Phi(r)}{kT}\right] \quad (\text{S8.6-3})$$

Substitution of the Boltzmann expression (S8.6-3) for ρ into (S8.6-1) leads to the *Poisson–Boltzmann* (PB) equation

$$\nabla^2 \Phi = (\kappa \epsilon_0)^{-1} \sum_i \frac{z_i n_i \mathcal{F}}{V} \exp\left(-\frac{z_i e \Phi}{kT}\right) \quad (\text{S8.6-4})$$

for the unknown $\Phi(r)$.

To obtain a useful approximate solution of the PB equation (S8.6-4), we consider the dilute limit in which the electric potential Φ is weak compared with the ambient thermal energy kT . In this limit, the Boltzmann exponential can be “linearized” by retaining only the leading term in the power series expansion

$$\exp\left(-\frac{z_i e \Phi}{kT}\right) \simeq 1 - \frac{z_i e \Phi}{kT} + \dots \quad (\Phi \ll kT) \quad (\text{S8.6-5})$$

Equation (8.6-4) is therefore replaced by the *linearized Poisson–Boltzmann* (LPB) equation

$$\nabla^2 \Phi = (\kappa \epsilon_0)^{-1} \left(\sum_i \frac{z_i n_i \mathcal{F}}{V} + \sum_i \frac{z_i^2 n_i e \mathcal{F} \Phi}{VkT} \right) \quad (\text{S8.6-6})$$

The first term on the right-hand side is seen to vanish due to overall charge balance,

$$\frac{\mathcal{F}}{\kappa \epsilon_0 V} \sum_i z_i n_i = 0 \quad (\text{S8.6-7})$$

and, by multiplying and dividing by solvent mass w_{solv} (in kg), the second term can be seen to be proportional to ionic strength I [cf. (8.68)]:

$$\sum_i \frac{z_i^2 n_i e \mathcal{F} \Phi}{\kappa \epsilon_0 V kT} = \frac{2w_{\text{solv}} e \mathcal{F} \Phi}{\kappa \epsilon_0 V kT} \left[\frac{1}{2} \sum_i (n_i / w_{\text{solv}}) z_i^2 \right] = \frac{2w_{\text{solv}} e \mathcal{F} \Phi}{\kappa \epsilon_0 V kT} I \quad (\text{S8.6-8})$$

The LPB equation (S8.6-6) can therefore be rewritten symbolically as

$$\nabla^2 \Phi = b^2 \Phi \quad (\text{S8.6-9})$$

where b is a “screening” constant,

$$b = \frac{1}{\lambda} = \left(\frac{2w_{\text{solv}} e \mathcal{F} I}{\kappa \epsilon_0 V kT} \right)^{1/2} \quad (\text{S8.6-10})$$

which is seen to be proportional to the square root of ionic strength. Physically, b has units of inverse length, and its inverse λ_D , the *Debye screening length*,

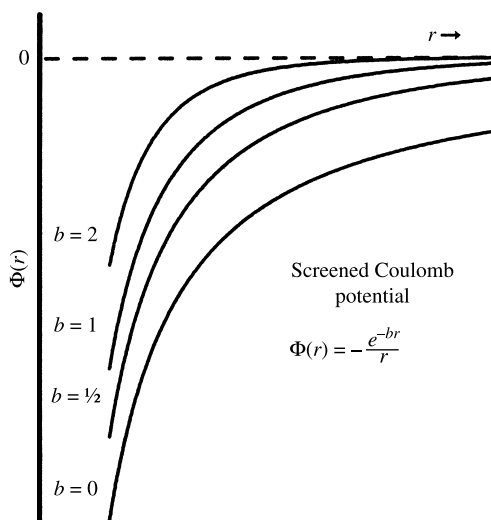
$$\lambda_D = \frac{1}{b} \quad (\text{S8.6-11})$$

gives the characteristic lengthscale “thickness” of the counterion cloud surrounding each ion.

The LPB differential equation (S8.6-9) can be solved analytically to give the potential $\Phi(r)$ as

$$\Phi(r) = -\frac{e^{-br}}{r} \quad (\text{S8.6-12})$$

(or any multiple), as may be verified by substitution. For $b = 0$, this reduces to the Coulomb potential energy of a bare (unscreened) charge, $\Phi_{b=0} = -1/r$. As shown in the graph below, the range of the potential is successively weakened by the screening effect of the counterion cloud for increasing $b > 0$ (or reduced Debye length λ_D). Only at small r does $\Phi(r)$ resemble that of a bare ion, whereas the potential is markedly weakened at distances of the order of the Debye length λ , consistent with the physical assumption of the theory.



To evaluate the effect of counterion screening on the chemical potential of the chosen solute ions, we first multiply (S8.6-12) by the appropriate prefactor for ions of charge Q ($= ze$) in a medium of dielectric constant κ :

$$\Phi(r) = \frac{Q}{4\pi\kappa\epsilon_0} \frac{e^{-br}}{r} \quad (\text{S8.6-13})$$

This potential can be separated into the contribution of the isolated ion (Φ_{ion} , the ideal limit) plus the “extra” contribution of the counterion atmosphere (Φ_{extra} , the nonideality correction):

$$\Phi = \Phi_{\text{ion}} + \Phi_{\text{extra}} \quad (\text{S8.6-14})$$

To estimate Φ_{extra} for small b , we expand the exponential on the right-hand side of (S8.6-13), retaining only the linear term:

$$\frac{Q}{4\pi\kappa\epsilon_0} \frac{e^{-br}}{r} = \frac{Q}{4\pi\kappa\epsilon_0} \left(\frac{1}{r} - b + \dots \right) \quad (\text{S8.6-15})$$

Recognizing that the first term on the right-hand side is the bare ion term Φ_{ion} , we can identify the “atmosphere” term Φ_{extra} with the leading correction:

$$\Phi_{\text{extra}} = \frac{-Qb}{4\pi\kappa\epsilon_0} \quad (\text{S8.6-16})$$

where $Q = z_+e$ or z_-e for the solute cation or anion, respectively.

To relate these expressions to Gibbs free energy changes, and thus to ion activities, we recall that ΔG is equivalent to electrical *work* under constant- T, P conditions. We therefore integrate dw_{emf} [cf. (3.16), where we used the symbol E for electrical potential Φ] to obtain

$$\Delta G = \int_0^{ze} dw_{\text{emf}} = \int_0^{ze} \Phi dQ = \int_0^{ze} (\Phi_{\text{ion}} + \Phi_{\text{extra}}) dQ = \Delta G_{\text{ideal}} + \Delta G_{\text{extra}} \quad (\text{S8.6-17})$$

In this expression, ΔG_{ideal} is the work to “charge up” the bare ion from 0 to ze , while ΔG_{extra} is the accompanying work of polarizing the surroundings (the nonideality effect). Employing (S8.6-16), we obtain ΔG_{extra} (on a per-ion basis) as

$$\Delta G_{\text{extra}} = \int_0^{ze} \Phi_{\text{extra}} dQ = -\frac{b}{4\pi\kappa\epsilon_0} \int_0^{ze} Q dQ = -\frac{bz^2e^2}{8\pi\kappa\epsilon_0} \quad (\text{S8.6-18})$$

From the association $\Delta G_{\text{extra}} = kT \ln \gamma$ with the ion activity coefficient γ , we can therefore write

$$\ln \gamma = \frac{\Delta G_{\text{extra}}}{kT} = -\frac{e^2}{8\pi\kappa\epsilon_0 kT} z^2 b \quad (\text{S8.6-19})$$

for either cation ($z = z_+$) or anion ($z = z_-$). Neither ion activity coefficient can be observed directly, but, with a little logarithmic arithmetic, we can verify that an analogous equation holds for the mean ionic activity coefficient $\gamma_{\pm}$, namely,

$$\ln \gamma_{\pm} = -\frac{e^2}{8\pi\kappa\epsilon_0 kT} |z_+ z_-| b \quad (\text{S8.6-20})$$

or, from (S8.6-10),

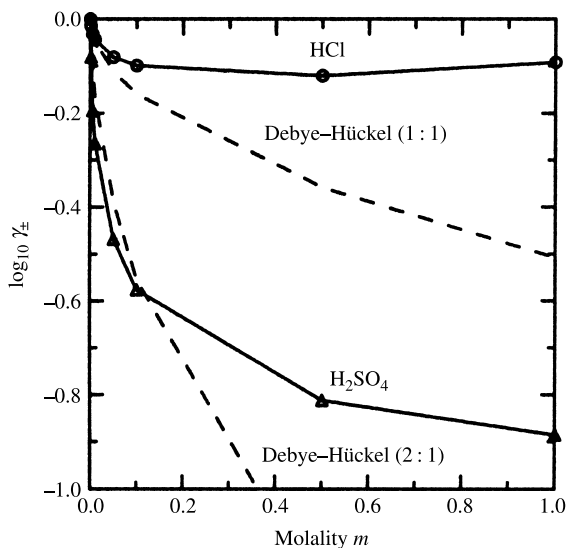
$$\log_{10} \gamma_{\pm} = -\frac{e^2}{2.303(8\pi\kappa\epsilon_0 kT)} \left(\frac{2w_{\text{solv}}e\mathcal{F}}{V\kappa\epsilon_0 kT} \right)^{1/2} |z_+ z_-| I^{1/2} \quad (\text{S8.6-21})$$

Evaluating the physical constants for water (with $\kappa \simeq 78.4$ and $w_{\text{solv}}/V \simeq 997 \text{ kg m}^{-3}$ at standard 25°C, 1 atm), we obtain the celebrated Debye–Hückel expression [cf. (8.69)]

$$\log_{10} \gamma_{\pm} = -0.509 |z_+ z_-| I^{1/2} \quad (\text{S8.6-22})$$

The following plot compares the Debye–Hückel theory (dashed) with experiment (solid) for HCl (circles) and H₂SO₄ (triangles), showing the generally satisfactory agreement at

low m but increasingly large deviations at larger m or I :



It is apparent from this plot that the Debye–Hückel theory is only “correct” in a narrow range of extreme dilution, showing (as might have been anticipated) that classical electrostatics is qualitatively inadequate to describe the realistic interactions in aqueous electrolyte solutions, except in the asymptotic limit of extremely large ion–ion separations.

8.9 CONCLUDING SYNOPSIS OF GIBBS' THEORY

In this concluding section of Part II, it is appropriate to summarize the major conceptual and mathematical features of the Gibbs formulation of thermodynamic equilibrium theory, as a preface to its geometrical reformulation in the ensuing Part III. The following equations (8.70)–(8.95) summarize the essential mathematical structure of the Gibbsian formalism, as erected on the historical foundation of Chapters 1–4 and exploited in the applications of Chapters 5–8.

The central focus of Gibbs' theory is the *equilibrium state* $\mathcal{S}$, a quiescent limiting condition of a sufficiently large (“macroscopic”) physical system that exhibits characteristically simple responses to attempted changes of the “control variables” that specify the state.

The control variables of the state may be selected rather arbitrarily from two classes:

The first class comprises *quantity-type* (“extensive”) properties X_i , such as the quantity of mass or spatial volume of the system. A distinctive characteristic of these properties is their additivity in subunits of the system, such that each is linearly proportional to overall “scale” (number of identical subunits), as measured, for example, by total mass. Arbitrary linear combinations of the independent X_i ,

$$X'_j = \sum_i a_i X_i \quad (8.70)$$

also share this additivity property and are equally suitable extensive descriptors of $\mathcal{S}$. The most fundamental extensity is the “internal energy” of the system, U , as physically defined through its differential relationship to heat q and work w measurements [cf. (3.37b)]:

$$dU \equiv dq + dw \quad (8.71)$$

The functional relationship that expresses how U depends on other extensities $\{X_i\}$ is called the “fundamental equation” of equilibrium thermodynamics:

$$U = U(X_1, X_2, \dots, X_i, \dots) \quad (8.72)$$

Knowledge of the first and second differentials of U is sufficient for *complete* thermodynamic description of the state.

The second class of control variables comprises derivative *strength*-type (“intensive”) properties R_i , such as temperature and pressure. Each R_i is related through the fundamental equation (8.72) to a “conjugate” extensity X_i by a derivative relationship (“equation of state”) of the form [cf. (3.32), (4.33)]

$$R_i = \left(\frac{\partial U}{\partial X_i} \right)_{\underline{X}} \quad (8.73)$$

so that

$$dU = \sum_i R_i dX_i \quad (8.74)$$

The distinctive characteristic of intensities R_i is their uniformity throughout all subunits of the system. Arbitrary linear combinations of the independent R_i ,

$$R'_j = \sum_i a_i R_i \quad (8.75)$$

also share this uniformity, and are equally suitable intensive descriptors of $\mathcal{S}$. Any equilibrium system sharing the same values of all the R_i is in the same equilibrium “state.” The state of equilibrium is therefore uniquely specified by the independent intensities $\{R_i\}$, whose numerical values $\{r_i\}$ in state $\mathcal{S}$,

$$r_i = \left(\frac{\partial U}{\partial X_i} \right)_{\underline{X}} \Big|_{\mathcal{S}} \quad (8.76)$$

can be regarded as fundamental coordinates of $\mathcal{S}$ in the thermodynamic space of all such states:

$$\mathcal{S} = \mathcal{S}(\{r_i\}) \quad (8.77)$$

Experimental investigation of a specified equilibrium state (8.77) involves measurement of its responses, such as the change in R_j resulting from a change in control variable X_j (all other control variables being constant). Such measurable responses (“response functions”)

can be expressed in differential language as

$$M_{ij} = \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} \quad (8.78)$$

with measured value m_{ij} (in $\mathcal{S}$),

$$m_{ij} \equiv \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} \Big|_{\mathcal{S}} \quad (8.79)$$

As in the preceding paragraph, we wish to notationally distinguish an intrinsic quality or property of $\mathcal{S}$ [such as (8.73) or (8.78)] from its measured value [such as (8.76) or (8.79)]. The latter is an ordinary *number* (associated with a chosen unit system), but the former refers to the complex interdependences between first and second derivatives of U . The experimental consequences of the theory are generally expressed as relationships between measured m_{ij} values, with each m_{ij} representing the limiting proportionality of response dR_i to stimulus dX_j (other control variables being fixed).

The “laws of thermodynamics” express *analytical properties* of the fundamental state descriptor U , which Gibbs identified as “criteria of equilibrium.”

The “first law” establishes that (8.71) is an *exact* differential, satisfying the Euler criterion [cf. (1.21)]

$$\left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} = \left(\frac{\partial R_j}{\partial X_i} \right)_{\underline{X}}, \quad \text{all } i, j \quad (8.80)$$

The criterion (8.80) guarantees that $\oint dU = 0$ [cf. (3.40e)] in all cyclic processes, assuring *conservation* of U in all processes returning to the same state. Similarly, each remaining differential extensity dX_i of U may be formally expressed as

$$dX_i = R_i^{-1} dU \quad (\text{other } X_j = \text{constant}) \quad (8.81)$$

assuring its exactness by the same criterion. This establishes that the “entropy” S (the conjugate of temperature, with $dS = T^{-1} dU$, at closed constant- V conditions) is also a conserved state property, equivalent to U itself as a fundamental equation for complete thermodynamic description.

The “second law” establishes the *stability* (restorative tendency) of equilibrium, as expressed by the *positivity* criterion

$$\left(\frac{\partial R_i}{\partial X_i} \right)_{\underline{X}} \geq 0, \quad \text{all } X_i \quad (8.82)$$

for an internal energy minimum. Although the extremal character of equilibrium was first established (with opposite sign!) for the entropy function [cf. (5.16)], the essential equivalence between U and S as fundamental equations (see previous paragraph) allows one to equivalently express the second law in U -based form (8.82). Note that “all X_i ” includes arbitrary linear combinations such as (8.70).

For chemical purposes, the internal energy (8.71) must include chemical work terms, one for each of the c independent *chemical components* participating in “active” equilibria. Together with the usual extensities X_i for heat (S) and pressure–volume work (V), the arguments of the internal energy function must be extended to include c additional chemical extensities (such as the mole numbers $n_1, n_2, \dots, n_c$, with conjugate “chemical potentials” $\mu_1, \mu_2, \dots, \mu_c$),

$$U = U(X_1, X_2, \dots, X_{c+2}) = U(S, V, n_1, \dots, n_c) \quad (8.83)$$

to describe the thermodynamics of chemical equilibria.

A chemical equilibrium system may generally consist of p distinct homogeneous subregions (“phases”) in coexistence. For such a composite equilibrium system, the total internal energy is additive in each subregion:

$$U = \sum_{\alpha=1}^p U^{(\alpha)} \quad (8.84)$$

Each uniform phase $U^{(\alpha)}(X_1^{(\alpha)}, \dots, X_{c+2}^{(\alpha)})$ satisfies a mathematical scaling property (Euler “first-order homogeneity”)

$$U^{(\alpha)}(\lambda X_1^{(\alpha)}, \lambda X_2^{(\alpha)}, \dots, \lambda X_{c+2}^{(\alpha)}) = \lambda U^{(\alpha)}(X_1^{(\alpha)}, X_2^{(\alpha)}, \dots, X_{c+2}^{(\alpha)}), \quad (\alpha = 1, \dots, p) \quad (8.85)$$

for all positive scale factors λ , where $X_i^{(\alpha)}$ is the amount of X_i in phase α . Each such homogeneity property leads to a corresponding “Euler condition” [cf. (6.29)]

$$U^{(\alpha)} = \sum_{i=1}^{c+2} X_i^{(\alpha)} R_i^{(\alpha)} \quad (8.86)$$

and thereby to an associated “Gibbs–Duhem equation”

$$0 = \sum_{i=1}^{c+2} X_i^{(\alpha)} dR_i, \quad \alpha = 1, \dots, p \quad (8.87)$$

where $X_i^{(\alpha)}$, the quantity of X_i in phase α , satisfies the overall conservation condition

$$X_i = \sum_{\alpha=1}^p X_i^{(\alpha)} \quad (8.88)$$

The $c + 2$ differential intensities dR_i are therefore not all independent, being subject to the p equations (8.87) of phase coexistence. The number of independent intensities (the dimensionality of state space), f , is therefore given by [cf. (7.6)]

$$f = c + 2 - p \quad (8.89)$$

the celebrated “Gibbs phase rule” of phase equilibrium theory.

As shown by (8.74), stationary values of U ($dU = 0$) are achieved only for equilibrium states of constant $X_1, \dots, X_{c+2}$ (i.e., closed adiabatic and isochoric conditions: $dS = dV =$

$dn_1 = \dots = dn_c = 0$). However, one may consider alternative “Legendre transforms” of U [cf. (5.45)], such as

$$L_i \equiv U - R_i X_i \quad (8.90a)$$

$$L_{i,j} \equiv U - R_i X_i - R_j X_j \quad (8.90b)$$

and so forth. From (8.90a), one can see, for example, that (with $X_1 = S$, $X_2 = V$)

$$dL_1 = dU - R_1 dX_1 - X_1 dR_1 = dU - T dS - S dT = -S dT - P dV + \dots = dA \quad (8.91)$$

so that

$$L_1 = L_1(R_1, X_2, \dots, x_{c+2}) = A \equiv U - TS \quad (8.92)$$

the “Helmholtz free energy” [cf. (5.36)]. Similarly, for L_2 ,

$$L_2 = L_2(X_1, R_2, X_3, \dots, X_{c+2}) = H \equiv U - (-P)V = U + PV \quad (8.93)$$

[“enthalpy”; cf. (3.49)], and for $L_{1,2}$,

$$L_{1,2} = L_{1,2}(R_1, R_2, X_3, \dots, X_{c+2}) = G \equiv U - TS + PV \quad (8.94)$$

[“Gibbs free energy”; cf. (5.40)], one obtains corresponding criteria of equilibrium under other conditions. Note that allowed Legendre transforms must have at least one remaining extensity in the argument list (as scale factor), because a “full” Legendre transform *vanishes* [cf. (8.86)]:

$$L_{1,2,\dots,c+2} = U - \sum_{i=1}^{c+2} X_i R_i = 0 \quad (8.95)$$

As Gibbs recognized, the analytical structure expressed by (8.70)–(8.95) is often best represented in graphical form. Indeed, the Euler criterion (8.80) guarantees that the fundamental equation (8.72) that underlies thermodynamic relationships is *graphable*, so that derivative equations based on this function can naturally be recast into graphical representations (cf. Fig. 1.1). Accurate recognition of how the abstract relationships (8.70)–(8.95) are manifested both in experimental context and in corresponding graphical depictions represents the central conceptual challenge to the student of Gibbsian thermodynamics.

METRIC GEOMETRY OF EQUILIBRIUM THERMODYNAMICS

Introduction to Vector Geometry and Metric Spaces

Although geometrical representations of propositions in the thermodynamics of fluids are in general use, and have done good service in disseminating clear notions in this science, yet they have by no means received the extension in respect to variety and generality of which they are capable.

—J. Willard Gibbs (1873)

Gibbs' first pair of published papers [*Trans. Conn. Acad.* **2**, 309, 382 (1873), from which the above opening sentence is taken] were entitled "Graphical Methods in the Thermodynamics of Fluids" and "A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces." These prescient papers scarcely hint at what was to come in Gibbs' *magnum opus* of 1876–78, but both accurately presage the deeply insightful exploitations of "graphical methods" and "geometric representation" that were to become a Gibbs trademark. Yet, despite its richly graphical and geometrical flavor, Gibbs' formulation does not yet capture a still deeper sense in which equilibrium thermodynamics is inherently "geometrical." (The concluding synopsis of Section 8.9 should be closely reviewed in preparation for Chapters 9 and 10.)

What is "geometry"? As its etymology suggests, geometry (*γεω-μετρία*: "Earth-measuring") is intimately associated with the quantification and characterization of spatial lengths, areas, and volumes of terrestrial landforms, as epitomized by Euclidean geometry. Such "metric" (measurement) geometry might be represented in terms of coordinates and axes that are rather arbitrarily chosen, but the *internal* relationships of its "points" are invariant to such arbitrary choices. In the metric geometry of a two-dimensional terrestrial surface, for example, the distances, angles, and areas of chosen surface points are *intrinsic*, not subject to arbitrary choices concerning their assigned coordinates. Such strong metric-type geometry is to be contrasted with weaker affine or topological "geometrical" concepts, which lack intrinsic metric significance.

A thermodynamic example may be illustrative. Consider Maxwell's model of the Gibbs *USV* surface for water (Fig. 1.1), as depicted schematically in Fig. 9.1. In this model, the physical (U , S , V) "coordinates" are associated with mutually perpendicular axes, and three chosen points on this surface form a triangle whose edges, angles, and area are as shown in Fig. 9.1a. However, the model might have been constructed (with equal thermodynamic justification) in a skewed *nonorthogonal* axis system (Fig. 9.1b) in which the

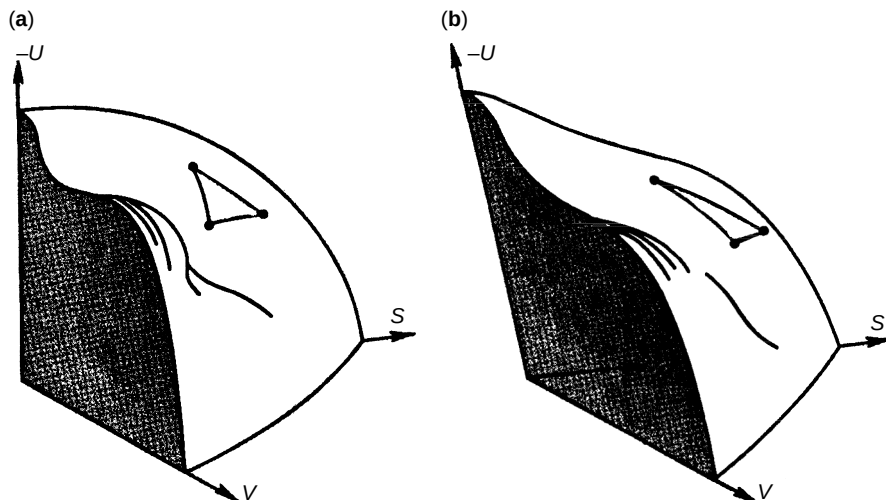


Figure 9.1 Schematic Gibbs-space USV model in orthogonal (a) and nonorthogonal (b) axis systems, showing distortion of surface metric properties (e.g., distances, angles in marked triangle) with arbitrary change of axes.

edges, angles, and area of the *same* three points *differ* from those of Fig. 9.1a. Either surface might be used to evaluate the derivatives and other thermodynamic relationships between states, and there appears to be no justification (other than mathematical convenience) for choosing one set of USV axes rather than another. This nonmetric aspect of Gibbs' original graphical representations was clearly recognized, and indeed it was claimed [L. Tisza. *Generalized Thermodynamics* (MIT Press, Cambridge MA, 1966), pp. 105, 235] that such concepts as orthogonality and metric cannot be defined in the thermodynamic Gibbs space. This claim is incorrect.

In the present context, we wish to employ geometrical constructs in a highly unusual and abstract fashion. For this purpose, it is convenient to employ *matrix algebra*, a mathematical framework that is sufficiently general to encompass description of both metric and nonmetric spaces. We shall first briefly introduce the general concepts and notation of matrix algebra (Section 9.1), particularly the “matrix representations” of *operators* and *vectors* in spaces of arbitrary dimension. We then introduce the powerful Dirac notation (Section 9.2) for these operators and vectors, anticipating that matrix algebra provides only one of several possible *equivalent* representations of the abstract geometry of the space. We finally introduce the general concept of *metric* (Section 9.3) and discuss the mathematical prerequisites (necessary and sufficient conditions) for its appearance in a general “spatial” framework, preparatory to the thermodynamic applications of Chapters 10–13. (Those wishing a preliminary glimpse of the “look and feel” of such strangely reformulated thermodynamic relationships may wish to glance forward to Fig. 11.2.)

As mentioned in the Preface, our goal in Part III is not merely to re-generate the material of Parts I and II (as summarized in Section 8.9) in new mathematical dress. We re-derive (rather trivially) many earlier thermodynamic identities and stability conditions to illustrate the geometrical techniques, but our primary emphasis is on thermodynamic extensions (particularly, to saturation properties, critical phenomena, multicomponent Gibbs–Konowalow-type relationships, higher-derivative properties, and general reversible changes

of state) that go considerably beyond Parts I and II. We also explore the statistical mechanical origins and near-equilibrium irreversible extensions of thermodynamic geometry, as well as the general statistical thermodynamical bridge to modern *ab initio* molecular and supramolecular interaction theory, touching on numerous topics of modern research beyond the scope of earlier chapters. Although the geometrical concepts needed to treat the basic thermodynamic problems of Parts I and II are rather elementary, we introduce the vector geometry in the more general framework of matrix-algebraic and eigenvalue concepts that prove useful in numerous areas of modern scientific and computational research beyond the advanced thermodynamic topics of Chapters 11–13.

9.1 VECTOR AND MATRIX ALGEBRA

“Matrix algebra,” as originally developed around 1845 by Sylvester and Cayley, provides a remarkably powerful and general mathematical framework for formulating physical relationships. The power and beauty of matrix-algebraic techniques was recognized by none other than J. Willard Gibbs, who advocated broader acceptance of “multiple algebra” (as he called it) in the scientific and mathematical curriculum [*Proc. Am. Assoc. Adv. Sci.* **35**, 37–66 (1886)]. Nowadays, matrix algebra underlies most technical applications of electronic computers and is considered an indispensable tool in practically every area of modern theoretical science. Hence, the student of thermodynamics, if not already familiar with matrix algebra, will find many rewards in mastering matrix concepts for the present purpose of understanding and applying the metric geometry of equilibrium thermodynamics, as well as for future scientific purposes. As Gibbs often observed, “the human mind has never invented a labor-saving machine equal to algebra.” *Matrix* algebra adds strong testimony to this truism.

To introduce the notation and concepts to be used below, let us first briefly recall some elementary aspects of the Euclidean geometry of a triangle of points $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ in ordinary three-dimensional physical space. Each point $\mathcal{P}_i$ can be “represented” by a *column vector* $\underline{v}_i$ (denoted with a single underbar) whose entries are the coordinates in a chosen Cartesian axis system at the origin of coordinates:

$$\underline{v}_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \quad \underline{v}_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}, \quad \underline{v}_3 = \begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} \quad (9.1)$$

Each vector has a direction (from the origin to the point) and length v_i , evaluated by the Pythagorean theorem as

$$v_i = |\underline{v}_i| = (x_i^2 + y_i^2 + z_i^2)^{1/2} \quad (9.2)$$

The vectors can be added and subtracted in arbitrary linear combinations to give new vectors $\underline{v}'$ (each with associated point $\mathcal{P}'$),

$$\underline{v}' = \sum_i c_i \underline{v}_i \quad (9.3)$$

the signature property of a “space.” Regardless of the arbitrariness in the choice of coordinate origin, lengths v_i , or individual coordinates (9.1) of these vectors, the intrinsic

geometry of each $\mathcal{P}_i\text{-}\mathcal{P}_j\text{-}\mathcal{P}_k$ triangle leads to invariance of such quantities as the connecting vector $\underline{d}_{ij}$ from $\mathcal{P}_j$ to $\mathcal{P}_i$,

$$\underline{d}_{ij} = \begin{pmatrix} x_i - x_j \\ y_i - y_j \\ z_i - z_j \end{pmatrix} \quad (9.4)$$

the distance d_{ij} between these points,

$$d_{ij} = d_{ji} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{1/2} \quad (9.5)$$

the angle θ_{i-j-k} at apex $\mathcal{P}_j$,

$$\cos \theta_{i-j-k} = \frac{(x_i - x_j)(x_k - x_j) + (y_i - y_j)(y_k - y_j) + (z_i - z_j)(z_k - z_j)}{d_{ij}d_{kj}} \quad (9.6)$$

the area of the $\mathcal{P}_i\text{-}\mathcal{P}_j\text{-}\mathcal{P}_k$ triangle, and so forth. Equations (9.2)–(9.6) of Euclidean 3-space are representative of surprising analogs to be encountered in the metric geometry of equilibrium thermodynamics.

Let us now see how such geometrical relationships can be expressed more generally and concisely in the language of matrices. Whereas ordinary algebra deals with individual numbers, matrix algebra deals with *tables of numbers* that are symbolized as a single algebraic “object” and manipulated with familiar arithmetic operations of addition, subtraction, etc., analogous in many respects to the corresponding operations of elementary algebra, as we now describe.

Each matrix $\underline{\underline{A}}$ (denoted with a double underbar) consists of a collection of numbers $\{a_{ij}\}$ (real or complex), labeled by row (i) and column (j) indices to indicate position in the table,

$$\underline{\underline{A}} = \{a_{ij}\} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n_c} \\ a_{21} & a_{22} & \cdots & a_{2n_c} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n_r1} & a_{n_r2} & \cdots & a_{n_rn_c} \end{pmatrix} \quad (9.7)$$

with n_r rows and n_c columns. If $\underline{\underline{A}} = \{a_{ij}\}$ and $\underline{\underline{B}} = \{b_{ij}\}$ denote $n_r \times n_c$ matrices, their *sum*, $\underline{\underline{A}} + \underline{\underline{B}}$, can be defined straightforwardly as

$$(\underline{\underline{A}} + \underline{\underline{B}})_{ij} = (\underline{\underline{A}})_{ij} + (\underline{\underline{B}})_{ij} = a_{ij} + b_{ij} \quad (9.8)$$

Similarly, the *scalar multiple* $\lambda \underline{\underline{A}}$ can be defined as

$$(\lambda \underline{\underline{A}})_{ij} = \lambda a_{ij} \quad (9.9)$$

for any scalar λ . The definitions (9.8), (9.9) allow one to straightforwardly define arbitrary *linear combinations* of matrices, such as

$$(\lambda \underline{\underline{A}} + \mu \underline{\underline{B}})_{ij} = \lambda a_{ij} + \mu b_{ij} \quad (9.10)$$

thereby establishing a simple “affine space” of matrices. For our present purposes, we shall usually consider the special case $n_r = n_c = f$ of *square* matrices of dimension f .

To further emulate ordinary scalar algebra, we require the operation of *matrix multiplication*. The product $\underline{\underline{AB}}$ of “conformable” matrices $\underline{\underline{A}}$, $\underline{\underline{B}}$ (i.e., with n_r of $\underline{\underline{B}}$ equal to n_c of $\underline{\underline{A}}$, as happens automatically for square matrices of dimension f) is defined by

$$\boxed{(\underline{\underline{AB}})_{ij} = \sum_k (\underline{\underline{A}})_{ik} (\underline{\underline{B}})_{kj}} \quad (9.11)$$

Whereas matrix addition (9.8) and scalar multiplication (9.9) have the usual associative and commutative properties of their scalar analogs, matrix multiplication (9.11), although associative [i.e., $\underline{\underline{A}}(\underline{\underline{B}}\underline{\underline{C}}) = (\underline{\underline{A}}\underline{\underline{B}})\underline{\underline{C}}$], is inherently *noncommutative* [i.e., $\underline{\underline{AB}} \neq \underline{\underline{BA}}$]. This noncommutativity leads to some of the most characteristic and surprising features of matrix algebra, and underlies the still more surprising matrix-algebraic features of quantum theory.

It is also possible to adapt the general matrix-algebraic operations (9.8)–(9.11) to describe the Euclidean geometry of (9.2)–(9.6). To do so, we note that each column vector $\underline{v}_i$ can be identified as a “matrix of one column” ($n_c = 1$), so that (9.3) becomes a special case of (9.10) to define a “space of column vectors.” We can now create an associated “space of row vectors” by defining, for any given column vector $\underline{v}$,

$$\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_f \end{pmatrix} \quad (9.12)$$

the associated *adjoint* vector $\underline{v}^\dagger$ (denoted by a dagger),

$$\underline{v}^\dagger = (v_1^* \quad v_2^* \quad \dots \quad v_f^*) \quad (9.13)$$

a “matrix of one row” (where asterisks denote complex conjugation). Definition (9.13) is a special case of the more general definition of *matrix adjoint* $\underline{\underline{A}}^\dagger$, namely

$$(\underline{\underline{A}}^\dagger)_{ij} \equiv (\underline{\underline{A}})_{ji}^* \quad (9.14)$$

the “complex-conjugate transpose” of $\underline{\underline{A}}$. Note that (9.13), (9.14) both provide for the general case of *complex* matrix elements, but in the remainder of this book we restrict attention to matrices with *real* matrix elements, so the complex-conjugation stars in (9.13), (9.14) are superfluous and “adjoint” is synonymous with “transpose” (interchange of rows and columns).

Note that the column vectors $\underline{v}_i$ and the adjoint row vectors $\underline{v}_i^\dagger$ “live” in *different* spaces, so it makes no sense, for example, to “add” $\underline{v}_i$ and $\underline{v}_i^\dagger$. However, according to the rule (9.11) of matrix multiplication, it makes perfect sense to *multiply* an adjoint $\underline{v}_i^\dagger$ row vector (with

elements $v_{i1}, v_{i2}, \dots, v_{if}$) and a $\underline{v}_j$ column vector (with elements $v_{1j}, v_{2j}, \dots, v_{fj}$) in either order:

$$\boxed{\underline{v}_i^\dagger \underline{v}_j = \sum_k v_{ik} v_{kj}} \quad (\text{a scalar}) \quad (9.15)$$

$$(\underline{v}_j \underline{v}_i^\dagger)_{rs} = v_{rj} v_{is} \quad (\text{a matrix}) \quad (9.16)$$

Equation (9.15) defines the “scalar product” (or “inner product,” “dot product”: $\underline{v}_i^\dagger \underline{v}_j$), and (9.16) defines the corresponding “dyadic product” (or “outer product”: $\underline{v}_j \underline{v}_i^\dagger$) of vectors $\underline{v}_i$ and $\underline{v}_j$, which are different *kinds* of mathematical object.

Whereas the scalar components v_{ki} of a given column vector $\underline{v}_i$ [labeled in accordance with the matrix convention (9.7)],

$$\underline{v}_i = \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{fi} \end{pmatrix} \quad (9.17)$$

and the corresponding components v_{jk} of an adjoint row vector $\underline{v}_j^\dagger$,

$$\underline{v}_j^\dagger = (v_{j1} \quad v_{j2} \quad \dots \quad v_{jf}) \quad (9.18)$$

have rather arbitrary values (dependent on the orientation of the Cartesian axis system), the scalars m_{ij} that result from a scalar product (9.15),

$$\boxed{m_{ij} = \underline{v}_i^\dagger \underline{v}_j} \quad (9.19)$$

are *intrinsic* to the space, independent of any orientation of coordinate system. A space $\mathcal{M}$ invested with the scalar product (9.19) therefore has intrinsic “metric structure” that distinguishes it from weaker affine-type manifolds that lack this structure. The Euclidean geometry of three-dimensional physical space is the prototype “metric geometry” of this type.

To see how the matrix-algebraic operations concisely recover the familiar Euclidean 3-space formulas, we need only note that (9.2)–(9.6) are expressed successively in matrix-algebraic language as:

$$v_i = |\underline{v}_i| = (\underline{v}_i^\dagger \underline{v}_i)^{1/2} \quad [\text{cf. (9.2)}] \quad (9.20a)$$

$$\underline{v}' = \sum_i c_i \underline{v}_i \quad [\text{cf. (9.3)}] \quad (9.20b)$$

$$\underline{d}_{ij} = \underline{v}_i - \underline{v}_j \quad [\text{cf. (9.4)}] \quad (9.20c)$$

$$d_{ij} = (\underline{d}_{ij}^\dagger \underline{d}_{ij})^{1/2} \quad [\text{cf. (9.5)}] \quad (9.20d)$$

$$\cos \theta_{i-j-k} = \frac{\underline{d}_{ij}^\dagger \underline{d}_{kj}}{d_{ij} d_{kj}} \quad [\text{cf. (9.6)}] \quad (9.20e)$$

Because the matrix-algebraic expressions can be easily extended to spaces of any dimension f , equations such as (9.20a–e) capture the essential mathematical structure of a space $\mathcal{M}$ in more general and incisive fashion. The use of a “matrix representation” of geometric points $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ of a Euclidean 3-space barely scratches the surface of possibilities inherent in matrix-algebraic equations such as (9.7)–(9.11). Further aspects of matrix algebra are outlined in Sidebar 9.1.

SIDEBAR 9.1: FURTHER ELEMENTS OF MATRIX ALGEBRA

It is important to recognize that the small subset of matrix equations introduced in the main text (typically, restricted to *real* matrix elements) will be found sufficient to exploit the geometrical simplicity that underlies equilibrium thermodynamics. Nevertheless, it is useful to introduce the thermodynamic vector geometry in the broader framework of matrix theory and Dirac notation that is broadly applicable to the advanced thermodynamic topics of Chapters 11–13, as well as to many other areas of modern physical chemistry research.

An “algebra” typically involves the operations of adding, subtracting, multiplying, or dividing the objects it describes, whether matrices or simple numbers. For completeness, we now summarize some other aspects of matrix algebra, built on the fundamental definitions of addition/subtraction (9.8), scalar multiplication (9.9), and matrix multiplication (9.11).

Two special elements of any algebra are the “zero” and “unit” elements. The “zero” $\underline{\underline{0}}$ of matrix algebra is defined simply as the matrix with all zero elements,

$$(\underline{\underline{0}})_{ij} = 0, \quad \text{all } i, j \quad (\text{S9.1-1})$$

and the “unit” (or “identity”) matrix $\underline{\underline{I}}$ is the matrix with 1’s on the “diagonal” ($i = j$) and 0’s otherwise,

$$(\underline{\underline{I}})_{ij} = \delta_{ij} \quad (\text{S9.1-2})$$

[where δ_{ij} is the Kronecker delta function: $\delta_{ij} = 1$ ($i = j$) or 0 ($i \neq j$)]. From the multiplication law (9.11), one can easily define positive powers $\underline{\underline{A}}^n$ of a matrix $\underline{\underline{A}}$ by repeated self-multiplication:

$$\underline{\underline{A}}^n = \underline{\underline{A}}^{n-1} \underline{\underline{A}} = \underline{\underline{A}} \underline{\underline{A}} \dots \underline{\underline{A}} \quad (n \text{ times}) \quad (\text{S9.1-3})$$

From these powers, one can also construct power series, including special functions such as the exponential function

$$\exp(\underline{\underline{A}}) = \sum_{n=0}^{\infty} \frac{\underline{\underline{A}}^n}{n!} \quad (\text{S9.1-4})$$

and so forth.

One can also define negative powers $\underline{\underline{A}}^{-n}$ starting from the “inverse” matrix $\underline{\underline{A}}^{-1}$, which is obtained as solution (either the “left inverse” $\underline{\underline{A}}_{\text{L}}^{-1}$ or “right inverse” $\underline{\underline{A}}_{\text{R}}^{-1}$) of the equations

$$\underline{\underline{A}}_{\text{L}}^{-1} \underline{\underline{A}} = \underline{\underline{I}} = \underline{\underline{A}} \underline{\underline{A}}_{\text{R}}^{-1} \quad (\text{S9.1-5})$$

Note that $\underline{\underline{A}}_{\text{L}}^{-1}$ and $\underline{\underline{A}}_{\text{R}}^{-1}$ need not be the same matrix (although they are for the important special case of real symmetric $\underline{\underline{A}}$ that we are most concerned with). Note also that $\underline{\underline{A}}^{-1}$ need not *exist*, even if $\underline{\underline{A}} \neq \underline{\underline{0}}$. A matrix for which $\underline{\underline{A}}^{-1}$ exists is called “nonsingular” (see below) and leads to many arithmetic extensions that are not permitted to singular matrices. The many varieties of singularity (not just $\underline{\underline{A}} = \underline{\underline{0}}$) and the (potentially) noncommutative aspects of multiplication distinguish matrix algebra from its scalar counterpart in interesting ways.

An important general concept of matrix algebra is the association of each matrix $\underline{\underline{A}}$ with a corresponding *adjoint* matrix $\underline{\underline{A}}^\dagger$, defined in (9.14):

$$(\underline{\underline{A}}^\dagger)_{ij} = (\underline{\underline{A}})^*_{ji} \quad (\text{S9.1-6})$$

As remarked in the text, the adjoint is synonymous with “transpose” (i.e., interchange of rows and columns by “flipping” the matrix around its diagonal) when $\underline{\underline{A}}$ is real. However, the ubiquity of complex numbers in physical applications usually requires us to distinguish the transpose $\underline{\underline{A}}^t$,

$$(\underline{\underline{A}}^t)_{ij} = (\underline{\underline{A}})_{ji} \quad (\text{S9.1-7})$$

from the adjoint $\underline{\underline{A}}^\dagger$, the latter serving as a unique “matrix-type complex conjugation.” Note that transposes, adjoints, and inverses all satisfy a characteristic product rule of the form

$$(\underline{\underline{AB}})^\dagger = \underline{\underline{B}}^\dagger \underline{\underline{A}}^\dagger \quad (\text{S9.1-8a})$$

$$(\underline{\underline{AB}})^t = \underline{\underline{B}}^t \underline{\underline{A}}^t \quad (\text{S9.1-8b})$$

$$(\underline{\underline{AB}})^{-1} = \underline{\underline{B}}^{-1} \underline{\underline{A}}^{-1} \quad (\text{S9.1-8c})$$

A matrix $\underline{\underline{A}}$ is further categorized according to whether it *commutes* with $\underline{\underline{A}}^\dagger$,

$$\underline{\underline{AA}}^\dagger = \underline{\underline{A}}^\dagger \underline{\underline{A}} \quad (\text{“normal”}) \quad (\text{S9.1-9})$$

or even *coincides* with its own adjoint,

$$\underline{\underline{A}}^\dagger = \underline{\underline{A}} \quad (\text{“self-adjoint”}) \quad (\text{S9.1-10})$$

The “self-adjoint” (“Hermitian”) matrices often play a particularly important role in representing physical phenomena (e.g., as the “observables” in quantum theory), and they also include the real symmetric matrices to be encountered in the metric geometry of equilibrium thermodynamics.

An important subset of normal matrices $\underline{\underline{U}}$ are the “unitary” matrices, which satisfy the property

$$\underline{\underline{UU}}^\dagger = \underline{\underline{U}}^\dagger \underline{\underline{U}} = \underline{\underline{I}} \quad (\text{S9.1-11})$$

i.e., with adjoint identical to inverse,

$$\underline{\underline{U}}^\dagger = \underline{\underline{U}}^{-1} \quad (\text{S9.1-12})$$

A unitary matrix may therefore be considered a kind of “square root of unity,” often complex-valued. Unitary matrices with all real elements are called “orthogonal” $\underline{\underline{O}}$, and satisfy a property analogous to (S9.1-12):

$$\underline{\underline{O}}^t = \underline{\underline{O}}^{-1} \quad (\text{S9.1-13})$$

(Note that unitary and orthogonal matrices are generally *non*-Hermitian, an exception being the trivial case $\underline{\underline{U}} = \underline{\underline{I}} = \underline{\underline{U}}^\dagger$, the identity matrix, which is also orthogonal.)

Unitary matrices allow one to transform matrices $\underline{\underline{A}}$ to closely related matrices $\underline{\underline{A}}'$ by “similarity transformations” of the form

$$\underline{\underline{A}}' = \underline{\underline{U}}\underline{\underline{A}}\underline{\underline{U}}^\dagger \quad (\text{S9.1-14})$$

Such transformations superficially change the individual elements of the matrix, but they leave invariant certain deep features of the matrix (the “eigenvalues”) that are its most important qualities. Matrices $\underline{\underline{A}}$, $\underline{\underline{A}}'$ that are related by (S9.1-14) (and thus share common eigenvalues) are indeed fundamentally “similar,” in a sense to be described below.

The concept of the *eigenvalue equation* associated with a matrix $\underline{\underline{A}}$ is perhaps the most profound extension of matrix algebra beyond the algebra of scalars. The eigenvalue equation for $\underline{\underline{A}}$ can generally be written as

$$\underline{\underline{A}}\underline{\underline{v}}_i = a_i\underline{\underline{v}}_i, \quad i = 1, 2, \dots, f \quad (\text{S9.1-15})$$

where each (column vector) $\underline{\underline{v}}_i$ is an “eigenvector” and each (scalar) a_i is an “eigenvalue” of $\underline{\underline{A}}$ (one for each “dimension” of $\underline{\underline{A}}$). Whereas multiplication of $\underline{\underline{A}}$ by an arbitrary vector $\underline{\underline{u}}$ (e.g., $\underline{\underline{A}}\underline{\underline{u}} = \underline{\underline{w}}$) usually gives a vector $\underline{\underline{w}}$ that differs from $\underline{\underline{u}}$ in both length and *direction*, the intrinsic *eigen* (“own,” “personal”) vectors $\underline{\underline{v}}_i$ “pass through” $\underline{\underline{A}}$ with only a change of *length* (by a_i), but not direction. In effect, each $\underline{\underline{v}}_i$ is an intrinsic “personal axis (direction)” of $\underline{\underline{A}}$ along which $\underline{\underline{A}}$ behaves like an ordinary *scalar multiplier* (by the eigenvalue a_i), with different “multiplicative strength” in different directions. For all normal $\underline{\underline{A}}$ satisfying (S9.1-9), these “eigen-axes” $\underline{\underline{v}}_i$ can be shown (or chosen, without loss of generality) to be *mutually perpendicular* [“orthogonal”; cf. (9.20e) for $\theta = 90^\circ$], and they can be chosen to be of *unit length* (“normalized”), namely,

$$\underline{\underline{v}}_i^\dagger \underline{\underline{v}}_j = (\underline{\underline{I}})_{ij} = \delta_{ij} \quad (\text{if } \underline{\underline{A}}, \underline{\underline{A}}^\dagger \text{ commute}) \quad (\text{S9.1-16})$$

completely analogous to the orthonormal unit vectors of a Cartesian axis system. Moreover, for self-adjoint matrices (S9.1-10), the “eigen-multipliers” a_i can be shown to be *real* numbers,

$$a_i = a_i^* \quad (\text{if } \underline{\underline{A}}^\dagger = \underline{\underline{A}}) \quad (\text{S9.1-17})$$

acting as “stretching” or “scale” factors for vectors oriented in proper (eigen) directions.

The eigenvectors $\{\underline{v}_i\}$ are also a “complete” set that “span the space,” in the sense that any vector $\underline{u}$ in the space can be written as a linear combination

$$\underline{u} = \sum_{i=1}^f c_i \underline{v}_i \quad (\text{any } \underline{u}) \quad (\text{S9.1-18})$$

By multiplying each side of (S9.1-18) by $\underline{v}_j^\dagger$ and using (S9.1-16), we evaluate the coefficients as

$$c_j = \underline{v}_j^\dagger \underline{u} \quad (\text{S9.1-19})$$

Equation (S9.1-18) can be rewritten as (putting the scalar coefficient after, rather than before, the vector it multiplies)

$$\underline{u} = \sum_{i=1}^f \underline{v}_i (\underline{v}_i^\dagger \underline{u}) \quad (\text{S9.1-20})$$

so that, in light of (S9.1-15),

$$\underline{\underline{A}} \underline{u} = \underline{w} = \sum_{i=1}^f \underline{v}_i (a_i \underline{v}_i^\dagger \underline{u}) \quad (\text{any } \underline{u}) \quad (\text{S9.1-21})$$

The eigenvalue equation (S9.1-15) therefore presents an intuitive geometrical picture of how a matrix $\underline{\underline{A}}$ “operates” on a general vector $\underline{u}$ by differentially “stretching” its components in different eigen-directions.

The deep connection between $\underline{\underline{A}}$ and its “spectrum” of eigenvalues $\{a_i\}$ and eigenvectors $\{\underline{v}_i\}$ is best exhibited by the “spectral theorem”

$$\underline{\underline{A}} = \sum_{i=1}^f a_i \underline{v}_i \underline{v}_i^\dagger \quad (\text{S9.1-22})$$

showing that knowledge of the eigenvalues $\{a_i\}$ and eigenvectors $\{\underline{v}_i\}$ is equivalent to *full* knowledge of $\underline{\underline{A}}$ itself. [Equation (S9.1-22) easily follows from (S9.1-21), which results from applying the matrix equation (S9.1-22) to an arbitrary $\underline{u}$.] If we perform the similarity transform (S9.1-14) on both sides of (S9.1-17), we see that [since $\underline{v}_i^\dagger \underline{\underline{U}}^\dagger = (\underline{\underline{U}} \underline{v}_i)^\dagger$, as may be easily verified]

$$\underline{\underline{A}}' = \underline{\underline{U}} \underline{\underline{A}} \underline{\underline{U}}^\dagger = \sum_{i=1}^f a_i (\underline{\underline{U}} \underline{v}_i) (\underline{\underline{U}} \underline{v}_i)^\dagger \quad (\text{S9.1-23})$$

which is just the spectral theorem for $\underline{\underline{A}}'$. This establishes that $\underline{\underline{A}}'$ satisfies an eigenvalue equation with *identical* eigenvalues a_i and *similar* eigenvectors $\underline{v}_i' = \underline{\underline{U}} \underline{v}_i$ to those of $\underline{\underline{A}}$, the latter merely being *rotated* by $\underline{\underline{U}}$ (i.e., as though represented in a rotated coordinate system). Thus, matrices related by a similarity transformation (S9.1-14) are representing essentially the *same* underlying “operator,” but in different (arbitrarily chosen) coordinate systems.

The spectral theorem can also be used to express many “functions” of $\underline{\underline{A}}$, by recognizing that all powers of $\underline{\underline{A}}$ have the same eigenvectors as $\underline{\underline{A}}$ and the associated eigenvalues are equivalent functions of the a_i :

$$f(\underline{\underline{A}}) = \sum_{i=1}^f f(a_i) \underline{\underline{v}}_i \underline{\underline{v}}_i^\dagger \quad (\text{S9.1-24})$$

Thus, for example, the inverse matrix $\underline{\underline{A}}^{-1}$ can be evaluated (if no $a_i = 0$) as

$$\underline{\underline{A}}^{-1} = \sum_{i=1}^f a_i^{-1} \underline{\underline{v}}_i \underline{\underline{v}}_i^\dagger \quad (\text{S9.1-25})$$

the exponential operator $\exp(\underline{\underline{A}})$ [cf. (S9.1-4)] can be evaluated as

$$\exp(\underline{\underline{A}}) = \sum_{i=1}^f e^{a_i} \underline{\underline{v}}_i \underline{\underline{v}}_i^\dagger \quad (\text{S9.1-26})$$

and other matrix relationships can be established with this theorem.

The powerful concepts of matrix algebra can also be further extended to “partitioned” matrices, whose elements are themselves *matrices* rather than scalars [cf. (9.7)]:

$$\underline{\underline{A}} = \{\underline{\underline{a}}_{ij}\} = \begin{pmatrix} \underline{\underline{a}}_{11} & \underline{\underline{a}}_{12} & \cdots & \underline{\underline{a}}_{1n_c} \\ \underline{\underline{a}}_{21} & \underline{\underline{a}}_{22} & \cdots & \underline{\underline{a}}_{2n_c} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\underline{a}}_{n_r1} & \underline{\underline{a}}_{n_r2} & \cdots & \underline{\underline{a}}_{n_r n_c} \end{pmatrix} \quad (\text{S9.1-27})$$

So long as dimensional conformability is maintained, such “super-matrices” (matrices of matrices) obey additive, scalar-multiplicative, and matrix-multiplicative equations analogous to (9.8)–(9.11), such as

$$(\underline{\underline{A}}\underline{\underline{B}})_{ij} = \sum_k \underline{\underline{a}}_{ik} \underline{\underline{b}}_{kj} \quad (\text{S9.1-28})$$

providing a rich mathematical framework for representing diverse physical phenomena (including quantum mechanics).

9.2 DIRAC NOTATION

The matrix-algebraic “representation” (9.20a–e) of Euclidean geometrical relationships has both conceptual and notational drawbacks. On the conceptual side, the introduction of an arbitrary Cartesian axis system (or alternatively, of an arbitrarily chosen set of “basis vectors”) to provide “vector representations” $\underline{\underline{v}}_j$ of geometric points $\mathcal{P}$ seems to detract from the intrinsic geometrical properties of the points themselves. On the notational side, typographical resources are strained by the need to carefully distinguish various *types* of

mathematical objects by distinctive symbols. For example, for given point $\mathcal{P}_i$ we may need to distinguish its symbolic acronym “name” P (Roman), its scalar numerical value P (italic), its associated Cartesian vector $\underline{P}$ (single underbar), or an associated dyadic matrix $\underline{\underline{P}}$ (double underbar). Both drawbacks call for a more effective notation that focuses on deeper (intrinsic and invariant) relationships between the $\mathcal{P}_i$ themselves, rather than details of their numerical representation.

Physicist P. A. M. Dirac suggested an inspired notation for the *Hilbert space* of quantum mechanics [essentially, the Euclidean space of (9.20a, b) for $f \rightarrow \infty$, which introduces some subtleties not required for the finite-dimensional thermodynamic geometry]. Dirac’s notation applies equally well to *matrix* equations [such as (9.7)–(9.19)] and to *differential* equations [such as Schrödinger’s equation] that relate “operators” (mathematical objects that change functions or vectors of the space) and “wavefunctions” in quantum theory. Dirac’s notation shows explicitly that the disparate-looking “matrix mechanical” vs. “wave mechanical” representations of quantum theory are actually *equivalent*, by exhibiting them in unified symbols that are free of the extraneous details of a particular mathematical representation. Dirac’s notation can also help us to recognize such commonality in alternative mathematical representations of equilibrium thermodynamics.

The essence of Dirac’s notation is to distinguish different types of mathematical objects by *enclosures* around symbols, rather than differences in the symbols themselves. We can associate different enclosure “brackets” (or portions thereof) with scalars, vectors (ordinary or adjoint), and matrices (operators) as shown in Table 9.1.

To further illustrate Dirac notation for some simple formulas in Euclidean 3-space, we can rewrite analogs of (9.20a–e) in Dirac notation, all in terms of underlying Dirac objects $|\mathbf{v}_i\rangle$ (using boldface symbols to stress the association with ordinary vectors):

$$v_i = \langle \mathbf{v}_i | \mathbf{v}_i \rangle \quad (\text{number}) \quad (9.21a)$$

$$|\mathbf{v}'\rangle = \sum_i c_i |\mathbf{v}_i\rangle \quad (\text{linear combination vector}) \quad (9.21b)$$

$$|\mathbf{v}_i - \mathbf{v}_j\rangle = |\mathbf{v}_i\rangle - |\mathbf{v}_j\rangle \quad (\text{difference vector}) \quad (9.21c)$$

$$d_{ij} = \langle \mathbf{v}_i - \mathbf{v}_j | \mathbf{v}_i - \mathbf{v}_j \rangle^{1/2} \quad (\text{number}) \quad (9.21d)$$

$$\cos \theta_{i-j-k} = \frac{\langle \mathbf{v}_i - \mathbf{v}_j | \mathbf{v}_k - \mathbf{v}_j \rangle}{d_{ij}d_{kj}} \quad (\text{ratio of numbers}) \quad (9.21e)$$

TABLE 9.1 Dirac Enclosure Symbols for Different Types of Mathematical Object, with Matrix-Algebraic Examples

Object	Enclosures	Matrix-Algebraic Example ^a
“ket” vector	$ \rangle$	$ \mathbf{v}\rangle \longleftrightarrow$ column vector (9.12)
“bra” vector	$\langle $	$\langle \mathbf{v} \longleftrightarrow$ adjoint row vector (9.13)
“bra–ket” scalar	$\langle $	$\langle \mathbf{v} \mathbf{v} \rangle \longleftrightarrow$ scalar product (9.15)
matrix operator	$ $	$ \mathbf{v}\rangle \langle \mathbf{v} \longleftrightarrow$ dyadic matrix (9.16)

^aNote that the double-headed arrow ($\longleftrightarrow$) means “maps onto” (rather than “equals”) because the referred-to equation is only one way of “representing” the deeper Dirac mathematical object.

Use of Dirac notation allows us to recognize at a glance that $|\mathbf{v}\rangle$ is a column vector, $\langle\mathbf{v}|$ is the adjoint row vector, $\langle\mathbf{v}|\mathbf{v}\rangle$ is the scalar product of these two vectors, and $|\mathbf{v}\rangle\langle\mathbf{v}|$ is a corresponding matrix dyadic, all referring to underlying object $\mathbf{v}$. Further examples of Dirac notation are shown in Sidebar 9.2.

SIDEBAR 9.2: DIRAC OPERATOR ALGEBRA

The basic concept of Dirac notation is that a matrix $\underline{\underline{A}}$ is only a “representation” of an underlying *operator* $\hat{A}$ in a chosen set of “basis vectors” $\{|\mathbf{u}_i\rangle\}$ (written as kets)

$$\underline{\underline{A}} \longleftrightarrow \hat{A} \text{ in basis } \{|\mathbf{u}_i\rangle\} \quad (\text{S9.2-1})$$

with matrix elements

$$(\underline{\underline{A}})_{ij} = \langle\mathbf{u}_i|\hat{A}|\mathbf{u}_j\rangle \quad (\text{S9.2-2})$$

For convenience, we take the basis vectors to be orthonormal,

$$\langle\mathbf{u}_i|\mathbf{u}_j\rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (\text{S9.2-3})$$

and the operator to be Hermitian ($\hat{A} = \hat{A}^\dagger$), so that the matrix equation (S9.1-6) becomes, in Dirac notation,

$$\langle\mathbf{u}_i|\hat{A}^\dagger|\mathbf{u}_j\rangle = \langle\mathbf{u}_j|\hat{A}|\mathbf{u}_i\rangle^* \quad (\text{“turn-over rule”}) \quad (\text{S9.2-4})$$

For the special case $\hat{A} = \hat{I}$ (the unit operator), (S9.2-4) becomes

$$\langle\mathbf{u}_i|\mathbf{u}_j\rangle = \langle\mathbf{u}_j|\mathbf{u}_i\rangle^* \quad (\text{S9.2-5})$$

the fundamental rule that reversing the bra and ket vectors gives the complex conjugate scalar product value.

Problem Write the eigenvalue problem (S9.1-15) in Dirac notation, and prove the fundamental properties (S9.1-16), (S9.1-17) of the eigenvectors and eigenvalues for Hermitian $\hat{A}$, assuming the eigenvalues are nondegenerate (unequal).

Solution The eigenvalue problem for $\hat{A}$ corresponding to eigenvalue a_i is

$$\hat{A}|\mathbf{v}_i\rangle = a_i|\mathbf{v}_i\rangle \quad (\text{S9.2-6a})$$

and that corresponding to a_j is

$$\hat{A}|\mathbf{v}_j\rangle = a_j|\mathbf{v}_j\rangle \quad (\text{S9.2-6b})$$

If we multiply (S9.2-6a) on the left by $\langle\mathbf{v}_j|$ and (S9.2-6b) by $\langle\mathbf{v}_i|$, we obtain

$$\langle\mathbf{v}_j|\hat{A}|\mathbf{v}_i\rangle = a_i\langle\mathbf{v}_j|\mathbf{v}_i\rangle \quad (\text{S9.2-7a})$$

$$\langle\mathbf{v}_i|\hat{A}|\mathbf{v}_j\rangle = a_j\langle\mathbf{v}_i|\mathbf{v}_j\rangle \quad (\text{S9.2-7b})$$

Taking the complex conjugate of (S9.2-7a), we obtain

$$\langle \mathbf{v}_j | \hat{A} | \mathbf{v}_i \rangle^* = a_i^* \langle \mathbf{v}_j | \mathbf{v}_i \rangle^* \quad (\text{S9.2-7c})$$

or, in view of (S9.2-4) and (S9.2-5),

$$\langle \mathbf{v}_i | \hat{A} | \mathbf{v}_j \rangle = a_i^* \langle \mathbf{v}_i | \mathbf{v}_j \rangle \quad (\text{S9.2-7d})$$

Subtracting (S9.2-7d) from (S9.2-7b), we obtain

$$(a_j - a_i^*) \langle \mathbf{v}_i | \mathbf{v}_j \rangle = 0, \quad \text{all } i, j \quad (\text{S9.2-8})$$

For the case $i = j$, Eq. (S9.2-8) becomes

$$(a_i - a_i^*) \langle \mathbf{v}_i | \mathbf{v}_i \rangle = 0 \quad (\text{S9.2-9})$$

and, since $\langle \mathbf{v}_i | \mathbf{v}_i \rangle \neq 0$ (and may be chosen as $\langle \mathbf{v}_i | \mathbf{v}_i \rangle = 1$ for convenience), we conclude that

$$a_i = a_i^* \quad (\text{S9.2-10})$$

[cf. (S9.1-17)], the reality property of Hermitian eigenvalues. We can therefore drop the superfluous complex conjugation to rewrite (S9.2-8) as

$$(a_j - a_i) \langle \mathbf{v}_i | \mathbf{v}_j \rangle = 0, \quad \text{all } i, j \quad (\text{S9.2-11})$$

Then, under the stated assumption of nondegeneracy, $a_i \neq a_j$, we conclude that

$$\langle \mathbf{v}_i | \mathbf{v}_j \rangle = 0 \quad (a_i \neq a_j) \quad (\text{S9.2-12})$$

the orthogonality of Hermitian eigenvectors, as claimed in (S9.1-16)—*QED*.

(Note that eigenvector orthonormality can also be safely assumed for the degenerate case, $a_i = a_j$, without loss of generality.)

In dealing with abstract vector or operator algebra, it is necessary to clarify the meaning of algebraic “equality” in equations such as

$$|\mathbf{v}\rangle = |\mathbf{w}\rangle \quad (\text{vector equality}) \quad (\text{S9.2-13})$$

$$\hat{A} = \hat{B} \quad (\text{operator equality}) \quad (\text{S9.2-14})$$

In general, we can relate vector or operator equality to ordinary scalar equality by requiring that (S9.2-13), (S9.2-14) remain true when multiplied by *any* vector $|\mathbf{u}\rangle$ (or $\langle \mathbf{u}|$, as appropriate) of the space. Thus, we say that the vector equation (S9.2-13) is true if (and only if)

$$\langle \mathbf{u} | \mathbf{v} \rangle = \langle \mathbf{u} | \mathbf{w} \rangle \quad (\text{for all } \langle \mathbf{u} |) \quad (\text{S9.2-15})$$

and that the operator equation (S9.2-14) is true if (and only if)

$$\hat{A}|\mathbf{u}\rangle = \hat{B}|\mathbf{u}\rangle \quad (\text{for all } |\mathbf{u}\rangle) \quad (\text{S9.2-16})$$

The following problem gives practice in Dirac operator equations.

Problem Write (S9.1-21) in Dirac form, and show that this equation implies both (S9.1-15) and (S9.1-22) in Dirac form.

Solution In Dirac form, (S9.1-21) is

$$\hat{A}|\mathbf{u}\rangle = \sum_{i=1}^f |\mathbf{v}_i\rangle a_i \langle \mathbf{v}_i | \mathbf{u} \rangle \quad (\text{any } |\mathbf{u}\rangle) \quad (\text{S9.2-17})$$

In view of (S9.2-16), this means, in operator-equation form,

$$\hat{A} = \sum_{i=1}^f |\mathbf{v}_i\rangle a_i \langle \mathbf{v}_i| \quad (\text{S9.2-18})$$

which is (S9.1-22), the spectral theorem for $\hat{A}$. Note that it is “legal” to move the scalar a_i to either side of a bra or ket vector (because scalars have commutative multiplication with vectors). Note also that the outermost enclosure symbols on the right-hand side of (S9.2-18) are $|\dots\rangle$ (cf. Table 9.1), implying that both sides of the equation are indeed *operator-valued*, as must be true for equation consistency.

If we multiply (S9.2-18) on the right by any chosen $|\mathbf{v}_j\rangle$, we obtain

$$\hat{A}|\mathbf{v}_j\rangle = \sum_{i=1}^f |\mathbf{v}_i\rangle a_i \langle \mathbf{v}_i | \mathbf{v}_j \rangle \quad (\text{S9.2-19})$$

which, in view of the orthonormality of eigenvectors [cf. (S9.1-16)],

$$\langle \mathbf{v}_i | \mathbf{v}_j \rangle = \delta_{ij} \quad (\text{S9.2-20})$$

becomes

$$\hat{A}|\mathbf{v}_j\rangle = \sum_{i=1}^f |\mathbf{v}_i\rangle a_i \delta_{ij} = |\mathbf{v}_j\rangle a_j \quad (\text{S9.2-21})$$

equivalent to the original eigenvalue equation (S9.1-15)—*QED*.

It is also useful to note the special form of the spectral theorem (S9.2-18) for $\hat{A} = \hat{I}$, the identity operator (with all $a_i = 1$),

$$\hat{I} = \sum_{i=1}^f |\mathbf{v}_i\rangle \langle \mathbf{v}_i| \quad (\text{S9.2-22})$$

which is called the “resolution of the identity” in terms of any complete orthonormal set $\{|\mathbf{v}_i\rangle\}$. If we multiply both sides of (S9.2-22) on the right by an arbitrary ket vector $|\mathbf{u}\rangle$, we obtain the general formula for expanding $|\mathbf{u}\rangle$ in the complete orthonormal set $\{|\mathbf{v}_i\rangle\}$ [cf. (S9.1-18) or (S9.1-20)]:

$$|\mathbf{u}\rangle = \sum_{i=1}^f |\mathbf{v}_i\rangle \langle \mathbf{v}_i | \mathbf{u} \rangle \quad (\text{S9.2-23})$$

It will be useful practice for the physical chemistry student to rewrite other matrix and vector equations of Sidebar 9.1 in Dirac notation, both for future applications to quantum theory as well as the intended present application to equilibrium thermodynamics.

9.3 METRIC SPACES

Euclidean geometry was originally deduced from Euclid's five axioms. However, it is now known that necessary and sufficient criteria for Euclidean spatial structure can be stated succinctly in terms of *distances*, *angles*, and *triangles*, or, alternatively, the *scalar product* of the space. We can express these criteria by employing Dirac notation for abstract ket vectors $|\mathbf{R}_i\rangle$ of a given space $\mathcal{M}$ with scalar product $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$.

Suppose that

$$r_i = \langle \mathbf{R}_i | \mathbf{R}_i \rangle^{1/2}, \quad r_j = \langle \mathbf{R}_j | \mathbf{R}_j \rangle^{1/2}, \quad r_k = \langle \mathbf{R}_k | \mathbf{R}_k \rangle^{1/2} \quad (9.22)$$

are the edges of any triangle in the space $\mathcal{M}$. Then $\mathcal{M}$ is a Euclidean space if, and only if, these edges satisfy the *triangle inequality*

$$r_i + r_j \geq r_k \quad (9.23)$$

for all possible triangles in $\mathcal{M}$.

Alternatively (and equivalently), we can say that if $|\mathbf{R}_i\rangle, |\mathbf{R}_j\rangle$ are any two vectors in the space, with scalar product $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$, then $\mathcal{M}$ is a Euclidean space if, and only if, they satisfy the *Schwarz inequality*

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle^2 \leq \langle \mathbf{R}_i | \mathbf{R}_i \rangle \langle \mathbf{R}_j | \mathbf{R}_j \rangle \quad (9.24)$$

for all possible vectors in $\mathcal{M}$.

We can also express (9.24) in terms of the *angle* θ_{ij} between vectors $|\mathbf{R}_i\rangle, |\mathbf{R}_j\rangle$, as defined by

$$\cos \theta_{ij} = \frac{\langle \mathbf{R}_i | \mathbf{R}_j \rangle}{(\langle \mathbf{R}_i | \mathbf{R}_i \rangle \langle \mathbf{R}_j | \mathbf{R}_j \rangle)^{1/2}} \quad (9.25)$$

In terms of (9.25), the Schwarz inequality (9.24) is equivalent to

$$-1 \leq \cos \theta_{ij} \leq +1 \quad (9.26)$$

for all possible angles in $\mathcal{M}$.

The criteria (9.23), (9.24), and (9.26) are all rather obvious properties of Euclidean geometry. All of these properties can be traced back to mathematical properties of the *scalar product* $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$, the key “structure-maker” of a metric space. We therefore wish to determine whether a proposed definition of scalar product satisfies these criteria, and thus guarantees that $\mathcal{M}$ is a Euclidean space.

The essential mathematical requirements for a Euclidean scalar product can be stated as follows (for all possible vectors $|\mathbf{R}_i\rangle, |\mathbf{R}_j\rangle, |\mathbf{R}_k\rangle$ of $\mathcal{M}$):

distributive: $\langle \mathbf{R}_i \lambda \mathbf{R}_j + \mu \mathbf{R}_k \rangle = \lambda \langle \mathbf{R}_i \mathbf{R}_j \rangle + \mu \langle \mathbf{R}_i \mathbf{R}_k \rangle$ (all λ, μ)	(9.27a)
symmetric: $\langle \mathbf{R}_i \mathbf{R}_j \rangle = \langle \mathbf{R}_j \mathbf{R}_i \rangle$	(9.27b)
positive: $\langle \mathbf{R}_i \mathbf{R}_i \rangle \geq 0$ (= 0 only if $\mathbf{R}_i \equiv 0$)	(9.27c)

The proof that the criteria (9.27a–c) are indeed equivalent to the Schwarz inequality (9.24), and thus to the other criteria (9.23), (9.26) for a Euclidean space, is sketched in Sidebar 9.3.

The notion of spatial “metric” in $\mathcal{M}$ can be expressed in another form. Suppose that $\xi_1, \xi_2, \dots, \xi_f$ are “coordinates” in $\mathcal{M}$ with respect to chosen axis vectors $|\mathbf{R}_1\rangle, |\mathbf{R}_2\rangle, \dots, |\mathbf{R}_f\rangle$. Then the length of any small displacement ds in the space can be expressed with respect to the corresponding coordinate changes $\{d\xi_i\}$ as

$$ds = \left(\sum_{i,j=1}^f M_{ij} d\xi_i d\xi_j \right)^{1/2} \quad (9.28)$$

where M_{ij} are the elements of the *metric matrix*

$$M_{ij} = (\underline{\mathbf{M}})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle, \quad i, j = 1, 2, \dots, f \quad (9.29)$$

defined by the scalar products $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$. If the metric matrix elements satisfy (9.27a, b), then $\underline{\mathbf{M}}$ is a *Euclidean metric*, and $\mathcal{M}$ is equivalent to a Euclidean space of f dimensions. Indeed, one can see that if the $|\mathbf{R}_i\rangle$ are chosen, as usual, as mutually perpendicular Cartesian unit vectors in 3-space, satisfying

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, 3 \quad (9.30)$$

[where δ_{ij} is the Kronecker delta function: $\delta_{ij} = 1$ ($i = j$) or 0 ($i \neq j$)] then (9.28) reduces to the familiar expression for an infinitesimal line segment in three-dimensional Euclidean geometry.

The general line-element expression (9.28) allows one to envision possible “geometries” with *non*-Euclidean metric [i.e., failing to satisfy one or more of the conditions (9.27a–c)] or with *variable* metric [i.e., with a matrix $\underline{\mathbf{M}}$ that *varies* with position $\{\xi_i\}$ in the space, $\underline{\mathbf{M}} = \underline{\mathbf{M}}(\{\xi_i\})$, a “Riemannian geometry” that is only “locally” Euclidean; cf. Section 13.1]. However, for the present equilibrium thermodynamic purposes (Chapters 9–12) we may consider only the simplest version of (9.28), with metric elements $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ satisfying the Euclidean requirements (9.27a–c).

The extensions of “geometry” considered here may seem daunting, especially when expressed in the somewhat unusual Dirac notation. However, it is important to recognize that the basic concepts and equations [such as (9.20a–e) or (9.21a–e)] required for the thermodynamic applications in later chapters were probably encountered in high-school geometry or trigonometry class. Thus, the necessary mathematical tools to master Chapters 10–12 were familiar long before you began studies in physical chemistry, and should present no additional barrier to conceptual reformulation of equilibrium thermodynamics in Euclidean geometrical terms.

SIDEBAR 9.3: CRITERIA FOR EUCLIDEAN SCALAR PRODUCT

Problem Prove that the mathematical criteria (9.27a–c) for a proposed “scalar product” $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ are sufficient for the general Schwarz inequality (9.24) in the space $\mathcal{M}$, thereby guaranteeing that $\mathcal{M}$ is *Euclidean*.

Proof Consider any two vectors $|\mathbf{R}_i\rangle, |\mathbf{R}_j\rangle$ of $\mathcal{M}$ that are not linearly dependent, i.e.,

$$|\mathbf{R}_i\rangle \neq \lambda |\mathbf{R}_j\rangle \quad (\text{for any } \lambda) \quad (\text{S9.3-1})$$

Then $|\mathbf{v}\rangle = |\mathbf{R}_i - \lambda\mathbf{R}_j\rangle$ is a nonzero vector whose norm (squared length) $\langle\mathbf{v}|\mathbf{v}\rangle$ is non-negative by (9.27c):

$$0 \leq \langle\mathbf{v}|\mathbf{v}\rangle = \langle\mathbf{R}_i - \lambda\mathbf{R}_j|\mathbf{R}_i - \lambda\mathbf{R}_j\rangle \quad (\text{S9.3-2})$$

Using the distributive and symmetric properties (9.27a, b), we obtain

$$0 \leq \langle\mathbf{R}_i|\mathbf{R}_i\rangle + \lambda^2\langle\mathbf{R}_j|\mathbf{R}_j\rangle - 2\lambda\langle\mathbf{R}_i|\mathbf{R}_j\rangle \quad (\text{S9.3-3})$$

Let us *minimize* the right-hand side of (S9.3-3) to make the inequality as tight as possible. Differentiating the right-hand side with respect to λ and setting the result equal to zero,

$$\frac{d(\text{r.h.s.})}{d\lambda} = 2\lambda\langle\mathbf{R}_j|\mathbf{R}_j\rangle - 2\langle\mathbf{R}_i|\mathbf{R}_j\rangle = 0 \quad (\text{S9.3-4})$$

we obtain the minimizing value $\lambda_{\min}$ as

$$\lambda_{\min} = \frac{\langle\mathbf{R}_i|\mathbf{R}_i\rangle}{\langle\mathbf{R}_j|\mathbf{R}_j\rangle} \quad (\text{S9.3-5})$$

Substituting $\lambda = \lambda_{\min}$ into (S9.3-3), we obtain the *strongest possible* form of the inequality,

$$0 \leq \langle\mathbf{R}_i|\mathbf{R}_i\rangle - \frac{\langle\mathbf{R}_i|\mathbf{R}_i\rangle^2}{\langle\mathbf{R}_j|\mathbf{R}_j\rangle} \quad (\text{S9.3-6})$$

which is the Schwarz inequality (9.24)—*QED*.

Aside Although we focus here on the case where all quantities in (9.27a–c) are *real*, the proof given above can be readily generalized to a complex-valued scalar product, with complex symmetric

$$\langle\mathbf{R}_i|\mathbf{R}_j\rangle = \langle\mathbf{R}_j|\mathbf{R}_i\rangle^* \quad (\text{S9.3-7})$$

in place of (9.27b). This establishes that *complex*-valued thermodynamic responses (e.g., of electromagnetic type) could make perfect sense in the metric geometrical framework to be established in Chapter 10. [For further aspects of metric spaces, see P. Dennery and A. Kryzwicki. *Mathematics for Physicists* (Harper & Row, New York, 1967), Chapter 2.]

Metric Geometry of Thermodynamic Responses

10.1 THE SPACE OF THERMODYNAMIC RESPONSE VECTORS

In what sense might a thermodynamic variable be considered as a “vector,” or the associated equilibrium state as a “geometrical space” of vectors?

In characterizing a thermodynamic state variable (e.g., temperature, pressure, or energy), we may choose to focus on either of two distinctive aspects of the variable: (i) its numerical *value*, which may serve as a scalar descriptor or identifier of the state, and (ii) its *variability* or “responsiveness” to some “stimulus” (i.e., an attempted change of a chosen control variable of the state). These distinct aspects may require distinct mathematical objects and symbols for proper description.

In the calculus-based description of thermodynamic phenomena, these two aspects of a variable are distinguished as the value (F) versus the variability (dF) of an underlying function $[F(\dots)]$, for example,

$$F(X_1, X_2, \dots, X_i, \dots) \quad (10.1)$$

where $\{X_i\}$ are chosen control parameters of the system under discussion. Whereas “value” is single-valued (scalar-like), “variability” or “responsiveness” has intrinsic *multivalued* (vector-like) character (“response to *what?*”). This multivariate quality is seen most clearly in the chain rule for dF ,

$$dF = \left(\frac{\partial F}{\partial X_1} \right)_{\underline{X}} dX_1 + \left(\frac{\partial F}{\partial X_2} \right)_{\underline{X}} dX_2 + \dots + \left(\frac{\partial F}{\partial X_i} \right)_{\underline{X}} dX_i + \dots \quad (10.2)$$

wherein we express the multiple possible “partial” contributions to dF from a specified control parameter X_i . This multivariate character signals an underlying vector-like aspect of thermodynamic responses.

The calculus-based symbolism (10.1), (10.2) provides dependable, if unwieldy, descriptions of thermodynamic responses. The differential aspect (10.2) often transcends the functional aspect (10.1) for thermodynamic purposes, and, indeed, special criteria may be invoked to test whether an underlying function (10.1) *exists* for measured differential

responses (10.2). Moreover, it is remarkable that the “laws of thermodynamics” and their associated domain of description exhibit explicit dependence only on a *single* underlying function [the fundamental equation (8.72)], and then only to *first and second derivatives* of this function. Indeed (as described in Section 12.6), mathematical assumptions concerning the existence (or other specifics) of third or higher derivatives of (8.72) commonly come into conflict with experiments. Accordingly, the full calculus-based machinery of differentiable functions seems somewhat ill-fitted to the mathematical needs of thermodynamic science.

Let us consider instead the possibility of a geometry-based description. In this framework, a thermodynamic variable is identified with a “vector” of the geometrical “space,” designated $\mathcal{M}_S$, describing the equilibrium state S of interest. To distinguish this vector aspect notationally, we employ the Dirac ket symbol $|\mathbf{F}\rangle$ to denote the abstract geometrical vector that “corresponds to” (double-headed arrow) the calculus-based descriptors of the variable,

$$|\mathbf{F}\rangle \longleftrightarrow \{dF, F(\dots)\} \quad (10.3)$$

A key mathematical requirement for “space” is the concept of *linear combination*, i.e., the requirement that if $|\mathbf{F}_1\rangle$ and $|\mathbf{F}_2\rangle$ are in the space ($|\mathbf{F}_1\rangle, |\mathbf{F}_2\rangle \in \mathcal{M}_S$), then any linear combination with real coefficients λ_1, λ_2 is also in the space:

$$\lambda_1|\mathbf{F}_1\rangle + \lambda_2|\mathbf{F}_2\rangle \in \mathcal{M}_S \quad (\text{for all } |\mathbf{F}_1\rangle, |\mathbf{F}_2\rangle \in \mathcal{M}_S) \quad (10.4)$$

This property is automatically guaranteed by the chain rule (10.2),

$$d(\lambda_1 F_1 + \lambda_2 F_2) = \lambda_1 dF_1 + \lambda_2 dF_2 = \sum_i \left[\lambda_1 \left(\frac{\partial F_1}{\partial X_i} \right)_{\underline{X}} + \lambda_2 \left(\frac{\partial F_2}{\partial X_i} \right)_{\underline{X}} \right] dX_i \quad (10.5)$$

based on the linear character of first-order differentials.

Given the property (10.4), one can identify the *dimension* f and a suitable set of *basis vectors* ($\{|\mathbf{R}_i\rangle\}$, $i = 1, 2, \dots, f$) to represent any possible vector in $\mathcal{M}_S$:

$$|\mathbf{F}\rangle = \sum_{i=1}^f \lambda_i |\mathbf{R}_i\rangle \quad (\text{any } |\mathbf{F}\rangle \in \mathcal{M}_S) \quad (10.6)$$

For thermodynamic purposes, the dimensionality of a system with c independent chemical components and p distinct phases is determined by the Gibbs phase rule (8.89):

$$f = c + 2 - p \quad (10.7)$$

The basis vectors $|\mathbf{R}_i\rangle$ are most conveniently chosen to correspond to *intensities* R_i given by (8.73),

$$|\mathbf{R}_i\rangle \longleftrightarrow \{dR_i, R_i(\dots)\} \quad (10.8)$$

since the Gibbs phase rule is intrinsically related to counting the number of such independent intensities R_i as the “degrees of freedom” of the system (Section 7.1). Choosing a set of

basis vectors is inherently a matter of great freedom, and a considerably broader set of choices involving extensive-type vectors will be described below. Sidebar 10.1 anticipates some features of extensive-type vectors and touches on reasons that intensive-type basis vectors (10.8) are safest and “best” for many purposes.

SIDEBAR 10.1: REMARKS ON THERMODYNAMIC BASIS VECTORS

A curious feature of the space $\mathcal{M}_S$ of thermodynamic variables in an equilibrium state S is that its dimensionality *varies* with the number of phases, p , even though the values of the intensive variables (which might be used to parametrize the state S) do not. The intensive-type ket vectors $|\mathbf{R}_i\rangle$ of (10.8) can actually be defined for all $c + 2$ intensities ($T, -P, \mu_1, \mu_2, \dots, \mu_c$) arising from the fundamental equation of a c -component system, $U(S, V, n_1, n_2, \dots, n_c)$, even if only f of these remain linearly independent when p phases are present.

As will be shown in subsequent sections, the thermodynamic space $\mathcal{M}_S$ includes vectors corresponding to both intensive-type ($|\mathbf{R}_i\rangle$) and extensive-type ($|\mathbf{X}_i\rangle$) variables. In contrast to the freedom and generality to choose any f of the $c + 2$ definable intensive-type vectors $|\mathbf{R}_i\rangle$ (or any nonsingular linear combinations of this set) as basis vectors, the choices of extensive-type $|\mathbf{X}_i\rangle$ basis vectors are much more restricted. It will be found that no more than f extensive-type ket vectors $|\mathbf{X}_i\rangle$ (or their nonsingular linear combinations) are safely definable in *any* phase region. Moreover, the “safe” $|\mathbf{X}_i\rangle$ depend on the phases present, and *no* extensive-type ket is “safe” near a critical region. To consistently describe the geometry of $\mathcal{M}_S$ through phase changes, it is therefore *obligatory* to revert to intensive basis kets $|\mathbf{R}_i\rangle$ when approaching critical regions. Away from criticality, one has the usual freedom to choose f independent basis kets rather arbitrarily, including mixed sets of intensive and extensive type. However, intensive-type basis vectors (10.8) are the most convenient, consistent, and reliable choice for general purposes.

10.2 THE METRIC OF THERMODYNAMIC RESPONSE SPACE

Although the linearity of the chain-rule differential expressions (10.5) confers primitive affine-type “spatial” structure on thermodynamic variables, it does not yet provide a sense of “distance” or “metric” on the space (other than what might be displayed in an arbitrarily chosen axis system). In order to bring intrinsic geometrical structure to the thermodynamic space, we need to define the scalar product $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ [(9.29)] that dictates the spatial metric on $\mathcal{M}_S$. The metric on $\mathcal{M}_S$ should reflect intrinsic physical properties of the thermodynamic responses, not merely generic chain rule-type mathematical properties of their differential representation. At the same time, we must exhibit how the space $\mathcal{M}_S$ is explicitly “connected” to the physical measurements of thermodynamic responses. Because such measurements assign *scalar values* to physical properties, it is natural to associate each “scalar product” of $\mathcal{M}_S$ with the “scalar value” of an experimental measurement. How can this be done?

Having made this long detour into vector geometry and metric spaces, the student of thermodynamics will naturally be impatient to learn the “missing link” that connects these disparate domains, i.e., that associates the scalar products of the geometry domain

with the experimental measurements of the thermodynamic domain. Such a linkage cannot be “derived” from either geometry or thermodynamics alone, because it merely serves as the “portal” between domains, or the “translator” between domains of parallel content but different language (or mathematical formalism). Whether the linkage “makes sense” in each domain can only be established after presentation of a specific *Ansatz*. This *Ansatz* must have the form of a general equation whose left-hand side is interpretable in the geometrical domain, while its right-hand side is interpretable in the thermodynamic domain, i.e., an equation of *isomorphism* (domain equivalence). Specifically, we introduce the isomorphic association between geometrical scalar products $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ and thermodynamic intensive properties $R_i = (\partial U / \partial X_i)_{\underline{X}}$, $R_j = (\partial U / \partial X_j)_{\underline{X}}$ as [F. Weinhold. *J. Chem. Phys.* **63**, 2479 (1975)]

$$\boxed{\langle \mathbf{R}_i | \mathbf{R}_j \rangle = \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}, S}} \quad (10.9)$$

where the right-hand side is the measured value of the specified response function in the state S [cf. (8.79)]. Equation (10.9) is the fundamental “Rosetta stone” that translates thermodynamic properties from the calculus-based language of partial derivatives (on the right) to the geometry-based language of scalar products (on the left). Given the validity of this isomorphism [to be established in (10.10a–d) below], the remainder of this book consists of verifying that our previous thermodynamic results of Parts I and II can be greatly simplified, while many notoriously difficult thermodynamic problems now become tractable (i.e., the geometrical domain has better technology!).

How can (10.9) “make sense” as a geometrical scalar product? From the chain-rule linearity property (10.4) of partial derivatives, one can see that the $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ values defined by (10.9) will automatically satisfy the distributive property (9.27a):

$$\langle \mathbf{R}_i | \lambda \mathbf{R}_j + \mu \mathbf{R}_k \rangle = \lambda \langle \mathbf{R}_i | \mathbf{R}_j \rangle + \mu \langle \mathbf{R}_i | \mathbf{R}_k \rangle, \quad \text{all } i, j, k \quad (10.10a)$$

Moreover, from the “first-law” (Maxwell-type exactness) relationship between mixed partial derivatives, as expressed by (8.80), we see that the $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ values also satisfy the symmetric property (9.27b):

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = \langle \mathbf{R}_j | \mathbf{R}_i \rangle, \quad \text{all } i, j \quad (10.10b)$$

Finally, from the “second law” (stability) property, as expressed by (8.82), we see that each $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ also satisfies the positivity condition (9.27c):

$$\langle \mathbf{R}_i | \mathbf{R}_i \rangle \geq 0, \quad \text{all } i \quad (10.10c)$$

Moreover, the stipulation that these conditions hold for *all* intensities requires satisfaction of (10.10a–c) for arbitrary nonsingular linear combinations of the R_i , and from such considerations (see Sidebar 10.2) we can also infer that

$$\langle \mathbf{R}_i | \mathbf{R}_i \rangle \longrightarrow 0 \quad \text{only if } |\mathbf{R}_i\rangle \longrightarrow 0 \quad (10.10d)$$

Thus, we can conclude that the thermodynamic laws, as expressed in Gibbsian form [(10.10a–d)], *precisely* guarantee that $\langle \mathbf{R}_i | \mathbf{R}_j \rangle$ satisfies the distributive, symmetric, and

positivity conditions (9.27a–c) of a *Euclidean* geometry, the simplest and most familiar form of metric space.

Equation (10.9) establishes the deep *isomorphism* between geometrical and thermodynamic measurements. Let us see what this means in more concrete terms. Imagine that a geometer and a thermodynamicist occupy booths isolated from one another and from the outside except for slots through which written inquiries can be passed. A questioner may request information from each booth through a computer that simultaneously translates questions into the separate languages of the geometer (“geometrical distances”) and thermodynamicist (“thermodynamic responses”); these questions are to be acceptable only if they can be answered in some common language (such as pure numbers) by each respondent.

The questioner will find that he can in no way distinguish which booth holds which scientist, for the replies of the geometer are always in accordance with the principles of thermodynamics, just as those of the thermodynamicist are always in accordance with Euclid. If each were given a particular object for study—perhaps a molten alloy for the thermodynamicist and an irregular tetrahedron for the geometer—the objects could be so chosen that requests for measured properties will always be answered by identical numerical replies from the two booths. In a sense, both scientists are studying the “same” object. Thermodynamics *is* geometry! In fact, thermodynamics is *Euclidean* geometry!

The geometer would find little incentive to adopt the cumbersome calculus-based formalism of equilibrium thermodynamics to solve his mensuration problems, but the thermodynamicist may be startled by the ease with which thorny thermodynamic problems yield to the geometer’s tools. To exploit this advantage, he will naturally use vector algebra in place of the partial differential equations, chain rules and cross-differentiation identities of the usual thermodynamic formalism. He may be initially disquieted that the “laws of thermodynamics” have somehow disappeared (they remain only implicitly in the mathematical structure being employed), but he will soon notice that his deductions are always fully consistent with those laws so long as he “does geometry” according to the usual Euclidean rules.

SIDEBAR 10.2: TRANSFORMATION TO NEW BASIS INTENSITIES

An intrinsic feature of the thermodynamic formalism is the freedom to consider general combinations of extensive or intensive variables [cf. (8.70), (8.75)] as alternatives to “standard” choices. This freedom is used, for example, in considering the Gibbs free energy $G = U - (T)S + (P)V$ as a linear combination of standard (U, S, V) extensivities, or the “phase-coexistence coordinate” σ [cf. (7.27), (7.28)] as a linear combination of standard (T, P) intensities.

A linear transformation such as

$$R'_i = \sum_{j=1}^f a_{ij} R_j, \quad i = 1, 2, \dots, f \quad (\text{S10.2-1})$$

is a special case of the more general functional coordinate transformation of the form

$$R'_i = R'_i(R_1, R_2, \dots, R_f), \quad i = 1, 2, \dots, f \quad (\text{S10.2-2})$$

However, one can see from the chain-rule equations for partial differentials dR'_i

$$dR'_i = \sum_{j=1}^f \left(\frac{\partial R'_i}{\partial R_j} \right) dR_j \quad (\text{S10.2-3})$$

that the transformation (S10.2-2) can be assumed to be of “locally” linear form (S10.2-1), with

$$a_{ij} = \left(\frac{\partial R'_i}{\partial R_j} \right)_{\underline{R}} \quad (\text{S10.2-4})$$

at the point of interest. Thus, without loss of generality, we may consider the general linear combination

$$|\mathbf{R}'_i\rangle = \sum_{j=1}^f a_{ij} |\mathbf{R}_j\rangle, \quad i = 1, 2, \dots, f \quad (\text{S10.2-5})$$

to form new basis ket vectors $\{|\mathbf{R}'_i\rangle\}$, emulating (S10.2-1).

It is easy to recognize that (10.10a, b) remain true under the linear transformations (S10.2-5). By substituting (S10.2-5) into each side of the scalar product, one can also prove that $\langle \mathbf{R}'_i | \mathbf{R}'_i \rangle \geq 0$ for any such $|\mathbf{R}'_i\rangle$. We now wish to examine the conditions under which the inequality (10.10c) becomes an *equality*, and, more specifically, whether $\langle \mathbf{R}'_i | \mathbf{R}'_i \rangle$ might possibly vanish for some *nonvanishing* vector $|\mathbf{R}'_i\rangle$ allowed by (S10.2-5). We shall show that this is impossible for any permissible (nonsingular) transformation of basis vectors, so that

$$\langle \mathbf{R}'_i | \mathbf{R}'_i \rangle = 0 \quad \text{only if } |\mathbf{R}'_i\rangle = 0 \quad (\text{S10.2-6})$$

The issue is whether it is possible that $(\partial R'_i / \partial X'_i)_{\underline{X}'} = 0$ for some *nonzero* intensive differential dR'_i , which could then be taken as one of the f independent intensive differentials (S10.2-5) to describe the system, with corresponding conjugate extensive differentials dX'_i . The usual condition of independence of these differentials is the nonvanishing of the *Jacobian determinant*:

$$\frac{\partial(R'_1, R'_2, \dots, R'_f)}{\partial(X'_1, X'_2, \dots, X'_f)} = \det \left| \left(\frac{\partial R'_i}{\partial X'_j} \right)_{\underline{X}'} \right| \neq 0 \quad (\text{S10.2-7})$$

However, (S10.2-7) is incompatible with the vanishing of any diagonal element ($i = j$) of the determinant, because it is well known* from general matrix theory that a positive-semidefinite matrix with a vanishing diagonal element is necessarily singular, and therefore has *vanishing* determinant, contrary to (S10.2-7). Thus, the only condition under which $\langle \mathbf{R}'_i | \mathbf{R}'_i \rangle = 0$ is the trivial case where dR'_i is *stationary*,

$$\langle \mathbf{R}'_i | \mathbf{R}'_i \rangle = \left(\frac{\partial R'_i}{\partial X'_i} \right)_{\underline{X}'} = 0 \quad \text{only if } dR'_i = 0 \quad (\text{S10.2-8})$$

as claimed in (S10.2-6).

*This follows directly from a theorem of Frobenius; see, e.g., L. Mirsky, *An Introduction to Linear Algebra* (Oxford University Press, New York, 1955), pp. 400ff.

The singular limit (S10.2-6), in which one of the f “independent” $|\mathbf{R}_i\rangle$ is a *zero* vector (implying that the original basis vectors $|\mathbf{R}_i\rangle$ were linearly dependent), occurs at a critical state $\mathcal{S}_c$, where the number of phases p and dimension f are changing. Critical state limits will be examined in Chapter 11.

10.3 LINEAR DEPENDENCE, DIMENSIONALITY, AND GIBBS–DUHEM EQUATIONS

The starting point for thermodynamic description, whether in the calculus-based or the geometry-based formalism, is the Gibbs “fundamental equation” for a given equilibrium state $\mathcal{S}$. In the “energy representation,” this is expressed as

$$U = U(S, V, n_1, n_2, \dots, n_c) \quad (10.11)$$

and in the dual “entropy representation” (Section 5.4) as

$$S = S(U, V, n_1, n_2, \dots, n_c) \quad (10.12)$$

Either could be taken (with equal justification) as the starting point for a *full* thermodynamic description. We have explicitly chosen the U -based starting point (10.11) for the geometrical description, but it is important to realize that an entirely *equivalent* description might have been constructed from the S -based starting point (10.12), based on

$$\left(\underline{\underline{\mathbf{M}}}^{(S)}\right)_{ij} = -\left(\frac{\partial^2 S}{\partial X_i \partial X_j}\right)_{\underline{X}} \quad (10.13)$$

instead of

$$\left(\underline{\underline{\mathbf{M}}}^{(U)}\right)_{ij} = -\left(\frac{\partial^2 U}{\partial X_i \partial X_j}\right)_{\underline{X}} \quad (10.14)$$

as underlying metric (with slightly altered list of “ X_i ” parameters).

From the chosen U -based starting point ($\underline{\underline{\mathbf{M}}} = \underline{\underline{\mathbf{M}}}^{(U)}$), each possible state $\mathcal{S} = \mathcal{S}(\underline{\xi})$ of a single-phase system of c independent chemical components can be parametrized by the numerical values (ξ_i) of the $c + 2$ extensive variables X_i in $\mathcal{S}$:

$$\xi_1 = S, \quad \xi_2 = V, \quad \xi_3 = n_1, \dots, \quad \xi_{c+2} = n_c \quad (10.15a)$$

or, in column-vector form,

$$\underline{\underline{\xi}} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_{c+2} \end{pmatrix} = \begin{pmatrix} S \\ V \\ n_1 \\ \vdots \\ n_c \end{pmatrix}_{\mathcal{S}} \quad (10.15b)$$

where, as indicated, all extensive variables X_i are to be evaluated in state $\mathcal{S}$. It is important to distinguish the $\underline{\xi}$ “state descriptors” (10.14a) from the intrinsic ket vectors of the metric space $\mathcal{M}_{\mathcal{S}}$. The state may be symbolized more completely as $\mathcal{M}_{\mathcal{S}(\underline{\xi})}$ to indicate how $\underline{\xi}$ parametrizes or “identifies” the state of interest.

For a multiphase system (p phases), the state descriptors should be extended to a list of values $(\xi^{(\alpha)})$ for each phase $\alpha = 1, 2, \dots, p$. These are conveniently represented as a *matrix* $(\underline{\xi})$ of $c + 2$ rows and p columns (for phases $\alpha, \beta, \dots, \pi$)

$$\underline{\underline{\xi}} = \begin{pmatrix} \xi_1^{(\alpha)} & \xi_1^{(\beta)} & \dots & \xi_1^{(\pi)} \\ \xi_2^{(\alpha)} & \xi_2^{(\beta)} & \dots & \xi_2^{(\pi)} \\ \xi_3^{(\alpha)} & \xi_3^{(\beta)} & \dots & \xi_3^{(\pi)} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{c+2}^{(\alpha)} & \xi_{c+2}^{(\beta)} & \dots & \xi_{c+2}^{(\pi)} \end{pmatrix} = \begin{pmatrix} S^{(\alpha)} & S^{(\beta)} & \dots & S^{(\pi)} \\ V^{(\alpha)} & V^{(\beta)} & \dots & V^{(\pi)} \\ n_1^{(\alpha)} & n_1^{(\beta)} & \dots & n_1^{(\pi)} \\ \vdots & \vdots & \ddots & \vdots \\ n_c^{(\alpha)} & n_c^{(\beta)} & \dots & n_c^{(\pi)} \end{pmatrix}_{\mathcal{S}} \quad (10.16)$$

or, in more compact (matrix-partitioned) notation, as

$$\underline{\underline{\xi}} = \begin{pmatrix} \underline{\xi}^{(\alpha)} & \underline{\xi}^{(\beta)} & \dots & \underline{\xi}^{(\pi)} \end{pmatrix} \quad (10.17)$$

In view of their extensive property, the state descriptors of each phase must sum to the total value in the system:

$$\underline{\xi} = \sum_{\alpha=1}^p \underline{\xi}^{(\alpha)} \quad (10.18)$$

Let us now turn to the ket vectors $|\mathbf{R}_i\rangle$ and metric matrix

$$(\underline{\underline{\mathbf{M}}}^{(c+2)})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle, \quad i, j = 1, 2, \dots, c + 2 \quad (10.19)$$

in the full space $\mathcal{M}_{\mathcal{S}}$ of $\mathcal{S}(\underline{\xi})$. As shown by (10.19), $\underline{\underline{\mathbf{M}}}^{(c+2)}$ is a type of “overlap matrix” (technically, a “Gramian matrix”) for the nonorthogonal vectors $|\mathbf{R}_1\rangle, |\mathbf{R}_2\rangle, \dots, |\mathbf{R}_{c+2}\rangle$ associated with chosen conjugate extensivities $X_1, X_2, \dots, X_{c+2}$. However, as shown by (8.73) and (10.9), $\underline{\underline{\mathbf{M}}}^{(c+2)}$ can also be described as a “Hessian matrix” (second-derivative matrix) of the underlying U :

$$(\underline{\underline{\mathbf{M}}}^{(c+2)})_{ij} = \left(\frac{\partial^2 U}{\partial X_i \partial X_j} \right)_X, \quad i, j = 1, 2, \dots, c + 2 \quad (10.20)$$

In these terms, the essence of the thermodynamic geometrical description can be succinctly stated as follows:

The Hessian of the internal energy is the Gramian of a metric space.

(10.21)

Both Hessian and Gramian aspects of $\underline{\underline{\mathbf{M}}}^{(c+2)}$ contribute to the richness of $\mathcal{M}_{\mathcal{S}}$.

The dimensionality f of $\mathcal{M}_{\mathcal{S}}$ was introduced in (10.7) to agree with the Gibbs phase rule. In general, this feature of a space is uniquely determined by the *rank* (number of

nonvanishing eigenvalues) of the metric matrix,

$$\text{rank}(\underline{\underline{\mathbf{M}}}^{(c+2)}) = f \quad (10.22)$$

or, equivalently, the maximum number of *linearly independent* $|\mathbf{R}_i\rangle$ (i.e., with no possible linear combination equal to the zero vector). Let the general eigenvalue problem for $\underline{\underline{\mathbf{M}}}^{(c+2)}$ be written as [cf. (S9.1-15)],

$$\underline{\underline{\mathbf{M}}}^{(c+2)} \underline{\mathbf{u}}_i = e_i \underline{\mathbf{u}}_i, \quad i = 1, 2, \dots, c+2 \quad (10.23)$$

Agreement of (10.7) and (10.22) then requires that the full metric matrix have exactly p independent *null eigenvectors* satisfying $\underline{\underline{\mathbf{M}}}^{(c+2)} \underline{\mathbf{u}}_i = 0$ (i.e., $e_i = 0$). We now wish to show that these null eigenvectors can be taken as the $\underline{\xi}^{(\alpha)}$ vectors of (10.17), namely,

$$\underline{\underline{\mathbf{M}}}^{(c+2)} \underline{\xi}^{(\alpha)} = 0, \quad \alpha = 1, 2, \dots, p \quad (10.24)$$

(or any linear combinations), so that $\text{rank}(\underline{\underline{\mathbf{M}}}^{(c+2)}) = c+2-p$, as advertised.

To prove (10.24), we first note that this equation is equivalent to a *linear dependence* condition among the $|\mathbf{R}_i\rangle$, namely,

$$\sum_{j=1}^{c+2} (\underline{\xi}^{(\alpha)})_j |\mathbf{R}_j\rangle = 0 \quad (10.25)$$

[To see that (10.25) is equivalent to (10.24), left-multiply each side of (10.25) by $\langle \mathbf{R}_i |$ and note that the result is just (10.24).] However, when (10.25) is written in terms of explicit elements of $\underline{\xi}^{(\alpha)}$,

$$\sum_{j=1}^{c+2} \xi_j^{(\alpha)} |\mathbf{R}_j\rangle = 0 \quad (10.26)$$

the result is just the geometrical equivalent of the *Gibbs–Duhem equation* (8.87) for homogeneous phase α . The p homogeneity conditions for phases $\alpha = 1, 2, \dots, p$ therefore lead directly to p null eigenvectors (10.25) that bring (10.22) into compliance with the Gibbs phase rule (8.89). Of course, this result is consistent with the manner in which the Gibbs–Duhem equation was originally used to derive the Gibbs phase rule in Section 7.1. The entire collection of Gibbs–Duhem equations for separate phases can also be expressed more compactly in matrix notation as [cf. (10.16)]

$$\boxed{\underline{\underline{\mathbf{M}}}^{(c+2)} \underline{\xi} = 0} \quad (10.27)$$

The fact that the full metric matrix (10.19) of order $c+2$ is singular (with p null eigenvalues) implies that its determinant vanishes,

$$\det |\underline{\underline{\mathbf{M}}}^{(c+2)}| = 0 \quad (10.28)$$

as does that of any submatrix of order higher than $c+2-p$. To obtain a *nonsingular* metric matrix associated with a linearly *independent* set of basis vectors $|\mathbf{R}_i\rangle$, we therefore

employ a submatrix $\underline{\underline{\mathbf{M}}}^{(f)}$ of order f ,

$$(\underline{\underline{\mathbf{M}}}^{(f)})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle, \quad i, j = 1, 2, \dots, f \quad (10.29)$$

with

$$\text{rank}(\underline{\underline{\mathbf{M}}}^{(f)}) = f = c + 2 - p, \quad \det |\underline{\underline{\mathbf{M}}}^{(f)}| \neq 0 \quad (10.30)$$

We may refer to the reduced-order $\underline{\underline{\mathbf{M}}}^{(f)}$ of (10.29) as the “internal metric” to distinguish it from the “full metric” of (10.19).

The physical meaning of a null metric eigenvector (10.24) is that the thermodynamic state $\mathcal{M}_S$ has no “response” in the specified “direction,” i.e., that an overall “change” dY of the form

$$dY = \sum_{i=1}^{c+2} \xi_i^{(\alpha)} dR_i \quad (10.31)$$

effects *no* change in system state. This is precisely the case if one carries out an overall change of system *size* by scaling all extensive variables by a uniform scale factor λ [cf. (6.24)–(6.27) or (8.85)]:

$$U(\lambda X_1, \lambda X_2, \dots, \lambda X_{c+2}) = \lambda U(X_1, X_2, \dots, X_{c+2}) \quad (10.32)$$

Sidebar 10.3 outlines the useful analogy to normal-mode analysis of molecular vibrations, where the null modes correspond to overall translations or rotations of the coordinate system that lead to spurious alterations of coordinate values, but no real *internal* changes of interatomic distances. For this reason, the internal metric $\underline{\underline{\mathbf{M}}}^{(f)}$ of (10.29) is the starting point for analyzing intrinsic *state*-related (as opposed to *size*-related) aspects of a given physical system of interest.

SIDEBAR 10.3: NORMAL-MODE ANALYSIS OF THERMODYNAMIC METRIC

In view of the Hessian character (10.20) of the thermodynamic metric matrix $\underline{\underline{\mathbf{M}}}^{(c+2)}$, the eigenvalue problem for $\underline{\underline{\mathbf{M}}}^{(c+2)}$ [(10.23)] can be usefully analogized with “normal-mode” analysis of *molecular vibrations* [E. B. Wilson, Jr, J. C. Decius, and P. C. Cross. *Molecular Vibrations* (McGraw-Hill, New York, 1955)]. The latter theory starts from a similar Hessian-type matrix, based on second derivatives of the mechanical potential energy V_{pot} (cf. Sidebar 2.8) rather than the thermodynamic internal energy U .

In the vibrational case, the $3N$ Cartesian-like coordinates (x_i , $i = 1, 2, \dots, 3N$) of N atoms may be chosen as the basis for the $3N \times 3N$ Hessian (“force constant”) matrix $\underline{\underline{\mathbf{H}}}$

$$(\underline{\underline{\mathbf{H}}})_{ij} = \frac{\partial^2 V_{\text{pot}}}{\partial x_i \partial x_j}, \quad i, j = 1, 2, \dots, 3N \quad (\text{S10.3-1})$$

The “normal modes of vibration” $\underline{\mathbf{u}}_i$ are obtained as solutions of the Hessian eigenvalue problem,

$$\underline{\underline{\mathbf{H}}} \underline{\mathbf{u}}_i = e_i \underline{\mathbf{u}}_i, \quad i, j = 1, 2, \dots, 3N \quad (\text{S10.3-2})$$

each with associated *vibrational frequency* v_i proportional to the square root of e_i ,

$$v_i \propto e_i^{1/2} \quad (\text{S10.3-3})$$

However, for a general nonlinear polyatomic molecule of N atoms, only $3N - 6$ of these modes correspond to physical vibrations with $e_i \neq 0$ (cf. Sidebar 3.8). The remaining six *null eigenvectors* $\underline{\eta}_i$ satisfying

$$\underline{\underline{H}} \underline{\eta}_i = 0, \quad i = 1, \dots, 6 \quad (\text{S10.3-4})$$

correspond to *unphysical* (“zero-frequency”) modes that represent only the arbitrary choice of Cartesian coordinate system with six degrees of freedom (three translational, three rotational). The unphysical “null modes” of the molecular vibration problem (S10.3-4) evidently correspond to the scaling-type null eigenvectors of the thermodynamic metric (10.24). The suggestive correspondence between vibrational “force constants” and thermodynamic “responses” conveys a useful verbal analogy.

Sidebar 10.4 briefly describes the related concept of “generalized homogeneity,” which has been hypothesized to underlie certain aspects of critical-state behavior but appears superfluous in the metric context.

SIDEBAR 10.4: “GENERALIZED HOMOGENEITY”

The concept of “generalized homogeneity” [see, e.g., H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971), pp. 178ff] refers to a more general class of functions that are envisioned to satisfy scaling equations of the form

$$F(\lambda^{a_1} x_1, \lambda^{a_2} x_2, \dots) = \lambda F(x_1, x_2, \dots) \quad (\text{S10.4-1})$$

rather than ordinary first-order homogeneity,

$$F(\lambda x_1, \lambda x_2, \dots) = \lambda F(x_1, x_2, \dots) \quad (\text{S10.4-2})$$

(i.e., the special case $a_1 = a_2 = \dots = 1$). An analogous generalization may be hypothesized for zeroth-order homogeneity of intensive functions, namely,

$$Q(\lambda^{a_1} x_1, \lambda^{a_2} x_2, \dots) = Q(x_1, x_2, \dots) \quad (\text{S10.4-3})$$

Generalized scaling behavior has been hypothesized to underlie certain observed critical-state limits (to be described in Chapter 11). We wish to show here how *apparent* generalized scaling behavior such as (S10.4-3) may be inferred from the existence of null eigenvectors ($\underline{\eta}$) of the metric matrix [cf. (10.24)],

$$\underline{\underline{M}}^{(c+2)} \underline{\eta} = 0 \quad (\text{S10.4-4})$$

without subsidiary hypotheses.

Let f be the number of independent intensities and X_{f+1} an additional extensity (e.g., system mass) that establishes the overall scale of the system. Instead of the nonsingular

$\underline{\underline{M}}^{(f)}$ [(10.29), (10.30)], we now consider the enlarged metric matrix of order $f + 1$ that incorporates the intensive $|\mathbf{R}_{f+1}\rangle$ conjugate to X_{f+1} ,

$$(\underline{\underline{M}}^{(f+1)})_{i,f+1} = \langle \mathbf{R}_i | \mathbf{R}_{f+1} \rangle = \left(\frac{\partial R_i}{\partial X_{f+1}} \right)_{X_1 \dots X_f}, \quad i = 1, 2, \dots, f + 1 \quad (\text{S10.4-5})$$

with a single null eigenvector $\underline{\eta}$.

We begin by introducing the ratios a_i ($i = 1, 2, \dots, f + 1$) defined by

$$a_i \equiv \frac{\eta_i}{X_i} = \frac{\eta_i}{\xi_i} \quad \text{at state } \mathcal{S} = \mathcal{S}(\underline{\xi}) \quad (\text{S10.4-6})$$

The i th row of (S10.4-4) can be rewritten, rather tortuously, as

$$\frac{\partial R_i}{\partial X_{f+1}} = \sum_{j=1}^f \left(-\frac{a_j}{a_{f+1}} \frac{X_j}{X_{f+1}} \right) \frac{\partial R_i}{\partial X_j} \quad \text{at } \mathcal{S} = \mathcal{S}(\underline{\xi}) \quad (\text{S10.4-7})$$

so long as $a_{f+1} \neq 0$. Suppose now that Y_j are f new variables,

$$Y_j = F_j(X_{f+1})X_j, \quad j = 1, 2, \dots, f \quad (\text{S10.4-8})$$

related to the old X_j by some (as yet unspecified) function $F_j = F_j(X_{f+1})$ of the proposed scale factor. Then, for each $j = 1, 2, \dots, f$,

$$X_j \left(\frac{\partial R_i}{\partial X_j} \right)_{X_1 \dots X_{j-1} X_{j+1} \dots X_{f+1}} = Y_j \left(\frac{\partial R_i}{\partial Y_j} \right)_{Y_1 \dots Y_{j-1} Y_{j+1} \dots Y_f} \quad (\text{S10.4-9})$$

and (S10.4-7) can be rewritten as

$$\frac{\partial R_i}{\partial X_{f+1}} = \sum_{j=1}^f \left(-\frac{a_j}{a_{f+1}} \frac{Y_j}{X_{f+1}} \frac{\partial R_i}{\partial Y_j} \right) = \sum_{j=1}^f \frac{\partial Y_j}{\partial X_{f+1}} \frac{\partial R_i}{\partial Y_j} \quad \text{at } \mathcal{S} = \mathcal{S}(\underline{\xi}) \quad (\text{S10.4-10})$$

provided that

$$\left(\frac{\partial Y_j}{\partial X_{f+1}} \right)_{X_1 \dots X_f} = -\frac{a_j}{a_{f+1}} \frac{Y_j}{X_{f+1}}, \quad j = 1, 2, \dots, f \quad (\text{S10.4-11})$$

Now (S10.4-6) is apparently the chain rule for a function having the special property (for some constant c_0)

$$R_i(X_1, X_2, \dots, X_{f+1}) = R_i(Y_1, Y_2, \dots, Y_f, c_0) \quad (\text{S10.4-12})$$

in some asymptotic sense about $\mathcal{S} = \mathcal{S}(\underline{\xi})$. The special functions F_j that lead to (S10.4-11) and (S10.4-12) must satisfy

$$\left(\frac{\partial \ln Y_i}{\partial \ln X_{f+1}} \right)_{X_1 \dots X_f} = \frac{d \ln F_i}{d \ln X_{f+1}} = -\frac{a_j}{a_{f+1}} \quad (\text{S10.4-13})$$

Functions F_j satisfying (S10.4-13) have the form

$$F_j = c_j X_{f+1}^{a_j/a_{f+1}}, \quad c_j = \text{constant} \quad (\text{S10.4-14})$$

Equation (S10.4-12) can therefore be written as

$$R_i(X_1, X_2, \dots, X_{f+1}) = R_i(\lambda^{a_1} X_1, \lambda^{a_2} X_2, \dots, \lambda^{a_{f+1}} X_{f+1}) \quad (\text{S10.4-15})$$

provided we define

$$\lambda^{a_{f+1}} = \frac{c_0}{X_{f+1}} \quad (\text{S10.4-16})$$

$$c_j = c_0^{a_j/a_{f+1}} \quad (\text{S10.4-17})$$

for any chosen c_0 . Apparent “generalized scaling-type behavior” (S10.4-15) therefore seems to represent only an indirect (and awkwardly expressed) aspect of null metric eigenvectors (S10.4-4) that are demanded by the Gibbs phase rule, and only the limit of conventional Gibbs–Duhem scaling

$$a_1 = a_2 = \dots = a_{f+1} = 1 \quad (\text{S10.4-18})$$

appears thermodynamically justifiable.

Geometrical Representation of Equilibrium Thermodynamics

11.1 THERMODYNAMIC VECTORS AND GEOMETRY

Building on the concepts introduced in Chapters 9 and 10, we now wish to describe the geometrical “essence” of thermodynamic intensive properties R_i in terms of an associated Dirac ket symbols $|\mathbf{R}_i\rangle$. The boldface ket label serves to distinguish this underlying geometrical character from the name (R_i), value (R_i), and variability (dR_i) of a given intensive property. The boldface label also distinguishes the Dirac $|\mathbf{R}_i\rangle$ from alternative symbols reserved for its matrix-algebraic representation, for example, in single-underline ($\underline{r}$) or double-underline ($\underline{\underline{R}}$) notation. (Note that heinous hybrid notation such as “ $|dR_i\rangle$ ” should be scrupulously avoided!)

In the abstract space $\mathcal{M}_S$ associated with a given equilibrium state S , the fundamental geometry for thermodynamic vectors $|\mathbf{R}_i\rangle$, $|\mathbf{R}_j\rangle$ is fixed by the scalar product

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} \quad (11.1)$$

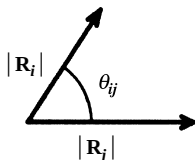
as evaluated in terms of the measured response function $(\partial R_i / \partial X_j)_{\underline{X}}$ for the system of interest. In terms of this scalar product, the “length” of $|\mathbf{R}_i\rangle$ is given by

$$|\mathbf{R}_i| = \langle \mathbf{R}_i | \mathbf{R}_i \rangle^{1/2} \quad (11.2)$$

and the “angle” θ_{ij} between vectors $|\mathbf{R}_i\rangle$ and $|\mathbf{R}_j\rangle$ is given by

$$\cos \theta_{ij} = \frac{\langle \mathbf{R}_i | \mathbf{R}_j \rangle}{|\mathbf{R}_i| |\mathbf{R}_j|} \quad (11.3)$$

as represented graphically by:



Similarly, we can state the general *Schwarz inequality*

$$|\langle \mathbf{R}_i | \mathbf{R}_j \rangle| \leq |\mathbf{R}_i| |\mathbf{R}_j| \quad (11.4)$$

and, for any triangle of points, the *triangle inequality*

$$|\mathbf{R}_i - \mathbf{R}_j| \leq |\mathbf{R}_i - \mathbf{R}_k| + |\mathbf{R}_k - \mathbf{R}_j| \quad (11.5)$$

We shall say that two vectors $|\mathbf{R}_i\rangle$ and $|\mathbf{R}_j\rangle$ are *equal* if and only if their separation $|\mathbf{R}_i - \mathbf{R}_j|$ vanishes,

$$|\mathbf{R}_i\rangle = |\mathbf{R}_j\rangle \quad \text{iff} \quad |\mathbf{R}_i - \mathbf{R}_j| = 0 \quad (11.6)$$

and that they are *orthogonal* (“perpendicular”; $\theta_{ij} = 90^\circ$) if

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = 0 \quad (11.7)$$

In short, the theorems, terminology, and working methods of Euclidean geometry can be carried over intact into this abstract thermodynamic domain.

With the association (11.1) to a measured value of a thermodynamic response function, we can characterize the *physical* significance of the abstract lengths and angles in $\mathcal{M}_S$. The length of $|\mathbf{R}_i\rangle$ measures the *responsiveness* of the system to a change in the associated extensive parameter X_i , i.e., the extent to which the system adjusts its value of R_i in response to a small change in X_i . (For example, the length of the temperature vector is related to the inverse heat capacity of the system, and that of the pressure vector to its inverse compressibility.) The angle θ_{ij} between vectors $|\mathbf{R}_i\rangle$ and $|\mathbf{R}_j\rangle$ measures the extent to which different responses are *coupled*, i.e., to what extent a small change in X_i will produce a response in R_j , and vice versa. (For example, the angle between the temperature and pressure vectors is related to the thermal expansion coefficient.) The metric parameters of $\mathcal{M}_S$ therefore have *intrinsic* physical significance, being uniquely associated with measured properties of a particular system under discussion. Note that lengths in $\mathcal{M}_S$ depend on the physical units in which the associated responses are measured, but the coupling angles θ_{ij} do not.

We noted in Chapter 10 that the dimensionality of the subspace spanned by any chosen set of n vectors $|\mathbf{R}_i\rangle$ is given by the rank of the associated *metric matrix* (or *Gram matrix*) $\underline{\underline{M}}^{(n)}$, whose elements are scalar products among the $|\mathbf{R}_i\rangle$. Because the intrinsic dimensionality f for a system of c independent chemical components and p phases is dictated by the Gibbs phase rule ($f = c - p + 2$), any metric matrix $\underline{\underline{M}}^{(f+1)}$ of $f + 1$ vectors is necessarily singular in the f -dimensional space, and the corresponding *metric determinant* (“Gramian” determinant) $M^{(f+1)}$ accordingly vanishes:

$$M^{(f+1)} \equiv \det |\underline{\underline{M}}^{(f+1)}| = 0 \quad (11.8)$$

Equation (11.8) expresses the geometrical necessity of linear dependence among any $f + 1$ vectors in an f -dimensional space. Such linear dependence (corresponding to a null

eigenvector of $\underline{\underline{M}}^{(f+1)}$) was recognized in Chapter 10 as a generalized form of the *Gibbs–Duhem equation*, to which it directly reduces when the internal energy U scales in the usual manner with the extensities X_i .

Suppose f independent reference intensities R_i have been selected, each with corresponding conjugate extensity X_i . The associated vectors $|\mathbf{R}_i\rangle$ are linearly independent, the corresponding metric matrix $\underline{\underline{M}}^{(f)}$ is nonsingular, and the associated metric determinant $M^{(f)}$ is nonzero. Because the intrinsic dimensionality f of $\underline{\underline{M}}^{(f)}$ is clear in context, we shall generally omit the superscript of this “internal metric,” writing $\underline{\underline{M}}$ (instead of $\underline{\underline{M}}^{(f)}$) to obtain

$$(\underline{\underline{M}})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle = \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}}, \quad i, j = 1, 2, \dots, f \quad (11.9a)$$

$$\text{rank}(\underline{\underline{M}}) = f = c - p + 2 \quad (11.9b)$$

$$M = \det |\underline{\underline{M}}| \neq 0 \quad (11.9c)$$

The reference vectors $\mathbf{R}_i$ span the f -dimensional space, and thus form a *basis* in $\mathcal{M}_S$. A general element of this space, denoted $|\mathbf{R}_{\underline{\alpha}}\rangle$, may be expanded as

$$|\mathbf{R}_{\underline{\alpha}}\rangle = \sum_{i=1}^f \alpha_i |\mathbf{R}_i\rangle \quad (11.10)$$

with “components” α_i that provide a unique label $\underline{\alpha}$ for $|\mathbf{R}_{\underline{\alpha}}\rangle$ in the set of (generally non-orthogonal) basis vectors $|\mathbf{R}_i\rangle$:

$$\alpha_i = (\underline{\alpha})_i = \left(\frac{\partial R_{\underline{\alpha}}}{\partial R_i} \right)_{\underline{R}} \quad (11.11)$$

The label (column-vector) $\underline{\alpha}$ is said to “represent” $|\mathbf{R}_{\underline{\alpha}}\rangle$ in the basis of $|\mathbf{R}_i\rangle$ ’s. A general scalar product of vectors $|\mathbf{R}_{\underline{\alpha}}\rangle$, $|\mathbf{R}_{\underline{\beta}}\rangle$ then becomes, in view of (11.9a) and (11.10),

$$\langle \mathbf{R}_{\underline{\alpha}} | \mathbf{R}_{\underline{\beta}} \rangle = \underline{\alpha}^t \underline{\underline{M}} \underline{\beta} \quad (11.12)$$

where superscript “t” denotes the transpose. Thus, knowledge of the metric $\underline{\underline{M}}$ for reference vectors $|\mathbf{R}_i\rangle$ is sufficient to determine *all possible* scalar products in $\mathcal{M}_S$, and thereby to specify the thermodynamic geometry completely.

The metric matrix $\underline{\underline{M}}$ is symmetric through (9.27b),

$$\underline{\underline{M}}^t = \underline{\underline{M}} \quad (11.13)$$

which natural symmetry can be recognized as summarizing the various *Maxwell relations* (Section 5.6). Because of this symmetry, one can see that no more than $f(f+1)/2$ elements

of $\underline{\underline{M}}$ are independent. This in turn establishes the “ $f(f+1)/2$ rule” (cf. the claim following Sidebars 5.4–5.6 in Chapter 5):

<u>$f(f+1)/2$ Rule:</u> Only $f(f+1)/2$ independent response functions need to be experimentally measured in order to fully characterize the thermodynamic geometry, and thus <i>all</i> possible thermodynamic responses of the system.	(11.14)

While such results can also be inferred from the classical formalism [F. H. Crawford, *Phys. Rev.* **72**, 521A (1947)], they have a particularly transparent basis in the metric space $\mathcal{M}_S$.

11.2 CONJUGATE VARIABLES AND CONJUGATE VECTORS

The f chosen reference axes $|\mathbf{R}_i\rangle$ of the thermodynamic space will not generally be mutually perpendicular unless some very special choice of thermodynamic variables has been made. But if $|\mathbf{R}_1\rangle$ (say) is not perpendicular to $|\mathbf{R}_2\rangle, |\mathbf{R}_3\rangle, \dots, |\mathbf{R}_f\rangle$, we can always find the unique direction that *is* perpendicular to these other axes. We label a vector along this direction as $|\bar{\mathbf{R}}_1\rangle$, and choose its length so that the scalar product of $|\mathbf{R}_1\rangle$ and $|\bar{\mathbf{R}}_1\rangle$ (the product of their lengths times the cosine of their separation angle) is unity. By similar reasoning, we can find vectors $|\bar{\mathbf{R}}_2\rangle, |\bar{\mathbf{R}}_3\rangle, \dots, |\bar{\mathbf{R}}_f\rangle$, where each $|\bar{\mathbf{R}}_i\rangle$ lies perpendicular to all the vectors $|\mathbf{R}_j\rangle$ (except $|\mathbf{R}_i\rangle$ itself), and the lengths are chosen such that the scalar product of $|\mathbf{R}_i\rangle$ and $|\bar{\mathbf{R}}_i\rangle$ is unity for each i . Figure 11.1 illustrates this process in three dimensions. It shows also that this new set of vectors bears a close geometrical relationship to the former set, in the sense that if we had begun with the vectors $|\bar{\mathbf{R}}_i\rangle$, we should have been led uniquely by the above procedure to the original vectors $|\mathbf{R}_i\rangle$, and vice versa.

Let us now describe this graphical construction in more mathematical terms. Although the reference axes $|\mathbf{R}_i\rangle$ are generally nonorthogonal, one can always construct the associated set of “conjugate” vectors $|\bar{\mathbf{R}}_i\rangle$ that are *biorthogonal* to the $|\mathbf{R}_i\rangle$, namely,

$$\langle \bar{\mathbf{R}}_i | \mathbf{R}_j \rangle = \delta_{ij}, \quad ij = 1, 2, \dots, f \quad (11.15)$$

Such vectors can be found whenever the metric matrix $\underline{\underline{M}}$ is nonsingular, and take the explicit form

$$|\bar{\mathbf{R}}_i\rangle = \sum_{k=1}^f (\underline{\underline{M}}^{-1})_{ik} |\mathbf{R}_k\rangle, \quad i = 1, 2, \dots, f \quad (11.16)$$

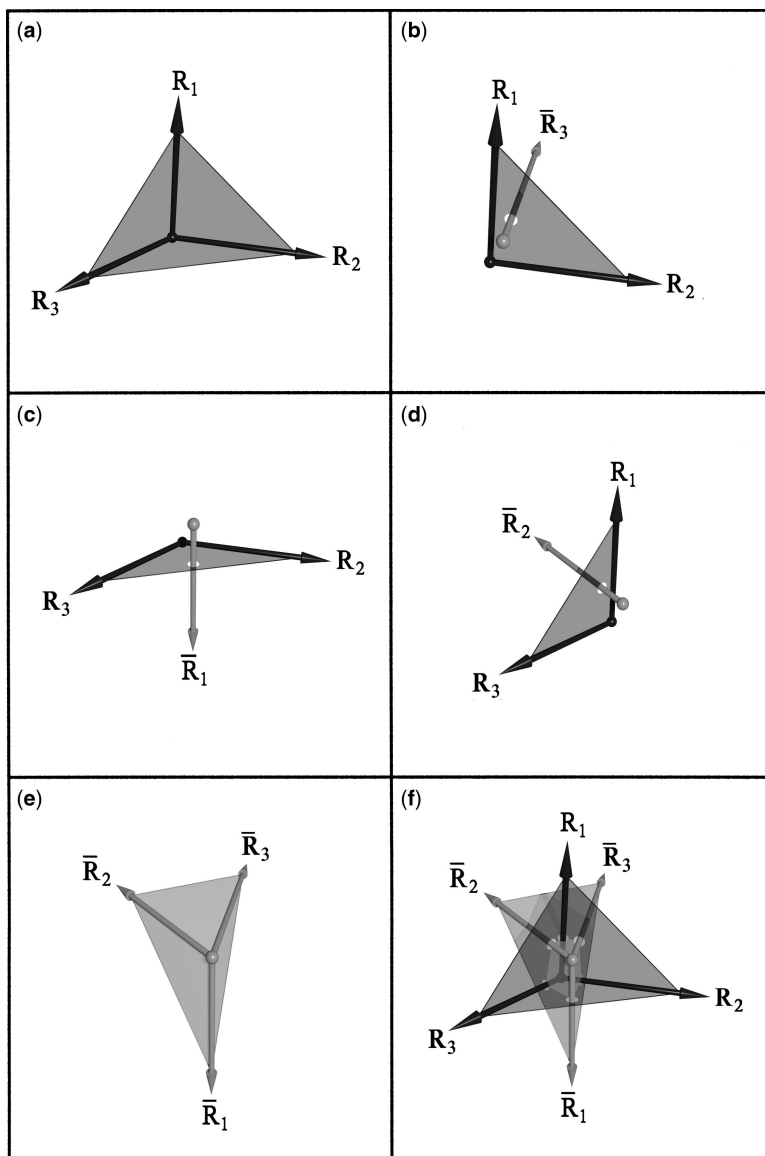


Figure 11.1 Construction of conjugate thermodynamic vectors is illustrated for a three-dimensional system, such as a binary alloy. The starting vectors $|\mathbf{R}_1\rangle$, $|\mathbf{R}_2\rangle$, $|\mathbf{R}_3\rangle$ might represent responses to thermal, mechanical and chemical disturbances; they form an irregular tetrahedron (a). The vector $|\bar{\mathbf{R}}_3\rangle$ conjugate to $|\mathbf{R}_3\rangle$ must be perpendicular to the plane of the other two vectors, as shown in (b); its length must be such that the scalar product of the conjugate vectors is unity. Conjugate vectors $|\bar{\mathbf{R}}_1\rangle$ (c) and $|\bar{\mathbf{R}}_2\rangle$ (d) are constructed similarly. The three new vectors form a conjugate tetrahedron (e). This is shown interpenetrating the original tetrahedron in (f), suggesting the complementary character of the construction.

By substituting (11.16) in (11.15) and using (11.9a), one easily verifies the biorthogonality property (11.15),

$$\langle \bar{\mathbf{R}}_i | \mathbf{R}_j \rangle = \sum_{k=1}^f (\underline{\underline{\mathbf{M}}}^{-1})_{ik} \langle \bar{\mathbf{R}}_k | \mathbf{R}_j \rangle = \sum_{k=1}^f (\underline{\underline{\mathbf{M}}}^{-1})_{ik} (\underline{\underline{\mathbf{M}}})_{kj} = \delta_{ij} \quad (11.17)$$

as claimed.

The new vectors must of course be associated with some corresponding thermodynamic variables $\{\bar{R}_i\}$, and it is then natural to ask how these new variables are related to the initial $\{R_i\}$. It turns out that $\bar{R}_i$ and R_i are indeed *conjugates* in the sense in which that term was used previously (8.73). Indeed, in a differential sense the variables $\bar{R}_i$ behave precisely as do the reference extensities X_i , as shown in Sidebar 11.1,

$$dX_i = d\bar{R}_i, \quad i = 1, 2, \dots, f \quad (11.18)$$

so that the symbols X_i and $\bar{R}_i$ are essentially interchangeable in any expressions involving changes of the extensive parameter, such as

$$R_i = \left(\frac{\partial U}{\partial \bar{R}_i} \right)_{\bar{\mathbf{R}}} \quad (11.19)$$

Note, however (as remarked in Sidebar 10.1) that only f of the $c + 2$ extensities X_i have $\bar{R}_i$ “images” in $\mathcal{M}_S$. Each conjugate vector $|\bar{\mathbf{R}}_i\rangle$ can therefore be thought of (and accordingly labeled) as corresponding to the extensive variable X_i ,

$$|\bar{\mathbf{R}}_i\rangle = |\mathbf{X}_i\rangle \longleftrightarrow \{dX_i, X_i(\dots)\}, \quad i = 1, 2, \dots, f \quad (11.20)$$

in much the same manner as $|\mathbf{R}_i\rangle$ is related to the corresponding intensive variables R_i in (10.8).

SIDEBAR 11.1: PROOF OF (11.18)

Equation (11.16) is the geometrical “image” of the differential equation

$$d\bar{R}_i = \sum_{k=1}^f (\underline{\underline{\mathbf{M}}}^{-1})_{ik} dR_k, \quad i = 1, 2, \dots, f \quad (\text{S11.1-1})$$

Let $\underline{\mathbf{u}}$, $\underline{\bar{\mathbf{u}}}$, $\underline{\mathbf{v}}$ be column vectors of differentials defined by

$$\underline{\mathbf{u}} = \begin{pmatrix} dR_1 \\ dR_2 \\ \vdots \\ dR_f \end{pmatrix}, \quad \underline{\bar{\mathbf{u}}} = \begin{pmatrix} d\bar{R}_1 \\ d\bar{R}_2 \\ \vdots \\ d\bar{R}_f \end{pmatrix}, \quad \underline{\mathbf{v}} = \begin{pmatrix} dX_1 \\ dX_2 \\ \vdots \\ dX_f \end{pmatrix} \quad (\text{S11.1-2})$$

Then (S11.1-1) is equivalent to the matrix-algebraic equation

$$\underline{\bar{u}} = \underline{\underline{M}}^{-1} \underline{u} \quad (\text{S11.1-3})$$

The general chain-rule expression for $dR_i = dR_i(X_1, X_2, \dots, X_f)$,

$$dR_i = \sum_{j=1}^f \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} dX_j, \quad i = 1, 2, \dots, f \quad (\text{S11.1-4})$$

can similarly be written as

$$\underline{u} = \underline{\underline{M}} \underline{v} \quad (\text{S11.1-5})$$

or equivalently as

$$\underline{v} = \underline{\underline{M}}^{-1} \underline{u} \quad (\text{S11.1-6})$$

From (S11.1-3) and (S11.1-6), we therefore obtain

$$\underline{v} = \underline{\bar{u}} \quad (\text{S11.1-7})$$

which is equivalent to (11.18)—*QED*.

The conjugacy relationship of the Gibbsian formalism therefore reappears as a kind of perpendicularity requirement in the abstract geometry. As in ordinary plane geometry, drawing the perpendiculars often turns out to be a useful device for analyzing the figures that arise in a thermodynamic context. Just as Gibbs was motivated to study the chemical potential because it is conjugate to the mass of a chemical component (a variable of interest), so one will generally find it convenient to introduce the conjugates $\bar{R}_i$ of whatever variables R_i have been chosen (out of convenience or necessity) to represent the state of the system.

The biorthogonality relations (11.15) make clear the far-reaching symmetry between intensive vectors $|\mathbf{R}_i\rangle$ and their conjugates $|\bar{\mathbf{R}}_i\rangle$ in the geometrical formalism. The formal symmetry is also seen in relations of the form

$$|\bar{\bar{\mathbf{R}}}_i\rangle = |\mathbf{R}_i\rangle \quad (\text{11.21})$$

which exhibits the *mutual* character of vector conjugacy. A general scalar product of two conjugate vectors becomes

$$\langle \bar{\mathbf{R}}_i | \bar{\mathbf{R}}_j \rangle = \left(\frac{\partial X_i}{\partial R_j} \right)_{\underline{R}} \quad (\text{11.22})$$

paralleling (11.1). In the conjugate basis $\{|\bar{\mathbf{R}}_i\rangle\}$, the roles of conjugate variables X_i and R_i are evidently reversed, and one deals with response functions in which the intensities R_i (rather than extensities X_i) play the role of independent variables.

Scalar products among conjugate vectors can also be evaluated from (11.9a) and (11.16) in the form

$$\langle \bar{\mathbf{R}}_i | \bar{\mathbf{R}}_j \rangle = (\underline{\underline{\mathbf{M}}}^{-1} \underline{\underline{\mathbf{M}}} \underline{\underline{\mathbf{M}}}^{-1})_{ij} = (\underline{\underline{\mathbf{M}}}^{-1})_{ij} \quad (11.23)$$

Taken together, (11.22) and (11.23) lead to various thermodynamic identities between measured response functions, as will be illustrated below. Equation (11.23) shows that the inverse metric matrix $\underline{\underline{\mathbf{M}}}^{-1}$ plays a role for conjugate vectors $|\bar{\mathbf{R}}_i\rangle$ that is highly analogous to the role played by $\underline{\underline{\mathbf{M}}}$ itself for the intensive vectors $|\mathbf{R}_i\rangle$. In view of this far-reaching relationship, we can define the *conjugate metric* $\bar{\underline{\underline{\mathbf{M}}}}$,

$$\bar{\underline{\underline{\mathbf{M}}}} = \underline{\underline{\mathbf{M}}}^{-1} = (\underline{\underline{\mathbf{M}}}^{-1})^t \quad (11.24)$$

so that (11.23) becomes the obvious “conjugate” of (11.9a):

$$(\bar{\underline{\underline{\mathbf{M}}}})_{ij} = \langle \bar{\mathbf{R}}_i | \bar{\mathbf{R}}_j \rangle = \left(\frac{\partial X_i}{\partial R_j} \right)_R, \quad i, j = 1, 2, \dots, f \quad (11.25)$$

Although $\mathcal{M}_S$ was initially constructed from intensive variables R_i , introduction of the conjugate vectors $|\bar{\mathbf{R}}_i\rangle = |\mathbf{X}_i\rangle$ finally permits intensive and extensive variables to be handled in a nearly symmetrical fashion in the geometrical formalism. Nevertheless, a fundamental asymmetry persists in the formalism between these two types of variables, as discussed further in Sidebar 11.2. For example, if X_{f+1} is a scale factor and R_{f+1} its conjugate intensity, it will be possible to give an expression for the vector $|\mathbf{R}_{f+1}\rangle$ representing R_{f+1} in $\mathcal{M}_S$ [its expansion is given by the Gibbs–Duhem equation; cf. (10.26)]. However, there is no possible vector representing X_{f+1} , because the inverse matrix that would be required in (11.16) does not exist, as shown by (11.8).

SIDEBAR 11.2: ASYMMETRY OF CONJUGATE INTENSITIES AND EXTENSITIES

As previously remarked (Sidebar 10.1), *intensive* vectors have privileged status in $\mathcal{M}_S$. The asymmetry between intensive and extensive variables can already be recognized in the U -based (or S -based) “fundamental equation” of Gibbs

$$U = U(\{X_i\}) \quad (\text{S11.2-1a})$$

in which extensives $\{X_i\}$ appear as “control parameters” (independent variables) of the state, whereas intensities $\{R_i\}$

$$R_i = \left(\frac{\partial U}{\partial X_i} \right)_X \quad (\text{S11.2-1b})$$

are the dependent experimental responses. We may refer to the fundamental description based on (S11.2-1a, b) as the “(U, X_i, R_i) construction.”

As described in Section 5.5, various alternative “ (U_L, R_i, X_i) constructions” might be considered in which the roles of R_i and X_i are exchanged and U is replaced by an associated Legendre transform U_L (such as Gibbs free energy G). However, such U_L -based descriptions are not “fundamental” in the sense of U -based or S -based descriptions (Section 5.4). Although they remain equivalent to U -based description in *noncritical* regions where the transformation equations between $\{R_i\}$ and $\{X_i\}$ remain invertible, they become indeterminate at critical-state singularities, corresponding to dimensional changes in $\mathcal{M}_S$.

The basic asymmetry between intensive and extensive vectors can also be recognized in the *Gibbs phase rule*. This establishes the dimensionality of $\mathcal{M}_S$ in terms of the number of independent *intensities*, as expressed in (11.9b) in terms of $\text{rank}(\underline{\underline{M}})$. An alternative extensity-based (or $\underline{\underline{M}}$ -based) description necessarily *diverges* at points where $\underline{\underline{M}}$ becomes singular, i.e., at critical limits, where dimensionality changes, as shown by (11.24).

In a deeper sense, the distinguished role of intensities was already apparent in the fundamental definition of “state” (Section 2.10). Although the high symmetry between (a subset of) the $|\mathbf{R}_i\rangle$ and the conjugate $|\overline{\mathbf{R}}_i\rangle$ largely disguises this distinction in states of nonsingular $\underline{\underline{M}}$, the asymmetry becomes evident at critical states where matrix elements of $\underline{\underline{M}}$ become ill-defined and divergent, whereas those of $\underline{\underline{M}}$ do not.

11.3 METRIC OF A HOMOGENEOUS FLUID

Let us now illustrate some of these ideas more concretely with the simple example of a one-component fluid, say a sample of water. In this case $c = p = 1$, leading to

$$f = c + 2 - p = 2 \quad (11.26)$$

so the abstract $\mathcal{M}_S$ is the ordinary two-dimensional space of plane geometry. The system will be discussed in terms of such standard experimental properties as the constant-pressure and constant-volume *heat capacities*,

$$C_P = T \left(\frac{\partial S}{\partial T} \right)_P, \quad C_V = T \left(\frac{\partial S}{\partial T} \right)_V \quad (11.27)$$

the constant-temperature (isothermal) and constant-entropy (adiabatic) *compressibilities*,

$$\beta_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T, \quad \beta_S = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_S \quad (11.28)$$

the *thermal expansion coefficient*,

$$\alpha_P = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P \quad (11.29)$$

and the constant-volume *heat of pressure variation*,

$$\Gamma_V = T \left(\frac{\partial S}{\partial P} \right)_V \quad (11.30)$$

As pointed out in the “ $f(f+1)/2$ rule” (11.14), only $2(3)/2 = 3$ of the above properties (and others that could be measured) are independent.

For the standard choice of thermodynamic intensities and extensities

$$R_1 = T, \quad R_2 = -P \quad (11.31a)$$

$$X_1 = S, \quad X_2 = V \quad (11.31b)$$

the corresponding vectors of $\mathcal{M}_S$ are

$$|\mathbf{R}_1\rangle = |\mathbf{T}\rangle, \quad |\mathbf{R}_2\rangle = |-\mathbf{P}\rangle = -|\mathbf{P}\rangle \quad (11.32a)$$

$$|\mathbf{X}_1\rangle = |\bar{\mathbf{R}}_1\rangle = |\mathbf{S}\rangle, \quad |\mathbf{X}_2\rangle = |\bar{\mathbf{R}}_2\rangle = |\mathbf{V}\rangle \quad (11.32b)$$

The lengths of these vectors can be readily evaluated from (11.2), together with (11.1) and the definitions (11.27)–(11.30) of measured properties; for the temperature vector, for example,

$$|\mathbf{T}| = \langle \mathbf{T} | \mathbf{T} \rangle^{1/2} = \left(\frac{\partial R_1}{\partial X_1} \right)_{X_2}^{1/2} = \left(\frac{\partial T}{\partial S} \right)_V^{1/2} = \left(\frac{T}{C_V} \right)^{1/2} \quad (11.33a)$$

and, similarly,

$$|\mathbf{P}| = (V\beta_S)^{-1/2} \quad (11.33b)$$

$$|\mathbf{S}| = (C_P/T)^{1/2} \quad (11.33c)$$

$$|\mathbf{V}| = (V\beta_T)^{1/2} \quad (11.33d)$$

The angles between vectors are similarly evaluated from (11.3). Two such characteristic angles are θ_{SV} , separating $|\mathbf{S}\rangle$ and $|\mathbf{V}\rangle$,

$$\cos \theta_{SV} = \alpha_P \left(\frac{TV}{C_P\beta_T} \right)^{1/2} \quad (11.34a)$$

and θ_{ST} , separating $|\mathbf{S}\rangle$ and $|\mathbf{T}\rangle$,

$$\cos \theta_{ST} = \left(\frac{C_V}{C_P} \right)^{1/2} \quad (11.34b)$$

and the remaining angles can be easily worked out by plane geometry. Table 11.1 summarizes the various scalar products that arise from the basis vectors (11.32a, b).

Figure 11.2 schematically exhibits these vector properties in a plane diagram. This figure depicts the temperature vector $|\mathbf{T}\rangle$ and pressure vector $|\mathbf{P}\rangle$, together with their conjugates,

TABLE 11.1 “Standard” Scalar Products for a Homogeneous Fluid

Bra/ket	$ \mathbf{T}\rangle$	$ \mathbf{-P}\rangle$	$ \mathbf{S}\rangle$	$ \mathbf{V}\rangle$
$\langle\mathbf{T} $	T/C_V	$-T/\Gamma_V$	1	0
$\langle\mathbf{-P} $	$-T/\Gamma_V$	$1/V\beta_S$	0	1
$\langle\mathbf{S} $	1	0	C_P/T	$V\alpha_P$
$\langle\mathbf{V} $	0	1	$V\alpha_P$	$V\beta_T$

the entropy vector $|\mathbf{S}\rangle$ and volume vector $|\mathbf{V}\rangle$, respectively, with the lengths and separation angles expressed in terms of the standard properties (11.27)–(11.30) defined above. Note that, in this case, the biorthogonality relationships (11.15) that define the conjugate vectors simply mean that $|\mathbf{S}\rangle$ is perpendicular to $|\mathbf{P}\rangle$, and $|\mathbf{V}\rangle$ to $|\mathbf{T}\rangle$.

An inspection of Fig. 11.2 readily reveals that the angles θ_{ST} and θ_{-PV} must satisfy

$$\cos^2 \theta_{ST} = \cos^2 \theta_{-PV} \quad (11.35a)$$

which translates to the thermodynamic identity

$$C_P/C_V = \beta_T/\beta_S \quad (11.35b)$$

Similarly, the obvious relationship

$$\sin^2 \theta_{ST} + \cos^2 \theta_{ST} = 1 \quad (11.36a)$$

becomes the thermodynamic identity (cf. Sidebar 5.5)

$$C_P = C_V + TV\alpha_P^2/\beta_T \quad (11.36b)$$

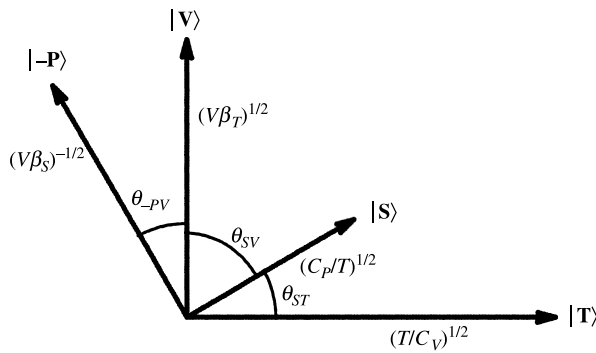


Figure 11.2 Thermodynamic vectors for a simple fluid, represented in a two-dimensional diagram in which lengths and angles are expressed in terms of experimental properties; for example, $\cos \theta_{ST} = (C_V/C_P)^{1/2}$ and $\cos \theta_{SV} = \alpha_P(TV/C_P\beta_T)^{1/2}$. The thermodynamically conjugate temperature and volume vectors $|\mathbf{T}\rangle$, $|\mathbf{V}\rangle$ are perpendicular, as are the pressure and entropy vectors $|\mathbf{P}\rangle$, $|\mathbf{S}\rangle$. A number of thermodynamic relationships among the experimental quantities can be read off directly from the diagram.

and other identities can be read off from the geometry of the figure. Moreover, the well-known thermodynamic stability condition

$$C_P \geq C_V \quad (11.37a)$$

(the impossibility of cooling a sample of water held at constant volume so as to produce an equivalent temperature increase in a similar sample held at constant pressure) can be recognized to follow from the elementary geometrical requirement

$$\cos^2 \theta_{ST} \leq 1 \quad (11.37b)$$

and so forth. Sidebar 11.3 illustrates numerical evaluation of these geometrical parameters for the elementary example of a monatomic ideal gas.

Let us return to (11.23) to obtain a group of thermodynamic identities simultaneously. From Table 11.1, the metric matrix $\underline{\underline{M}}$ and its inverse are

$$\underline{\underline{M}} = \begin{pmatrix} \langle \mathbf{T} | \mathbf{T} \rangle & \langle \mathbf{T} | -\mathbf{P} \rangle \\ \langle -\mathbf{P} | \mathbf{T} \rangle & \langle -\mathbf{P} | -\mathbf{P} \rangle \end{pmatrix} = \begin{pmatrix} T/C_V & -T/\Gamma_V \\ -T/\Gamma_V & 1/V\beta_S \end{pmatrix} \quad (11.38)$$

$$\underline{\underline{M}}^{-1} = \begin{pmatrix} \langle \mathbf{S} | \mathbf{S} \rangle & \langle \mathbf{S} | \mathbf{V} \rangle \\ \langle \mathbf{V} | \mathbf{S} \rangle & \langle \mathbf{V} | \mathbf{V} \rangle \end{pmatrix} = \begin{pmatrix} C_P/T & V\alpha_P \\ V\alpha_P & V\beta_T \end{pmatrix} \quad (11.39)$$

If we invert the 2×2 matrix of (11.39),

$$(\underline{\underline{M}}^{-1})^{-1} = \underline{\underline{M}} = M \begin{pmatrix} V\beta_T & -V\alpha_P \\ -V\alpha_P & C_P/T \end{pmatrix} \quad (11.40)$$

and compare element-by-element with (11.38), we read off the equations

$$M = \frac{T/C_V}{V\beta_T} = \frac{T/\Gamma_V}{V\alpha_P} = \frac{1/V\beta_S}{C_P/T} \quad (11.41)$$

from which follow the identities

$$C_V\beta_T = \alpha_P\Gamma_V = C_P\beta_S = \frac{T}{VM} \quad (11.42)$$

where $M = \det |\underline{\underline{M}}|$ is the Gramian. The three independent equations in (11.42) are evidently sufficient to express any of the six properties C_P , C_V , β_T , β_S , α_P , Γ_V in terms of any chosen set of *three* independent responses, as the “ $f(f+1)/2$ rule” led us to expect.

SIDEBAR 11.3: THERMODYNAMIC GEOMETRY OF A MONATOMIC IDEAL GAS

For a monatomic ideal gas, we know that (cf. Sidebar 2.2)

$$\alpha_P = \frac{1}{T} \quad (\text{S11.3-1})$$

$$\beta_T = \frac{1}{P} \quad (\text{S11.3-2})$$

and that (cf. Section 3.6.4)

$$C_P = \frac{5}{2} nR, \quad C_V = \frac{3}{2} nR \quad (\text{S11.3-3})$$

At standard-state conditions ($T = 298.15\text{K}$, $P = 1\text{ atm}$), we can therefore use (11.33a–d) and (11.35d) to evaluate (in SI units, for $n = 1\text{ mol}$):

$$|\mathbf{T}| = (T/C_V)^{1/2} = 4.889\text{ J}^{-1/2}\text{K} \quad (\text{S11.3-4a})$$

$$|\mathbf{S}| = (C_P/T)^{1/2} = 0.264\text{ J}^{1/2}\text{K}^{-1} \quad (\text{S11.3-4b})$$

$$|\mathbf{V}| = (V\beta_T)^{1/2} = 4.913 \times 10^{-4}\text{ J}^{1/2}\text{Pa}^{-1} \quad (\text{S11.3-4c})$$

$$|\mathbf{P}| = (V\beta_S)^{-1/2} = 2.627 \times 10^3\text{ J}^{-1/2}\text{Pa} \quad (\text{S11.3-4d})$$

Similarly, from (11.34b) and the geometry of Fig. 11.2, we find

$$\theta_{ST} = \theta_{-PV} = \cos^{-1}(C_V/C_P)^{1/2} = \cos^{-1}(3/5)^{1/2} = 39.2^\circ \quad (\text{S11.3-5a})$$

$$\theta_{SV} = 90^\circ - \theta_{ST} = 50.8^\circ \quad (\text{S11.3-5b})$$

As shown by these values, the pressure vector dominates the thermodynamic responsiveness of the ideal gas, conferring near “one-dimensional” character on its thermodynamic geometry.

11.4 GENERAL TRANSFORMATION THEORY IN THERMODYNAMIC METRIC SPACE

Thus far the discussion has been largely confined to a specific set of reference intensities and extensities. However, equations such as (11.10) make clear the possibility of treating more general types of thermodynamic variations. Transformations among thermodynamic variables will correspond to ordinary Euclidean vector transformations, which can therefore be treated simply and systematically in $\mathcal{M}_S$.

An application of (11.10) was already seen in (11.16), where a specified linear combination of intensive variables was found to be associated with variations of an extensive coordinate. Such linear combinations are also necessary to represent variations along a coexistence curve, or along other paths in a phase diagram that are not parallel to one of the axes. Additional incentives to describe more general variations may arise from purely experimental considerations, where the variables under practical experimental control may involve simultaneous changes of two or more “reference” variables. It is therefore desirable that general expressions be available to allow easy transformation from one thermodynamic “coordinate system” to another.

To develop such transformation properties of $\mathcal{M}_S$, we now consider the simultaneous transformations of the f reference intensities R_i that are generated by an arbitrary (but non-singular) real $f \times f$ matrix $\underline{\underline{A}}$:

$$\underline{\underline{A}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1f} \\ a_{21} & a_{22} & \cdots & a_{2f} \\ \vdots & \vdots & \vdots & \vdots \\ a_{f1} & a_{f2} & \cdots & a_{ff} \end{pmatrix} \quad (11.43a)$$

$$\det |\underline{\underline{A}}| \neq 0 \quad (11.43b)$$

The i th column of this matrix is a collection of numbers denoted by column vector $\underline{a}_i$:

$$(\underline{\underline{A}})_{ji} = (\underline{\underline{A}}^t)_{ij} = (\underline{a}_i)_j \quad (11.44)$$

The transformed intensive variables generated by this matrix are denoted $R_{\underline{a}_i}$ [cf. (11.10)] and take the form

$$R_{\underline{a}_i} = \sum_{j=1}^f (\underline{a}_i)_j R_j = \sum_{j=1}^f R_j A_{ji}, \quad i = 1, 2, \dots, f \quad (11.45)$$

The reference intensities R_i and transformed fields $R_{\underline{a}_i}$ may be gathered into column vectors $\underline{R}$ and $\underline{R}_{\underline{\underline{A}}}$, respectively,

$$\underline{R} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_f \end{pmatrix}, \quad \underline{R}_{\underline{\underline{A}}} = \begin{pmatrix} R_{\underline{a}_1} \\ R_{\underline{a}_2} \\ \vdots \\ R_{\underline{a}_f} \end{pmatrix} \quad (11.46)$$

in such a manner that the overall transformation takes the form

$$\underline{R}_{\underline{\underline{A}}} = \underline{\underline{A}}^t \underline{R} \quad (11.47)$$

In this notation, for example, (11.16) becomes [with the help of (11.24)]

$$\overline{\underline{R}} = \overline{\underline{\underline{M}}} \underline{R} \quad (11.48a)$$

that is,

$$\overline{\underline{R}} = \underline{R} \overline{\underline{\underline{M}}} \quad (11.48b)$$

where $\underline{\bar{R}}$ is the column vector of conjugate variables:

$$\underline{\bar{R}} = \begin{pmatrix} \bar{R}_1 \\ \bar{R}_2 \\ \vdots \\ \bar{R}_f \end{pmatrix} \quad (11.49)$$

Given a set $R_{\underline{a}_i}$ of transformed intensive variables, we should naturally seek expressions for the corresponding *conjugate* variables $\bar{R}_{\underline{a}_i}$, which must satisfy

$$R_{\underline{a}_i} = \left(\frac{\partial U}{\partial \bar{R}_{\underline{a}_i}} \right)_{\underline{\bar{R}}_{\underline{A}}} , \quad i = 1, 2, \dots, f \quad (11.50)$$

The considerations that led to (11.16) and (11.48) now give

$$\underline{\bar{R}}_{\underline{A}} = \underline{\bar{M}}_{\underline{A}} \underline{R}_{\underline{A}} \quad (11.51)$$

where $\underline{\bar{R}}_{\underline{A}}$ is the column vector of the $\bar{R}_{\underline{a}_i}$ and $\underline{\bar{M}}_{\underline{A}}$ is the metric matrix for the transformed intensities

$$(\underline{\bar{M}}_{\underline{A}})_{ij} = \langle R_{\underline{a}_i} | R_{\underline{a}_j} \rangle, \quad i, j = 1, 2, \dots, f \quad (11.52)$$

with corresponding inverse $\underline{\bar{M}}_{\underline{A}}$. From (11.12), we can see that the transformed metric matrix is simply

$$\underline{\bar{M}}_{\underline{A}} = \underline{A}^t \underline{M} \underline{A} \quad (11.53)$$

Let us now introduce the *conjugate* (inverse transpose) matrix $\underline{\bar{A}}$,

$$\underline{\bar{A}} = (\underline{A}^{-1})^t = (\underline{A}^t)^{-1} \equiv \underline{A}^{-t} \quad (11.54)$$

which must exist for any nonsingular $\underline{A}$. In general, either a left-conjugate $\underline{\bar{A}}_L$ or a right-conjugate $\underline{\bar{A}}_R$ can be defined [cf. (S9.1-5)]. From the product rule (S9.1-8c) for inverse matrices, we can rewrite (11.53) as

$$\underline{\bar{M}}_{\underline{A}} = (\underline{A}^t \underline{M} \underline{A})^{-1} = \underline{A}^{-1} \underline{M}^{-1} (\underline{A}^t)^{-1} = \underline{\bar{A}}_R^t \underline{\bar{M}}_{\underline{A}} \underline{\bar{A}}_L \quad (11.55)$$

so that (11.51) becomes, with (11.47),

$$\underline{\bar{R}}_{\underline{A}} = \underline{\bar{A}}_R^t \underline{\bar{M}}_{\underline{A}} \underline{\bar{A}}_L \underline{A}^t \underline{R} = \underline{\bar{A}}_R^t \underline{\bar{M}}_{\underline{A}} \underline{A}^{-t} \underline{A}^t \underline{R} = \underline{\bar{A}}_R^t \underline{\bar{M}} \underline{R} \quad (11.56)$$

From (11.48a), we therefore obtain

$$\underline{\underline{\overline{R}}}_{\underline{\underline{A}}} = \underline{\underline{\overline{A}}}_{\underline{\underline{R}}}^t \underline{\underline{\overline{R}}} \quad (11.57)$$

which is apparently the “conjugate” of (11.47).

The conjugacy relation (11.54) for matrices is readily seen to have the properties [cf. (11.21)]

$$\underline{\underline{\overline{\overline{A}}}} = \underline{\underline{A}} \quad (11.58)$$

$$(\underline{\underline{\overline{A} \overline{B}}}) = \underline{\underline{\overline{A} \overline{B}}} \quad (11.59)$$

The property (11.59) then permits one to easily rewrite thermodynamic equations in their “conjugate” form; for example, (11.53) becomes

$$\underline{\underline{\overline{M}}}_{\underline{\underline{A}}} = \underline{\underline{\overline{A}}}_{\underline{\underline{M} \overline{A}}}^t \underline{\underline{\overline{M} \overline{A}}} \quad (11.60)$$

and so forth.

For completeness, we remark finally that scalar products among the transformed variables become completely analogous to the corresponding equations (11.9a), (11.15), (11.22), namely,

$$\langle \underline{\underline{\mathbf{R}}}_{\underline{\underline{a}}_i} | \underline{\underline{\mathbf{R}}}_{\underline{\underline{a}}_j} \rangle = \left(\frac{\partial \underline{\underline{R}}_{\underline{\underline{a}}_i}}{\partial \underline{\underline{R}}_{\underline{\underline{a}}_j}} \right)_{\underline{\underline{\overline{R}}}_{\underline{\underline{A}}}} = (\underline{\underline{\overline{M}}}_{\underline{\underline{A}}})_{ij} \quad (11.61a)$$

$$\langle \underline{\underline{\overline{\mathbf{R}}}}_{\underline{\underline{a}}_i} | \underline{\underline{\overline{\mathbf{R}}}}_{\underline{\underline{a}}_j} \rangle = \left(\frac{\partial \underline{\underline{\overline{R}}}_{\underline{\underline{a}}_i}}{\partial \underline{\underline{\overline{R}}}_{\underline{\underline{a}}_j}} \right)_{\underline{\underline{\mathbf{R}}}_{\underline{\underline{A}}}} = (\underline{\underline{\overline{\overline{M}}}}_{\underline{\underline{A}}})_{ij} \quad (11.61b)$$

$$\langle \underline{\underline{\overline{\mathbf{R}}}}_{\underline{\underline{a}}_i} | \underline{\underline{\mathbf{R}}}_{\underline{\underline{a}}_j} \rangle = \delta_{ij} \quad (11.61c)$$

11.5 SATURATION PROPERTIES ALONG THE VAPOR-PRESSURE CURVE

Let us illustrate the general ideas of the previous section with a specific example. Suppose that in place of the standard isobaric properties C_P and α_P we now wish to consider the analogous *saturation properties* C_σ and α_σ ,

$$C_\sigma = T \left(\frac{\partial S}{\partial T} \right)_\sigma \quad (11.62a)$$

$$\alpha_\sigma = \frac{1}{V} \left(\frac{\partial S}{\partial T} \right)_\sigma \quad (11.62b)$$

for a saturated fluid along its vapor-pressure curve [$\sigma = \text{constant}$; cf. (7.28a)].

To deal with such properties, we wish to transform from the former reference intensities T , $-P$,

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} T \\ -P \end{pmatrix} \quad (11.63)$$

to a new set

$$\begin{pmatrix} R'_1 \\ R'_2 \end{pmatrix} = \begin{pmatrix} T \\ \sigma \end{pmatrix} \quad (11.64)$$

in which the “coexistence coordinate” σ is itself a variable. If $\underline{\underline{A}}$ is the transformation matrix from old intensities to new,

$$\begin{pmatrix} R'_1 \\ R'_2 \end{pmatrix} = \underline{\underline{A}}^t \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \quad (11.65)$$

then the new metric matrix $\underline{\underline{M'}}$ is obtained from the old by (11.53):

$$\underline{\underline{M'}} = \underline{\underline{A}}^t \underline{\underline{M}} \underline{\underline{A}} \quad (11.66)$$

Suppose that γ_σ represents the slope of the coexistence curve in a conventional PT phase diagram, so that, by definition,

$$\left(\frac{\partial P}{\partial T} \right)_\sigma = \gamma_\sigma \quad (11.67)$$

The thermodynamic vector $|\sigma\rangle$ for the coexistence coordinate σ is therefore composed from the old intensive vectors $|\mathbf{T}\rangle$, $|\mathbf{-P}\rangle$ by the equation

$$|\sigma\rangle = \gamma_\sigma |\mathbf{T}\rangle + |\mathbf{-P}\rangle \quad (11.68)$$

(See Sidebar 11.4 for some dimensional aspects of this and similar vector equations.) The transformation matrix $\underline{\underline{A}}$ of (11.65) is therefore

$$\underline{\underline{A}} = \begin{pmatrix} 1 & \gamma_\sigma \\ 0 & 1 \end{pmatrix} \quad (11.69a)$$

with conjugate matrix $\overline{\underline{\underline{A}}} (= \overline{\underline{\underline{A}}}_L = \overline{\underline{\underline{A}}}_R)$ given by

$$\overline{\underline{\underline{A}}} = \begin{pmatrix} 1 & 0 \\ -\gamma_\sigma & 1 \end{pmatrix} \quad (11.69b)$$

The extensities X'_1 and X'_2 conjugate to R'_1 and R'_2 are evaluated from (11.57) as

$$\begin{pmatrix} \bar{T} \\ \bar{\sigma} \end{pmatrix} = \begin{pmatrix} X'_1 \\ X'_2 \end{pmatrix} = \bar{\underline{\underline{A}}}^t \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \bar{\underline{\underline{A}}}^t \begin{pmatrix} S \\ V \end{pmatrix} = \begin{pmatrix} S - \gamma_\sigma & V \\ V & \end{pmatrix} \quad (11.70)$$

Of course, γ_σ is merely a numerical constant for the particular thermodynamic state under consideration.

The properties (11.62) of interest involve derivatives with respect to T , σ (the R'_i), which occur in the conjugate metric $(\bar{\underline{\underline{M}}}')^{-1} = \bar{\underline{\underline{M}}}'$. According to (11.61b), this matrix is evaluated as

$$\begin{aligned} \bar{\underline{\underline{M}}}' &= \begin{pmatrix} \left(\frac{\partial X'_1}{\partial R'_1} \right)_{R'_2} & \left(\frac{\partial X'_2}{\partial R'_1} \right)_{R'_2} \\ \left(\frac{\partial X'_1}{\partial R'_2} \right)_{R'_1} & \left(\frac{\partial X'_2}{\partial R'_2} \right)_{R'_1} \end{pmatrix} \\ &= \begin{pmatrix} C_\sigma/T - V\gamma_\sigma\alpha_\sigma & V\alpha_\sigma \\ V\alpha_\sigma & \left(\frac{\partial V}{\partial \sigma} \right)_T \end{pmatrix} \end{aligned} \quad (11.71)$$

However, $\bar{\underline{\underline{M}}}'$ can also be written from the conjugate of (11.66) [using the standard form (11.39) of $\bar{\underline{\underline{M}}} = \underline{\underline{M}}^{-1}$] as

$$\begin{aligned} \bar{\underline{\underline{M}}}' &= \bar{\underline{\underline{A}}}^t \bar{\underline{\underline{M}}} \bar{\underline{\underline{A}}} = \begin{pmatrix} 1 & -\gamma_\sigma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} C_P/T & V\alpha_P \\ V\alpha_P & V\beta_T \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\gamma_\sigma & 1 \end{pmatrix} \\ &= \begin{pmatrix} C_P/T - 2V\gamma_\sigma\alpha_P + V\gamma_\sigma^2\beta_T & V\alpha_P - V\gamma_\sigma\beta_T \\ V\alpha_P - V\gamma_\sigma\beta_T & V\beta_T \end{pmatrix} \end{aligned} \quad (11.72)$$

Element-by-element comparison of (11.71), (11.72) then gives the desired identities

$$\alpha_\sigma = \alpha_P - \gamma_\sigma\beta_T \quad (11.73a)$$

$$C_\sigma = C_P - TV\gamma_\sigma\alpha_P \quad (11.73b)$$

$$(\partial V/\partial \sigma)_T = V\beta_T \quad (11.73c)$$

The thermodynamic identities (11.73a, b) can be derived by more conventional means [see, e.g., J. T. Rowlinson. *Liquids and Liquid Mixtures* (Academic Press, New York, 1959), Chap. 2], but their derivation here illustrates rather general and systematic matrix-algebraic procedures that remain effective when traditional methods are unduly cumbersome.

SIDEBAR 11.4: DIMENSIONALITY OF VECTOR EQUATIONS

Certain dimensional questions may be raised by vector equations such as (11.68), which seem to “mix apples and oranges.” Each vector $|\mathbf{R}_i\rangle$ carries physical units, because its “length” as

given by (11.2) depends on the units in which the response function $\partial R_i / \partial X_i$ is measured]. It may appear superficially that the general vector equation (11.10) [or its representative example, (11.68)] is therefore rendered meaningless on dimensional grounds. That this is incorrect can be seen from (11.11), where the coefficients of the linear combination of vectors are themselves seen to “carry units” in a manner that removes any potential inconsistency.

One can appreciate in (11.68) the intrinsic physical significance that the coexistence coordinate σ possesses in an actual phase diagram, even though it involves simultaneous changes of coordinates T, P having different units (so that “units” of σ would be difficult to discuss). If $|\mathbf{T}\rangle$ and $|\mathbf{P}\rangle$ were measured in different units, the numerical value of the coefficient γ_σ would readjust to represent the same *physical* coordinate, so that no mathematical or physical inconsistency could result. As remarked previously, lengths in $\mathcal{M}$ s depend (consistently) on the chosen units of measured response functions, whereas internal angles are dimensionless invariants of the system.

Note also that (11.67) determines the coefficients in (11.68) only up to a multiplicative factor, which has been chosen so that V (rather than a multiple thereof) becomes the remaining conjugate variable in (11.70).

11.6 SELF-CONJUGATE AND NORMAL RESPONSE MODES

Arbitrarily chosen intensive vectors $|\mathbf{R}_{\underline{a}_i}\rangle$ are in general neither orthogonal nor normalized. It is therefore of interest to identify a particular choice of $\underline{\underline{A}}$ for which these vectors become an *orthonormal* set. Such “unit-like” response vectors, denoted as $|\mathbf{E}_i\rangle$, can be expressed as usual as linear transformations of standard reference intensities

$$|\mathbf{E}_i\rangle = \sum_{j=1}^f (\underline{\underline{A}}^t)_{ij} |\mathbf{R}_j\rangle \quad (11.74)$$

and satisfy the characteristic orthonormality relationship

$$\langle \mathbf{E}_i | \mathbf{E}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (11.75)$$

We note first that the matrix $\underline{\underline{A}}$ leading to vectors with the property (11.75) is not uniquely defined, since if $\underline{\underline{Q}}$ is any *orthogonal* matrix satisfying [cf. (S9.1-13)]

$$\underline{\underline{Q}}^t = \underline{\underline{Q}}^{-1} \quad (11.76)$$

then the substitution

$$\underline{\underline{A}}^t \longrightarrow \underline{\underline{Q}} \underline{\underline{A}}^t \quad (11.77)$$

in (11.74) leaves the orthonormality property (11.75) unaffected. When the equations (11.74) are inverted for the reference intensities $|\mathbf{R}_i\rangle$,

$$|\mathbf{R}_i\rangle = \sum_{j=1}^f (\underline{\underline{A}}^{-1})_{ji} |\mathbf{E}_j\rangle \quad (11.78)$$

it is apparent that the substitution (11.77) cannot affect the scalar products of the $|\mathbf{R}_i\rangle$, which are the only physical observables in the formalism. The orthogonal matrix $\underline{\underline{Q}}$ can be chosen for convenience in simplifying $\underline{\underline{A}}$, without loss of generality; it merely reflects the usual freedom to arbitrarily rotate Cartesian axes in an ordinary Euclidean space.

The orthonormality property (11.75) establishes, in conjunction with (11.15), that the normal intensities are *self-conjugate*,

$$|\mathbf{E}_i\rangle = |\overline{\mathbf{E}}_i\rangle \quad (11.79a)$$

$$E_i = \left(\frac{\partial U}{\partial E_i} \right)_E, \quad i = 1, 2, \dots, f \quad (11.79b)$$

i.e., they correspond to particular thermodynamic variables $\{R_{\underline{a}i}\}$ for which $R_{\underline{a}i}$ and $\overline{R}_{\underline{a}i}$ coincide for each i ,

$$\underline{R}_{\underline{A}} = \overline{R}_{\underline{A}} \quad (11.80)$$

Such self-conjugate response “modes” are no longer associated preferentially with intensive or extensive character and will be found to have special utility in expressing the Euclidean structure of $\mathcal{M}_S$ as direct analogs of its “Cartesian axes.” Equation (11.51) shows that the desired matrix $\underline{\underline{A}}$ leading to self-conjugacy (11.80) must satisfy

$$\overline{\underline{\underline{M}}}_{\underline{\underline{A}}} = \underline{\underline{M}}_{\underline{\underline{A}}} = \underline{\underline{1}} \quad (11.81)$$

or, from (11.53) and (11.55),

$$\underline{\underline{A}}_{\underline{\underline{L}}}^{-1} \underline{\underline{A}}_{\underline{\underline{R}}}^{-1} = \underline{\underline{M}} \quad (11.82a)$$

$$\underline{\underline{A}} \underline{\underline{A}}^t = \overline{\underline{\underline{M}}} \quad (11.82b)$$

The invariance of (11.82a, b) under the orthogonal transformation (11.77) is again manifest. As shown by (11.82b), the desired $\underline{\underline{A}}$ is a kind of “square root” of $\overline{\underline{\underline{M}}}$, or any “rotated” (orthogonally transformed) version thereof, closely associated with intrinsic *eigenvectors* of $\overline{\underline{\underline{M}}}$ and $\underline{\underline{M}}$ (Sidebar 9.1).

Let us now introduce the orthogonal matrix $\underline{\underline{Q}}$, whose columns $\underline{q}_i$,

$$(\underline{\underline{Q}})_{ji} = (\underline{q}_i)_j, \quad (\underline{\underline{Q}}^t = \underline{\underline{Q}}^{-1}) \quad (11.83)$$

are *eigenvectors* of $\underline{\underline{M}}$, satisfying

$$\underline{\underline{M}} \underline{q}_i = e_i \underline{q}_i, \quad i = 1, 2, \dots, f \quad (11.84)$$

The orthogonal matrix $\underline{\underline{Q}}$ transforms the real symmetric metric matrix $\underline{\underline{M}}$ to its diagonal matrix of *eigenvalues* $\underline{\underline{e}}$:

$$\underline{\underline{Q}}^t \underline{\underline{M}} \underline{\underline{Q}} = \underline{\underline{e}} \quad (11.85a)$$

$$(\underline{\underline{e}})_{ij} = e_i \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (11.85b)$$

It follows from (11.85) [cf. (S9.1-25)] that

$$\underline{\underline{M}}^{-1} = (\underline{\underline{Q}} \underline{\underline{e}} \underline{\underline{Q}}^t)^{-1} = (\underline{\underline{e}}^{-1/2} \underline{\underline{Q}})(\underline{\underline{e}}^{-1/2} \underline{\underline{Q}})^t \quad (11.86)$$

where the (diagonal) square-root matrix $\underline{\underline{e}}^{-1/2}$ is defined as

$$(\underline{\underline{e}}^{-1/2})_{ij} = e_i^{-1/2} \delta_{ij} \quad (11.87)$$

and where the positive square root of each real non-negative e_i is to be taken. Equations (11.82), (11.86) evidently have the solution

$$\underline{\underline{A}} = \underline{\underline{Q}} \underline{\underline{e}}^{-1/2} \quad (11.88)$$

which is unique up to orthogonal transformation (11.77).

In terms of the eigenvalues $\{e_i\}$ and eigenvectors $\{\underline{\underline{q}}_i\}$ of the metric matrix $\underline{\underline{M}}$, the “normal-mode” vectors are therefore

$$|\mathbf{E}_i\rangle = \sum_{j=1}^f e_i^{-1/2} (\underline{\underline{q}}_i)_j |\mathbf{R}_j\rangle, \quad i = 1, 2, \dots, f \quad (11.89)$$

that is,

$$\underline{\underline{a}}_i = e_i^{-1/2} \underline{\underline{q}}_i \quad (11.90)$$

Each reference $|\mathbf{R}_i\rangle$ or $|\bar{\mathbf{R}}_i\rangle$ can also be expanded in normal modes $|\mathbf{E}_i\rangle$:

$$|\mathbf{R}_i\rangle = \sum_{j=1}^f e_j^{1/2} (\underline{\underline{q}}_j)_i |\mathbf{E}_j\rangle \quad (11.91a)$$

$$|\bar{\mathbf{R}}_i\rangle = \sum_{j=1}^f e_j^{-1/2} (\underline{\underline{q}}_j)_i |\mathbf{E}_j\rangle \quad (11.91b)$$

Equations (11.91a, b) provide useful starting points for obtaining explicit numerical representations of the abstract $|\mathbf{R}_i\rangle$ or $|\bar{\mathbf{R}}_i\rangle$ as ordinary column vectors. For this purpose, the normal vectors $|\mathbf{E}_i\rangle$ are conveniently represented by unit vectors of an orthogonal

Cartesian coordinate system,

$$|\mathbf{E}_1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |\mathbf{E}_2\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, |\mathbf{E}_f\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad (11.92)$$

In terms of these coordinates, the reference $|\mathbf{R}_i\rangle$ or $|\overline{\mathbf{R}}_i\rangle$ vectors are easily constructed from eigenvalues and eigenvectors of the metric $\underline{\underline{\mathbf{M}}}$ according to the prescription (11.91a, b).

As suggested in Sidebar 9.1 (cf. Sidebar 10.3), the deepest characterization of a matrix is in terms of its invariant *eigen*properties. Hence, the most fundamental descriptors of $\mathcal{M}_S$ are expected to originate from eigenvalues $\{e_i\}$ and eigenvectors $\{\underline{\mathbf{q}}_i\}$ of the underlying thermodynamic metric $\underline{\underline{\mathbf{M}}}$, particularly through the fundamental “normal modes of thermodynamic response” $\{|\mathbf{E}_i\rangle\}$, as expressed (in a chosen set of response variables) by (11.89). The metric eigenvalues $\{e_i\}$ are most important in this respect, being invariant to *all* arbitrary theoretical or experimental choices (except selection of a physical unit system) for representing $\mathcal{M}_S$. In contrast, the vectors chosen to represent the $\underline{\mathbf{q}}_i$ (or equivalently, to represent the $|\mathbf{E}_i\rangle$) remain open to details of selecting and vectorially representing the measurement variables of $\mathcal{M}_S$, corresponding to overall freedom (11.77) of orthogonal transformation. Such freedom leads merely to overall rotation of the Cartesian “*eigen*coordinate” axis system chosen in (11.92), which seems a preferred (but not obligatory) coordinate choice for many theoretical purposes. We are thereby led to focus on eigenvalues $\{e_i\}$ or other invariant “coupling” parameters between chosen thermodynamic responses [e.g., the angles θ_{ij} of (11.3)] as preferred geometrical descriptors of thermodynamic responses in $\mathcal{M}_S$.

11.7 GEOMETRICAL CHARACTERIZATION OF COMMON FLUIDS

To illustrate the abstract geometrical concepts of previous sections, let us consider a collection of common laboratory liquids:

acetone (CH_3COCH_3)
 benzene (C_6H_6)
 carbon tetrachloride (CCl_4)
 ethanol ($\text{C}_2\text{H}_5\text{OH}$)
 diethyl ether ($\text{C}_2\text{H}_5\text{OC}_2\text{H}_5$)
 water (H_2O)
 liquid mercury (Hg)

for which experimental C_P , β_T , α_P values at 20°C , 1 atm are available, as shown in Table 11.2. For comparison, we also include values for the (physically unrealistic) monatomic ideal gas limit, as previously described in Sidebar 11.3.

TABLE 11.2 Measured Thermodynamic Properties (in SI Units) of Some Common Fluids at 20°C, 1 atm: Molar Heat Capacity C_P , Isothermal Compressibility β_T , Coefficient of Thermal Expansion α_P , and Molar Volume $\bar{V}$, with Monatomic Ideal Gas Values (cf. Sidebar 11.3) Shown for Comparison

Fluid	C_P (J mol ⁻¹ K ⁻¹)	β_T (Pa ⁻¹)	α_P (K ⁻¹)	$\bar{V}$ (m ³)
Acetone	128.31	1.275×10^{-9}	1.487×10^{-3}	46.00×10^{-6}
Benzene	132.69	0.944×10^{-9}	1.237×10^{-3}	68.66×10^{-6}
Carbon tetrachloride	129.35	1.034×10^{-9}	1.236×10^{-3}	245.37×10^{-6}
Ethanol	110.26 ^a	1.098×10^{-9}	1.120×10^{-3}	36.36×10^{-6}
Ether	167.77 ^a	1.865×10^{-9}	1.656×10^{-3}	52.88×10^{-6}
Water	75.29	0.459×10^{-9}	0.207×10^{-3}	18.05×10^{-6}
Mercury	27.90	3.850×10^{-11}	0.182×10^{-3}	14.81×10^{-6}
Ideal gas ^b	20.79	9.869×10^{-6}	3.411×10^{-3}	24.05×10^{-3}

^aInterpolated from nearby temperature values.

^bMonatomic ideal gas (cf. Sidebar 11.3).

Table 11.3 displays a variety of geometrical descriptors for each fluid (extending results presented in Sidebar 11.3 for a monatomic ideal gas). These descriptors include the lengths of various thermodynamic vectors ($|\mathbf{T}|$, $|\mathbf{P}|$, $|\mathbf{S}|$, $|\mathbf{V}|$), θ_{SV} coupling angle ($=\theta_{TP}$; cf. Fig. 11.2), Gramian M , and “minor” eigenvalue e_2 of the metric $\underline{M}$. Sidebar 11.5 describes numerical evaluation of e_2 .

As shown in Table 11.2, the unfortunate scale of SI units leads to rather wide ranges of numerical values for elementary C_P , β_P , α_P properties, with the scale of mechanical compressibility responses differing by about nine orders of magnitude from that for thermal heat capacity responses. These scale disparities lead in turn to gross mismatches of geometrical magnitudes shown in Table 11.3, with, e.g., vectors $|\mathbf{P}|$ and $|\mathbf{V}|$ differing by about 13 orders of magnitude!

Although it is impractical to attempt graphical representations of $\mathcal{M}_S$ in any single plot that employs SI-valued coordinate values, we can observe that the SI-valued thermal $|\mathbf{T}|$, $|\mathbf{S}|$ vectors are sensibly “sized” for plotting purposes, as is also the minor eigenvalue e_2 of $\underline{M}$. Furthermore, the coupling angle θ_{SV} is completely independent of SI-based numerics, as are also the angles (such as θ_{ST} , θ_{PV} , θ_{TP}) that may be readily inferred from θ_{SV} by elementary vector geometry. We shall therefore focus on graphical representations that depict pure angular relationships, scale-independent comparisons of a single response mode for different fluids, or thermal $|\mathbf{S}|$, $|\mathbf{T}|$ vectors of comparable SI-based lengths, in order to minimize the geometrical imbalances associated with vastly disparate SI-based scales for thermal and mechanical responses.

SIDEBAR 11.5: EIGENVALUES OF A 2×2 MATRIX

For a 2×2 real symmetric matrix $\underline{A}$

$$\underline{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \quad (\text{S11.5-1})$$

TABLE 11.3 Geometrical Descriptors for the Fluids of Table 11.2, Showing the Thermodynamic Vector Lengths [(11.33a–d)], Separation Angle θ_{SV} [(11.34a)], Gramian M [(11.41)], and “Minor” Eigenvalue e_2 of $\underline{\underline{M}}$ [(11.84)] (θ_{SV} in Degrees, Others in SI Units)

Fluid	$ T $ ($J^{-1/2}K$)	$ P $ ($J^{-1/2}Pa$)	$ S $ ($J^{1/2}K^{-1}$)	$ V $ ($J^{1/2}Pa^{-1}$)	θ_{SV} ($^{\circ}$)	M ($J^2K^{-2}Pa^{-2}$)	e_2 ($JK^{-1}Pa^{-1}$)
Acetone	1.672	4.565×10^6	0.662	2.422×10^{-7}	64.73	4.764×10^{13}	2.284
Benzene	1.712	4.523×10^6	0.673	2.546×10^{-7}	60.27	4.520×10^{13}	2.210
Carbon tetrachloride	3.564	4.700×10^6	0.664	5.037×10^{-7}	24.98	5.006×10^{13}	2.268
Ethanol	1.729	5.306×10^6	0.613	1.998×10^{-7}	70.59	7.487×10^{13}	2.659
Ether	1.422	3.425×10^6	0.757	3.140×10^{-7}	68.37	2.050×10^{13}	1.747
Water	1.980	11.027×10^6	0.507	0.910×10^{-7}	85.35	4.734×10^{14}	3.893
Mercury	3.483	44.999×10^6	0.309	0.239×10^{-7}	68.56	2.127×10^{16}	10.510
Ideal gas ^a	4.849	2.649×10^3	0.264	4.873×10^{-4}	50.77	9.902×10^7	14.102

^aMonatomic ideal gas (cf. Sidebar 11.3).

the eigenvalue problem [cf. Sidebar 9.1, (S9.1-15)] for an unknown eigenvalue λ and eigenvector $\underline{y}$ can be written as

$$(\underline{A} - \lambda \underline{1})\underline{y} = \begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{pmatrix} \underline{y} = 0 \quad (\text{S11.5-2})$$

which is only possible if [cf. Eqs. (10.27), (10.28)]

$$\det \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} = 0 \quad (\text{S11.5-3})$$

Expansion of the determinant (S11.5-3) leads to a quadratic equation of the unknown λ ,

$$\lambda^2 - \lambda(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}^2 = 0 \quad (\text{S11.5-4})$$

with general solution

$$\lambda_{\pm} = \frac{1}{2} \left\{ (a_{11} + a_{22}) \pm [(a_{11} + a_{22})^2 - 4a_{11}a_{22} + 4a_{12}^2]^{1/2} \right\} \quad (\text{S11.5-5})$$

which can be written as

$$\lambda_{\pm} = \frac{a_{11} + a_{22}}{2} \pm \frac{a_{11} - a_{22}}{2} \left[1 + \left(\frac{2a_{12}}{a_{11} - a_{22}} \right)^2 \right]^{1/2} \quad (\text{S11.5-6})$$

For small $|x|$, we can approximate $(1 + x^2)^{1/2}$ with a Taylor series expansion [cf. Sidebar 1.7, Exercise (d)]:

$$(1 + x^2)^{1/2} \simeq 1 + \frac{1}{2}x^2 - \frac{1}{8}x^4 + \cdots \quad (x^2 \ll 1) \quad (\text{S11.5-7})$$

For $|a_{12}| \ll |a_{11} - a_{22}|$, we can therefore approximate (S11.5-6) as

$$\lambda_{\pm} \simeq \frac{a_{11} + a_{22}}{2} \pm \frac{a_{11} - a_{22}}{2} \pm \frac{1}{4} \left[\frac{(2a_{12})^2}{(a_{11} - a_{22})} \right] \cdots \quad (\text{S11.5-8})$$

In the extreme case discussion, with

$$a_{11} \gg |a_{12}| \gg a_{22} > 0 \quad (\text{S11.5-9})$$

the lower (λ_-) eigenvalue is well approximated by the leading correction

$$\lambda_- \simeq a_{22} - \frac{a_{12}^2}{a_{11} - a_{22}} \quad (\text{S11.5-10})$$

which was used to estimate the e_2 values of Tables 11.3.

Let us first compare the thermodynamic vector lengths for these fluids. Because these lengths “carry units,” we discuss each vector type in turn, comparing different liquids on the common SI-based scale of responsiveness for the specified thermodynamic variable.

Figure 11.3 compares the *temperature* vectors for each fluid, ranging in length from 1.4 (ether) to 4.9 (ideal gas) in SI units. The length $|\mathbf{T}|$ measures the “ T -sensitivity” of the fluid, i.e., the extent to which a slight addition of heat causes a sharp change of temperature (under isochoric conditions). As is common experience, a certain quantity of heating has a much sharper heating effect on dry air than on ocean water, as reflected in the larger $|\mathbf{T}|$ value for ideal gas than for water. Liquid mercury and carbon tetrachloride are also seen to be anomalously heat-sensitive compared with many common organic liquids.

Figure 11.4 compares the length of *pressure* vectors for the fluids, which exhibit values ranging from about 2600 (ideal gas) up to about 11 million (H_2O) or about 45 million (Hg) in SI units. The length $|\mathbf{P}|$ measures the pressure sensitivity of the fluid, i.e., the extent to which a slight volume decrease causes a sharp pressure increase (under adiabatic conditions). As expected, this vector is much shorter for a highly compressible gas than for a highly incompressible liquid such as water or liquid mercury. As mentioned above, the great elongation of $|\mathbf{P}|$ vectors compared with $|\mathbf{T}|$ vectors has no intrinsic physical significance, but results from the awkwardly small SI pressure unit (pascal: approximately the pressure exerted by a mosquito on take-off). Some further aspects of dimensional “imbalance” in the SI representation of thermodynamic relationships are discussed in Sidebar 11.6.

Figure 11.5 compares the fluid *entropy* vectors, whose lengths range from about 0.25 (ideal gas) to about 0.75 (ether). As expected, the entropy vectors exhibit an approximate “inverted” or complementary (conjugate) relationship to the corresponding $|\mathbf{T}|$ vectors of Fig. 11.3. The length of each $|\mathbf{S}|$ vector reflects resistance to attempted temperature change (under isobaric conditions), i.e., the “capacity” to absorb heat with little temperature response. The lack of strict inversion order with respect to the $|\mathbf{T}|$ lengths of Table 11.3 reflects subtle heat-capacity variations between isochoric and isobaric conditions, as quantified in the heat-capacity or compressibility ratio

$$\gamma \equiv C_P/C_V = \beta_T/\beta_S \quad (11.93a)$$

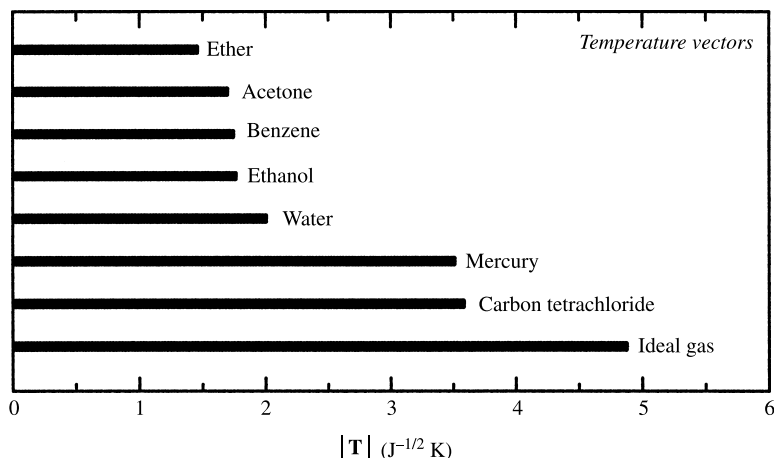


Figure 11.3 Temperature vectors (SI units) of common laboratory liquids at 20°C, 1 atm.

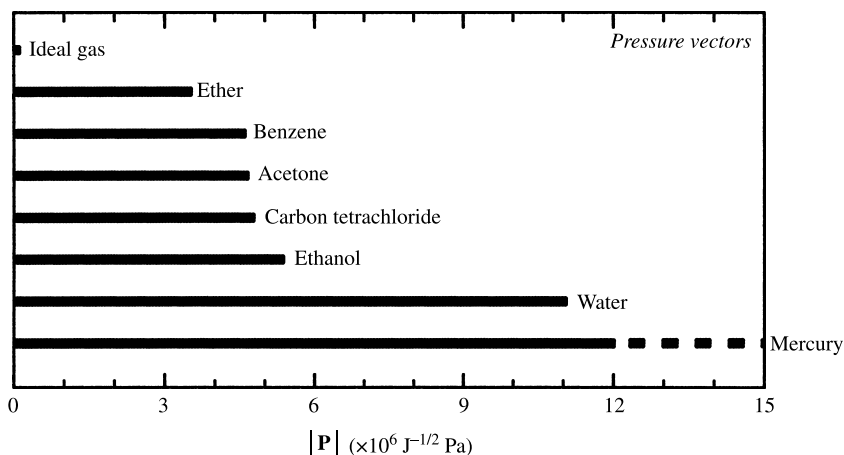


Figure 11.4 Pressure vectors (SI units) of common laboratory liquids at 20°C, 1 atm. The mercury vector extends off scale to 45 SI units, and the ideal gas vector is 0.003 SI units.

This ratio is easily seen to be given by a thermodynamic angle relationship [cf. (11.36), (11.37)]

$$\gamma = \frac{1}{\sin^2 \theta_{SV}} \quad (11.93b)$$

to be discussed below.

From a molecular viewpoint, we know that heat capacity is closely connected to internal modes of molecular vibration. According to the classical equipartition theorem (Sidebar 3.8), a nonlinear polyatomic molecule of N_{at} atoms has $n_{\text{modes}} = 3N_{\text{at}} - 6$ independent internal “modes” of vibration, each of which would contribute *equally* to heat capacity

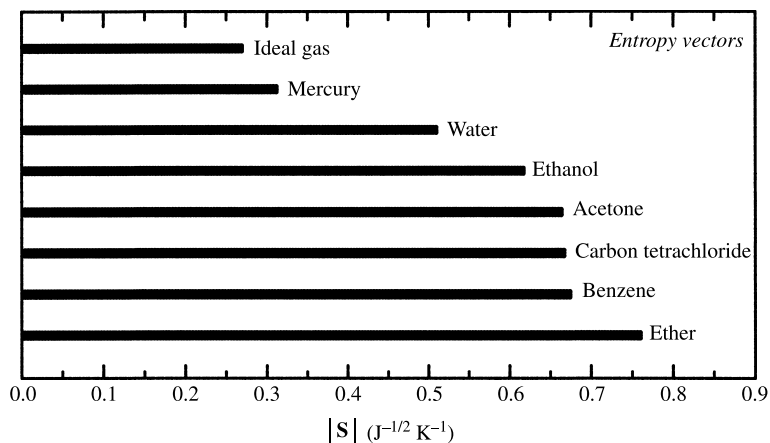


Figure 11.5 Entropy vectors (SI units) of common laboratory liquids at 20°C, 1 atm.

(if classical mechanics were true!). From this molecular viewpoint, it is more incisive to compare heat capacities on a “per mode” basis,

$$|\mathbf{S}_{\text{mode}}| \equiv \frac{|\mathbf{S}|}{3N_{\text{at}} - 6} \quad (11.94)$$

to remove the intrinsic bias toward higher heat capacity (and greater $|\mathbf{S}|$ length) in larger molecules. In terms of this descriptor, water (0.169) has conspicuously high $|\mathbf{S}_{\text{mode}}|$ capacity compared with carbon tetrachloride (0.074), ethanol (0.029), acetone (0.028), benzene (0.022), and ether (0.019).

Figure 11.6 compares the fluid *volume* vectors, whose lengths are assigned extremely small SI values ranging from about 2×10^{-8} for mercury up to about 5×10^{-4} for an ideal gas. As expected the length of the $|\mathbf{V}|$ vector corresponds to the volume sensitivity (“compressibility”) with respect to small pressure changes (under isothermal conditions). The order of lengths of $|\mathbf{V}|$ vectors thus tends to be the inverse (complement) of that of the conjugate $|\mathbf{P}|$ vectors in Fig. 11.3, but with ordered variations that again reflect subtle compressibility variations between adiabatic and isothermal conditions, as expressed by γ [(11.93a, b)].

Figure 11.7 graphically compares the smaller eigenvalue e_2 of the metric matrix, which signals the approach to metric singularity and dimensional collapse. This eigenvalue plays the role of an “order parameter” or “criticality index” with respect to approach to a critical point. Numerical SI values of the e_2 -criticality index range from 1.747 (ether) up to 10.510 (Hg), with ether evidently in closest proximity to its critical limit among these liquids. [Interpretation of the ideal gas value (14.102) is problematic, because such a gas has *no* critical point; however, the large value correctly suggests that this mythical species is indeed “farther” from its critical point than any real substance.]

Figure 11.8 illustrates the direct correlation of the metric eigenvalue e_2 with the critical temperature T_c for each fluid. A similar correlation (not shown) is found to relate e_2 to

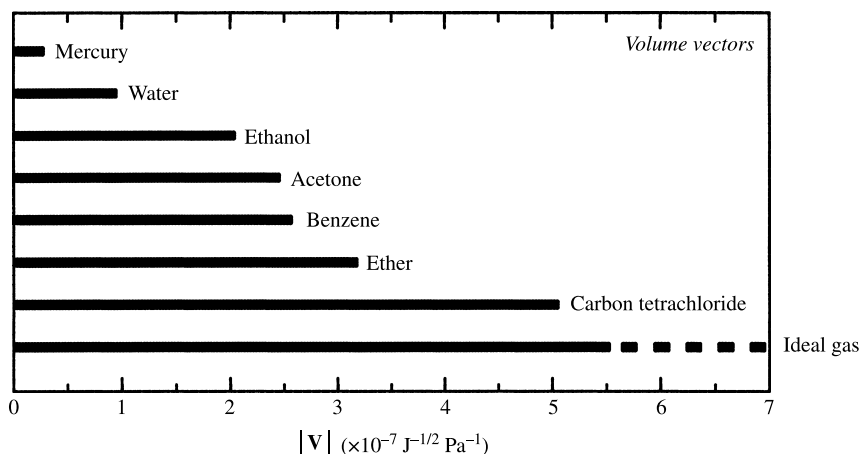


Figure 11.6 Volume vectors (SI units) of common laboratory liquids at 20°C, 1 atm. The ideal gas vector extends off scale to 4873 SI units.

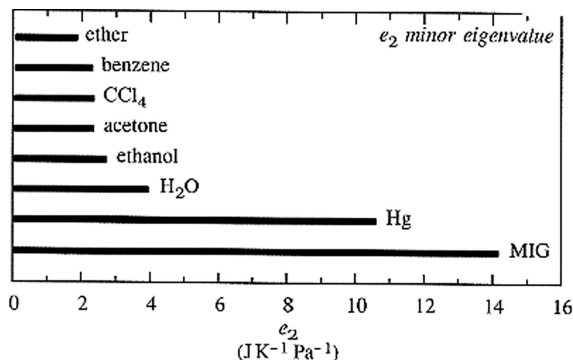


Figure 11.7 Minor metric eigenvalue e_2 (SI units) of common laboratory liquids at 20°C, 1 atm.

the critical pressure P_c . Such correlations demonstrate that thermodynamic measurements at 20°C, 1 atm are already signaling the approximate location of the critical state through metric geometrical properties.

In contrast to the unit dependence of the thermodynamic vector lengths and metric eigenvalues, the thermodynamic angles are pure dimensionless numbers. Figure 11.9 exhibits the angle θ_{SV} that each entropy vector $|\mathbf{S}\rangle$ makes with respect to the volume (abscissa) and temperature (ordinate) axes.

As before (Fig. 11.5), the lengths of entropy vectors in Fig. 11.9 are proportional to heat capacities. The angle θ_{SV} (or $\theta_{TS} = 90^\circ - \theta_{SV}$) that each entropy vector makes with the $|\hat{\mathbf{V}}\rangle$ (or $|\hat{\mathbf{T}}\rangle$) axis depicts the *coupling* between thermal and mechanical modes of the liquid. A large θ_{SV} angle, as for water (85.35°), corresponds to near-*normal* mode character, with $|\mathbf{S}\rangle$ and $|\mathbf{T}\rangle$ vectors nearly coincident (“self-conjugate”). In such a case, the heat-capacity ratio γ approaches unity, consistent with the geometrical identity (11.93b). A small $\theta_{SV}(=\theta_{TP})$ angle, as for CCl₄ (24.98°), signals stronger “mixing” of thermal and mechanical modes, leading to increasing differences in C_P versus C_V (or β_T versus

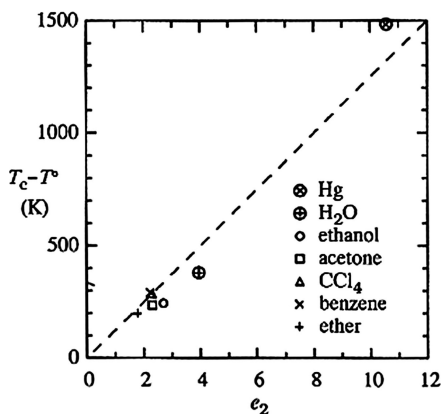


Figure 11.8 Correlation of metric eigenvalue e_2 (at 20°C, 1 atm) with critical temperature T_c for common laboratory fluids.

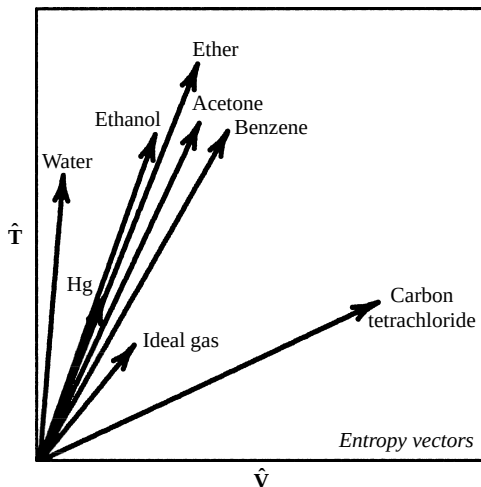


Figure 11.9 Entropy vectors $|\mathbf{S}|$ of common laboratory fluids at 20°C, 1 atm, showing each vector in its proper orientation with respect to orthogonal $|\hat{\mathbf{V}}|$ and $|\hat{\mathbf{T}}|$ directions.

β_S). Such angles evidently provide useful descriptors pertaining to sound propagation, heat transfer, suitability as a heat engine working fluid, and other thermomechanical properties. However, further aspects of the molecular origins or practical applications of distinctive geometrical features for specific fluids will not be pursued here.

SIDEBAR 11.6: CHOICE OF UNITS FOR REPRESENTING THERMODYNAMIC RELATIONSHIPS

Whereas SI units tend to make the $|\mathbf{P}|$ vectors about a million-fold too large for convenient plotting (compared, for example, to thermal $|\mathbf{T}|$, $|\mathbf{S}|$ vectors with lengths of order unity), these units correspondingly force the conjugate $|\mathbf{V}|$ vectors to be about a million-fold too *small*. This results in a *million-million-fold* disparity between lengths of $|\mathbf{P}|$ and $|\mathbf{V}|$ vectors, preventing effective SI-based graphical representations of thermodynamic relationships.

Much of this problem results from imbalance between what may be perceived as useful “units” by the chemist versus other scientists. Whereas a chemist tends to focus on *molar* quantities, corresponding to masses of a few grams or volumes of a few cubic centimeters, the SI system adopts large multiples of these quantities (10^3 for mass, 10^6 for volume) as its “units.” Only *one* such mismatch could be “repaired” by rescaling molar units (e.g., from g-mole to kg-mole).

How can the choice of units promote (or inhibit) effective representation of scientific relationships? This question was addressed by none other than J. W. Gibbs in his first scientific presentation [“The proper magnitude of the units of length and of other quantities used in mechanics.” Read before the Connecticut Academy of Sciences, March 21, 1866; reprinted in Appendix II of L. P. Wheeler. *Josiah Willard Gibbs: The History of a Great Mind* (Archon Books, Hamden, CT, 1970)]. Although the choice of units may superficially appear arbitrary, Gibbs recognized that one can rationally address the question of the

“conditions which it is most necessary for these units to fulfil for the convenience both of men of science and of the multitude.” Gibbs’ primary focus was on mechanics, but one can equally well apply his considerations to thermodynamic or electromagnetic phenomena.

Gibbs noted that “in causing a constant factor (or divisor) to disappear, the solution of mechanical problems is simplified by the choice of appropriate units.” Accordingly, one should select appropriate units that “greatly simplify the relations of the numerical representatives of the quantities and expedite the calculation.” Expressed another way, we should select units that cause the most important relationships to appear in their simplest and most transparent form.

An informative example of such unit-dependent representation of basic relationships is provided by electrical phenomena. Arguably, the most fundamental equation of the electrical sciences is *Coulomb’s law* for the interaction energy ($\mathcal{V}_{\text{elec}}$) of charges q_1 , q_2 at distance R . As recognized by Gibbs, each choice of unit system leads to a different expressions for Coulomb’s law, all containing the basic physical ratio $q_1 q_2 / R$ but differing by a unit-dependent constant factor K_{units} :

$$\mathcal{V}_{\text{elec}} = K_{\text{units}} \frac{q_1 q_2}{R} \quad (\text{S11.6-1})$$

Adopting atomic units (see below) or Gaussian electrostatic units and thereby “causing a constant factor (or divisor) to disappear,”

$$K_{\text{a.u.}} = K_{\text{Gauss}} = 1 \quad (\text{S11.6-2})$$

should therefore be considered preferable to the unfortunate consequence of selecting SI units,

$$K_{\text{SI}} = \frac{1}{4\pi\epsilon_0} \quad (\text{S11.6-3})$$

where the strange numerical constants 4π and “vacuum permittivity” ($\epsilon_0 \simeq 8.854 \times 10^{-12}$ in SI units) seem to unnecessarily obfuscate the simplicity of the underlying law.

A physical *unit system* is essentially defined by three chosen *base quantities* and corresponding *base units*, which suffice to determine dimensionally consistent units for other measurable physical quantities. In the *Système International d’Unités* (SI) framework, the three base quantities and their units are as follows:

Base Quantity	Base Unit	Symbol	Physical Identification
Mass	Kilogram	kg	A platinum–iridium bar kept in a guarded chateau near Paris ^a
Length	Meter	m	Originally, a metal stick whose length represented the King’s arm or a certain fraction of the Earth’s circumference (now the distance traveled by light in vacuum in a time interval of $1/299,792,458$ of a second)
Time	Second	s	Originally, a certain fraction of a day (now $9,192,631,770$ Cs hyperfine oscillations)

^a*New York Times* (May 27, 2003): “Scientists Struggling to Make the Kilogram Right Again: For mysterious reasons, a platinum–iridium cylinder that defines the kilogram has been losing weight. So scientists are looking for other ways to set the standard.”

In the alternative *atomic units* (a.u.) favored by electronic structure theorists, the “base units” ($m_e = e = \hbar = 1$) are instead identified with fundamental physical objects of the

science under discussion:

Base Quantity	Physical Identification of Base Unit	Symbol
Mass	Mass of an electron	m_e
Charge	Charge of an electron	e
Action ^a	Planck's constant	h

^aEquivalent to units of angular momentum.

The latter choice insures that the values of atomic and molecular properties are of “balanced” magnitudes (typically of order *unity* on a per-atom basis). Atomic units are therefore well adapted to exhibit simple numerical and graphical relationships between molecular properties, whereas the corresponding comparisons often require strange multiplicative factors (such as those defining the meter and second) if expressed in SI units.

Because thermodynamics is inherently concerned with describing the interplay of mechanical, thermal, chemical, electrical, and other forms of energy, the representation of thermodynamic relationships offers challenges to many possible choices of unit system. Geometrical relationships in thermodynamics present new demands in this respect, and SI units are quickly seen to be particularly *disadvantageous* for displaying such relationships effectively. Indeed, adoption of SI units practically *prevents* application of geometrical and graphical tools that are routinely used to display geometrical relationships in other contexts. This strongly supports consideration of alternative unit systems for thermodynamic purposes.

For thermodynamic purposes, it would be highly desirable to have mass and volume units corresponding to physically reasonable “unit density.” This could be most easily effected by connecting mass and volume units to the density of an actual physical system, such as that of liquid water under triple-point conditions (or any other easily replicated standard state). Even reinstatement of the former c.g.s. (centimeter–gram–second-based) system would be vastly (of order a million-fold!) preferable in this respect. A particularly attractive prospect is to associate *both* mass and volume units with a countable number of *atoms*, for example, in an isotopically pure crystal of silicon containing Avogadro's number of atoms. If combined with a base charge e [or other base unit yielding $K_{\text{units}} = 1$ in (S11.6-1)], one might better approach the ideal of “Gibbs Units” that meet the criteria expressed in Gibbs' first presentation.

11.8 STABILITY CONDITIONS AND THE “THIRD LAW” FOR HOMOGENEOUS PHASES

The stability conditions associated with the second law of thermodynamics appear as *metric-positivity conditions* in the thermodynamic geometry. The nonsingularity condition that is essential for stability of any homogeneous phase,

$$\det |\underline{\underline{\mathbf{M}}}| \neq 0 \quad (11.95)$$

must now be augmented by the condition of *positive-definiteness* of the metric matrix $\underline{\underline{\mathbf{M}}}$. This is expressed symbolically by the matrix inequality

$$\underline{\underline{\mathbf{M}}} > 0 \quad (11.96)$$

which summarizes all possible inequalities among thermodynamic response functions that can be inferred from stability constraints on a homogeneous phase.

An alternative way to express (11.96) (essentially, the *meaning* of this matrix inequality) is to say that

$$\underline{v}^t \underline{\underline{M}} \underline{v} > 0 \quad (\text{for any real vector } \underline{v}) \quad (11.97)$$

By the product rule for matrix multiplication, (11.97) is equivalent to the algebraic requirement

$$\sum_{i,j=1}^f \underline{\underline{M}}_{ij} v_i v_j > 0 \quad (11.98)$$

for all possible real numbers v_i , $i = 1, 2, \dots, f$. One obvious necessary requirement is that the diagonal elements all be positive,

$$M_{ii} > 0, \quad i = 1, 2, \dots, f \quad (11.99)$$

but it is also necessary (as well as sufficient) for positive-definiteness that all *principal sub-determinants* of $\underline{\underline{M}}$ be positive:

$$\det |\underline{\underline{M}}_r| > 0, \quad r = 1, 2, \dots, f \quad (11.100)$$

where $\underline{\underline{M}}_r$ is any $r \times r$ submatrix drawn from along the principal diagonal. Condition (11.100) includes (11.99) ($r = 1$) and positivity of the full $f \times f$ metric determinant M [consistent with (11.95)],

$$M = \det |\underline{\underline{M}}| > 0 \quad (11.101)$$

as well as all intermediate 2×2 , 3×3 , $\dots$ determinants

$$\det \begin{vmatrix} M_{ii} & M_{ij} \\ M_{ij} & M_{jj} \end{vmatrix} > 0, \quad \det \begin{vmatrix} M_{ii} & M_{ij} & M_{ik} \\ M_{ij} & M_{jj} & M_{jk} \\ M_{ik} & M_{jk} & M_{kk} \end{vmatrix} > 0, \dots \quad (11.102)$$

However, the necessary and sufficient conditions for positive-definiteness of $\underline{\underline{M}}$ can be expressed most concisely in terms of the positivity of its *eigenvalues* $\{e_i\}$,

$$e_i > 0, \quad i = 1, 2, \dots, f \quad (11.103)$$

and we can take (11.103) as the fundamental feature of $\underline{\underline{M}}$ that underlies the special cases (11.97)–(11.102).

The nonsingularity (11.95) of $\underline{\underline{M}}$ also implies the existence of the conjugate (inverse) matrix $\overline{\underline{\underline{M}}}$ and an associated set of positivity and finiteness conditions for its diagonal elements, principal subdeterminants, and eigenvalues. For the eigenvalues of $\overline{\underline{\underline{M}}}$

(reciprocals of those of $\underline{\underline{M}}$), these conditions take the form

$$\infty > e_i > 0, \quad i = 1, 2, \dots, f \quad (11.104)$$

and for the principal subdeterminants, the analogous conditions are

$$\infty > \det |\underline{\underline{M}}_r| > 0 \quad (11.105)$$

Special cases of (11.105) include such inequalities as

$$\infty > \underline{\underline{M}}_{ii} > 0 \quad (11.106a)$$

$$\infty > \det \begin{vmatrix} \overline{M}_{ii} & \overline{M}_{ij} \\ \overline{M}_{ij} & \overline{M}_{jj} \end{vmatrix} > 0, \dots \quad (11.106b)$$

$$\infty > \overline{M} = \det |\underline{\underline{M}}| > 0 \quad (11.106c)$$

Note that (11.95)–(11.106) must all be understood as *strict* inequalities as a consequence of the fundamental nonsingularity condition (11.95).

Let us now consider the consequences of these conditions for any pure substance, for which the metric and its inverse take the form (cf. Table 11.1)

$$\underline{\underline{M}} = \begin{pmatrix} T/C_V & -T/\Gamma_V \\ -T/\Gamma_V & 1/V\beta_S \end{pmatrix}, \quad \underline{\underline{M}} = \begin{pmatrix} C_P/T & V\alpha_P \\ V\alpha_P & V\beta_T \end{pmatrix} \quad (11.107)$$

Condition (11.106a) therefore becomes (for $i = 1$)

$$\infty > \frac{C_P}{T} > 0 \quad (11.108)$$

which must remain true for all physically possible T values. If we consider the approach to absolute zero ($T \rightarrow 0$), we can see from the near-zero limit of (11.108) that

$$\lim_{T \rightarrow 0} \frac{C_P}{T} > 0 \quad (11.109a)$$

or, equivalently,

$$\boxed{C_P \rightarrow 0 \text{ faster than } T \rightarrow 0} \quad (11.109b)$$

which is the essential empirical observation underlying the Nernst “heat theorem” and “third law” [cf. (5.77a, b)]. As noted by Fowler and Guggenheim (cf. Section 5.8.2), this may be summarized as the “unattainability of absolute zero,” the only valid “third-law” statement of rigorous and universal validity. However, this statement is seen to be merely another consequence of *second-law* stability requirements, and thus requires no new thermodynamic “law.” As stated in Section 5.8.2, all known statements of the “third law” are superfluous, tautologous, or invalid.

Similarly, from the case $i = 2$ of (11.106a), we obtain analogous stability constraints on the isothermal compressibility β_T :

$$\infty > V\beta_T > 0 \quad (11.110)$$

If we consider the general volume dependence of β_T in the limits of large or small V , we can see that (11.110) requires, in the small- V limit,

$$1/\beta_T \rightarrow 0 \quad \text{faster than} \quad V \rightarrow 0 \quad (11.111a)$$

and, in the large- V limit,

$$\beta_T \rightarrow 0 \quad \text{faster than} \quad 1/V \rightarrow 0 \quad (11.111b)$$

If we consider the 2×2 determinantal constraint (11.106b), we can similarly infer the inequality

$$\infty > \frac{C_P \beta_T}{TV \alpha_p^2} > 0 \quad (11.112)$$

which can be restated as the asymptotic requirement

$$\frac{C_P \beta_T}{\alpha_p^2} \rightarrow 0 \quad \text{faster than} \quad T \rightarrow 0, \quad V \rightarrow 0, \quad \text{or} \quad TV \rightarrow 0 \quad (11.113)$$

Many similar asymptotic restrictions can be obtained from other matrix elements of $\underline{\underline{\mathbf{M}}}$ or $\underline{\underline{\mathbf{M}}}$ in standard or transformed coordinates.

11.9 THE CRITICAL INSTABILITY LIMIT

In many respects, the most dramatic, beautiful, and profound feature of a thermodynamic system is associated with the critical instability limit

$$\det|\underline{\underline{\mathbf{M}}}| \rightarrow 0 \quad (11.114)$$

This limit marks the singular exit from the domain (11.95) of single-phase stability, the onset of “spontaneous symmetry breaking,” and the spectacular ambiguity of emergent two-phase coexistence (Section 2.5). From the geometrical perspective [cf. (11.9b)], the condition (11.114) is also the limit of dimensional collapse, the mind-bending “portal” between two-dimensional ($p = 1$) and one-dimensional ($p = 2$) geometries. In this section, we wish to rigorously characterize the thermodynamic critical state in terms of its metric geometrical properties, making contact with modern experimental and statistical mechanical studies of critical-point phenomena.

The general features of critical dimensional collapse in $\mathcal{M}_S$ are fairly evident. As previously remarked [cf. (10.21)], the key feature of $\underline{\underline{\mathbf{M}}}$ is its *Gramian* character, i.e., its composition from scalar products $\underline{\mathbf{R}}_i^\dagger \underline{\mathbf{R}}_j = (\underline{\underline{\mathbf{M}}})_{ij}$ of underlying vectors $\{\underline{\mathbf{R}}_i\}$. Sidebar 11.7 outlines the general mathematical theory relating these Gramian vectors

to intrinsic *eigen*properties of $\underline{\underline{M}}$ (eigenvectors $\{\underline{u}_i\}$ and eigenvalues $\{e_i\}$), with brief discussion of the associated geometrical picture. We can then employ these generic properties of Gramian geometry in the present context to evaluate and interpret the key mathematical descriptors of dimensional collapse in the thermodynamic space $\mathcal{M}_S$ at the critical limit (T_c, P_c) .

**SIDEBAR 11.7: EIGENVALUES AND EIGENVECTORS OF
A GRAM MATRIX**

We consider a 2×2 real symmetric Gram matrix $\underline{\underline{M}}$ associated with an underlying matrix $\underline{\underline{R}} = (\underline{R}_1 \underline{R}_2)$ of real vectors $\underline{R}_i$ whose scalar products are its elements:

$$\underline{\underline{M}} = \underline{\underline{R}}^t \underline{\underline{R}} = \begin{pmatrix} \underline{R}_1^t \underline{R}_2 & \underline{R}_1^t \underline{R}_2 \\ \underline{R}_2^t \underline{R}_1 & \underline{R}_2^t \underline{R}_2 \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{pmatrix} \quad (\text{S11.7-1})$$

Let us denote by $\underline{\underline{U}}$,

$$\underline{\underline{U}} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} = (U_1 \ U_2) \quad (\text{S11.7-2})$$

the orthogonal matrix of eigenvectors $\underline{u}_i$ satisfying (cf. Sidebar 11.5)

$$\underline{\underline{M}} \underline{u}_i = e_i \underline{u}_i, \quad i = 1, 2 \quad (\text{S11.7-3})$$

with eigenvalues $e_1 \geq e_2 \geq 0$. The matrix $\underline{\underline{M}}$ is therefore diagonalized by orthogonal transformation with $\underline{\underline{U}}$ to give the diagonal matrix $\underline{\underline{e}}$:

$$\underline{\underline{U}}^t \underline{\underline{M}} \underline{\underline{U}} = \underline{\underline{e}}, \quad \text{with } (\underline{\underline{e}})_{ij} = e_i \delta_{ij} \quad (\text{S11.7-4})$$

Multiplying (S11.7-4) on the left by $\underline{\underline{U}}$ and on the right by $\underline{\underline{U}}^t$, and using the property $\underline{\underline{U}}^t \underline{\underline{U}} = \underline{\underline{U}} \underline{\underline{U}}^t = \underline{\underline{I}}$, we obtain

$$\underline{\underline{M}} = \underline{\underline{U}} \underline{\underline{e}} \underline{\underline{U}}^t \quad (\text{S11.7-5})$$

Let us now “factorize” $\underline{\underline{e}} = \underline{\underline{e}}^{1/2} \underline{\underline{e}}^{1/2}$ in terms of its (real, non-negative) square-root matrix to rewrite (S11.7-5) as

$$\underline{\underline{M}} = (\underline{\underline{U}} \underline{\underline{e}}^{1/2}) (\underline{\underline{e}}^{1/2} \underline{\underline{U}}^t) = (\underline{\underline{e}}^{1/2} \underline{\underline{U}}^t)^t (\underline{\underline{e}}^{1/2} \underline{\underline{U}}^t) \quad (\text{S11.7-6})$$

Comparing (S11.7-6) with (S11.7-1), we can see that $\underline{\underline{R}}$ and $\underline{\underline{U}}$ are related simply by

$$\underline{\underline{R}} = \underline{\underline{e}}^{1/2} \underline{\underline{U}}^t = \begin{pmatrix} e_1^{1/2} & 0 \\ 0 & e_2^{1/2} \end{pmatrix} \begin{pmatrix} U_{11} & U_{21} \\ U_{12} & U_{22} \end{pmatrix} = \begin{pmatrix} e_1^{1/2} U_{11} & e_1^{1/2} U_{21} \\ e_2^{1/2} U_{12} & e_2^{1/2} U_{22} \end{pmatrix} \quad (\text{S11.7-7})$$

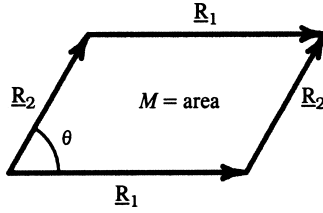
(or any orthogonal transformation, which merely gives a different $\underline{\underline{e}}^{1/2}$ matrix).

The equations relating Gramian vectors $\underline{R}_i$ to eigenvectors $\underline{u}_i$ have a simple geometrical interpretation. One can show [see, e.g., G. E. Shilov. *An Introduction to the Theory of*

Linear Spaces (Prentice-Hall, Englewood Cliffs, NJ, 1961), pp. 167–73] that the Gramian determinant

$$M = \det |\underline{\underline{M}}| = \text{Area}(\underline{R}_1, \underline{R}_2) = |\underline{R}_1| |\underline{R}_2| \sin \theta \quad (\text{S11.7-8})$$

is equal to the area of the parallelogram formed by $\underline{R}_1, \underline{R}_2$:



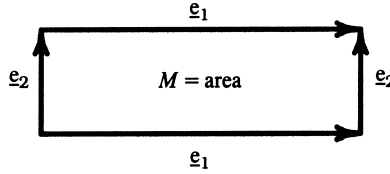
Because determinants of matrices $\underline{\underline{A}}, \underline{\underline{B}}$ satisfy the general product rule

$$\det |\underline{\underline{A}} \underline{\underline{B}}| = \det |\underline{\underline{B}} \underline{\underline{A}}| = \det |\underline{\underline{A}}| \det |\underline{\underline{B}}| \quad (\text{S11.7-9})$$

we can recognize from (S11.7-5) that

$$M = \det |\underline{\underline{M}}| = \det |\underline{\underline{U}}^t \underline{\underline{e}} \underline{\underline{U}}| = \det |\underline{\underline{e}}| \det |\underline{\underline{U}}^t \underline{\underline{U}}| = \det |\underline{\underline{e}}| = e_1 e_2 \quad (\text{S11.7-10})$$

This shows that M is equally the area of the *eigen*-rectangle with e_1, e_2 as its edges:



Although we have focused on the 2×2 case, the results are easily generalized to Gram matrices of any dimension f . If we write $\text{Vol}\{\underline{R}_1, \underline{R}_2, \dots, \underline{R}_f\}$ for the volume of the parallelepiped spanned by the vectors $\{\underline{R}_i\}$, then we can show analogously that

$$M = \text{Vol}\{R_1, R_2, \dots, R_f\} = \prod_{i=1}^f e_i = e_1 e_2 \dots e_f \quad (\text{S11.7-11})$$

i.e., *equal* volumes are enclosed in the hyperparallelepiped of nonorthogonal vectors $\underline{R}_1, \underline{R}_2, \dots, \underline{R}_f$ or the corresponding orthogonal *eigen*-parallelepiped with edges $e_1, e_2, \dots, e_f$. Thus, we are free to study the Gramian “collapse” $M \rightarrow 0$ in whatever geometrical picture is convenient.

As shown in Sidebar 11.7, the determinant M of the 2×2 matrix in (11.114) can quite generally be written in terms of the metric eigenvalues e_1, e_2 as

$$M = \det |\underline{\underline{M}}| = e_1 e_2 \quad (11.115)$$

The geometrical significance of (11.114) thereby becomes clear: M is the area of the *eigen*-rectangle with long-edge e_1 and short-edge e_2 . The necessary and sufficient condition for dimensional collapse (11.114) can therefore be stated as the condition that the critical eigen-rectangle must collapse to *zero* area at (T_c, P_c) , i.e., that its short edge, the minor eigenvalue e_2 , *vanishes* at the critical limit (identified with subscript “c”):

$$e_{2c} = 0 \quad \text{at } T_c, P_c \quad (11.116)$$

As shown in Sidebar 11.7, the array of vectors $\underline{\underline{R}} = (\underline{R}_1 \underline{R}_2)$ whose scalar products define $\underline{\underline{M}}$ can be represented as

$$\underline{\underline{R}} = \begin{pmatrix} e_1^{1/2} U_{11} & e_1^{1/2} U_{21} \\ e_2^{1/2} U_{12} & e_2^{1/2} U_{22} \end{pmatrix} \quad (11.117)$$

In view of (11.116), the critical limit $\underline{\underline{R}}_c$ of this array must therefore be

$$\underline{\underline{R}}_c = \begin{pmatrix} e_{1c}^{1/2} U_{11c} & e_{1c}^{1/2} U_{21c} \\ 0 & 0 \end{pmatrix} \quad (11.118)$$

Equation (11.118) establishes the proportionality of column vectors $\underline{R}_{1c}, \underline{R}_{2c}$ at the critical limit

$$\underline{R}_{1c} = k \underline{R}_{2c}, \quad k = U_{11c}/U_{21c} \quad (11.119)$$

Under this condition, the vectors are co-aligned (with critical angle $\theta_{12c} = 0$), so the Schwarz inequality (11.4) necessarily becomes an *equality*

$$\cos \theta_{12c} = 1 = \frac{M_{12}}{(M_{11}M_{22})^{1/2}} = \frac{V_c \alpha_{Pc}}{(C_{Pc} \beta_{Tc} V_c / T_c)^{1/2}} \quad \text{at } T_c, P_c \quad (11.120a)$$

equivalent to the critical identity

$$\alpha_{Pc} = \left(\frac{C_{Pc} \beta_{Tc}}{T_c V_c} \right)^{1/2} \quad (11.120b)$$

To evaluate the unknowns e_{1c}, U_{11c}, U_{21c} , we use (11.118) to obtain the *critical metric* $\underline{\underline{M}}_c$ as

$$\underline{\underline{M}}_c = \underline{\underline{R}}_c^t \underline{\underline{R}}_c = \begin{pmatrix} e_1 U_{11}^2 & e_1 U_{11} U_{21} \\ e_1 U_{11} U_{21} & e_1 U_{21}^2 \end{pmatrix}_c = \begin{pmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{pmatrix}_c \quad (11.121)$$

where all quantities are evaluated at (T_c, P_c) . If we write the unknown U-matrix elements in terms of the characteristic eigenangle ω defined by

$$U_{11} = \cos \omega, \quad U_{21} = \sin \omega \quad (11.122)$$

we obtain from (11.121)

$$e_{1c} \begin{pmatrix} \cos^2 \omega & \cos \omega \sin \omega \\ \cos \omega \sin \omega & \sin^2 \omega \end{pmatrix} = \begin{pmatrix} M_{11c} & M_{12c} \\ M_{12c} & M_{22c} \end{pmatrix} \quad (11.123)$$

From (11.123), we obtain

$$e_{1c}(\cos^2 \omega + \sin^2 \omega) = e_{1c} = M_{11c} + M_{22c} \quad (11.124a)$$

$$e_{1c} \cos \omega \sin \omega = \frac{1}{2} e_{1c} \sin 2\omega = M_{12} \quad (11.124b)$$

from which the unknown critical eigenangle ω_c can be evaluated as

$$\sin 2\omega_c = \frac{2M_{12c}}{M_{11c} + M_{22c}} \quad (11.125)$$

Alternatively, in view of (11.120), we can express the solution as

$$\sin 2\omega_c = \frac{(M_{11c}M_{22c})^{1/2}}{(M_{11c} + M_{22c})/2} = \frac{\text{geometric mean response}}{\text{arithmetic mean response}} \quad (11.126)$$

Equation (11.126) shows that the critical angle ω_c relating base vectors $\underline{R}_1, \underline{R}_2$ to eigenvectors $\underline{u}_1, \underline{u}_2$ at (T_c, P_c) is determined by the ratio of geometrical and arithmetic means of the diagonal metric elements. Inserting the specific metric expressions from Table 11.1, we obtain finally

$$\cos \theta_c = 1 = - \frac{(T_c V_c C_{Vc} \beta_{Sc})^{1/2}}{\Gamma_{Vc}} \quad (11.127)$$

$$\sin 2\omega_c = \frac{2(T_c V_c C_{Vc} \beta_{Sc})^{1/2}}{C_{Vc} + T_c V_c \beta_{Sc}} \quad (11.128)$$

$$e_{1c} = \frac{T_c}{C_{Vc}} + \frac{1}{V_c \beta_{Sc}} \quad (11.129)$$

For completeness, we also give the expression for the tangent of the critical eigenangle ω_c , which identifies the “direction” of the surviving eigenvector at the critical limit

$$\tan \omega_c = \tan \left\{ \frac{1}{2} \sin^{-1} \left[\frac{2(T_c V_c C_{Vc} \beta_{Sc})^{1/2}}{C_{Vc} + T_c V_c \beta_{Sc}} \right] \right\} \quad (11.130)$$

Equations (11.127)–(11.130) allow us to construct the critical-point geometry in detail from measured response functions at (T_c, V_c) .

11.10 CRITICAL DIVERGENCE AND EXPONENTS

The collapse of dimensionality associated with the critical singularity (11.114) has many dramatic consequences in $\mathcal{M}_S$. In this limit, *all* conjugate vectors and response functions become mathematically ill-defined (divergent), corresponding to “*infinities*” in physical properties associated with the conjugate metric $\underline{\underline{M}}$ (cf. Table 11.1):

$$C_P \rightarrow \infty, \quad \beta_T \rightarrow \infty, \quad \alpha_P \rightarrow \infty \quad (11.131)$$

Although the critical extensive vectors $|\mathbf{S}\rangle_c, |\mathbf{V}\rangle_c$ and associated responses $C_{Pc}, \beta_{Tc}, \alpha_{Pc}$ are strictly undefined *at* the critical state, we can consider this state as a “target” for approach along a chosen thermodynamic path. This will enable us to characterize physical and mathematical details of the asymptotic divergences (11.131) that are the hallmark of critical phenomena.

The natural “path variable” or “order parameter” to characterize proximity to the critical limit is the minor eigenvalue e_2 of the thermodynamic metric. This suggests that we examine the functional dependence of conjugate responses on e_2 ,

$$C_P = C_P(e_2), \quad \beta_T = \beta_T(e_2), \quad \alpha_P = \alpha_P(e_2) \quad (11.132)$$

in order to formally characterize the divergences (11.131) in terms of the asymptotic limiting behavior:

$$\lim_{e_2 \rightarrow 0} C_P(e_2) = “C_{Pc}” \quad (11.133a)$$

$$\lim_{e_2 \rightarrow 0} \beta_T(e_2) = “\beta_{Tc}” \quad (11.133b)$$

$$\lim_{e_2 \rightarrow 0} \alpha_P(e_2) = “\alpha_{Pc}” \quad (11.133c)$$

Of course, we should identify the particular path chosen to approach the critical limit, and for this purpose it is convenient to introduce a dimensionless path parameter such as

$$\tau \equiv \frac{T - T_c}{T_c} \quad (11.134)$$

stipulated to follow, for example, the critical isochore V_c .

A common *mathematical assumption* is that the dependence of the order parameter e_2 on the path parameter τ can be described by a “critical exponent” relationship of the form [see, e.g., H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971)]

$$e_2(\tau) \simeq A_\epsilon \tau^\epsilon, \quad \tau \rightarrow 0 \quad (11.135)$$

for sufficiently small $\tau \rightarrow 0$. Here A_ϵ denotes the “critical amplitude” and ϵ the “critical exponent” that describes the asymptotic $e_2(\tau)$ behavior. We emphasize that such an assumption has no thermodynamic basis, and, indeed, experimental fits of A_ϵ , ϵ often seem to depend on what is considered “sufficiently small.” Nevertheless, (11.135), if valid in some experimental sense, can be conveniently used to describe experimental critical divergences (11.131). We shall employ this assumption in the present section to study metric restrictions on the critical asymptotics of the conjugate response functions (11.133a–c).

For this purpose, we start from the expressions (cf. Sidebar 11.7) for $\underline{\underline{M}}$ and $\underline{\underline{M}}$ in terms of their eigenvectors and eigenvalues,

$$\underline{\underline{M}} = \underline{U} \underline{e} \underline{U}^t \quad (11.136a)$$

$$\underline{\underline{M}} = \underline{U} \underline{e}^{-1} \underline{U}^t \quad (11.136b)$$

noting that $\underline{\underline{M}}$ and $\underline{\underline{M}}$ have the same eigenvectors but inverted eigenvalues. The matrix equation (11.136b) can be written explicitly as

$$\begin{aligned} \underline{\underline{M}} &= \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} e_1^{-1} & 0 \\ 0 & e_2^{-1} \end{pmatrix} \begin{pmatrix} U_{11} & U_{21} \\ U_{12} & U_{22} \end{pmatrix} \\ &= \begin{pmatrix} U_{11}^2/e_1 + U_{12}^2/e_2 & U_{11}U_{21}/e_1 + U_{12}U_{22}/e_2 \\ U_{11}U_{21}/e_1 + U_{12}U_{22}/e_2 & U_{21}^2/e_1 + U_{22}^2/e_2 \end{pmatrix} \\ &= \begin{pmatrix} \overline{M}_{11} & \overline{M}_{12} \\ \overline{M}_{12} & \overline{M}_{22} \end{pmatrix} \end{aligned} \quad (11.137)$$

Introducing (11.135) for e_2 , and treating all other terms as finite constants at the critical limit, we obtain

$$\lim_{\tau \rightarrow 0} \overline{M}_{11}(\tau) = \text{const} + (U_{12}^2/A_\epsilon) \tau^{-\epsilon} \quad (11.138a)$$

$$\lim_{\tau \rightarrow 0} \overline{M}_{22}(\tau) = \text{const}' + (U_{22}^2/A_\epsilon) \tau^{-\epsilon} \quad (11.138b)$$

$$\lim_{\tau \rightarrow 0} \overline{M}_{12}(\tau) = \text{const}'' + (U_{12}U_{22}/A_\epsilon) \tau^{-\epsilon} \quad (11.138c)$$

Furthermore, from the geometrical identity

$$\overline{\theta}_{12} = \theta_{12} \quad (11.139)$$

which holds everywhere along the path, we also infer the limit [cf. (11.120)]

$$\lim_{\tau \rightarrow 0} \frac{\overline{M}_{12}^2}{\overline{M}_{11}\overline{M}_{22}} = 1 \quad (11.140)$$

From these equations, we conclude that the conjugate responses C_P , β_T , α_P must all diverge with the *same* critical exponent $-\epsilon$, but with varying critical amplitudes that depend on the numerical U_{12} , U_{22} values. The latter are in turn easily obtained from (11.128), together with the unitarity condition that relates them to U_{21} , U_{22} . The equivalencies (11.138a–c) of critical exponents for C_P , β_T , α_P were previously recognized [R. B. Griffiths and J. C. Wheeler. *Phys. Rev. A* **2**, 1047 (1970)], but the quantitative relationships between their critical amplitudes seem not to have been previously suggested or tested.

Note that metric collapse requires that stability *inequalities* that hold elsewhere in $\mathcal{M}_S$ must become strict *equalities* at the critical limit [e.g., (11.120) and (11.140)]. This in turn implies that “critical exponent inequalities” inferred from such stability conditions [see, e.g., G. S. Rushbrooke. *J. Chem. Phys.* **39**, 842 (1963)] are necessarily *equalities*, if indeed the critical exponent assumption is valid at all. Such equality in critical-point relationships seems to be supported by all available experimental data, and its justification is straightforward in the metric framework, independent of subsidiary scaling hypotheses (cf. Sidebar 10.4).

11.11 PHASE HETEROGENEITY AND CRITICALITY

As discussed in Section 10.3, equilibrium phase homogeneity is associated with existence of a null eigenvector $\underline{\eta}$ of the full $(c + 2)$ -dimensional metric matrix

$$\underline{\underline{M}}^{(c+2)} \underline{\eta} = 0 \quad (11.141a)$$

$$\underline{\eta} = \begin{pmatrix} S \\ V \\ n_1 \\ \vdots \\ n_c \end{pmatrix} \quad (11.141b)$$

equivalent to the Gibbs–Duhem equation (Section 6.3) for thermodynamic intensities. We shall refer to (11.141b) as a *Gibbs–Duhem vector* (GD vector) for the homogeneous region in question. Geometrically, (11.141a) expresses the fact that the GD vector for a homogeneous phase region *vanishes* in $\mathcal{M}_S$,

$$\underline{\eta} \text{ “} = \text{” } 0 \quad \text{in } \mathcal{M}_S \quad (11.142)$$

as discussed in Sidebar 11.8.

Let us now consider the case of a heterogeneous system of distinct phases α, β . Each phase is associated with its own GD vector

$$\underline{\eta}_\alpha = \begin{pmatrix} S_\alpha \\ V_\alpha \\ n_{1\alpha} \\ \vdots \\ n_{c\alpha} \end{pmatrix}, \quad \underline{\eta}_\beta = \begin{pmatrix} S_\beta \\ V_\beta \\ n_{1\beta} \\ \vdots \\ n_{c\beta} \end{pmatrix} \quad (11.143)$$

satisfying

$$\underline{\underline{M}}^{(c+2)} \underline{\eta}_\alpha = \underline{\underline{M}}^{(c+2)} \underline{\eta}_\beta = 0 \quad (\text{i.e., } \underline{\eta}_\alpha = \underline{\eta}_\beta \text{ "}" } 0 \text{ in } \mathcal{M}_S) \quad (11.144)$$

Any linear combination of GD vectors,

$$\tilde{\underline{\eta}} = c_\alpha \underline{\eta}_\alpha + c_\beta \underline{\eta}_\beta \text{ "}" } 0 \text{ in } \mathcal{M}_S \quad (11.145)$$

is also a null (vanishing) vector in $\mathcal{M}_S$. If we let n_α, n_β denote the total number of moles in each phase, for example,

$$\eta_\alpha = \sum_{i=1}^c \eta_{i\alpha} \quad (11.146)$$

and choose the phase sizes to distribute one component equally between phases, for example,

$$\eta_{c\alpha} = \eta_{c\beta} \quad (11.147)$$

then we can choose the coefficients in (11.145) as

$$c_\alpha = -\frac{1}{n_\alpha}, \quad c_\beta = \frac{1}{n_\beta} \quad (11.148)$$

so as to define the difference vector

$$\underline{\delta}_{\alpha\beta}^{(c+2)} = \begin{pmatrix} \bar{S}_\beta - \bar{S}_\alpha \\ \bar{V}_\beta - \bar{V}_\alpha \\ \bar{n}_{1\beta} - \bar{n}_{1\alpha} \\ \vdots \\ \bar{n}_{(c-1)\beta} - \bar{n}_{(c-1)\alpha} \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta \bar{S}_{\alpha\beta} \\ \Delta \bar{V}_{\alpha\beta} \\ \Delta n_{1\alpha\beta} \\ \vdots \\ \Delta n_{(c+1)\alpha\beta} \\ 0 \end{pmatrix} \quad (11.149)$$

which gives the molar extensity changes between the two phases. The vector $\underline{\delta}_{\alpha\beta}^{(c+2)}$ may be designated as the *two-phase Clapeyron vector* for the $\alpha \rightarrow \beta$ phase transition, in accordance with the nomenclature discussed below. The characteristic metric singularity for $\alpha\beta$ -phase coexistence (cf. Sidebar 7.2),

$$\underline{\underline{M}}^{(c+2)} \underline{\delta}_{\alpha\beta}^{(c+2)} = 0 \quad (11.150)$$

can thereby be expressed completely in the $(c + 1)$ -dimensional subspace (ignoring the final row and column of $\underline{\underline{\mathbf{M}}}^{(c+2)}$). The reduced-dimensional analog of (11.150) will be written for simplicity as

$$\underline{\underline{\mathbf{M}}}^{(c+1)} \underline{\underline{\delta}}_{\alpha\beta}^{(c+1)} \equiv \underline{\underline{\mathbf{M}}}^{(\alpha\beta)} \underline{\underline{\delta}}_{\alpha\beta} = 0 \quad (11.151)$$

where $\underline{\underline{\mathbf{M}}}^{(\alpha\beta)}$ is the metric and $\underline{\underline{\delta}}_{\alpha\beta}$ is the difference of GD vectors in the appropriate $(c + 1)$ -dimensional space.

SIDEBAR 11.8: “ZERO” VECTORS IN $\mathcal{M}_S$

According to the positivity property (9.27c) of a Euclidean scalar property, a vector $|\boldsymbol{\eta}\rangle$ is considered “zero” if and only if it is of zero *length*, namely,

$$|\boldsymbol{\eta}\rangle = 0 \quad \text{if and only if} \quad \langle \boldsymbol{\eta} | \boldsymbol{\eta} \rangle = 0 \quad (\text{S11.8-1})$$

Alternatively, for a column vector $\underline{\boldsymbol{\eta}}$ such as (11.141b), the equivalent condition is

$$\underline{\boldsymbol{\eta}} = 0 \quad \text{if and only if} \quad \underline{\boldsymbol{\eta}}^t \underline{\underline{\mathbf{M}}} \underline{\boldsymbol{\eta}} = 0 \quad (\text{S11.8-2})$$

Thus, any vector satisfying $\underline{\underline{\mathbf{M}}} \underline{\boldsymbol{\eta}} = 0$ (“null eigenvector” of $\underline{\underline{\mathbf{M}}}$) is a zero vector of $\mathcal{M}_S$, even if some or all of its elements η_i are nonvanishing. Indeed, the trivial case of all $\eta_i = 0$ in (11.141b) is never considered (no system!). We may therefore write $\underline{\boldsymbol{\eta}}$ “=” 0 to distinguish such a vector of zero length from a vector with all zero elements.

Let us now focus on the simplest case of a two-phase system for a pure substance ($c = 1$), where $\underline{\underline{\mathbf{M}}}^{(\alpha\beta)}$, and $\underline{\underline{\delta}}_{\alpha\beta}$ are given by simple 2-space objects

$$\underline{\underline{\mathbf{M}}}^{(\alpha\beta)} = \begin{pmatrix} T/C_V^{(\alpha\beta)} & -T/\Gamma_V^{(\alpha\beta)} \\ -T/\Gamma_V^{(\alpha\beta)} & 1/V\beta_S^{(\alpha\beta)} \end{pmatrix} \quad (11.152)$$

$$\underline{\underline{\delta}}_{\alpha\beta} = \begin{pmatrix} \Delta \bar{S}_{\alpha\beta} \\ \Delta \bar{V}_{\alpha\beta} \end{pmatrix} \quad (11.153)$$

The vanishing of the two-phase Clapeyron vector $\underline{\underline{\delta}}_{\alpha\beta}$ (11.151) is now equivalent to

$$\Delta \bar{S}_{\alpha\beta} |\mathbf{T}\rangle_{\alpha\beta} - \Delta \bar{V}_{\alpha\beta} |\mathbf{P}\rangle_{\alpha\beta} = 0 \quad \text{in } \mathcal{M}_S(\alpha\beta) \quad (11.154)$$

Writing out the matrix elements of (11.151) in detail, we obtain

$$\begin{aligned}\underline{\underline{\mathbf{M}}}^{(\alpha\beta)}\underline{\eta}_{\alpha\beta} &= \begin{pmatrix} T/C_V^{(\alpha\beta)} & -T/\Gamma_V^{(\alpha\beta)} \\ -T/\Gamma_V^{(\alpha\beta)} & 1/V\beta_S^{(\alpha\beta)} \end{pmatrix} \begin{pmatrix} \Delta\bar{\mathcal{S}}_{\alpha\beta} \\ \Delta\bar{V}_{\alpha\beta} \end{pmatrix} \\ &= \begin{pmatrix} T\Delta\bar{\mathcal{S}}_{\alpha\beta}/C_V^{(\alpha\beta)} - T\Delta\bar{V}_{\alpha\beta}/\Gamma_V^{(\alpha\beta)} \\ -T\Delta\bar{V}_{\alpha\beta}/\Gamma_V^{(\alpha\beta)} + \Delta\bar{V}_{\alpha\beta}/V\beta_S^{(\alpha\beta)} \end{pmatrix} = 0\end{aligned}\quad (11.155)$$

From this equation, we obtain the identities

$$\Gamma_V^{(\alpha\beta)} = C_V^{(\alpha\beta)} \frac{\Delta\bar{V}_{\alpha\beta}}{\Delta\bar{\mathcal{S}}_{\alpha\beta}} = TV\beta_S^{(\alpha\beta)} \frac{\Delta\bar{\mathcal{S}}_{\alpha\beta}}{\Delta\bar{V}_{\alpha\beta}} \quad (11.156)$$

from which also

$$V\beta_S^{(\alpha\beta)} = \frac{C_V^{(\alpha\beta)}}{T} \left(\frac{\Delta\bar{V}_{\alpha\beta}}{\Delta\bar{\mathcal{S}}_{\alpha\beta}} \right)^2 \quad (11.157)$$

Equations (11.156), (11.157) allow us to rewrite the metric matrix as

$$\underline{\underline{\mathbf{M}}}^{(\alpha\beta)} = \frac{T}{C_V^{(\alpha\beta)}} \begin{pmatrix} 1 & -\Delta\bar{\mathcal{S}}_{\alpha\beta}/\Delta\bar{V}_{\alpha\beta} \\ -\Delta\bar{\mathcal{S}}_{\alpha\beta}/\Delta\bar{V}_{\alpha\beta} & (\Delta\bar{\mathcal{S}}_{\alpha\beta}/\Delta\bar{V}_{\alpha\beta})^2 \end{pmatrix} \quad (11.158)$$

As expected from the phase rule ($f = 1$), this 2×2 matrix is singular,

$$\det |\underline{\underline{\mathbf{M}}}^{(\alpha\beta)}| = 0 \quad (11.159)$$

and the surviving metric “matrix” is in this case only the scalar element

$$\mathbf{M}_{11}^{(\alpha\beta)} = \frac{T}{C_V^{(\alpha\beta)}} \quad (11.160)$$

requiring the two-phase heat capacity under isochoric conditions.

The conjugate (inverse) of (11.160) gives the saturation two-phase heat capacity $C_{\sigma}^{(\alpha\beta)}$, which in this case coincides with $C_V^{(\alpha\beta)}$ (with each saturated phase adjusting to maintain constant total V):

$$\bar{M}_{11}^{(\alpha\beta)} = \frac{C_V^{(\alpha\beta)}}{T} \quad (11.161)$$

However, from the extensive property of heat capacity, we can recognize that the molar two-phase heat capacity $C_V^{(\alpha\beta)}$ must just be the mole fraction-weighted sum of the *saturation* heat capacities of the individual phases:

$$C_V^{(\alpha\beta)} = C_{\sigma}^{(\alpha\beta)} = x_{\alpha}C_{\sigma}^{(\alpha)} + x_{\beta}C_{\sigma}^{(\beta)} \quad (11.162)$$

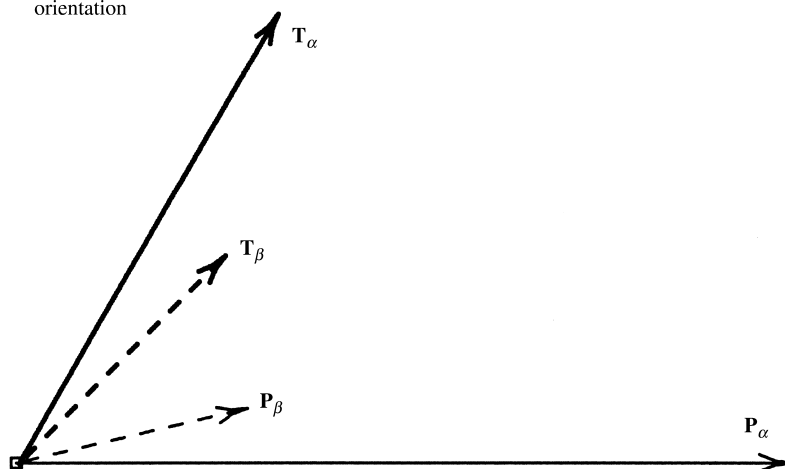
In this case, we can use the identity (11.73b) previously established for the saturation heat capacity of each phase to obtain finally [cf. J. S. Rowlinson. *Liquids and Liquid Mixtures*,

2nd edn (Butterworths, London, 1969), equation (2.92)]

$$C_V^{(\alpha\beta)} = x_\alpha[C_P^{(\alpha)} - TV\gamma_\sigma\alpha_P^{(\alpha)}] + x_\beta[C_P^{(\beta)} - TV\gamma_\sigma\alpha_P^{(\beta)}] \quad (11.163)$$

Equation (11.163) shows how the isochoric heat capacity of a heterogeneous two-phase system can be evaluated from known isobaric properties (C_P, α_P) of the individual phases and the “direction” γ_σ of the coexistence coordinate σ .

- (a) Coexisting base vectors $\mathbf{T}, \mathbf{P}$ in proper co-Clapeyron ($\alpha\beta$) orientation



- (b) Coincident Clapeyron vectors Σ for coexisting phases in co-Clapeyron ($\alpha\beta$) orientation

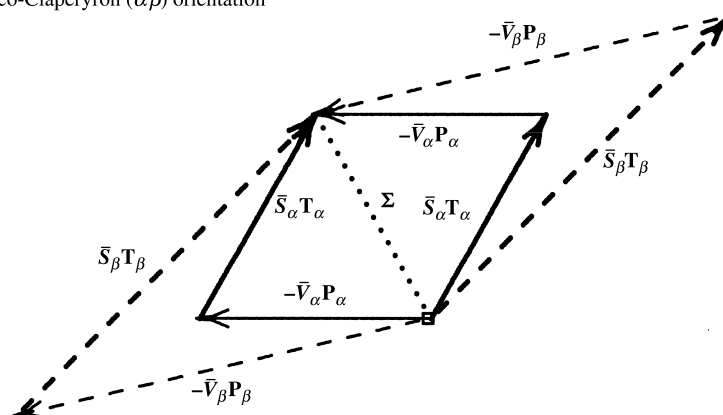


Figure 11.10 Schematic representation of “co-Clapeyron” vector orientations of (a) base $|\mathbf{T}\rangle, |\mathbf{P}\rangle$ vectors and (b) resultant Clapeyron vectors $|\Sigma\rangle$ (dotted), illustrating the geometrical condition (11.165a) for coexistence of α (solid line) and β (dashed line) phases. The coordinate origin is marked with a small square in each panel. [Vector lengths, angles, γ_σ , and $\Delta\bar{S}_{\alpha\beta}, \Delta\bar{V}_{\alpha\beta}$, values are chosen purely for illustrative purposes; the Clapeyron-matching condition cannot be sensibly illustrated in SI-based units for any real system (Sidebar 11.6).]

The geometrical picture of phase coexistence is most easily sketched in terms of what may be called the *Clapeyron vector* $|\Sigma\rangle$ of each phase:

$$|\Sigma\rangle \equiv \bar{S}|\mathbf{T}\rangle - \bar{V}|\mathbf{P}\rangle \quad (11.164)$$

The Clapeyron vector (evidently closely related to the coexistence coordinate $|\sigma\rangle$ at saturation; cf. Section 11.5) is merely the projection of the GD vector (11.141b) onto the non-singular $\underline{\mathbf{M}}^{(2)}$ space spanned by $|\mathbf{T}\rangle$, $|\mathbf{P}\rangle$. In terms of this vector, the coexistence condition (11.154) can be written as the “Clapeyron matching condition”

$$\boxed{|\Sigma\rangle_\alpha = |\Sigma\rangle_\beta \equiv |\Sigma\rangle_{\alpha\beta}} \quad (11.165a)$$

between Clapeyron vectors of the two individual phases, as illustrated in Fig. 11.10. In effect, (11.165) is equivalent to a matching condition on the *lengths* of the Clapeyron vectors

$$|\Sigma_\alpha| = |\Sigma_\beta| \quad (11.165b)$$

because one is always free to orthogonally transform (“rotate”) the overall space $\mathcal{M}_S(\alpha)$ of pure phase α to match the orientation of $|\Sigma\rangle_\beta$ in phase β . In the common two-phase space $\mathcal{M}_S(\alpha\beta)$ in which the Clapeyron vectors $|\Sigma\rangle_\alpha$, $|\Sigma\rangle_\beta (= |\Sigma\rangle_{\alpha\beta})$ are coincident, the two-phase temperature and pressure vectors are given by

$$|\mathbf{T}\rangle_{\alpha\beta} = \frac{1}{\Delta\bar{S}_{\alpha\beta}} |\Sigma\rangle_{\alpha\beta} \quad (11.166)$$

$$|\mathbf{P}\rangle_{\alpha\beta} = \frac{1}{\Delta\bar{V}_{\alpha\beta}} |\Sigma\rangle_{\alpha\beta} \quad (11.167)$$

and the two-phase Clapeyron-type vector condition (11.154) is obviously satisfied.

It only remains to follow the heterogeneous phase-coexistence boundary to its critical limit, where the two phases become *identical*. In this limit, (11.163) becomes

$$C_{Vc} = C_{Pc} - T_c V_c \gamma_{\sigma c} \alpha_{Pc} \quad (11.168)$$

or, in view of (11.120),

$$C_{Vc} = C_{Pc} - \gamma_{\sigma c} (T_c V_c C_{Pc} \beta_{Tc})^{1/2} \quad (11.169)$$

The limiting critical slope $\gamma_{\sigma c}$ on the heterogeneous side must therefore be identified with the critical eigenangle ω_c [(11.130)] of the homogeneous side, namely,

$$\boxed{\gamma_{\sigma c} = \tan \omega_c = \lim_{T \rightarrow T_c} \left(\frac{\Delta \bar{S}_{\alpha\beta}}{\Delta \bar{V}_{\alpha\beta}} \right) = \tan \left\{ \frac{1}{2} \sin^{-1} \left[\frac{2(T_c V_c C_{Vc} \beta_{Sc})^{1/2}}{C_{Vc} + T_c V_c \beta_{Sc}} \right] \right\}} \quad (11.170)$$

consistent with the Clapeyron equation (7.27b). Equation (11.170) shows how the slope of the emergent phase boundary on the heterogeneous “broken symmetry” approach to (T_c, P_c) is already fixed by the direction (eigenangle ω_c) of the surviving major metric eigenvector along any path of homogeneous approach.

Geometrical Evaluation of Thermodynamic Derivatives

The systematic derivation of thermodynamic partial derivatives is of long-standing scientific and engineering importance. In a typical case, one wishes to evaluate some partial derivative that cannot be obtained directly from experiment, but instead must be derived from other known properties. Even for the homogeneous fluid ($f=2$) case, the number of possible equations among partial derivatives of the common thermodynamic functions is astronomical [of order 10^{10} , as estimated by H. Margenau and G. M. Murphy. *The Mathematics of Physics and Chemistry* (Van Nostrand, New York, 1943), p. 15] and grows explosively with increasing degrees of freedom f . A more systematic and general procedure is therefore essential to isolate a needed formula as thermodynamic complexity increases.

Numerous attacks on this problem have been presented over the years. Bridgman developed elaborate formula tables that became widely known through incorporation in early textbooks [P. W. Bridgman. *Phys. Rev.* **3**, 273 (1914); P. W. Bridgman. *A Condensed Collection of Thermodynamic Formulas* (Harvard University Press, Cambridge MA, 1925)]. However, these tables become increasingly unwieldy when extended to more complex systems [R. W. Goranson. *Thermodynamic Relations in Multicomponent Systems* (Carnegie Institution, Washington, DC, 1930)]. Shaw developed a more general and elegant method that makes use of algebraic properties of *Jacobian determinants* [A. N. Shaw. *Phil. Trans. R. Soc. Lond. A* **234**, 299 (1935)]. The tables required for the Jacobian method are both simpler and more easily constructed than those of Bridgman, which they largely supplanted. Various extensions of Jacobian methods beyond $f=2$ have been attempted, although they become increasingly unwieldy and lack the flexibility to deal with any but the most common thermodynamic properties [see, e.g., F. H. Crawford. *Am. J. Phys.* **17**, 1, (1949); *Proc. Am. Acad. Arts Sci.* **78**, 165 (1950); **83**, 191 (1955); F. S. Manning and W. P. Manning. *J. Chem. Phys.* **33**, 1554 (1960)]. Other table-free procedures have also been suggested, often involving indirect use of Jacobian-type manipulations [see, e.g., F. Lerman. *J. Chem. Phys.* **5**, 792 (1937); A. Tobolsky. *J. Chem. Phys.* **10**, 644 (1942); B. Carroll and A. Lehrman. *J. Chem. Ed.* **24**, 389 (1947); H. A. Bent. *J. Chem. Phys.* **21**, 1408 (1953); C. W. Carroll. *Am. J. Phys.* **27**, 302 (1959); J. S. Thomsen. *Am. J. Phys.* **32**, 666 (1964); L. Tisza. *Generalized Thermodynamics* (MIT Press, Cambridge, MA, 1966), p. 70].

In this chapter, we describe general geometrical (vector-algebraic) techniques for solving this problem and show their relationship to Shaw's method and other special cases. We initially consider the simplest case of a homogeneous fluid (Sections 12.1 and 12.2), then extend these methods to higher-dimensional multicomponent and multiphase systems and higher-order derivatives in Sections 12.3–12.6. Throughout this chapter, we exploit the general freedom to expand an arbitrary thermodynamic function Y and the associated vector $|\mathbf{Y}\rangle$ in a set of chosen basis vectors $\{|\mathbf{R}_i\rangle\}$,

$$|\mathbf{Y}\rangle = \sum_{i=1}^f c_i |\mathbf{R}_i\rangle \quad (12.1)$$

with coefficients c_i given by

$$c_i = \langle \bar{\mathbf{R}}_i | \mathbf{Y} \rangle = \left(\frac{\partial Y}{\partial R_i} \right)_R \quad (12.2)$$

as follows easily from the general completeness theorem

$$1 = \sum_{i=1}^f |\mathbf{R}_i\rangle \langle \bar{\mathbf{R}}_i| = \sum_{i=1}^f |\bar{\mathbf{R}}_i\rangle \langle \mathbf{R}_i| \quad (12.3)$$

for biorthogonal vectors satisfying

$$\langle \mathbf{R}_i | \bar{\mathbf{R}}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (12.4)$$

12.1 THERMODYNAMIC VECTORS AND DERIVATIVES

Let us consider the simplest case of a homogeneous, single-component fluid of fixed mass, with two degrees of freedom. We seek to evaluate a general partial derivative $\mathcal{D}$ of the form

$$\mathcal{D} = \left(\frac{\partial X}{\partial Y} \right)_Z \quad (12.5)$$

where X, Y, Z are some chosen set of state properties. In order to obtain a general expression for such derivatives, we suppose that a pair of basis intensities R_1 and R_2 with associated conjugates $\bar{R}_1$ and $\bar{R}_2$ have been chosen. Because of the *mutual* nature of the conjugacy relation (Section 11.2), we can write

$$\bar{\bar{R}}_i = (\bar{\bar{R}}_i) = R_i \quad (12.6)$$

In this two-dimensional space, we can also describe the chosen basis intensities as *complementary*, and let the prime symbol on R'_i denote the *complement* of R_i , so that

$$R'_1 \equiv R_2, \quad R'_2 \equiv R_1 \quad (12.7)$$

With these notational conventions, we can uniquely identify the state variables Z' (complement), $\bar{Z}$ (conjugate) and $\bar{\bar{Z}}'$ (conjugate complement) for any chosen variable Z .

For example, when temperature T and negative pressure $-P$ are the chosen R_i ,

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} T \\ -P \end{pmatrix} \quad (12.8a)$$

with entropy S and volume V the corresponding conjugate $\bar{R}_i$,

$$\begin{pmatrix} \bar{R}_1 \\ \bar{R}_2 \end{pmatrix} = \begin{pmatrix} S \\ V \end{pmatrix} \quad (12.8b)$$

then $Z, \bar{Z}, Z', \bar{Z}'$ have the significance shown in Table 12.1. This choice of notation also allows us to express the biorthogonality condition (12.4) in the form

$$\langle Z | \bar{Z} \rangle = 1 \quad (12.9a)$$

$$\langle Z | \bar{Z}' \rangle = 0 \quad (12.9b)$$

for a general variable Z .

The standard scalar products among vectors $|\mathbf{T}\rangle, |-\mathbf{P}\rangle, |\mathbf{S}\rangle, |\mathbf{V}\rangle$ are gathered for convenience in Table 12.2, expressed in terms of standard response functions $C_P, C_V, \beta_T, \beta_S, \alpha_P, \Gamma_V$ [see (11.27)–(11.30)]. As described previously, scalar products for $|\mathbf{S}\rangle$ and $|\mathbf{V}\rangle$ (i.e., involving properties C_P, β_T, α_P) are obtained by matrix inversion from those for $|\mathbf{T}\rangle$ and $|\mathbf{P}\rangle$ (i.e., involving C_V, β_S, Γ_V). The vector-algebraic procedure to be described will automatically express any desired derivative in terms of the six properties in Table 12.2, and these expressions may subsequently be reduced (if desired) to involve only *three* independent properties by identities previously introduced [cf. (11.39)–(11.42)], consistent with the “ $f(f+1)/2$ rule.”

In order to evaluate a general partial derivative (12.5), one might attempt to express dZ in the form

$$dZ = \lambda dX + \mu dY \quad (12.10)$$

so that the desired derivative is

$$\mathcal{D} = \left(\frac{\partial X}{\partial Y} \right)_Z = -\frac{\mu}{\lambda} \quad (12.11)$$

In the geometrical representation, (12.10) is expressed as the vector relationship

$$|\mathbf{Z}\rangle = \lambda |\mathbf{X}\rangle + \mu |\mathbf{Y}\rangle \quad (12.12)$$

TABLE 12.1 “Standard” Variables for a Simple Fluid

Z	$\bar{Z}$	Z'	$\bar{Z}'$
T	S	$-P$	V
$-P$	V	T	S
S	T	V	$-P$
V	$-P$	S	T

TABLE 12.2 “Standard” Scalar Products for a Simple Fluid

	$ \mathbf{T}\rangle$	$ \mathbf{-P}\rangle$	$ \mathbf{S}\rangle$	$ \mathbf{V}\rangle$
$\langle\mathbf{T} $	T/C_V	$-T/\Gamma_V$	1	0
$\langle\mathbf{-P} $	$-T/\Gamma_V$	$1/V\beta_S$	0	1
$\langle\mathbf{S} $	1	0	C_P/T	$V\alpha_P$
$\langle\mathbf{V} $	0	1	$V\alpha_P$	$V\beta_T$

and the coefficients λ , μ are readily evaluated by ordinary vector methods. For example, if we take the scalar product of (12.12) with the conjugate complement vector $|\bar{\mathbf{Z}}'\rangle$ and use (12.9b), the result is

$$0 = \lambda\langle\mathbf{X}|\bar{\mathbf{Z}}'\rangle + \mu\langle\mathbf{Y}|\bar{\mathbf{Z}}'\rangle \quad (12.13)$$

so that $\mathcal{D}$ is evaluated simply as a ratio of projections of the vector $|\bar{\mathbf{Z}}'\rangle$ onto the vectors $|\mathbf{X}\rangle$ and $|\mathbf{Y}\rangle$:

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{\langle\mathbf{X}|\bar{\mathbf{Z}}'\rangle}{\langle\mathbf{Y}|\bar{\mathbf{Z}}'\rangle} \quad (12.14)$$

The application of this equation, to be used in conjunction with Tables 12.1 and 12.2, is illustrated with simple examples in Sidebars 12.1 and 12.2.

SIDEBAR 12.1: “MIXED” ELEMENTARY DERIVATIVES

Let us consider the application of (12.14) to evaluate such derivatives as

$$\left(\frac{\partial S}{\partial V}\right)_T, \quad \left(\frac{\partial S}{\partial V}\right)_P, \quad \left(\frac{\partial P}{\partial T}\right)_S, \quad \left(\frac{\partial P}{\partial T}\right)_V \quad (\text{S12.1-1})$$

in which mixtures of intensive and extensive properties serve as independent variables. Noting from Table 12.1 that $\bar{T}' = V$, and, using the “standard” scalar products defined in Table 12.2, we obtain

$$\left(\frac{\partial S}{\partial V}\right)_T = \frac{\langle\mathbf{S}|\bar{\mathbf{T}}'\rangle}{\langle\mathbf{V}|\bar{\mathbf{T}}'\rangle} = \frac{\langle\mathbf{S}|\mathbf{V}\rangle}{\langle\mathbf{V}|\mathbf{V}\rangle} = \frac{V\alpha_P}{V\beta_T} = \frac{\alpha_P}{\beta_T} \quad (\text{S12.1-2})$$

Similarly, because $\bar{S}' = -P$,

$$\left(\frac{\partial P}{\partial T}\right)_S = -\frac{\langle\mathbf{-P}|\mathbf{-P}\rangle}{\langle\mathbf{T}|\mathbf{-P}\rangle} = -\frac{(1/V\beta_S)}{(-T/\Gamma_V)} = \frac{\Gamma_V}{TV\beta_S} \quad (\text{S12.1-3})$$

The reader may verify in analogous manner the identities

$$\left(\frac{\partial S}{\partial V}\right)_P = \frac{C_P}{TV\alpha_P} \quad (\text{S12.1-4})$$

$$\left(\frac{\partial S}{\partial V}\right)_V = \frac{C_V}{\Gamma_V} \quad (\text{S12.1-5})$$

SIDEBAR 12.2: SATURATION DERIVATIVES

The application of (12.14) to variables other than the standard S , V , T and $-P$ can be illustrated with the saturation properties C_σ and α_σ (Section 11.5):

$$C_\sigma = T \left(\frac{\partial S}{\partial T}\right)_\sigma \quad (\text{S12.2-1})$$

$$\alpha_\sigma = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_\sigma \quad (\text{S12.2-2})$$

These involve the “coexistence coordinate” σ ,

$$|\sigma\rangle = \gamma_\sigma |\mathbf{T}\rangle + |-\mathbf{P}\rangle \quad (\text{S12.2-3})$$

where γ_σ is the slope of the coexistence curve. With T and σ as chosen basis variables, the conjugate variables were found to be (11.70)

$$|\bar{\mathbf{T}}\rangle = |\mathbf{S}\rangle - \gamma_\sigma |\mathbf{V}\rangle \quad (\text{S12.2-4})$$

$$|\bar{\sigma}\rangle = |\mathbf{V}\rangle \quad (\text{S12.2-5})$$

from which we recognize, for example, that

$$|\bar{\sigma}'\rangle = |\bar{\mathbf{T}}\rangle = |\mathbf{S}\rangle - \gamma_\sigma |\mathbf{V}\rangle \quad (\text{S12.2-6})$$

The saturation heat capacity C_σ is therefore found from

$$\frac{C_\sigma}{T} = \left(\frac{\partial S}{\partial T}\right)_\sigma = \frac{\langle \mathbf{S} | \bar{\sigma}' \rangle}{\langle \mathbf{T} | \bar{\sigma}' \rangle} = \frac{\langle \mathbf{S} | \mathbf{S} \rangle - \gamma_\sigma \langle \mathbf{S} | \mathbf{V} \rangle}{\langle \mathbf{T} | \mathbf{S} \rangle - \gamma_\sigma \langle \mathbf{T} | \mathbf{V} \rangle} = \frac{(C_P/T) - \gamma_\sigma (V\alpha_P)}{(1) - \gamma_\sigma(0)} \quad (\text{S12.2-7})$$

which is the identity

$$C_\sigma = C_P - TV\gamma_\sigma\alpha_P \quad (\text{S12.2-8})$$

The reader may verify the analogous derivation for the thermal expansion coefficient α_σ

$$\alpha_\sigma = \alpha_P - \gamma_\sigma\beta_T \quad (\text{S12.2-9})$$

If the scalar products appearing in Eq. (12.14) are somehow not convenient (e.g., because the vector $|\bar{\mathbf{Z}}'\rangle$ is not readily available), one can return to solve (12.12) in other forms. For instance, if we take the scalar product of (12.12) successively with the vectors $|\bar{\mathbf{X}}'\rangle$ and $|\bar{\mathbf{Y}}'\rangle$, the result is

$$\langle \bar{\mathbf{X}}' | \mathbf{Z} \rangle = \lambda(0) + \mu \langle \bar{\mathbf{X}}' | \mathbf{Z} \rangle \quad (12.15a)$$

$$\langle \bar{\mathbf{Y}}' | \mathbf{Z} \rangle = \lambda \langle \bar{\mathbf{Y}}' | \mathbf{X} \rangle + \mu(0) \quad (12.15b)$$

Equation (12.11) then becomes

$$\boxed{\left(\frac{\partial X}{\partial Y} \right)_Z = - \frac{\langle \mathbf{X} | \bar{\mathbf{Y}}' \rangle \langle \bar{\mathbf{X}}' | \mathbf{Z} \rangle}{\langle \mathbf{Y} | \bar{\mathbf{X}}' \rangle \langle \bar{\mathbf{Y}}' | \mathbf{Z} \rangle}} \quad (12.16)$$

which is again rather easily remembered or rederived as the occasion warrants. Some applications of (12.16) are illustrated in Sidebars 12.3 and 12.4.

SIDEBAR 12.3: IDENTITIES AMONG BASIC RESPONSE FUNCTIONS

Equation (12.16) is useful in recovering various identities that connect the entries of Table 12.2. For example, from the definition $C_V = T(\partial S / \partial T)_V$, we deduce

$$\begin{aligned} \frac{C_V}{T} &= \left(\frac{\partial S}{\partial T} \right)_V = - \frac{\langle \mathbf{S} | \bar{\mathbf{T}}' \rangle \langle \bar{\mathbf{S}}' | \mathbf{V} \rangle}{\langle \mathbf{T} | \bar{\mathbf{S}}' \rangle \langle \bar{\mathbf{T}}' | \mathbf{V} \rangle} \\ &= - \frac{\langle \mathbf{S} | \mathbf{V} \rangle \langle -\mathbf{P} | \mathbf{V} \rangle}{\langle \mathbf{T} | -\mathbf{P} \rangle \langle \mathbf{V} | \mathbf{V} \rangle} = - \frac{(V\alpha_P)(1)}{(-T/\Gamma_V)(V\beta_T)} \end{aligned} \quad (\text{S12.3-1})$$

which is the identity [cf. (11.42)]

$$C_V \beta_T = \alpha_P \Gamma_V \quad (\text{S12.3-2})$$

and so forth.

SIDEBAR 12.4: MORE SATURATION PROPERTIES

In treating saturation properties $(\partial X / \partial Y)_\sigma$ with (12.16), we note that only σ itself (but not $\bar{\sigma}$ or $\bar{\sigma}'$) is needed. Hence, we can return to the “standard” choice of basis variables from Table 12.1,

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} T \\ -P \end{pmatrix}, \quad \begin{pmatrix} \bar{R}_1 \\ \bar{R}_2 \end{pmatrix} = \begin{pmatrix} S \\ V \end{pmatrix} = \begin{pmatrix} \bar{T} \\ -\bar{P} \end{pmatrix} \quad (\text{S12.4-1})$$

rather than the special set

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} T \\ \sigma \end{pmatrix}, \quad \begin{pmatrix} \bar{R}_1 \\ \bar{R}_2 \end{pmatrix} = \begin{pmatrix} S - \gamma_\sigma V \\ V \end{pmatrix} = \begin{pmatrix} \bar{T} \\ \bar{\sigma} \end{pmatrix} \quad (\text{S12.4-2})$$

that was employed in (11.70) and following. For example, α_σ is found from

$$\begin{aligned} V\alpha_\sigma &= \left(\frac{\partial V}{\partial T} \right)_\sigma = \frac{\langle \mathbf{V} | \bar{\mathbf{T}}' \rangle \langle \bar{\mathbf{V}}' | \sigma \rangle}{\langle \mathbf{T} | \bar{\mathbf{V}}' \rangle \langle \bar{\mathbf{T}}' | \sigma \rangle} = - \frac{\langle \mathbf{V} | \mathbf{V} \rangle \langle \mathbf{T} | \sigma \rangle}{\langle \mathbf{T} | \mathbf{T} \rangle \langle \mathbf{V} | \sigma \rangle} \\ &= - \frac{\langle \mathbf{V} | \mathbf{V} \rangle [\gamma_\sigma \langle \mathbf{T} | \mathbf{T} \rangle + \langle \mathbf{T} | -\mathbf{P} \rangle]}{\langle \mathbf{T} | \mathbf{T} \rangle [\gamma_\sigma \langle \mathbf{V} | \mathbf{T} \rangle + \langle \mathbf{V} | -\mathbf{P} \rangle]} \\ &= - \frac{(V\beta_T) [\gamma_\sigma (T/C_V) + (-T/\Gamma_V)]}{(T/C_V) [\gamma_\sigma(0) + (1)]} \\ &= -V\beta_T(\gamma_\sigma - C_V/\Gamma_V) \end{aligned} \quad (\text{S12.4-3})$$

which, in conjunction with (S12.3-2), is again (S12.2-9). Equation (S12.2-8) for C_σ is recovered in an analogous manner. Note that α_σ could have been evaluated as easily in the special basis set (S12.4-2) [because V is a “natural” variable in both (S12.4-1) and (S12.4-2)], but C_σ could not.

Equations (12.14) and (12.16) retain their validity even when X , Y and Z do not belong to a single set of base variables and conjugates. (For example, each might be expressed as a linear combination of two other base variables.) In such cases, one may select the complementary variables rather arbitrarily. For example, it is often convenient to define X and Y to be complementary variables:

$$X' = Y, \quad Y' = X \quad (12.17)$$

With this choice, (12.16) becomes [in view of (12.9a)]

$$\boxed{\left(\frac{\partial X}{\partial Y} \right)_Z = - \frac{\langle \bar{\mathbf{Y}} | \mathbf{Z} \rangle}{\langle \bar{\mathbf{X}} | \mathbf{Z} \rangle}} \quad (12.18)$$

which is again a compact and easily remembered form. Of course, the conjugate variables $\bar{X}$ and $\bar{Y}$ of (12.18) must now be defined specifically with respect to the choice of complementary fields implied by (12.17), because, for example, “ $\bar{X}$ ” is wholly undefined until X' has been specified. Some applications of (12.18) are illustrated in Sidebars 12.5 and 12.6.

SIDEBAR 12.5: MORE IDENTITIES

Because S and V are complementary variables in Table 12.1, we can use (12.18) to evaluate, for example,

$$\left(\frac{\partial S}{\partial V} \right)_T = - \frac{\langle \bar{\mathbf{V}} | \mathbf{T} \rangle}{\langle \bar{\mathbf{S}} | \mathbf{T} \rangle} = - \frac{\langle -\mathbf{P} | \mathbf{T} \rangle}{\langle \mathbf{T} | \mathbf{T} \rangle} = - \frac{(-T/\Gamma_V)}{(T/C_V)} = \frac{C_V}{\Gamma_V} \quad (\text{S12.5-1})$$

In conjunction with the corresponding identity (S12.1-2), this is again equivalent to the identity (S12.3-2). Analogous saturation properties such as $(\partial S/\partial V)_\sigma$ are also handled easily with (12.18).

SIDEBAR 12.6: JOULE–THOMSON COEFFICIENT

Analysis of the Joule–Thomson experiment (Section 3.6.3) requires evaluation of the derivative $(\partial T/\partial P)_H$ at constant *enthalpy* ($H = U + PV$). The differential expression for dH ,

$$dH = T dS + V dP \quad (\text{S12.6-1})$$

gives the corresponding expression for the enthalpy vector $|\mathbf{H}\rangle$ in terms of base vectors $|\mathbf{T}\rangle$, $|\mathbf{-P}\rangle$ as

$$|\mathbf{H}\rangle = T|\mathbf{S}\rangle - V|\mathbf{-P}\rangle \quad (\text{S12.6-2})$$

In terms of Table 12.1, T and $-P$ are already standard complementary variables, so the Joule–Thomson coefficient is conveniently evaluated from (12.18), with $X = T$, $Y = -P$, $Z = H$:

$$\begin{aligned} \left(\frac{\partial T}{\partial P}\right)_H &= + \frac{\langle -\mathbf{P}|\mathbf{H}\rangle}{\langle \mathbf{T}|\mathbf{H}\rangle} = \frac{\langle \mathbf{V}|\mathbf{H}\rangle}{\langle \mathbf{S}|\mathbf{H}\rangle} = \frac{T\langle \mathbf{V}|\mathbf{S}\rangle - V\langle \mathbf{V}|\mathbf{-P}\rangle}{T\langle \mathbf{S}|\mathbf{S}\rangle - V\langle \mathbf{S}|\mathbf{-P}\rangle} \\ &= \frac{T(V\alpha_P) - V(1)}{T(C_P/T) - V(0)} = \frac{V}{C_P}(T\alpha_P - 1) \end{aligned} \quad (\text{S12.6-3})$$

In a similar manner, the reader may readily verify analogous derivatives involving other common thermodynamic potentials (G , A , U , S):

$$\left(\frac{\partial T}{\partial P}\right)_G = \frac{V}{S} \quad (\text{S12.6-4})$$

$$\left(\frac{\partial T}{\partial P}\right)_A = \frac{PV\beta_T}{(S + PV\alpha_P)} \quad (\text{S12.6-5})$$

$$\left(\frac{\partial T}{\partial P}\right)_U = \frac{V(T\alpha_P - P\beta_T)}{(C_P - PV\alpha_P)} \quad (\text{S12.6-6})$$

$$\left(\frac{\partial T}{\partial P}\right)_S = \frac{TV\alpha_P}{C_P} \quad (\text{S12.6-7})$$

It may finally be remarked that equations such as (12.14), (12.16), (12.18), which represent partial derivatives as simple ratios of geometrical projections, make it easy (as well as largely superfluous) to recover various identities among partial derivatives that are often taken as starting points in thermodynamic manipulations (cf. Section 1.2).

12.2 GENERAL SOLUTION FOR TWO DEGREES OF FREEDOM AND RELATIONSHIP TO JACOBIAN METHODS

In order to obtain the ratio $\mathcal{D} = -\mu/\lambda$ of coefficients from the vector equation (12.12) in its most general form, let $|\mathbf{A}\rangle$ and $|\mathbf{B}\rangle$ represent arbitrary vectors in the two-dimensional space, corresponding to thermodynamic variables A , B , respectively. The successive scalar products of these vectors with (12.12) then lead to the linear equations

$$\langle \mathbf{A} | \mathbf{Z} \rangle = \lambda \langle \mathbf{A} | \mathbf{X} \rangle + \mu \langle \mathbf{A} | \mathbf{Y} \rangle \quad (12.19a)$$

$$\langle \mathbf{B} | \mathbf{Z} \rangle = \lambda \langle \mathbf{B} | \mathbf{X} \rangle + \mu \langle \mathbf{B} | \mathbf{Y} \rangle \quad (12.19b)$$

These equations can be solved in the usual manner

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \langle \mathbf{A} | \mathbf{X} \rangle & \langle \mathbf{A} | \mathbf{Y} \rangle \\ \langle \mathbf{B} | \mathbf{X} \rangle & \langle \mathbf{B} | \mathbf{Y} \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle \mathbf{A} | \mathbf{Z} \rangle \\ \langle \mathbf{B} | \mathbf{Z} \rangle \end{pmatrix} \quad (12.20)$$

for the unknown coefficients λ , μ , provided that

$$\langle \mathbf{A} | \mathbf{X} \rangle \langle \mathbf{B} | \mathbf{Y} \rangle - \langle \mathbf{A} | \mathbf{Y} \rangle \langle \mathbf{B} | \mathbf{X} \rangle \neq 0 \quad (12.21)$$

Solving for the 2×2 inverse matrix in (12.20), we obtain the desired ratio in the form

$$\boxed{\left(\frac{\partial X}{\partial Y} \right)_Z = \frac{\langle \mathbf{X} | \mathbf{A} \rangle \langle \mathbf{B} | \mathbf{Z} \rangle - \langle \mathbf{X} | \mathbf{B} \rangle \langle \mathbf{A} | \mathbf{Z} \rangle}{\langle \mathbf{Y} | \mathbf{A} \rangle \langle \mathbf{B} | \mathbf{Z} \rangle - \langle \mathbf{Y} | \mathbf{B} \rangle \langle \mathbf{A} | \mathbf{Z} \rangle}} \quad (12.22)$$

Representative applications of this equation are illustrated in Sidebar 12.7.

SIDEBAR 12.7: IDENTITIES FOR C_V

From the definition $C_V = T(\partial S/\partial T)_V$ and (12.22), we obtain the general expression

$$\frac{C_V}{T} = \frac{\langle \mathbf{S} | \mathbf{A} \rangle \langle \mathbf{B} | \mathbf{V} \rangle - \langle \mathbf{S} | \mathbf{B} \rangle \langle \mathbf{A} | \mathbf{V} \rangle}{\langle \mathbf{T} | \mathbf{A} \rangle \langle \mathbf{B} | \mathbf{V} \rangle - \langle \mathbf{T} | \mathbf{B} \rangle \langle \mathbf{A} | \mathbf{V} \rangle} \quad (\text{S12.7-1})$$

for any chosen variables A and B . (Note, however, that one merely recovers the definition $C_V/T = \langle \mathbf{T} | \mathbf{T} \rangle^{-1}$ if either A or B is T .) With the choice

$$A = S, \quad B = V \quad (\text{S12.7-2})$$

this becomes

$$\frac{C_V}{T} = \frac{(C_P/T)(V\beta_T) - (V\alpha_P)(V\alpha_P)}{(1)(V\beta_T) - (0)(V\alpha_P)} \quad (\text{S12.7-3})$$

which is the identity

$$C_V = C_P - TV\alpha_P^2/\beta_T \quad (\text{S12.7-4})$$

Similarly, the choice

$$A = S, \quad B = -P \quad (\text{S12.7-5})$$

leads to

$$C_V = \frac{C_P}{1 + TV\alpha_P} \quad (\text{S12.7-6})$$

while the choice

$$A = V, \quad B = -P \quad (\text{S12.7-7})$$

leads to

$$C_V = \alpha_P \Gamma_V / \beta_T \quad (\text{S12.7-8})$$

which was previously obtained in Sidebar 12.3. Other examples of such identities [all of which are special instances of (11.42)] can be readily obtained by treating the other standard response functions in analogous manner.

The general solution (12.22) reduces to the simpler forms considered previously when we introduce special choices for the variables A and B . Thus, the particular choices

$$A = \bar{Z}, \quad B = \bar{Z}' \quad (12.23a)$$

$$A = \bar{X}', \quad B = \bar{Y}' \quad (12.23b)$$

$$A = \bar{Y}, \quad B = \bar{X} = A' \quad (12.23c)$$

lead to (12.14), (12.16), (12.18), respectively.

Another interesting special form of the solution (12.22) arises if we write the general scalar product $\langle \mathbf{X} | \mathbf{Y} \rangle$ in the form

$$\langle X | Y \rangle = \left(\frac{\partial X}{\partial \bar{Y}} \right)_{\bar{Y}'} \quad (12.24)$$

which is merely the general scalar product formula [cf. (10.9)]

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = \left(\frac{\partial R_i}{\partial \bar{R}_j} \right)_{\bar{R}} \quad (12.25)$$

for a two-dimensional metric space, re-expressed in the notation of (12.6), (12.7) and Table 12.1. Equation (12.22) then becomes

$$\left(\frac{\partial X}{\partial \bar{Y}} \right)_Z = \frac{(\partial X / \partial \bar{A})_{\bar{A}'} (\partial Z / \partial \bar{B})_{\bar{B}'} - (\partial Z / \partial \bar{A})_{\bar{A}'} (\partial X / \partial \bar{B})_{\bar{B}'}}{(\partial Y / \partial \bar{A})_{\bar{A}'} (\partial Z / \partial \bar{B})_{\bar{B}'} - (\partial Z / \partial \bar{A})_{\bar{A}'} (\partial Y / \partial \bar{B})_{\bar{B}'}} \quad (12.26)$$

If $\bar{A}$ and $\bar{B}$ are now chosen as *complementary* variables in each term,

$$\bar{A}' = \bar{B}, \quad \bar{B}' = \bar{A} \quad (12.27)$$

then (12.26) becomes

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{(\partial X/\partial \bar{A})_{\bar{B}}(\partial Z/\partial \bar{B})_{\bar{A}} - (\partial Z/\partial \bar{A})_{\bar{B}}(\partial X/\partial \bar{B})_{\bar{A}}}{(\partial Y/\partial \bar{A})_{\bar{B}}(\partial Z/\partial \bar{B})_{\bar{A}} - (\partial Z/\partial \bar{A})_{\bar{B}}(\partial Y/\partial \bar{B})_{\bar{A}}} \quad (12.28)$$

which can be recognized as a ratio of *Jacobian determinants*,

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{J(X, Y)}{J(Y, Z)} \quad (12.29)$$

where, for example,

$$J(X, Z) = \frac{\partial(X, Z)}{\partial(\bar{A}, \bar{B})} = \det \begin{vmatrix} (\partial X/\partial \bar{A})_{\bar{B}} & (\partial Z/\partial \bar{A})_{\bar{B}} \\ (\partial X/\partial \bar{B})_{\bar{A}} & (\partial Z/\partial \bar{B})_{\bar{A}} \end{vmatrix} \quad (12.30)$$

and $J(Y, Z)$ is defined correspondingly (the mutual dependence on fixed $\bar{A}$ and $\bar{B}$ being understood implicitly). Equation (12.29) can be recognized as the basic equation underlying Shaw's tables [A. N. Shaw. *Phil. Trans. R. Soc. Lond. A* **234**, 299 (1935)]. Thus, the relative simplicity of Jacobian methods can be understood in terms of their rather close connection with the underlying thermodynamic geometry. Nevertheless, (12.22), (12.23) and the examples of Section 12.1 show that the Jacobian solution is neither the simplest nor the most general algebraic representation of a general thermodynamic derivative.

Let us now examine the deeper geometrical significance of (12.26). For this purpose, it is convenient to select A, B as *self-conjugate* variables (corresponding to orthonormal thermodynamic vectors), such as the metric *eigenmodes* E_1, E_2 :

$$A = E_1, \quad B = E_2 \quad (12.31)$$

These eigen-coordinates satisfy self-conjugacy and orthonormality properties that are particularly convenient for geometrical representation:

$$E_1 = \bar{E}_1, \quad E_2 = \bar{E}_2 \quad (12.32a)$$

$$\langle \mathbf{E}_1 | \mathbf{E}_1 \rangle = \langle \mathbf{E}_2 | \mathbf{E}_2 \rangle = 1 \quad (12.32b)$$

$$\langle \mathbf{E}_1 | \mathbf{E}_2 \rangle = 0 \quad (12.32c)$$

In this case, we can choose $|\mathbf{E}_1\rangle$ and $|\mathbf{E}_2\rangle$ as the mutually perpendicular "unit axes" of a Cartesian xy coordinate system, with abscissa E_1 and ordinate E_2 :

$$x = E_1, \quad y = E_2 \quad (12.33)$$

In this coordinate system, the scalar products in (12.22) are expressed geometrically as

$$\langle \mathbf{X} | \mathbf{A} \rangle = \langle \mathbf{A} | \mathbf{X} \rangle = \langle \mathbf{X} | \mathbf{E}_1 \rangle = |\mathbf{X}| \cos \theta_X \quad (12.34a)$$

$$\langle \mathbf{X} | \mathbf{B} \rangle = \langle \mathbf{B} | \mathbf{X} \rangle = \langle \mathbf{X} | \mathbf{E}_2 \rangle = |\mathbf{X}| \sin \theta_X \quad (12.34b)$$

$$\langle \mathbf{Y} | \mathbf{A} \rangle = \langle \mathbf{A} | \mathbf{Y} \rangle = \langle \mathbf{Y} | \mathbf{E}_1 \rangle = |\mathbf{Y}| \cos \theta_Y \quad (12.34c)$$

$$\langle \mathbf{Y} | \mathbf{B} \rangle = \langle \mathbf{B} | \mathbf{Y} \rangle = \langle \mathbf{Y} | \mathbf{E}_2 \rangle = |\mathbf{Y}| \sin \theta_Y \quad (12.34d)$$

$$\langle \mathbf{Z} | \mathbf{A} \rangle = \langle \mathbf{A} | \mathbf{Z} \rangle = \langle \mathbf{Z} | \mathbf{E}_1 \rangle = |\mathbf{Z}| \cos \theta_Z \quad (12.34e)$$

$$\langle \mathbf{Z} | \mathbf{B} \rangle = \langle \mathbf{B} | \mathbf{Z} \rangle = \langle \mathbf{Z} | \mathbf{E}_2 \rangle = |\mathbf{Z}| \sin \theta_Z \quad (12.34f)$$

where θ_X , θ_Y , θ_Z are the angles and $|\mathbf{X}|$, $|\mathbf{Y}|$, $|\mathbf{Z}|$ are the lengths of the thermodynamic $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$, $|\mathbf{Z}\rangle$ vectors in a two-dimensional Cartesian representation as shown in Fig. 12.1. Inserting (12.34a–f) into (12.22), we obtain

$$\begin{aligned} \left(\frac{\partial X}{\partial Y}\right)_Z &= \frac{(|\mathbf{X}| \cos \theta_X)(|\mathbf{Z}| \sin \theta_Z) - (|\mathbf{X}| \sin \theta_X)(|\mathbf{Z}| \cos \theta_Z)}{(|\mathbf{Y}| \cos \theta_Y)(|\mathbf{Z}| \sin \theta_Z) - (|\mathbf{Y}| \sin \theta_Y)(|\mathbf{Z}| \cos \theta_Z)} \\ &= \frac{|\mathbf{X}| \cos \theta_X \sin \theta_Z - \sin \theta_X \cos \theta_Z}{|\mathbf{Y}| \cos \theta_Y \sin \theta_Z - \sin \theta_Y \cos \theta_Z} \end{aligned} \quad (12.35)$$

or, with the familiar trigonometric law for subtracting angles,

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{|\mathbf{X}| \sin(\theta_Z - \theta_X)}{|\mathbf{Y}| \sin(\theta_Z - \theta_Y)} \quad (12.36)$$

Equation (12.36) has a particularly simple geometrical interpretation, as shown in Fig. 12.1. If we drop perpendiculars from each vector $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$ to the vector $|\mathbf{Z}\rangle$ to obtain the “normal distances” of each vector $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$ from the common $\hat{\mathbf{Z}}$ direction [denoted X_Z and Y_Z (dashed lines) in Fig. 12.1], then $(\partial X/\partial Y)_Z$ is seen to be given by

$$\left(\frac{\partial X}{\partial Y}\right)_Z = \frac{X_Z}{Y_Z} = \text{ratio of dropped perpendiculars} \quad (12.37)$$

As can be seen from (12.35)–(12.37), the partial derivative value does not depend on the length $|\mathbf{Z}|$ of $|\mathbf{Z}\rangle$, but only on the angles that $\hat{\mathbf{Z}}$ makes with $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$. The choice of axes is also immaterial; any self-conjugate set satisfying (12.32a–c) is convenient for evaluating the scalar products (12.34a–f), but the final result (12.38) depends only on the internal geometry of the $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$, $|\mathbf{Z}\rangle$ vectors, not on their orientation in chosen axes. Thus, given the *geometry* of the $|\mathbf{X}\rangle$, $|\mathbf{Y}\rangle$ vectors with respect to the chosen $\hat{\mathbf{Z}}$ axis, the thermodynamic derivative $(\partial X/\partial Y)_Z$ is rather trivial to evaluate.

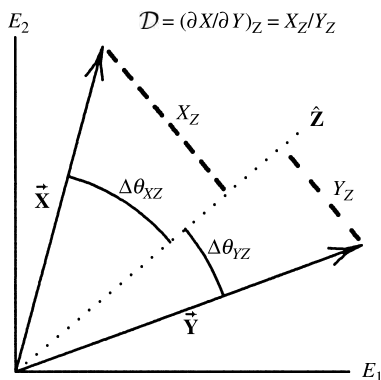


Figure 12.1 Geometrical evaluation of the thermodynamic derivative $\mathcal{D} = (\partial X/\partial Y)_Z$ according to (12.36), showing the “normal distances” (X_Z , Y_Z , dashed lines) of dropped perpendiculars from the vectors $\vec{X}$, $\vec{Y}$ to the $\hat{\mathbf{Z}}$ -axis, whose ratio gives the numerical value ($\mathcal{D} = X_Z/Y_Z$) of the derivative.

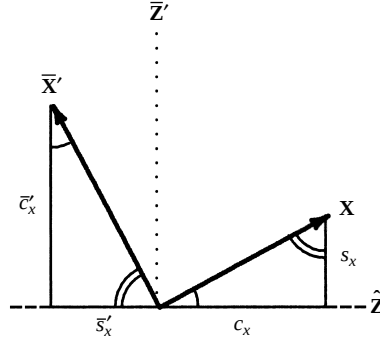


Figure 12.2 Schematic representation of conjugate complement vectors $\mathbf{X}$, $\bar{\mathbf{X}}'$ in the $(\mathbf{Z}, \bar{\mathbf{Z}}')$ reference frame, showing congruent angles leading to the geometrical identities of (12.38). Projections of $\mathbf{X}$, $\bar{\mathbf{X}}'$ onto the $\bar{\mathbf{Z}}'$ ordinate give additional congruencies used in other identities.

The geometry underlying the thermodynamic identity (12.18) is shown in Fig. 12.2. Taking the $\hat{\mathbf{Z}}$ axis (dashed line) along the abscissa, we draw the conjugate (perpendicular) vectors $\mathbf{X}$ and $\bar{\mathbf{X}}'$ from the origin to form the triangles with edges $(|\mathbf{X}|, s_x, c_x)$ and $(|\bar{\mathbf{X}}'|, \bar{s}_x', c_x')$, respectively. The angle θ_x between $\hat{\mathbf{Z}}$ and $\mathbf{X}$ is marked with a single arc, and the complementary angle $(90^\circ - \theta_x)$ with a double arc. Elementary geometry shows the congruency with the corresponding angles marked in the $(|\bar{\mathbf{X}}'|, \bar{s}_x', c_x')$ triangle, giving the geometrical identities between triangle edges:

$$\frac{s_x}{\bar{s}_x'} = \frac{c_x}{\bar{c}_x'} = \frac{|\mathbf{X}|}{|\bar{\mathbf{X}}'|} \quad (12.38)$$

The analogous construction for the vectors $\mathbf{Y}$, $\bar{\mathbf{Y}}'$ gives similarly

$$\frac{s_y}{\bar{s}_y'} = \frac{c_y}{\bar{c}_y'} = \frac{|\mathbf{Y}|}{|\bar{\mathbf{Y}}'|} \quad (12.39)$$

The ratio s_x/s_y (cf. X_Z/Y_Z of Fig. 12.1) therefore satisfies

$$\frac{s_x}{s_y} = \frac{\bar{s}_x'}{\bar{s}_y'} = \frac{\bar{c}_x'}{\bar{c}_y'} = \left(\frac{\partial X}{\partial Y} \right)_Z \quad (12.40)$$

which leads to (12.18) and other identities.

12.3 GENERAL PARTIAL DERIVATIVES IN HIGHER-DIMENSIONAL SYSTEMS

The extension of vector-algebraic techniques to multicomponent systems of higher dimensionality (degrees of freedom $f > 2$) can be carried out straightforwardly, even though one loses the convenience of mutually complementary pairs (X, X') and orthogonal complementary conjugates $(X, \bar{X}')$ that are a special feature of $f = 2$. In a space of f dimensions, a

general thermodynamic derivative $\mathcal{D}$ will have the form

$$\mathcal{D} = \left(\frac{\partial X}{\partial Y} \right)_{\underline{Z}} \quad (12.41)$$

where now $\underline{Z} = (Z_1, Z_2, \dots, Z_{f-1})$ are the $f - 1$ variables held constant during differentiation.

As before, we assume that the nonsingular metric matrix $\underline{\underline{M}}$ of order f is known in terms of a chosen basis set $\{R_i\}$ with conjugates $\{\bar{R}_i\}$, such that

$$(\underline{\underline{M}})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle \quad (12.42a)$$

$$(\underline{\underline{M}})_{ij} = \langle \bar{\mathbf{R}}_i | \bar{\mathbf{R}}_j \rangle, \quad i, j = 1, 2, \dots, f \quad (12.42b)$$

We furthermore assume that the variables $X, Y, \{Z_i\}$ of the chosen $\mathcal{D}$ are known (i.e., given by expansions) in terms of the basis variables $\{R_i\}$ or $\{\bar{R}_i\}$; for example, X is assumed to be given in either of the vector forms

$$|\mathbf{X}\rangle = \sum_{i=1}^f c_i |\mathbf{R}_i\rangle, \quad \text{with } c_i = \langle \bar{\mathbf{R}}_i | \mathbf{X} \rangle \quad (12.43)$$

$$|\mathbf{X}\rangle = \sum_{i=1}^f c'_i |\bar{\mathbf{R}}_i\rangle, \quad \text{with } c'_i = \langle \mathbf{R}_i | \mathbf{X} \rangle \quad (12.44)$$

with similar expansions for $|\mathbf{Y}\rangle, |\mathbf{Z}\rangle$.

To sketch the general strategy for evaluating derivatives in multidimensional geometry, we first note that the derivative $\mathcal{D}$ in (12.41) would become rather simple if we had made a shrewd choice of basis variables. Specifically, if new basis variables $\{R'_i\}$ (with conjugates $\{\bar{R}'_i\}$) were chosen such that

$$Y = \bar{R}'_f \quad (12.45a)$$

$$Z_i = \bar{R}'_i, \quad i = 1, 2, \dots, f - 1 \quad (12.45b)$$

then the new conjugate of Y in this basis is given by

$$\bar{Y}_{\underline{Z}} = R'_f \quad (12.46)$$

denoted $\bar{Y}_{\underline{Z}}$ to indicate that the $\{Z_i\}$ variables are “complementary” to Y in the new basis. With this choice of variables, the derivative (12.41) becomes simply

$$\mathcal{D} = \left(\frac{\partial X}{\partial Y} \right)_{\underline{Z}} = \langle \mathbf{X} | \mathbf{R}'_f \rangle = \langle \mathbf{X} | \bar{Y}_{\underline{Z}} \rangle \quad (12.47)$$

which can be readily evaluated from the expansion analogous to (12.43) in the new $\{R'_i\}$ basis. Thus, the only difficulty is to find the transformation to the special basis (12.45a, b) that enables easy evaluation of $\mathcal{D}$.

To achieve the desired simple form (12.47), it is only necessary to transform from the old conjugate basis $\{\bar{R}_i\}$ to the special conjugate basis $\{\bar{R}'_i\}$ with the appropriate transformation matrix $\underline{\underline{\bar{A}}}$ that satisfies

$$\begin{pmatrix} \bar{R}'_1 \\ \bar{R}'_2 \\ \vdots \\ \bar{R}'_{f-1} \\ \bar{R}'_f \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{f-1} \\ Y \end{pmatrix} = \underline{\underline{\bar{A}}} \begin{pmatrix} \bar{R}_1 \\ \bar{R}_2 \\ \vdots \\ \bar{R}_{f-1} \\ \bar{R}_f \end{pmatrix} \quad (12.48)$$

The matrix $\underline{\underline{\bar{A}}}$ is known from expansions analogous to (12.44) for $|\mathbf{Y}\rangle$, $|\mathbf{Z}\rangle$, and is given explicitly by

$$(\underline{\underline{\bar{A}}})_{ij} = \langle \bar{\mathbf{R}}_j | \mathbf{Z}_i \rangle \quad \text{for } i = 1, 2, \dots, f-1 \quad (12.49a)$$

$$(\underline{\underline{\bar{A}}})_{ij} = \langle \bar{\mathbf{R}}_j | \mathbf{Y} \rangle \quad (12.49b)$$

In the transformed basis $\{\bar{R}'_i\}$, the new metric matrix $\underline{\underline{\mathbf{M}}}'$ is obtained in the usual manner as

$$\underline{\underline{\mathbf{M}}}' = \underline{\underline{\bar{A}}} \underline{\underline{\mathbf{M}}} \underline{\underline{\bar{A}}}^t = \underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}} \quad (12.50)$$

The vector of transformed variables $\{\bar{R}'_i\}$ is thereby obtained as

$$\underline{\mathbf{R}}' = \underline{\underline{\mathbf{M}}}' \underline{\underline{\bar{R}}}' = \underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}} \underline{\underline{\bar{R}}}' \quad (12.51)$$

Only the final entry of the vector $\underline{\mathbf{R}}'$ in (12.51) is needed for the desired $Y_{\underline{\mathbf{Z}}}$ coordinate vector:

$$|\underline{\mathbf{Y}}_{\underline{\mathbf{Z}}}\rangle = \sum_{i=1}^{f-1} (\underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}})_{if} |\mathbf{Z}_i\rangle + (\underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}})_{ff} |\mathbf{Y}\rangle \quad (12.52)$$

With this substitution, (12.47) becomes finally

$$\boxed{\mathcal{D} = \left(\frac{\partial X}{\partial Y} \right)_{\underline{\mathbf{Z}}} = (\underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}})_{ff} \langle \mathbf{X} | \mathbf{Y} \rangle + \sum_{i=1}^{f-1} (\underline{\underline{\mathbf{M}}}_{\bar{\mathbf{A}}})_{if} \langle \mathbf{X} | \mathbf{Z}_i \rangle} \quad (12.53)$$

Equation (12.53) gives the desired evaluation of the general thermodynamic derivative $\mathcal{D}$ in a system of f degrees of freedom, expressed in terms of known geometrical quantities. As in the two-dimensional case, other expressions for $\mathcal{D}$ would be possible in other special choices of basis. Equation (12.53) is suitable for machine computation in multicomponent thermodynamic systems of arbitrary complexity.

12.4 PHASE-BOUNDARY DERIVATIVES IN MULTICOMPONENT SYSTEMS

Among the most important thermodynamic derivatives are those associated with phase boundaries. In the PT phase diagram for a heterogeneous system of c components and p phases, each phase boundary σ has characteristic slope $(\partial T/\partial P)_\sigma$ determined by the coexistence conditions for the adjacent homogeneous single-phase regions. Aspects of phase coexistence and saturation properties along two-phase boundaries were previously considered for single-component systems (see, e.g., Sections 7.2, 11.5, and 11.11), but we now wish to employ powerful vector-algebraic tools to address the problem in full generality.

The phase-boundary conditions are closely associated with the singularity conditions for the full $(c+2)$ -dimensional metric $\underline{\underline{M}}^{(c+2)}$ of a c -component system. In each case, a metric singularity is associated with a null eigenvector of $\underline{\underline{M}}^{(c+2)}$,

$$\underline{\underline{M}}^{(c+2)} \underline{\eta} = 0 \quad (12.54a)$$

and an associated linear dependence condition among the underlying $|\mathbf{R}_i\rangle$ vectors,

$$\sum_{i=1}^{c+2} \eta_i |\mathbf{R}_i\rangle = 0 \quad (12.54b)$$

Each such null vector may be considered an *invariant* or *symmetry* of the thermodynamic system, because it corresponds to an operation (change of extensive variables X_i) that produces *no* response in any intensive state variable and thus leaves the thermodynamic *state* unaltered (Sidebar 7.2). As described in Sidebar 10.3, these invariants also play a role somewhat analogous to overall rotations and translations (“null eigenmodes” of the Hessian matrix) in the theory of molecular vibrations.

Of course, any linear combination of null eigenvectors also satisfies (12.54a, b), so we are able to choose individual vectors of this “null manifold” with considerable freedom. We know in general that a p -dimensional manifold of linearly independent null eigenvectors must exist, in order that the rank of the metric matrix,

$$\text{rank}(\underline{\underline{M}}^{(c+2)}) = f = c + 2 - p \quad (12.55)$$

be consistent with the Gibbs phase rule. One such singularity of the metric matrix was identified in Section 10.3 with a *Gibbs–Duhem* null vector for single-phase homogeneity, and another in Section 11.11 with a *Clapeyron* null vector for two-phase coexistence. We now wish to consider the more general case that includes these two limits as special cases.

In accordance with (12.55), it is always possible to choose a nonsingular principal submatrix $\underline{\underline{M}} = \underline{\underline{M}}^{(f)}$ of order f from $\underline{\underline{M}}^{(c+2)}$:

$$(\underline{\underline{M}})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle, \quad i, j = 1, 2, \dots, f \quad (12.56a)$$

$$M = \det |\underline{\underline{M}}| \neq 0 \quad (12.56b)$$

The p extensivities $\hat{X}_\kappa$ ($\kappa = 1, 2, \dots, p$) thereby deleted from consideration may be regarded as *scale factors* having fixed values in all thermodynamic derivatives, and the corresponding

intensities $\hat{R}_\kappa = (\partial U / \partial \hat{X}_\kappa)_{\underline{X}}$ are considered “redundant” or “excess” variables, mere linear combinations of the f independent vectors $\{\mathbf{R}_i\}$ that form a nonsingular *axis* system for $\mathcal{M}_S$. The latter vectors permit the construction of conjugate basis vectors $|\bar{\mathbf{R}}_i\rangle$ that satisfy the biorthogonality property

$$\langle \bar{\mathbf{R}}_i | \mathbf{R}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (12.57)$$

and give rise to the conjugate metric matrix $\underline{\underline{\mathbf{M}}}$,

$$(\underline{\underline{\mathbf{M}}})_{ij} = (\underline{\underline{\mathbf{M}}}^{-1})_{ij} = \langle \bar{\mathbf{R}}_i | \bar{\mathbf{R}}_j \rangle = \left(\frac{\partial X_i}{\partial R_j} \right)_R \quad (12.58)$$

of a valid (nonsingular) axis system in $\mathcal{M}_S$. The subscript R in (12.58) should be understood to denote constancy of the $f-1$ R_k ($k \neq j$) and the p scale factors $X_{c-p+1}, \dots, X_{c+2}$ during the partial differentiation.

To formalize these distinctions between excess and axis variables and vectors

$$\text{“excess” } \{\hat{R}_\kappa\} \quad (\kappa = 1, 2, \dots, p) \quad (12.59a)$$

$$\text{“axis” } \{\mathbf{R}_i\} \quad (i = 1, 2, \dots, f) \quad (12.59b)$$

we shall henceforth distinguish the former by Greek indices and caret symbol (namely, $\hat{R}_\kappa$; $|\hat{\mathbf{R}}_\kappa\rangle$) and the latter, as before, with Latin indices (namely, R_i , $|\mathbf{R}_i\rangle$ or conjugate $\bar{R}_i$, $|\bar{\mathbf{R}}_i\rangle$). The location of each excess $|\hat{\mathbf{R}}_\kappa\rangle$ in the full list of $c+2$ $|\mathbf{R}_i\rangle$ can be specified by the convention

$$|\hat{\mathbf{R}}_\kappa\rangle = |\mathbf{R}_{f+\kappa}\rangle, \quad \kappa = 1, 2, \dots, p \quad (12.60)$$

Let us now introduce the Gibbs–Duhem (GD) vector of extensive values for each phase $\lambda = 1, 2, \dots, p$:

$$\underline{\underline{\xi}}^{(\lambda)} = \begin{pmatrix} S^{(\lambda)} \\ V^{(\lambda)} \\ n_1^{(\lambda)} \\ \vdots \\ n_c^{(\lambda)} \end{pmatrix} = \begin{pmatrix} \xi_1^{(\lambda)} \\ \xi_2^{(\lambda)} \\ \xi_3^{(\lambda)} \\ \vdots \\ \xi_{c+2}^{(\lambda)} \end{pmatrix} \quad (12.61)$$

Each such GD vector is a null vector satisfying the linear dependence condition [cf. 12.54b]

$$\sum_{i=1}^{c+2} \xi_i^{(\lambda)} |\mathbf{R}_i\rangle = 0 \quad (12.62)$$

To parallel (12.60), we can identify the excess elements $\hat{\xi}_\kappa^{(\lambda)}$ in each GD vector as

$$\hat{\xi}_\kappa^{(\lambda)} = \xi_{f+\kappa}^{(\lambda)}, \quad \kappa = 1, 2, \dots, p \quad (12.63)$$

The p “excess extensivities” $\hat{\xi}_\kappa^{(\lambda)}$ are scale factors of the composite system, which may be exhibited in a $p \times p$ “scale-factor matrix” $\hat{\underline{\underline{D}}}$ defined by

$$(\hat{\underline{\underline{D}}})_{\lambda\kappa} \equiv \hat{\xi}_\kappa^{(\lambda)}, \quad \kappa, \lambda = 1, 2, \dots, p \quad (12.64)$$

Similarly, the f “axis extensivities” $\xi_i^{(\lambda)}$ can be arranged in a $p \times f$ matrix $\underline{\underline{D}}$ defined by

$$(\underline{\underline{D}})_{\lambda i} \equiv \xi_i^{(\lambda)}, \quad \lambda = 1, 2, \dots, p; \quad i = 1, 2, \dots, f \quad (12.65)$$

With these definitions, (12.62) becomes

$$\sum_{i=1}^f (\underline{\underline{D}})_{\lambda i} |\mathbf{R}_i\rangle + \sum_{\kappa=1}^p (\hat{\underline{\underline{D}}})_{\lambda\kappa} |\hat{\mathbf{R}}_\kappa\rangle = 0 \quad (12.66)$$

for each phase $\lambda = 1, 2, \dots, p$.

As shown in Sidebar 12.8, the scale-factor matrix $\hat{\underline{\underline{D}}}$ is necessarily nonsingular, with

$$\det |\hat{\underline{\underline{D}}}| \neq 0 \quad (12.67)$$

Accordingly, we can solve the p homogeneous linear vector equations (12.66) for the p excess intensities $|\hat{\mathbf{R}}_\kappa\rangle$ in terms of the chosen independent set of f -axis intensities $|\mathbf{R}_i\rangle$. When each of the vector equations (12.66) is multiplied by $(\hat{\underline{\underline{D}}}^{-1})_{\kappa',\lambda}$ and summed over λ , the resulting solution is

$$|\hat{\mathbf{R}}_{\kappa'}\rangle = - \sum_{i=1}^f (\underline{\underline{\gamma}})_{\kappa' i} |\mathbf{R}_i\rangle, \quad \kappa' = 1, 2, \dots, p \quad (12.68)$$

where the *coexistence coefficients* $(\underline{\underline{\gamma}})_{\kappa, i}$ are given by

$$\underline{\underline{\gamma}} = \hat{\underline{\underline{D}}}^{-1} \underline{\underline{D}} \quad (12.69)$$

The explicit solution (12.69) can be brought to a useful alternative form after introducing the expression for the inverse scale-factor matrix,

$$(\hat{\underline{\underline{D}}}^{-1})_{\lambda\kappa} = \frac{\hat{D}_{\kappa\lambda}^{\text{cof}}}{\det |\hat{\underline{\underline{D}}}|} \quad (12.70)$$

where $\hat{D}_{\kappa\lambda}^{\text{cof}}$ denotes the cofactor (signed minor) of the $\kappa\lambda$ element of $\hat{\underline{\underline{D}}}$. With this substitution and the definition (12.65), we obtain

$$(\underline{\underline{\gamma}})_{\lambda i} = \det |\hat{\underline{\underline{D}}}|^{-1} \sum_{\kappa=1}^p \xi_i^{(\kappa)} \hat{D}_{\lambda\kappa}^{\text{cof}} \quad (12.71)$$

The expansion on the right-hand side of (12.71) can then be recognized as the expansion (down the λ th column) of a determinant that is identical to $\det |\hat{\underline{\underline{D}}}|$ except that each $(\hat{\underline{\underline{D}}})_{\kappa\lambda}$ is replaced by $\xi_i^{(\kappa)}$ in column λ .

We can better exhibit the form of (12.71) by defining p -component column vectors $\underline{d}_i$ and $\hat{\underline{d}}_\kappa$ that are drawn from the matrices $\underline{\underline{D}}$ and $\hat{\underline{\underline{D}}}$, respectively:

$$\underline{d}_i = \begin{pmatrix} \xi_i^{(1)} \\ \xi_i^{(2)} \\ \vdots \\ \xi_i^{(p)} \end{pmatrix}, \quad \hat{\underline{d}}_\kappa = \begin{pmatrix} \hat{\xi}_\kappa^{(1)} \\ \hat{\xi}_\kappa^{(2)} \\ \vdots \\ \hat{\xi}_\kappa^{(p)} \end{pmatrix} \quad (12.72)$$

As shown in (12.72), the successive elements of $\hat{\underline{d}}_i$ and $\hat{\underline{d}}_\kappa$ are labeled by the (superscript) phase numbers. The scale-factor determinant, $\det|\hat{\underline{\underline{D}}}|$, can be written in an obvious notation as

$$\det|\hat{\underline{\underline{D}}}| = \det|\hat{\underline{d}}_1 \hat{\underline{d}}_2 \cdots \hat{\underline{d}}_p| \quad (12.73)$$

Similarly, the summation on the right-hand side of (12.71) can be written as

$$\sum_{\kappa=1}^p \xi_i^{(\kappa)} \hat{D}_{\lambda\kappa}^{\text{cof}} = \det|\hat{\underline{d}}_1 \cdots \hat{\underline{d}}_{\lambda-1} \underline{d}_i \hat{\underline{d}}_{\lambda+1} \cdots \hat{\underline{d}}_p| \quad (12.74)$$

Such expressions allow us to represent (12.71) as

$$(\underline{\underline{\gamma}})_{\lambda i} = \frac{\det|\hat{\underline{d}}_1 \cdots \hat{\underline{d}}_{\lambda-1} \underline{d}_i \hat{\underline{d}}_{\lambda+1} \cdots \hat{\underline{d}}_p|}{\det|\hat{\underline{d}}_1 \hat{\underline{d}}_2 \cdots \hat{\underline{d}}_p|} \quad (12.75)$$

which is the desired solution. According to this expression, the coexistence coefficient $(\underline{\underline{\gamma}})_{\lambda i}$ is found by replacing the λ th column of the scale-factor matrix $\hat{\underline{\underline{D}}}$ by the i th column of $\underline{\underline{D}}$ and evaluating a ratio of substituted and unsubstituted determinants.

The coexistence conditions (12.68) that relate each excess intensive vector $|\hat{\mathbf{R}}_\kappa\rangle$ to the chosen axis intensities $|\mathbf{R}_i\rangle$ can also be written in terms of the conjugate extensive vectors $|\bar{\mathbf{R}}_i\rangle = |\mathbf{X}_i\rangle$. With the usual metric relationship between intensive and extensive vectors,

$$|\mathbf{R}_i\rangle = \sum_{j=1}^f (\underline{\underline{M}}_{ij}) |\mathbf{X}_j\rangle \quad (12.76)$$

one obtains from (12.68)

$$|\hat{\mathbf{R}}_\kappa\rangle = - \sum_{i=1}^f (\underline{\underline{\gamma}} \underline{\underline{M}})_{\kappa i} |\mathbf{X}_i\rangle \quad (12.77)$$

The biorthogonality property (12.4) then allows one to easily evaluate scalar products of each excess intensity $|\hat{\mathbf{R}}_\kappa\rangle$ with basis axis vectors $|\mathbf{R}_i\rangle$ or $|\bar{\mathbf{R}}_i\rangle$. The scalar product of extensive $|\mathbf{X}_i\rangle$ with (12.76) gives

$$\langle \mathbf{X}_i | \hat{\mathbf{R}}_\kappa \rangle = \left(\frac{\partial R_{f+\kappa}}{\partial R_i} \right)_R = -(\underline{\underline{\gamma}})_{\kappa i} \quad (12.78)$$

and the corresponding scalar product of intensive $|\mathbf{R}_i\rangle$ with (12.76) gives

$$\langle \mathbf{R}_i | \hat{\mathbf{R}}_\kappa \rangle = \left(\frac{\partial R_{f+\kappa}}{\partial X_i} \right)_X = -(\underline{\underline{\gamma}} \underline{\underline{M}})_{\kappa i} \quad (12.79)$$

for all $\kappa = 1, 2, \dots, p$ and $i = 1, 2, \dots, f$. From the scalar product of (12.68) with (12.77), we also obtain

$$\langle \hat{\mathbf{R}}_\kappa | \hat{\mathbf{R}}_\lambda \rangle = \left(\frac{\partial R_{f+\kappa}}{\partial X_{f+\lambda}} \right)_X = \left(\frac{\partial R_{f+\lambda}}{\partial X_{f+\kappa}} \right)_X = (\underline{\underline{\gamma}} \underline{\underline{M}} \underline{\underline{\gamma}}^t)_{\kappa \lambda} \quad (12.80)$$

for all $\kappa, \lambda = 1, 2, \dots, p$.

With the explicit formula (12.75) for the $\underline{\underline{\gamma}}$ coefficients, (12.78)–(12.80) become convenient formulas for the slopes of coexistence boundaries in various phase-diagram representations (including those with an extensive axis). Notice in particular that the derivatives (12.78) involving only intensive variables (as plotted in conventional phase diagrams) can be evaluated solely in terms of the $\underline{\underline{\gamma}}$ coefficients (i.e., in terms of extensive properties $\xi_i^{(\lambda)}$ of the separate phases), without reference to response functions of the metric matrix $\underline{\underline{M}}$. Sidebar 12.9 describes special cases of these equations that were previously known.

SIDEBAR 12.8: PROOF THAT THE SCALE-FACTOR MATRIX IS NONSINGULAR

The nonsingularity of the scale-factor matrix $\hat{\underline{\underline{D}}}$ is a necessary consequence of the linear independence of the chosen axis vectors $|\mathbf{R}_i\rangle$. To see this, suppose instead that $\hat{\underline{\underline{D}}}$ has a null eigenvector $\underline{\underline{\eta}}$, satisfying

$$\hat{\underline{\underline{D}}} \underline{\underline{\eta}} = 0 \quad (\text{S12.8-1})$$

or, equivalently,

$$\sum_{\lambda=1}^p (\hat{\underline{\underline{D}}})_{\kappa \lambda} \eta_\lambda = 0, \quad \kappa = 1, 2, \dots, p \quad (\text{S12.8-2})$$

If we multiply (12.66) by η_λ and sum over λ , using (S12.8-2), the result is

$$\sum_{i=1}^f (\underline{\underline{\eta}} \underline{\underline{D}})_i |\mathbf{R}_i\rangle = 0 \quad (\text{S12.8-3})$$

which says that the $|\mathbf{R}_i\rangle$ are linearly dependent, contrary to (12.56b). Thus, if condition (12.67) seems to fail, it is only necessary to renumber the vectors (i.e., make some new division into “axis” and “excess” vectors) to make (12.67)–(12.69) valid.

SIDEBAR 12.9: SPECIAL CASES OF (12.75), (12.78)

In simple cases, the coexistence coefficient matrix (12.75) and phase boundary equation (12.78) reduce to forms that were previously recognized.

Homogeneous System ($p = 1; f = c + 1$) In this case, (12.75) reduces rather trivially to

$$\underline{(\gamma)}_{1i} = \frac{\xi_i}{\xi_{f+1}} \quad (\text{S12.9-1})$$

In this case, (12.68) merely recovers the vector form of the Gibbs–Duhem equation for the single phase (cf. Section 10.3), as we should anticipate.

Two-Phase State of a Pure Substance ($p = 2; c = f = 1$) In this case, we may choose the usual $c + 2$ base variables as

$$R_1 = T, \quad R_2 = -P, \quad R_3 = \mu \quad (\text{S12.9-2a})$$

$$X_1 = S, \quad X_2 = V, \quad X_3 = N \quad (\text{S12.9-2b})$$

Equation (12.75) then becomes

$$\underline{(\gamma)}_{11} = \frac{\begin{vmatrix} S^{(1)} & N^{(1)} \\ S^{(2)} & N^{(2)} \end{vmatrix}}{\begin{vmatrix} V^{(1)} & N^{(1)} \\ V^{(2)} & N^{(2)} \end{vmatrix}} \quad (\text{S12.9-3a})$$

$$\underline{(\gamma)}_{21} = \frac{\begin{vmatrix} V^{(1)} & S^{(1)} \\ V^{(2)} & S^{(2)} \end{vmatrix}}{\begin{vmatrix} V^{(1)} & N^{(1)} \\ V^{(2)} & N^{(2)} \end{vmatrix}} \quad (\text{S12.9-3b})$$

and (12.78) are therefore simply

$$\frac{dp}{dT} = \frac{S^{(1)}N^{(2)} - N^{(1)}S^{(2)}}{V^{(1)}N^{(2)} - N^{(1)}V^{(2)}} \quad (\text{S12.9-4a})$$

$$\frac{d\mu}{dT} = \frac{S^{(1)}V^{(2)} - V^{(1)}S^{(2)}}{V^{(1)}N^{(2)} - N^{(1)}V^{(2)}} \quad (\text{S12.9-4b})$$

The first of these can be recognized as the ordinary *Clapeyron equation* for a pure two-phase system (usually written for equimolar phases $N^{(1)} = N^{(2)} = 1$; cf. Sections 7.2.2 and 11.11), and the second is an analogous equation determining the slope of the coexistence curve in the μ – T plane. These equations in turn determine the slope of the coexistence line in the μ – P plane:

$$\frac{d\mu}{dP} = \frac{S^{(1)}V^{(2)} - V^{(1)}S^{(2)}}{S^{(1)}N^{(2)} - N^{(1)}S^{(2)}} \quad (\text{S12.9-4c})$$

Univariant Systems ($f = 1$; $p = c + 1$) For this case, Gibbs discovered the explicit determinantal solution [J. W. Gibbs. *Collected Works* (Longmans Green, New York, 1928), Vol. I, pp. 98ff; cf. also J. G. Kirkwood and I. Oppenheim. *Chemical Thermodynamics* (McGraw-Hill, New York, 1961), p. 118]:

$$\frac{dP}{dT} = \frac{\begin{vmatrix} S^{(1)} & n_1^{(1)} & \cdots & n_c^{(1)} \\ S^{(2)} & n_1^{(2)} & \cdots & n_c^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ S^{(p)} & n_1^{(p)} & \cdots & n_c^{(p)} \end{vmatrix}}{\begin{vmatrix} V^{(1)} & n_1^{(1)} & \cdots & n_c^{(1)} \\ V^{(2)} & n_1^{(2)} & \cdots & n_c^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ V^{(p)} & n_1^{(p)} & \cdots & n_c^{(p)} \end{vmatrix}} \quad (\text{S12.9-5})$$

which is precisely the form given by (12.75), (12.78). The Clapeyron equation (S12.9-4a) is itself the simplest instance of this result for $c = 1$.

12.5 STATIONARY POINTS OF PHASE DIAGRAMS: GIBBS-KONOWALOW LAWS

Equations (12.68) establish how any $f + 1$ intensive vectors must be related to one another in the heterogeneous equilibrium of p phases. Let us now assume for definiteness that these are labeled

$$|\mathbf{R}_1\rangle, |\mathbf{R}_2\rangle, \dots, |\mathbf{R}_{f+1}\rangle, \quad (12.81)$$

and write for η_i the $p \times p$ determinant

$$\eta_i = \begin{vmatrix} \xi_i^{(1)} & \xi_{f+2}^{(1)} & \xi_{f+3}^{(1)} & \cdots & \xi_{c+2}^{(1)} \\ \xi_i^{(2)} & \xi_{f+2}^{(2)} & \xi_{f+3}^{(2)} & \cdots & \xi_{c+2}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \xi_i^{(p)} & \xi_{f+2}^{(p)} & \xi_{f+3}^{(p)} & \cdots & \xi_{c+2}^{(p)} \end{vmatrix} \quad (12.82)$$

This permits us to write each of (12.68) in the more symmetrical form

$$0 = \sum_{i=1}^{f+1} \eta_i |\mathbf{R}_i\rangle \quad (12.83)$$

as can be easily verified from the general form (12.75) of the coexistence coefficients. Equation (12.83) is obviously related closely to the ordinary Gibbs–Duhem equation, to which it formally reduces when $p = 1$. By virtue of their additional invariants, multiphase

systems can exhibit an interesting variety of behavior that is not possible in homogeneous systems. In particular, the determinantal coefficients η_i can take on either sign, or can *vanish* identically when the phase compositions have certain special values.

According to (12.82), η_i must vanish whenever the determinantal row or column vectors become linearly dependent, i.e., when there exists some linear combination Y of the variables $X_{f+2}, X_{f+3}, \dots, X_{c+2}$ such that $Y^{(\lambda)} = \xi_i^{(\lambda)}$ for every phase λ . For example, when $p = 2$, η_i must vanish when values of ξ_i and ξ_{c+2} are proportional in the two phases. The vanishing of η_i signals a type of “redundancy” of the extensive variable X_i , as though the system could be prepared from one fewer chemical component than had been supposed. However, the vanishing of an η_i could arise from other special linear relationships connecting entropies or volumes, as well as composition variables.

To see the consequences of some vanishing η_i , let us introduce a new basis set of f vectors $|\tilde{\mathbf{R}}_i\rangle$ that is identical to the starting set, except for the final member $|\mathbf{Z}\rangle$:

$$|\tilde{\mathbf{R}}_i\rangle = |\mathbf{R}_i\rangle, \quad i = 1, 2, \dots, f-1 \quad (12.84a)$$

$$|\tilde{\mathbf{R}}_f\rangle = |\mathbf{Z}\rangle \quad (12.84b)$$

where $|\mathbf{Z}\rangle$ is arbitrary so long as the resulting set is complete. By “complete” we mean of course that the $|\tilde{\mathbf{R}}_i\rangle$ should span all f dimensions, as may be confirmed from the rank of the associated Gram (metric $\underline{\tilde{\mathbf{M}}}$) matrix:

$$\text{rank}\{\langle \tilde{\mathbf{R}}_i | \tilde{\mathbf{R}}_j \rangle\} = f \quad (12.85)$$

This insures that the $f-1$ vectors $|\mathbf{R}_1\rangle, |\mathbf{R}_2\rangle, \dots, |\mathbf{R}_{f-1}\rangle$ must themselves span $f-1$ dimensions. In view of (12.85), we may now introduce the associated conjugate vectors $|\tilde{\mathbf{X}}_i\rangle$ satisfying

$$\langle \tilde{\mathbf{R}}_i | \tilde{\mathbf{X}}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (12.86)$$

Multiplying (12.83) on the left by a chosen $|\tilde{\mathbf{X}}_j\rangle$, we obtain

$$0 = \sum_{i=1}^{f-1} \eta_i \delta_{ij} + \eta_f \langle \tilde{\mathbf{X}}_j | \mathbf{R}_f \rangle + \eta_{f+1} \langle \tilde{\mathbf{X}}_j | \mathbf{R}_{f+1} \rangle \quad (12.87)$$

In particular, for $j = f$, this becomes

$$0 = \eta_f \langle \bar{\mathbf{Z}} | \mathbf{R}_f \rangle + \eta_{f+1} \langle \bar{\mathbf{Z}} | \mathbf{R}_{f+1} \rangle \quad (12.88)$$

where $|\bar{\mathbf{Z}}\rangle$ denotes the conjugate of $|\mathbf{Z}\rangle$. Equation (12.88) may also be written in the form

$$0 = \eta_f \left(\frac{\partial R_f}{\partial \underline{Z}} \right) + \eta_{f+1} \left(\frac{\partial R_{f+1}}{\partial \underline{Z}} \right) \quad (12.89)$$

where the underlined subscripts denote constancy of $R_1, \dots, R_{f-1}$ in the usual manner. Note that “ R_f ” and “ R_{f+1} ” could be *any* two base intensities, because the numbering in (12.88) was chosen only for convenience in introducing Z as a new differentiating variable in (12.84b).

With the help of (12.89), it is now possible to establish some simple theorems concerning the possibility of *stationary points* (e.g., maxima, minima, or horizontal inflections) in thermodynamic phase diagrams. In each case, we suppose that R_i, R_j are chosen from any set of $f+1$ intensive variables (spanning at least $f-1$ dimensions), and that η_i, η_j are defined as in (12.82) for a system of p coexisting phases.

Theorem 1 If $\eta_i = 0$, then any R_j for which $\eta_j \neq 0$ is necessarily *stationary* with respect to any thermodynamic change in which the remaining $f-1$ intensities (i.e., all but R_i, R_j) are held constant.

Proof If we choose $\eta_f = \eta_i = 0$ and $\eta_{f+1} = \eta_j \neq 0$, then the only allowed solution of (12.89) is

$$\left(\frac{\partial R_j}{\partial Z}\right)_R = 0 \quad \text{if } \eta_j \neq 0 \quad (12.90)$$

for any chosen Z that leaves the remaining $f-1$ base variables (all but R_i, R_j) fixed—*QED*.

Theorem 2 If R_j is stationary with respect to any change that alters R_i but holds constant the remaining $f-1$ intensities (i.e., all but R_i and R_j), then $\eta_i = 0$.

Proof Again taking $i = f, j = f+1$ in (12.89), we see that the stationary condition $(\partial R_i / \partial Z)_R$ requires that

$$\eta_j = 0 \quad \left(\text{if } \left(\frac{\partial R_j}{\partial Z}\right)_R \neq 0 \right) \quad (12.91)$$

for any allowed Z —*QED*.

Note that the thermodynamic variable Z could be chosen quite arbitrarily (for example, as a composition variable for one of the phases) so long as it is independent of the chosen $R_1, \dots, R_{f-1}$ (i.e., all but the variables R_i, R_j under discussion). Various special cases of these theorems were first deduced theoretically by Gibbs [J. W. Gibbs. *Collected Works* (Longmans Green, New York, 1928), Vol. I, pp. 99–100], then rediscovered empirically by Konowalow [D. Konowalow. *Ann. Phys. Chem.* **14**, 34 (1881)], and are often now referred to as Gibbs–Konowalow laws [see, e.g., H. F. Franzen and B. C. Gerstein. *Am. Inst. Chem. Eng. J.* **12**, 364 (1966); I. Prigogine and R. Defay. *Chemical Thermodynamics* (Longmans Green, New York, 1954), pp. 278ff; S. Glasstone. *Textbook of Physical Chemistry*, 2nd edn (Van Nostrand, New York, 1946), pp. 714–15; J. A. V. Butler, in F. G. Donnan and A. Haas (eds). *A Commentary on the Scientific Writings of J. Willard Gibbs* (Yale University Press, New Haven, 1936), Vol. I, pp. 111ff]. As expressed in Theorem 2, Gibbs–Konowalow laws relate the maxima or minima in boiling-point diagrams to special relationships in the compositions of liquid and vapor phases (cf. Section 7.3.3), which enter through the determinantal coefficients η_i . As such, these laws provide the framework for analyzing distillation processes, azeotropy, and other properties of liquid mixtures. The general vector-algebraic equations (12.82), (12.83) make it feasible to pursue such relationships in liquid mixtures of arbitrary complexity.

12.6 HIGHER-ORDER DERIVATIVES AND STATE CHANGES

A striking feature of Gibbsian thermodynamics is its sole dependence on first and second derivatives of U or S (cf. Inductive Law IL-2, Table 2.1). Indeed, it can be shown that higher U''' -type derivatives become ill-defined (multivalued) or do not *exist* in certain thermodynamic states. [For example, the assumption of a low-order Taylor series expansion of the Gibbs or Helmholtz free energy in the neighborhood of the critical point (as in the Landau phenomenological theory of phase transitions) leads to disagreements with observed critical properties; see M. E. Fisher. *Rep. Prog. Phys.* **30**, 615–730 (1967), pp. 659ff.] Attempts to “extend” thermodynamics by assumptions involving such higher-order derivative properties are therefore fraught with danger, and must be justified (if at all) on some nonthermodynamic basis.

Nevertheless, on physical grounds, we expect (and can verify experimentally) that “safe” low-order derivative response functions typically exhibit smooth variations with respect to changes of state *if* maintained safely away from criticality. With care, we can even approach the critical limit arbitrarily closely (cf. Section 11.10). Thus, within the range of contiguous states for which the metric $\underline{\underline{M}}$ remains nonsingular ($\det |\underline{\underline{M}}| \neq 0$), and for which such higher-order derivative properties have a well-established experimental basis, it is sensible to incorporate such properties into the thermodynamic description and to investigate their metric geometrical characteristics, as we wish to do in this section.

We first observe that higher-order derivatives of U or S are implicitly related to *changes of state* from the initial state to a “nearby” equilibrium state along a reversible path. Suppose that we parameterize the thermodynamic state $\mathcal{M}_S$

$$\mathcal{M}_S = \mathcal{M}_S(\underline{\xi}) \quad (12.92)$$

by some chosen manifold $\underline{\xi}$ of state properties (denoted as a generic column vector),

$$(\underline{\xi})_k = \xi_k, \quad i = 1, 2, \dots, f \quad (12.93)$$

whose values uniquely identify $\mathcal{M}_S$. [For example, we might choose the ξ_i to be the f numerical intensive values $R_i(T, P, \mu_1, \dots, \mu_{c-p})$ or extensive values $X_i(S, V, n_1, \dots, n_{c-p})$ that uniquely identify the state of the system in an experimental context.] Then, we can write (rather symbolically) a formal “Taylor series expansion” (Section 1.4) for $\mathcal{M}_S$,

$$\mathcal{M}_S(\xi_k + d\xi_k) = \mathcal{M}_S(\xi_k) + \Delta_k^{(1)} d\xi_k + \frac{1}{2} \Delta_k^{(2)} (d\xi_k)^2 + \dots \quad (12.94a)$$

with coefficients of displacement

$$\Delta_k^{(1)} = \left(\frac{\partial \mathcal{M}_S}{\partial \xi_k} \right)_{\underline{\xi}}, \quad \Delta_k^{(2)} = \left(\frac{\partial^2 \mathcal{M}_S}{\partial \xi_k^2} \right)_{\underline{\xi}}, \dots \quad (12.94b)$$

that allows us to define “ $d\mathcal{M}_S$ ” (the small change of state associated with $d\xi_k$) for any small change of state variable. Because the individual $d\xi_k$ can be assigned metric significance through associated ket vectors $|\xi_k\rangle$, formal expansions such as (12.94) allow us to similarly assign metric significance to changes of state “ $d\mathcal{M}_S$ ” described by the leading term in (12.94).

To be specific, let us first consider the metric matrix elements

$$M_{ij} = (\underline{\underline{\mathbf{M}}})_{ij} = \langle \mathbf{R}_i | \mathbf{R}_j \rangle = \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{\underline{\mathbf{X}}}} \quad (12.95)$$

that are fundamental descriptors of $\mathcal{M}_S$. The Taylor series for M_{ij} under a small displacement of variable X_k can be written as

$$M_{ij}(X_k + dX_k) = M_{ij}(X_k) + m_{ij;k} dX_k + \cdots \quad (12.96)$$

where

$$m_{ij;k} \equiv \left(\frac{\partial M_{ij}}{\partial X_k} \right)_{\underline{\underline{\mathbf{X}}}} \equiv (\underline{\underline{\mathbf{m}}}_{ij})_k \quad (12.97)$$

Similarly, for the entire manifold of displacements $\underline{\underline{\mathbf{X}}} + \underline{\underline{\mathbf{dX}}}$, we obtain,

$$M_{ij}(\underline{\underline{\mathbf{X}}} + \underline{\underline{\mathbf{dX}}}) = M_{ij}(\underline{\underline{\mathbf{X}}}) + \sum_{k=1}^f m_{ij;k} dX_k + \cdots \quad (12.98)$$

from which we infer, to leading order,

$$dM_{ij} = \sum_{k=1}^f m_{ij;k} dX_k + \cdots \quad (12.99)$$

In metric geometrical terms, this corresponds to the vector relationship

$$|\mathbf{M}_{ij}\rangle = \sum_{k=1}^f (\underline{\underline{\mathbf{m}}}_{ij})_k |\mathbf{X}_k\rangle \quad (12.100)$$

Of course, we might also consider the changes dM_{ij} with respect to changes in the intensive variables dR_k . The resulting geometrical relationship can be inferred most directly by substituting the usual expression for extensive $|\mathbf{X}_k\rangle$ in terms of intensive $|\mathbf{R}_\ell\rangle$ ket vectors,

$$|\mathbf{X}_k\rangle = \sum_{\ell=1}^f (\underline{\underline{\mathbf{M}}})_{k\ell} |\mathbf{R}_\ell\rangle \quad (12.101)$$

into (12.100) to obtain

$$|\mathbf{M}_{ij}\rangle = \sum_{k=1}^f (\underline{\underline{\mathbf{m}}}_{ij})_k \sum_{\ell=1}^f (\underline{\underline{\mathbf{M}}})_{k\ell} |\mathbf{R}_\ell\rangle \quad (12.102)$$

which can be recognized as

$$|\mathbf{M}_{ij}\rangle = \sum_{k=1}^f (\underline{\underline{\mathbf{M}}} \underline{\underline{\mathbf{m}}}_{ij})_k |\mathbf{R}_k\rangle \quad (12.103)$$

From (12.100), (12.103), we can evaluate the scalar product

$$\langle \mathbf{M}_{rs} | \mathbf{M}_{ij} \rangle = \sum_{k=1}^f (\underline{\mathbf{m}}_{rs})_k (\overline{\mathbf{M}} \underline{\mathbf{m}}_{ij})_\ell \langle \mathbf{X}_k | \mathbf{R}_\ell \rangle \quad (12.104)$$

and use the biorthogonality property ($\langle \mathbf{X}_k | \mathbf{R}_\ell \rangle = \delta_{k\ell}$) to obtain

$$\boxed{\langle \mathbf{M}_{rs} | \mathbf{M}_{ij} \rangle = \underline{\mathbf{m}}_{rs}^t \overline{\mathbf{M}} \underline{\mathbf{m}}_{ij}} \quad (12.105)$$

Alternatively, we might have started by considering the conjugate metric elements

$$\overline{M}_{ij} = (\overline{\mathbf{M}})_{ij} = \langle \mathbf{X}_i | \mathbf{X}_j \rangle = \left(\frac{\partial X_i}{\partial R_j} \right)_{\underline{X}} \quad (12.106)$$

and their displacements with respect to small changes $d\underline{\mathbf{R}}$ in the intensive properties $\underline{\mathbf{R}}$. Analogous manipulations to those given above lead to the corresponding conjugate results,

$$|\overline{\mathbf{M}}_{ij}\rangle = \sum_{k=1}^f (\underline{\mathbf{m}}_{ij})_k |\mathbf{R}_k\rangle \quad (12.107)$$

$$|\mathbf{M}_{ij}\rangle = \sum_{k=1}^f (\underline{\mathbf{M}} \underline{\mathbf{m}}_{ij})_k |\mathbf{X}_k\rangle \quad (12.108)$$

$$\langle \overline{\mathbf{M}}_{rs} | \mathbf{M}_{ij} \rangle = \underline{\mathbf{m}}_{rs}^t \underline{\mathbf{M}} \underline{\mathbf{m}}_{ij} \quad (12.109)$$

with conjugate derivative elements defined as

$$(\underline{\mathbf{m}}_{ij})_k = m_{ij;k} = \left(\frac{\partial \overline{M}_{ij}}{\partial R_k} \right)_{\underline{R}} \quad (12.110)$$

Thus, the metric elements M_{ij} that “underlie” the geometry of $\mathcal{M}_S$ themselves *become* geometrical vectors $|\mathbf{M}_{ij}\rangle$ of $\mathcal{M}_S$, if the higher-order derivative vectors $\underline{\mathbf{m}}_{ij}$ (or conjugate $\underline{\mathbf{m}}_{ij}$) are known. This testifies to the rather mind-bending mathematical richness of thermodynamic geometry.

The derivative vectors $\underline{\mathbf{m}}_{ij}$ of (12.97) can obviously be displayed in partitioned *matrix* form $\underline{\underline{\mathbf{m}}}$:

$$\underline{\underline{\mathbf{m}}} = \{m_{ij;k}\} = \begin{pmatrix} \underline{\mathbf{m}}_{11} & \underline{\mathbf{m}}_{12} & \cdots & \underline{\mathbf{m}}_{1f} \\ \underline{\mathbf{m}}_{12} & \underline{\mathbf{m}}_{22} & \cdots & \underline{\mathbf{m}}_{2f} \\ \vdots & \vdots & \vdots & \vdots \\ \underline{\mathbf{m}}_{1f} & \underline{\mathbf{m}}_{2f} & \cdots & \underline{\mathbf{m}}_{ff} \end{pmatrix} \quad (12.111)$$

This $f^2 \times f$ matrix (matrix of vectors), or the conjugate $\underline{\underline{\mathbf{m}}}$ matrix, contains the higher-order response functions needed to fully incorporate higher-order (U''') derivatives into the thermodynamic geometry. Geometrical identities for higher-order response functions can then be obtained in analogy to Sections 12.1–12.5.

The key difficulty in extending thermodynamic geometry to higher-order response functions is the additional experimental data required. To obtain, for example, the $\bar{m}_{11;1}$ element of a homogeneous fluid,

$$\bar{m}_{11;1} = \left(\frac{\partial(C_P/T)}{\partial T} \right)_P \quad (12.112)$$

one requires the thermal dependence of the isobaric heat capacity. This is often obtained in the form of a power series, for example,

$$C_P(T) = C_P(T^\circ) + a(T - T^\circ) + b(T - T^\circ)^2 + \dots \quad (12.113)$$

from which one obtains (at $T = T^\circ$)

$$m_{11;1} = -C_P/T^2 + a/T \quad (12.114)$$

Similar thermal and pressure dependence of $\alpha_P(T, P)$ and $\beta_T(T, P)$ is needed to complete the metric derivative matrix $\underline{\underline{m}}$ that would enable complete characterization of U''' -type derivatives in $\mathcal{M}_S$.

Some formal aspects of “ $d\mathcal{M}_S$ ” changes of state will be discussed in Section 13.1. However, further discussion of higher-order derivatives in $\mathcal{M}_S$ is beyond the scope of the present treatment.

Further Aspects of Thermodynamic Geometry

Chapters 9–12 have focused on the geometry of an individual equilibrium state, consistent with the Gibbs state-based perspective. In this concluding chapter, we now wish to examine some aspects of the “bigger picture,” including reversible changes of state (Section 13.1), irreversible state evolution from a nonequilibrium precursor (Section 13.2), and the origins of macroscopic thermodynamic geometry in the underlying atomic and molecular interactions (Section 13.3). Each aspect leads toward a frontier area of physical chemical research, hinting at fruitful areas for future study.

We also wish to briefly cite other research areas of science, engineering, and mathematics that relate to thermodynamic geometry, but are outside the scope of present discussion. A partial bibliography pertaining to interpretation and applications of thermodynamic geometry is given in Sidebar 13.1.

The metric geometry of equilibrium thermodynamics provides an unusual prototype in the rich spectrum of possibilities of differential geometry. Just as Einstein’s general relativistic theory of gravitation enriched the classical Riemann theory of curved spaces, so does its thermodynamic manifestation suggest further extensions of powerful Riemannian concepts. Theorems and tools of the differential geometer may be sharpened or extended by application to the unique Riemannian features of equilibrium chemical and phase thermodynamics.

Thermodynamic geometry also suggests novel approaches to certain conceptual problems of current physical theory. In the large-scale domain, metric thermodynamic concepts play a role in current theoretical analyses of black hole thermodynamics and cosmological space–time structure. In the small-scale limit, attempts to unify quantum mechanics with Einstein’s theory of gravitation often suggest analogies with “spontaneous symmetry breaking” in the thermodynamic domain; the surprising manner in which variable-dimension Euclidean geometrical structure *originates* in thermodynamic behavior might allow such analogies to be extended.

Geometrical tools prove useful in addressing various problems of finite-time thermodynamics and optimal control theory. These methods also have potential applicability to thermodynamic-type applications in subjects ranging from the chemical, biological, and materials sciences to information theory. Efficient vector-algebraic tools allow such applications to be extended to systems of virtually unlimited complexity, beyond realistic reach of classical methods.

The rich metric structure of macroscopic thermodynamics also presents unusually stringent tests of theoretical models. Attempts to understand thermodynamic phenomena at a molecular level seem to demand improved dynamical and quantum statistical thermodynamic models that adequately incorporate the subtleties of quantum-mechanical valency and bonding interactions. Development of such models is an active area of modern physical chemistry research, but a more complete survey of the current molecular theory of gases and liquids is beyond the scope of the present work.

SIDEBAR 13.1: SOME LEADING REFERENCES TO THERMODYNAMIC GEOMETRY

Some representative literature references pertaining to the thermodynamic geometry topics mentioned above, but not discussed elsewhere in this book, are listed below:

Mathematical Characterization and Interpretation

- G. E. Crooks. Measuring thermodynamic length. *Phys. Rev. Lett.* **99**, 100602 (2007).
- B. Andresen, R. S. Berry, E. Ihrig, and P. Salamon. Inducing Weinhold's metric from Euclidean and Riemannian metrics. *Phys. Rev. A* **37**, 849–51 (1988).
- P. Salamon, E. Ihrig, and R. S. Berry. A group of coordinate transformation which preserve the metric of Weinhold. *J. Math. Phys.* **24**, 2515–20 (1983).
- P. Salamon, B. Andresen, P. D. Gait, and R. S. Berry. The significance of Weinhold's length. *J. Chem. Phys.* **73**, 1001 (1980).
- P. Salamon and R. S. Berry. Thermodynamic length and dissipated availability. *Phys. Rev. Lett.* **51**, 1127–30 (1983).

Black Hole Thermodynamics

- J. Shen, R.-G. Cai, B. Wang, and R.-K. Su. Thermodynamic geometry and critical behavior of black holes. *Int. J. Mod. Phys. A* **22**, 11–27 (2007).
- T. Sarkar, G. Sengupta, and B. N. Tiwari. On the thermodynamic geometry of BTZ black holes. *JHEP* **11**, 15 (2006).
- J. E. Aman, I. Bengtsson, and N. Pidokrajt. Flat information geometries in black hole thermodynamics. *Gen. Rel. Grav.* **38**, 1305–15 (2006).
- R.-G. Cai and J.-H. Cho. Thermodynamic curvature of the BTZ black hole. *Phys. Rev. D* **60**, 067502 (1999).
- S. Ferrara, G. W. Gibbons, and R. Kallosh. Black holes and critical points in moduli space. *Nucl. Phys. B* **500**, 75–93 (1997).

Finite-Time Processes and Optimal Control Theory

- J. D. Nulton and P. Salamon. Optimality in multi-stage operations with asymptotically vanishing cost. *J. Non-Equil. Thermodyn.* **27**, 271 (2002).
- B. Andresen. Finite-time thermodynamics and simulated annealing. In J. S. Shiner (ed.), *Entropy and Entropy Generation* (Springer, New York, 2002), pp. 111–28.

- S. Sieniutycz and A. De Vos (eds). *Thermodynamics of Energy Conversion and Transport* (Springer, New York, 2000), Part III: Energy in Geometrical Thermodynamics, pp. 255–331.
- B. Andresen. Minimizing losses—tools of finite-time thermodynamics. In A. Bejan and E. Mamut (eds). *Thermodynamic Optimization of Complex Energy Systems* (Springer, New York, 1999), p. 411.
- B. Andresen, P. Salamon and R. S. Berry. Thermodynamics in finite time. *Phys. Today* **37**, 62–70 (1984).

Complex Chemical Equilibria

- E. Di Cera. *Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules* (Cambridge University Press, New York, 1995).
- E. Di Cera and J. Wyman. Global and local metric geometry of ligand binding thermodynamics. *Proc. Natl Acad. Sci. USA* **88**, 3494–7 (1991).

Information and Statistical Theory

- M. Portesi, A. Plastino, and F. Pennini. Information measures based on Tsallis' entropy and geometrical considerations for thermodynamic systems. *Physica A* **365**, 173–6 (2006).
- T. Feldmann, R. D. Levine, and P. Salamon. A geometrical measure for entropy changes. *J. Stat. Phys.* **42**, 1127–34 (1986).

Alternative Formulations

In addition, several alternative formulations of thermodynamic geometry have been presented, starting from entropy-based (or other) fundamental equations (see Sections 5.4 and 5.5). From the equilibrium thermodynamics viewpoint, these alternative formulations are completely equivalent, and each could be considered a special case of the general transformations outlined in Section 11.4. Nevertheless, each alternative may suggest distinct statistical-mechanical origins, Riemannian paths, or other connotations that make it preferable for applications outside the equilibrium thermodynamics framework.

- G. Ruppeiner. Riemannian geometry in thermodynamic fluctuation theory. *Rev. Mod. Phys.* **67**, 605–59 (1995).
 - D. Brody and N. Rivier. Geometrical aspects of statistical mechanics. *Phys. Rev. E* **51**, 1006–11 (1995).
 - R. Mrugala, J. D. Nulton, J. C. Schön, and P. Salamon. Statistical approach to the geometrical structure of thermodynamics. *Phys. Rev. A* **41**, 3156–60 (1990).
 - F. Schlögl. Thermodynamic metric and stochastic measures. *Z. Phys. B* **59**, 449–54 (1985).
 - K. Horn. Weinhold's metric geometry and the second law of thermodynamics. *Phys. Rev. A* **32**, 3142–3 (1985).
 - P. Salamon, J. Nulton, and E. Ihrig. On the relation between energy and entropy versions of thermodynamic length. *J. Chem. Phys.* **80**, 436–7 (1984).
-

13.1 REVERSIBLE CHANGES OF STATE: RIEMANNIAN GEOMETRY

Previously (Section 9.3), we have alluded briefly to various distinguishable types of mathematical “spaces” with differing degrees of algebraic and geometrical structure. We now wish to clarify these distinctions more carefully, in order to introduce a *new* type of space that differs profoundly from more familiar Euclidean-like varieties.

The definition of a mathematical space begins with the set of *objects* $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots\}$ that occupy the space (an intrinsically “empty space” being a physically problematic concept). Among the simplest algebraic structures that can characterize such objects is that of a *linear manifold*, also called a linear vector space, affine space, etc. By definition, such a manifold has only two operations—“addition” ($\mathbf{X} + \mathbf{Y}$) and “multiplication by a scalar” ($\lambda\mathbf{X}$)—resulting in each case in another element of the manifold. These operations have the usual distributive,

$$\lambda(\mathbf{X} + \mathbf{Y}) = \lambda\mathbf{X} + \lambda\mathbf{Y} \quad (13.1a)$$

$$(\lambda + \mu)\mathbf{X} = \lambda\mathbf{X} + \mu\mathbf{X} \quad (13.1b)$$

associative,

$$\mathbf{X} + (\mathbf{Y} + \mathbf{Z}) = (\mathbf{X} + \mathbf{Y}) + \mathbf{Z} \quad (13.2a)$$

$$\lambda(\mu\mathbf{X}) = (\lambda\mu)\mathbf{X} \quad (13.2b)$$

and commutative

$$\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X} \quad (13.3a)$$

$$\lambda\mu\mathbf{X} = \mu\lambda\mathbf{X} \quad (13.3b)$$

properties that one commonly assumes in ordinary arithmetic. A linear manifold therefore possesses the qualities of connectivity or continuity implied by the word “space,” but lacks the quality of *measurability*, i.e., the possibility of assigning numbers to each pair of objects that could reflect their “proximity.” The geometry of such a linear manifold therefore has only affine or topological (rather than metric) significance, and cannot be associated with geometry in its historical Euclidean sense as applied to distances, angles, areas, and other metric characteristics of spatial objects.

As discussed in Section 9.3, a higher level of mathematical structure is achieved by defining an additional “multiplication” ($\mathbf{X} \cdot \mathbf{Y}$) operation, that is, a rule that associates a (real) scalar with every pair of objects $\mathbf{X}, \mathbf{Y}$ in the manifold. For Euclidean-like spaces, the scalar product has distributive, commutative, and positivity properties given by

$$\mathbf{X} \cdot (\lambda\mathbf{Y} + \mu\mathbf{Z}) = \lambda\mathbf{X} \cdot \mathbf{Y} + \mu\mathbf{X} \cdot \mathbf{Z} \quad (13.4a)$$

$$\mathbf{X} \cdot \mathbf{Y} = \mathbf{Y} \cdot \mathbf{X} \quad (13.4b)$$

$$\mathbf{X} \cdot \mathbf{X} \geq 0 (= 0 \text{ only if } \mathbf{X} = 0) \quad (13.4c)$$

Each side of (13.4a–c) is an ordinary number, able to represent measurement of a spatial object (i.e., as some multiple of a chosen “unit”).

It is a remarkable fact that properties (13.4a–c) are necessary and sufficient to give Euclidean geometry. In other words, if any rule can be found that associates a number (say, $\langle \mathbf{X} | \mathbf{Y} \rangle$) with each pair of abstract objects (“vectors” $|\mathbf{X}\rangle, |\mathbf{Y}\rangle$) of the manifold in a way that satisfies (13.4a–c), then the manifold is isomorphic to a corresponding Euclidean vector space. We introduced a rather unconventional rule for the scalar products $\langle \mathbf{X} | \mathbf{Y} \rangle$ [recognizing that (13.4a–c) are guaranteed by the laws of thermodynamics] to construct the abstract Euclidean metric space $\mathcal{M}_S$ for an equilibrium state of a system S , characterized by a metric matrix $\underline{\underline{M}}$. This geometry allows the thermodynamic state description to be considerably simplified, as demonstrated in Chapters 9–12.

We now wish to introduce a still deeper form of geometry as first suggested by Bernhard Riemann (Sidebar 13.2). Riemann’s formalism makes possible a distinction between the space of vectors whose metrical relationships are specified by the metric $\underline{\underline{M}}$ and an associated linear manifold by which the vectors and metric are parametrized. Let $\underline{\xi}$ be an element of a linear manifold (in general, having no metric character) that can uniquely identify the “state” of a collection of metrical objects $\{|\mathbf{X}\rangle\}$. The Riemannian geometry permits the associated metric $\underline{\underline{M}}$ to itself be a function of the state,

$$\underline{\underline{M}} = \underline{\underline{M}}(\underline{\xi}) \quad (13.5)$$

so that the geometry of the objects $\{|\mathbf{X}\rangle\}$ varies with changes in $\underline{\xi}$. For any particular $\underline{\xi}$, the Riemannian metric $\underline{\underline{M}}(\underline{\xi})$ satisfies the usual conditions (13.4a–c) for Euclidean vectors, so the geometry is always “locally Euclidean.” However, one may now consider the multiply continuous family of such states, related by (13.5), that form the overall Riemannian geometry $\mathcal{M}_S(\underline{\xi})$. Although the Riemannian geometry is locally “flat” (Euclidean), its *global* structure can be curved in surprising ways, depending on the functional form of (13.5). Notice that “curvature” of a Riemannian space is profoundly different from that of the related concept in Euclidean space. As Gauss instinctively realized upon hearing the brilliant proposal of his student Riemann, the theory to describe surfaces in curved Riemann space is *not* equivalent to the “Gaussian curvature theory” of curved surfaces embedded in a flat space.

In the application of Riemann’s formalism to geometry, the manifold $\underline{\xi}$ is identified with a set of positions in a global space, while the metrical relationships governed by $\underline{\underline{M}}$ represent infinitesimal displacements in the same space. However, it has proved useful, as in Einstein’s gravitation theory [see, e.g., C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation* (Freeman, San Francisco, 1973)] to regard the metric as depending on physical properties $\underline{\xi}$ (e.g., the masses of physical objects) quite unrelated to the spatial properties that the geometry was intended to describe.

In a similar vein, Riemann’s formalism finds useful application in expressing the global thermodynamic behavior of a system S . The metric geometry governed by $\underline{\underline{M}}(\underline{\xi})$ represents thermodynamic responses (as before), while $\underline{\xi}$ labels distinct *states* of equilibrium, each exhibiting its own local geometry of responses. The state-specifier “manifold” $\underline{\xi}$ may actually be chosen rather freely, for example, as any f independent intensive variables (such as $\xi_1 = T, \xi_2 = P, \xi_3 = \mu_1, \dots, \xi_f = \mu_{c-p}$). For our purposes, it is particularly convenient to

choose the ξ_i as the *conjugates* $\{X_i\}$ of the chosen base intensities $\{R_i\}$ of $\underline{\underline{M}}$, namely,

$$\xi_1 = S, \xi_2 = V, \xi_3 = n_1, \dots, \xi_f = n_{c-p} \quad (13.6)$$

so that the ξ_i have the significance of Gibbs-space X_i variables,

$$d\xi_i = dX_i, \quad i = 1, 2, \dots, f \quad (13.7)$$

Note that these ξ_i exhibit only a subset of the usual properties (13.1)–(13.3) of an affine manifold (for example, any volume intermediate between two “allowed” volumes is also allowed, but this variable cannot be sensibly extended to negative values), so one should not put undue stress on this aspect.

The key feature of Riemannian geometry is the concept of a line element ds whose length is given by (Riemann’s only equation in his 1854 *Habilitationsvortrag*)

$$ds = \left[\sum_{i,j=1}^f (\underline{\underline{M}})_{ij} dX_i dX_j \right]^{1/2} \quad (13.8a)$$

which, in view of the choice (13.7), can also be written as

$$ds = \left[\sum_{i,j=1}^f (\underline{\underline{M}})_{ij} d\xi_i d\xi_j \right]^{1/2} \quad (13.8b)$$

We envision $s = s(t)$ as the “path length” from initial state $\underline{\underline{\xi}}'$ to final state $\underline{\underline{\xi}}''$ along a chosen *path*

$$\underline{\underline{\xi}} = \underline{\underline{\xi}}(t) \quad (13.9a)$$

parametrized by a scalar “progress variable” t that varies from 0 (initial state) to τ (final state):

$$\underline{\underline{\xi}}(t=0) = \underline{\underline{\xi}}', \quad \underline{\underline{\xi}}(t=\tau) = \underline{\underline{\xi}}'' \quad (13.9b)$$

The total distance $\mathcal{L} = s(\tau) - s(0)$ between initial and final states can then be evaluated as the *path integral*

$$\mathcal{L} = \int_0^\tau dt \left[\sum_{i,j} (\underline{\underline{M}})_{ij} \frac{dX_i}{dt} \frac{dX_j}{dt} \right]^{1/2} \quad (13.10)$$

where the metric $\underline{\underline{M}}(t)$ as well as extensities $X_i(t)$ carry path dependence. Of course, $ds = d\mathcal{L}$ as defined by the square-root relationship (13.8b) is an imperfect differential, and the distance $\mathcal{L}$ in (13.10) is path-dependent. The shortest possible distance between states is given by a *geodesic* path, the nearest equivalent to a “straight-line path” of Euclidean space. The physical meaning of $\mathcal{L}$ has engendered considerable discussion [see, e.g., Salamon *et al.*

(1980), Salamon and Berry (1983), Crooks (2007) in Sidebar 13.1], but is intrinsically dependent on the chosen path.

In principle, the integrand in (13.10) might be evaluated with Taylor series expansions such as (12.96), based on successively higher derivatives of the initial state. In practice, however, direct experimental evaluation of the functional dependence of each M_{ij} on path variables would be needed to evaluate $\mathcal{L}$ along extended paths. Further discussion of global curvature or other descriptors of the Riemannian geometry of real substances therefore awaits acquisition of appropriate experimental data, well beyond that required to describe individual points on a reversible path.

Although the Riemannian geometry framework is both mathematically apt and conceptually suggestive for treating reversible changes of thermodynamic state, it is important to note certain *differences* in the application of Riemannian concepts to thermodynamics as opposed to spatial geometry or gravitational theory. The thermodynamic Riemannian geometry has no “tensorial” aspect in the ordinary sense, because the theory has no dependence on spatial transformations, the isotropy of three-dimensional coordinate space, or related tensor-type concepts. For this reason, there is little justification (except perhaps for bringing out suggestive analogies) to adopt tensor-like notational conventions, such as covariant versus contravariant indices, contractions, and the like. Indeed, the deep way in which dimensional changes pervade the thermodynamic framework makes the “Einstein summation convention” a particularly dangerous source of ambiguity and confusion. Similarly, the thermodynamic Riemannian description has no apparent relationship to conventional “field”-type mathematical objects (i.e., presumed functional forms assigning scalar-, vector-, or tensor-valued properties to each point of an underlying coordinate space). Many of the elegant descriptors of Riemannian curvature in field-based frameworks such as Einstein’s general theory of relativity therefore seem ineffective in the thermodynamic context. Furthermore, the thermodynamic metric is *strictly* Riemannian (i.e., with strictly *non-negative* eigenvalues), rather than “pseudo-Riemannian” (i.e., with one negative-valued “time-like” metric component) as assumed in Einstein’s adaptation of Riemannian geometry to space–time description. Thus, Riemannian geometry as applied to thermodynamics should reflect the unique features that *distinguish* thermodynamic theory from ordinary dynamical field-type theories of matter in three-dimensional Cartesian or four-dimensional Minkowski spaces.

As described in Section 2.10, it is remarkable that equilibrium thermodynamics makes practically *no* reference to “position,” “shape,” or other geometry-like variables that are usually taken as starting points for physical description. Nor does equilibrium theory admit the *time-like* coordinate assumed necessary for space–time description of dynamical systems. In the conventional macroscopic equilibrium thermodynamics treated throughout this book, only the volume variable has appreciable geometrical quality (contributing “1” to the Gibbs phase rule, which otherwise appears sensitive only to chemical components c and phases p of the system). However, in the submacroscopic range in which shape or surface area become appreciable state variables, the energy variations dU must be enlarged to include possible independent contributions from each Cartesian direction, and the Gibbs–Duhem and phase rule relationships must be generalized accordingly. Such “non-extensive thermodynamics” [see, e.g., C. Tsallis. Non-extensive thermostatics: brief review and comments. *Physica A* **221**, 277 (1995)] seems increasingly necessary to describe current nanoscale applications, where the pronounced anisotropies of molecular-level valency and bonding interactions lead to richly anisotropic structures of crystals and polymorphic phase domains. A metric formulation of such a generalization could

reveal interesting relationships between thermo-geometrical versus spatio-geometrical concepts of “degrees of freedom.” However, extensions of thermodynamics to include richer details of three-dimensional spatial variability lie beyond the scope of the present treatment.

SIDEBAR 13.2: BERNHARD RIEMANN (1826–1866)



Georg Friedrich Bernhard Riemann was the second of six children born into the family of an impoverished Lutheran minister in lower Saxony, near modern Hannover, Germany. Riemann suffered from childhood malnutrition and was prone to excessive timidity, sickness, and nervous breakdowns throughout his difficult life, which ended in Italy while he was attempting recuperation from chronic tuberculosis (with a wife and young child at his side) at age 39.

Young Bernhard was home-schooled by his widowed father until age 14. Blessed with remarkable computational skills at an early age, Riemann juggled his mathematical proclivities with intense biblical studies (even attempting to establish Genesis by mathematical proof!) until sent away to his grandmother for public schooling in Hannover and Lüneburg. He soon attracted favorable attention from his high school mathematics teachers, one of whom allowed Riemann to borrow life-changing books of Legendre and Gauss from his private library. Riemann began university studies at Göttingen as a theology student, intending to follow his father into the ministry. But after succumbing to the spell of a course with Gauss (perhaps the greatest mathematician of all time), he was allowed thereafter to pursue mathematics. After a stay in Berlin, where he came under the influence of Jacobi, Dirichlet, and other leading mathematicians, Riemann returned to Göttingen to complete his doctoral thesis in 1851 under Gauss, who praised his “gloriously fertile originality.” Gauss also helped Riemann procure an initial (unpaid) *Privatdozent* lectureship position at Göttingen, where he remained throughout his academic career.

Riemann’s *Habilitation* and appointment required both a written essay and a public lecture, based on distinct topics to be selected by Gauss from a list prepared by the candidate. To Riemann’s surprise, Gauss selected “Fourier Series” for the essay and “Foundations of Geometry” for the lecture, both far removed from his path-breaking thesis studies on complex analysis. The written essay of 1853, introducing what is now called “Riemann integrability,” was itself considered a masterpiece. However, Riemann’s public lecture of June 10, 1854 [*Über die Hypothesen, welche der Geometrie zu Grunde liegen*,” published posthumously in 1868] is considered a pivotal moment of mathematical history. This lecture included but one equation (13.8a), and was probably not fully comprehended by any member of the audience except Gauss (in his last year of life), who saw his own theory of differential geometry being eclipsed.

Riemann’s lecture indeed shook geometry to its foundations. He was the first to propose extending Euclidean geometry concepts beyond three dimensions. More importantly, Riemann showed how one could entirely reject Euclid’s fifth postulate (“through a point

not on a given line there is but one line parallel to the given line”) and modify the second (“any straight line segment can be extended indefinitely as a straight line”) to create a consistent “curved” geometry with hitherto unimagined possibilities (*no* parallel lines, *all* triangles with sum of angles $>180^\circ$, etc.). Riemann clearly recognized that such curved non-Euclidean geometry may well be more than just a bizarre mathematical possibility, and that the true nature of physical space must ultimately be decided by *experiment* rather than abstract philosophical logic. As stated in his lecture:

It is quite conceivable that the geometry of space in the very small does not satisfy the axioms of [Euclidean] geometry . . . The properties which distinguish space from other conceivable triply-extended magnitudes are only to be deduced from experience.

Riemann’s brilliant insights touched many other areas of mathematics, and he was eventually rewarded (after 1857) with a salaried position and (after 1859) chairmanship of Göttingen mathematics. However, he remained fascinated by the idea that physical forces might be understood as “wrinkles” in a curved space of higher dimension. Virtually up to his untimely death, Riemann worked intensively to unify electromagnetism and gravitational theory on this basis, coming tantalizingly close to a geometrical theory of gravity. More than a half-century later, Riemann’s mathematical methods and physical insights were to animate Einstein’s successful development of the general theory of relativity. As Einstein himself later remarked of Riemann’s contribution:

Physicists were still far removed from such a way of thinking: space was still, for them, a rigid, homogeneous something, susceptible of no change or conditions. Only the genius of Riemann, solitary and uncomprehended, had already won its way by the middle of the last century to a new conception of space, in which space was deprived of its rigidity, and in which its power to take part in physical events was recognized as possible.

[For excellent insight into Riemann and his mathematics (primarily focusing on the Riemann zeta function and the “Riemann hypothesis”), see J. Derbyshire. *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics* (Penguin, New York, 2004).]

13.2 NEAR-EQUILIBRIUM IRREVERSIBLE THERMODYNAMICS: DIFFUSIONAL GEOMETRY

As pointed out in Section 5.1, Gibbs’ formulation of equilibrium thermodynamics emerged as a brilliant alternative to previous Clausius-type formulations in terms of *nonequilibrium* engines, cycles, and processes. The latter necessarily involve aspects of time evolution and *irreversibility* that are beyond the domain of the equilibrium theory. Although the Gibbs equilibrium theory (whether in classical or geometrical formulation) provides a powerful tool for characterizing ideal limiting equilibrium states and reversible processes, it is also of interest to consider extensions of this theory to the irreversible nonequilibrium aspects of real natural events.

Extensions of thermodynamic concepts beyond the equilibrium limit bring new opportunities and challenges to the metric geometrical formalism. Whereas the metric geometry of equilibrium thermodynamics is, in deepest sense, *equivalent* to earlier Gibbs-type or

Clausius-type formulations of the equilibrium limit, this need not be the case for suggested extensions to nonequilibrium phenomena. Indeed, intuitive assumptions about how a *geometry* might be “continued” into the time-dependent domain are likely to differ appreciably from the corresponding assumptions of a calculus-based approach. We therefore wish to investigate how geometrical concepts might be extended to the domain of irreversible phenomena, along the lines first recognized by G. Nathanson and O. Sinanoğlu [Metric geometry of irreversible thermodynamics. *J. Chem. Phys.* **72**, 3127–9 (1980)].

Attempts to formulate a more general “nonequilibrium thermodynamics” have a checkered history. Although consensus exists on certain rigorous results in the *near-equilibrium* regime (e.g., the Onsager reciprocity relations to be discussed below), more ambitious extensions to systems far from equilibrium have often led to controversy or inordinate dependence on microscopic assumptions that lack the compelling rigor of the equilibrium limit. Accordingly, the present discussion is confined to near-equilibrium limiting behavior, anchored in inductive foundations akin to those of the equilibrium theory itself.

Our search for nonequilibrium extension of Gibbsian thermostatics can be usefully guided by Gibbs’ own analysis of the equilibrium limit. As described in Section 5.2, the “striving” toward equilibrium is associated with nonequilibrium exchanges of heat, work, and chemical species between subregions of *local* equilibrium, leading to eventual *uniformity* of temperature, pressure, and chemical potential across the system—the signature of final overall equilibrium. Consistent with this Gibbsian picture, we now consider a spatial network of infinitesimal subregions (“cells”) in local thermodynamic equilibrium, continuously distributed throughout the evolving system. The natural mathematical object to describe local values of an evolving property Π is a *field* $\Pi(x, y, z, t)$, which assigns a property value (of scalar, vector, or higher tensorial character) to each spatial point (x, y, z) of a system at time t . The goal of a near-equilibrium theory is to describe the long-time asymptotic approach of nonuniform fields toward the equilibrium limit $\Pi(\text{eq})$, where each property becomes stationary (independent of t) and uniform throughout the system (independent of x, y, z). Accordingly, the properties of interest will be chosen as the f independent intensities R_i (T , $-P$, and $c - 2$ of the chemical potentials) whose field-type functional values $R_i(x, y, z, t)$ in the near-equilibrium regime “strive toward” limiting values $R_i(\text{eq})$ of the final equilibrium state description:

$$\lim_{t \rightarrow \infty} R_i(x, y, z, t) \longrightarrow R_i(\text{eq}), \quad i = 1, 2, \dots, f \quad (13.11)$$

Assuming that (13.11) makes sense in the context of the system under investigation (i.e., that physical relaxation times are in the appropriate range for the condition of local equilibrium to be satisfactorily approximated), we seek the field-type differential equation that describes asymptotic t -evolution of fields $R_i(x, y, z, t)$ toward the known metric geometrical limit. Solutions of this equation are expected to describe a wide variety of thermal, acoustic, and diffusion phenomena in nonequilibrium conditions where “local thermodynamic variables” retain experimental meaning.

In essence, we wish to follow the Gibbs entropy maximization procedure of Section 5.2 “backward in time.” Specifically, we seek to characterize the final stage of equilibration when effective local equilibrium has been achieved in each small cell n , but cell intensities are not yet equalized throughout the system. To make direct contact with Section 5.2, we shall temporarily revert to the *entropy representation* (Section 5.3), which generates a scalar product and metric geometry that is conformally equivalent to the U -based metric

adopted elsewhere in this book, namely,

$$S = S(U, V, n_1, \dots, n_c) = S(\{X_i\}); \quad R_i = \left(\frac{\partial S}{\partial X_i} \right)_{\underline{X}} \quad (13.12a)$$

$$\langle \mathbf{R}_i | \mathbf{R}_j \rangle = - \left(\frac{\partial R_i}{\partial X_j} \right)_{\underline{X}} = - \left(\frac{\partial^2 S}{\partial X_i \partial X_j} \right)_{\underline{X}} \quad (13.12b)$$

Note that the Hessian in (13.12b) refers to *negative* entropy, as required by the opposite curvatures of S versus U (Section 5.4). However, in other respects, the S -based concepts and equations closely parallel those introduced previously and can be employed without further comment in the present section.

In the entropy representation, the fundamental differential dS can be written as

$$dS = \sum_{i=1}^f R_i d\bar{R}_i \quad (13.13)$$

where

$$\{R_i\} = \{1/T, -P/T, \mu_1/T, \dots\} \quad (13.14a)$$

$$\{\bar{R}_i\} = \{U, V, N_1, \dots\} \quad (13.14b)$$

As in Section 11.6, it is convenient to transform to the self-conjugate *eigenmodes* E_i , in which

$$dS = \sum_{i=1}^f E_i dE_i \quad (13.15)$$

We can also partition dS into local *cell* (n) contributions

$$dS = \sum_n^{\text{cells}} \sum_{i=1}^f E_i^{(n)} dE_i^{(n)} \quad (13.16)$$

where

$$E_i = \sum_n^{\text{cells}} E_i^{(n)}, \quad \text{for all } t \quad (13.17)$$

$$E_i^{(1)} = E_2^{(2)} = \dots = E_i^{(n)} = \dots \quad \text{at equilibrium } (t \rightarrow \infty) \quad (13.18)$$

We now wish to envision the “de-equilibration” to times t when the condition (13.18) is not yet satisfied and values of E_i vary from cell to cell (i.e., from point to point in space) in a “smooth” fashion to be described. In this near-equilibrium regime, E_i can be properly

described as a mathematical *field*,

$$E_i \longrightarrow E_i(x, y, z, t) \quad (13.19)$$

that is, a function having well-defined value at each spatial position (x, y, z) and time t . Compared to the corresponding equilibrium expression (13.16), $E_i^{(n)}$ is altered by a small $\delta E_i^{(n)}$ that varies both spatially and temporally, leading to the t -dependent near-equilibrium expression

$$dS_{\text{ne}} = \sum_n \sum_{i=1}^f (E_i^{(n)} + \delta E_i^{(n)}) dE_i^{(n)} \quad (13.20)$$

To be specific, suppose that E_i has a spatial gradient in the z direction, as shown in Fig. 13.1. Then the cell centered at $z^{(n)}$ has the local value shifted by

$$\delta E_i^{(n)} = \left. \frac{\partial E_i}{\partial z} \right|_{z^{(n)}} \equiv E_{i;z}(z^{(n)}) \quad (13.21)$$

relative to the equilibrium value at $z = 0$. The notation $E_{i;z}$ denotes the z -gradient function of E_i , and $E_{i;z}(z^{(n)})$ denotes the value of this function at $z^{(n)}$.

With the notation of (13.21), the total differential of $\Delta S = S_{\text{ne}} - S_{\text{eq}}$ becomes

$$d(\Delta S) = \sum_n \sum_{i=1}^f E_{i;z}(z^{(n)}) dE_i^{(n)} \quad (13.22)$$

and its local contribution from cell n is

$$d(\Delta S^{(n)}) = \sum_{i=1}^f E_{i;z}(z^{(n)}) dE_i^{(n)} \quad (13.23)$$

The partial t -derivative of this expression therefore gives the rate of local entropy production ($\sigma^{(n)}$) as

$$\frac{\partial(\Delta S^{(n)})}{\partial t} = \sigma^{(n)} = \sum_{i=1}^f E_{i;z}(z^{(n)}) \left. \frac{\partial E_i}{\partial t} \right|_{z^{(n)}} = \sum_{i=1}^f E_{i;z}(z^{(n)}, t) E_{i;t}(z^{(n)}, t) \quad (13.24)$$

where we have similarly denoted $E_{i;t} = \partial E_i / \partial t$ as the partial time-derivative function.

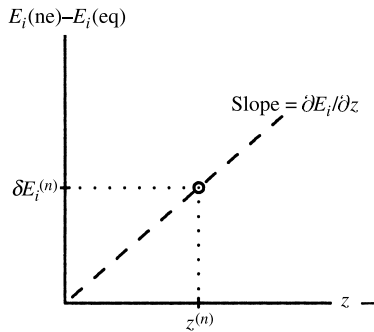


Figure 13.1 Indexing of cell n to a numerical position $z^{(n)}$ and associated nonequilibrium shift $\delta E_i^{(n)}$ in a gradient $\partial E_i / \partial z$ of eigenmode E_i .

The key feature of the near-equilibrium limit is the linear proportionality of “flows” (or “fluxes”) ($J_i^e = \partial E_i / \partial t$) and “forces” (or “affinities”) ($F_i^e \equiv -\partial E_i / \partial z$), which can be expressed in the present case as

$$\boxed{\frac{\partial E_i}{\partial t} = D_i \frac{\partial E_i}{\partial z}} \quad (13.25a)$$

or as

$$J_{iz}^e(z, t) = E_{i,t}(z, t) = D_i E_{i,z}(z, t) = -D_i F_i^e(z, t) \quad (13.25b)$$

Here D_i is a positive proportionality constant (“diffusion constant” for E_i), J_{iz}^e is z -ward flow induced by the gradient, and superscript “e” denotes *eigenmode* character of the associated force or flow. The proportionality (13.25) corresponds to Fick’s first law of diffusion when E_i is dominated by mass transport or to Fourier’s heat theorem when E_i is dominated by heat transport, but it applies here more deeply to the metric eigenvalues that control all transport phenomena. In the near-equilibrium limit (13.25), the local entropy production rate (13.24) is evaluated as

$$\sigma^{(n)} = \sum_{i=1}^f D_i [E_{i,z}(z^{(n)}, t)]^2 = \sum_{i=1}^f D_i^{-1} [E_{i,t}(z^{(n)}, t)]^2 \geq 0 \quad (13.26)$$

leading to *positive* entropy production (as expected) until topping out at the final $t \rightarrow \infty$ limit.

The one-dimensional force-flow equations (13.21)–(13.26) can be readily generalized for three-dimensional gradients and flows. In this case, $E_{i,z} = \partial E_i / \partial z$ is replaced by the three-dimensional gradient $\vec{\nabla}$,

$$E_{i,\vec{r}} = \vec{\nabla} E_i = \left(\frac{\partial E_i}{\partial x}, \frac{\partial E_i}{\partial y}, \frac{\partial E_i}{\partial z} \right) \quad (13.27)$$

leading to the three-dimensional force vector

$$\vec{F}_i^e = -\vec{\nabla} E_i, \quad i = 1, 2, \dots, f \quad (13.28)$$

The directional $J_{ix}^e, J_{iy}^e, J_{iz}^e$ flows correspondingly become the $\vec{J}_i^e$ flow vector (directional time derivative of E_i):

$$\vec{J}_i^e = (J_{ix}^e, J_{iy}^e, J_{iz}^e) = \vec{\partial}_t E_i \quad (13.29)$$

The local entropy production rate (13.24) is then expressed in terms of *eigenforces* and *eigenflows* as

$$\sigma^{(n)} = \sum_{i=1}^f (\vec{\nabla} E_i) \cdot (\vec{\partial}_t E_i) = - \sum_{i=1}^f \vec{F}_i^e \cdot \vec{J}_i^e \quad (13.30)$$

with the near-equilibrium assumption (13.25) now expressed as

$$-\vec{F}_i^e = \vec{\nabla} E_i = D_i^{-1}(\vec{\partial}_i E_i) = D_i^{-1} \vec{J}_i^e \quad (13.31)$$

where all quantities are evaluated at $(\vec{r}, t)$ for cell n at time t .

The eigenmode expansion (13.30) for the local entropy production rate can be expressed in terms of usual laboratory variables $R_i, \bar{R}_i$ (13.14a, b) using transformation equations analogous to those of Section 11.6. In the present S -based framework, the expansion of dE_i in intensities [cf. (11.89)] becomes

$$dE_i = \sum_{j=1}^f i |e_i|^{-1/2} (\underline{q}_i)_j dR_j \quad (13.32)$$

where imaginary i (the square-root of -1) accounts for the change of sign in the S -based metric (13.12). From (13.32), we obtain the expansion in base intensive forces $-\vec{\nabla} R_j$:

$$-\vec{F}_i^e = \vec{\nabla} E_i = i |e_i|^{-1/2} \sum_{j=1}^f (\underline{q}_i)_j \vec{\nabla} R_j \quad (13.33)$$

From the corresponding expansion of dE_i in conjugate extensities $\bar{R}_k$,

$$dE_i = \sum_{k=1}^f i |e_i|^{+1/2} (\underline{q}_i)_k d\bar{R}_k \quad (13.34)$$

we obtain similarly for the conjugate flow derivatives

$$\vec{J}_i^e = \vec{\partial}_i E_i = i |e_i|^{+1/2} \sum_{k=1}^f (\underline{q}_i)_k (\vec{\partial}_i \bar{R}_k) \quad (13.35)$$

Inserting (13.33) and (13.35) into the local entropy rate expression (13.30), we obtain

$$\begin{aligned} \sigma^{(n)} &= i^2 \sum_{j=1}^f \sum_{k=1}^f (\underline{q}_i)_j (\vec{\nabla} R_j) (\underline{q}_i)_k (\vec{\partial}_i \bar{R}_k) \\ &= - \sum_{j=1}^f \sum_{k=1}^f (\vec{\nabla} R_j) (\vec{\partial}_i \bar{R}_k) \sum_{i=1}^f (\underline{Q})_{ji} (\underline{Q}^t)_{ik} \\ &= + \sum_{j=1}^f \sum_{k=1}^f (\vec{F}_j \cdot \vec{J}_k) \delta_{jk} \end{aligned} \quad (13.36)$$

This leads to the final expression for the local entropy production rate

$$\sigma^{(n)} = \sum_{j=1}^f \vec{F}_j \cdot \vec{J}_j \quad (13.37)$$

in terms of conjugate forces and flows (referring now to biorthogonal R_i , $\bar{R}_i$ rather than orthogonal eigenmodes E_i):

$$\vec{F}_j \equiv -\vec{\nabla} R_j, \quad \vec{J}_j \equiv \vec{\partial}_t \bar{R}_j \quad (13.38)$$

where all quantities are evaluated at the chosen cell position $\vec{r}^{(n)}$ and time t .

The common connection to underlying metric eigenmodes [cf. Eqs. (13.32), (13.34)] implies a deep-seated geometrical relationship between the conjugate forces and flows. To find this relationship, we can use the inverse of (13.34) [cf. the U -based (11.91b)]

$$d\bar{R}_i = \sum_{j=1}^f i|e_j|^{-1/2} (\underline{q}_j)_i dE_j \quad (13.39)$$

to write, say, for the z component of $\vec{J}_i$,

$$J_{iz} \equiv \frac{\partial \bar{R}_i}{\partial t} = \frac{\partial}{\partial t} \left(\sum_{k=1}^f i|e_k|^{-1/2} (\underline{q}_k)_i dE_k \right) = \sum_{k=1}^f i|e_k|^{-1/2} (\underline{q}_k)_i J_{kz}^e \quad (13.40)$$

Equation (13.31) then gives, with help from (13.32),

$$\begin{aligned} J_{iz} &= \sum_{k=1}^f i|e_k|^{-1/2} (\underline{q}_k)_i D_k \frac{\partial E_k}{\partial z} \\ &= \sum_{k=1}^f i|e_k|^{-1/2} (\underline{q}_k)_i D_k \sum_{j=1}^f i|e_k|^{-1/2} (\underline{q}_k)_j \frac{\partial R_j}{\partial z} \\ &= \sum_{j=1}^f \left(\sum_{k=1}^f \frac{D_k}{|e_k|} (\underline{q}_k)_i (\underline{q}_k)_j \right) F_{jz} \end{aligned} \quad (13.41)$$

Equation (13.41) can be written in vector form as

$$\vec{J}_i = \sum_{j=1}^f (\underline{\underline{L}})_{ij} \vec{F}_j \quad (13.42)$$

This equation corresponds to the starting point of the phenomenological theory of near-equilibrium transport properties first developed by Lars Onsager (Sidebar 13.3).

The “Onsager coefficients” $(\underline{\underline{L}})_{ij}$ are here evaluated in terms of the real symmetric matrix

$$(\underline{\underline{L}})_{ij} \equiv \sum_{k=1}^f \frac{D_k}{|e_k|} (\underline{q}_k)_i (\underline{q}_k)_j \quad (13.43)$$

whose elements satisfy the obvious index symmetry

$$(\underline{\underline{L}})_{ij} = (\underline{\underline{L}})_{ji}, \quad i, j = 1, 2, \dots, f \quad (13.44)$$

Equations (13.44) are the celebrated *Onsager reciprocity relations* [L. Onsager. Reciprocal relations in irreversible processes. *Phys. Rev.* **37**, 405–26; **38**, 2265–79 (1931)] that relate off-diagonal $\vec{F}_j \cdot \vec{J}_i$ and $\vec{F}_i \cdot \vec{J}_j$ coupling and serve as a cornerstone of modern transport theory. They are derived here without reference to statistical mechanical or fluctuation assumptions.

As may easily be confirmed from (13.42), the Onsager matrix $\underline{\underline{L}}$ is a positive-semidefinite matrix satisfying the eigenvalue equation

$$\underline{\underline{L}} \underline{q}_k = (D_k/|e_k|) \underline{q}_k, \quad k = 1, 2, \dots, f \quad (13.45)$$

Thus, $\underline{\underline{L}}$ has the same eigenvectors $\underline{q}_k$ as the equilibrium metric matrix $\underline{\underline{M}}$ (or $\overline{\underline{M}}$), and its eigenvalues $D_k/|e_k|$ are closely related to those of the conjugate $\overline{\underline{\underline{L}}}$, each being merely “rescaled” by the corresponding diffusion constant D_k (>0) for eigenmode k . Moreover, in the nonsingular region ($\det|\underline{\underline{L}}| \neq 0$) where $\underline{\underline{L}}$ is invertible, with conjugate

$$\overline{\underline{\underline{L}}} \equiv \underline{\underline{L}}^{-1} \quad (13.46)$$

one will have the usual conjugate symmetry between diffusional forces and fluxes. It is therefore clear that $\underline{\underline{L}}$ can be taken as the metric for an abstract geometrical “diffusion space” $\mathcal{M}_{\mathcal{D}}$, with biorthogonal force ($|\mathbf{F}_i\rangle$) and flux ($|\mathbf{J}_j\rangle$) ket vectors satisfying

$$\langle \mathbf{F}_i | \mathbf{F}_j \rangle = (\overline{\underline{\underline{L}}})_{ij} = \left(\frac{\partial F_i}{\partial J_j} \right)_{\underline{L}} \quad (13.47)$$

$$\langle \mathbf{J}_i | \mathbf{J}_j \rangle = (\underline{\underline{L}})_{ij} = \left(\frac{\partial J_i}{\partial F_j} \right)_{\underline{E}} \quad (13.48)$$

$$\langle \mathbf{F}_i | \mathbf{J}_j \rangle = \delta_{ij}, \quad i, j = 1, 2, \dots, f \quad (13.49)$$

perfectly analogous to the equilibrium state space $\mathcal{M}_{\mathcal{S}}$. Of course, it should be unnecessary to emphasize that ket vectors $|\mathbf{F}_i\rangle$, $|\mathbf{J}_j\rangle$ “live” in an entirely different space than do the three-dimensional Cartesian vectors $\vec{F}_i$, $\vec{J}_j$ that were employed in writing field equations such as (13.36). The spaces $\mathcal{M}_{\mathcal{S}}$ and $\mathcal{M}_{\mathcal{D}}$ (having distinct metrics) are also distinct.

Figure 13.2 shows the schematic diffusional flows in a system of varying chemical potential $\mu(z)$. At z_1 , where the gradient force $\partial\mu/\partial z$ is strong in magnitude, diffusional flow (arrows) is rapid and in the direction to *oppose* the gradient (“yield” to the force), thereby diminishing the gradient more rapidly than at z_2 . Thus, diffusional flow acts

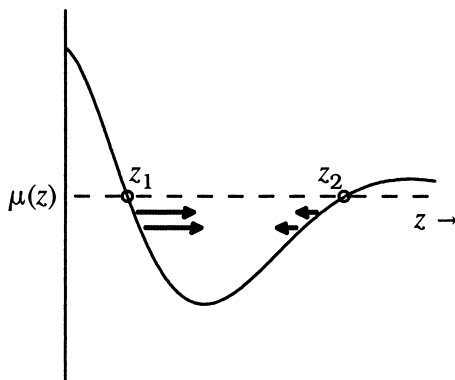


Figure 13.2 Schematic depiction of chemical diffusion (arrows) induced by chemical potential variations at points of steeply negative gradient (z_1) or weakly positive gradient (z_2), suggesting the proportionality (13.25a) between chemical diffusion rate and chemical potential gradient that characterizes the near-equilibrium state.

everywhere to diminish $\mu(t)$ with increasing t , but the t -evolution is faster (consistent with faster diffusional outflow) at points where the μ -gradient is strongest. Analogous relationships between intensive forces and extensive fluxes characterize heat flow, pressure wave attenuation, chemical diffusion, and other transport phenomena.

Recognition of the diffusional metric geometry $\mathcal{M}_D$ allows one to apply powerful vector-geometrical techniques to the derivation of complex identities between measured transport properties, analogous to those discussed in Chapter 12. However, further discussion of the rich science and applications of near-equilibrium transport phenomena is beyond the scope of the present treatment [see, e.g., R. B. Bird, W. E. Stewart, and E. N. Lightfoot. *Transport Phenomena*, revised 2nd edn (Wiley, New York, 2007)].

SIDEBAR 13.3: LARS ONSAGER (1903–76)



Lars Onsager was born in 1903 into a barrister's family in Oslo, Norway. His eclectic interests in music, classical Norwegian literature, horticulture, and other subjects were cultivated by diverse educational experiences from an early age (including an interval of home schooling by his mother). However, his interests turned increasingly to mathematics, spurred by intensive study of Whittaker and Watson's *Course of Modern Analysis*. Years later, colleagues marveled at Onsager's tattered copy of this classic monograph, everywhere bearing the markings of his annotations, generalizations, and extensions.

Following high school, Onsager embarked on chemical engineering studies at the Norwegian Institute of Technology in Trondheim. Onsager's

freshman chemistry lab awakened early interest in electrolyte solutions, particularly the emerging Debye–Hückel theory of ionic activity and conductivity (Sidebar 8.6). The conductance equation seemed to exhibit incorrect concentration dependence, and, based on deep mathematical and physical intuitions, Onsager recognized that the theory failed to exhibit a symmetry demanded by microscopic reversibility. This insight was to foreshadow his Nobel prize-winning work on reciprocal relations.

Fresh from completing studies at Trondheim in 1925 (where his work was judged insufficient to earn a doctoral degree), young Onsager boldly travelled to the Swiss Federal Institute at Zürich to meet with Peter Debye. Arriving unannounced at Debye's office, the tall young visitor strode solemnly to Debye's desk and said, "Professor Debye, your theory of electrolytes is incorrect!" To his credit, Debye invited Onsager to sit down and explain his objections, and was quickly impressed by the genius and accuracy of his young critic. Debye then invited Onsager to return to Zürich as his assistant, where he remained until 1928.

Thereafter, Onsager emigrated to the United States to take up a position at Johns Hopkins University. Although ebullient and well-intentioned, Onsager was utterly incomprehensible to the freshman chemistry students of his first assigned class, and he was summarily fired after one semester. According to his supervisor, "I won't say he was the world's worst lecturer, but he was certainly in contention." Onsager's legendary difficulty in communicating with lesser intellects was to be an enduring feature of his career.

After leaving Johns Hopkins, Onsager found a research instructor position at Brown University, where his seminal papers on the reciprocal relations were completed and published in 1931. However, this work attracted practically no attention at the time, and Onsager's teaching fared no better than before, his single graduate course becoming known as "Sadistical Mechanics." With the deepening of the Depression, Onsager's appointment at Brown was terminated in 1933.

Fortunately, Onsager was then able to procure a postdoctoral appointment at Yale University. However, the Chemistry Department was chagrined to learn that he lacked the PhD degree nominally required for such appointment. It was suggested that he submit any published paper as his "thesis" for formal completion of a Yale PhD, but, not wishing to use previously published work in this manner, Onsager rummaged in his desk for an unpublished draft of a mathematical paper on solutions of the Mathieu equation. This was quickly brought to finished form, but no member of the chemistry or physics faculty felt competent to recommend it for acceptance as a chemistry thesis. Finally, the Mathematics chairman (also an enthusiast for Whittaker and Watson) suggested that his department could happily recommend the degree if Chemistry were unwilling to do so. So the chemists relented and the matter was resolved. Onsager's Yale PhD was awarded in 1935, one year after his appointment as Assistant Professor.

Onsager spent the bulk of his academic career at Yale, becoming J. Willard Gibbs Professor of Theoretical Chemistry in 1945. There he obtained his remarkable closed-form solution of the two-dimensional Ising problem (1942), which revolutionized statistical mechanical understanding of critical phenomena. The importance of this work is indicated by Wolfgang Pauli's remark (to Dutch physicist H. B. G. Casimir, who had been isolated from Allied contacts) that nothing much of scientific interest occurred during World War II, "apart from Onsager's solution of the Ising problem."

The importance of Onsager's earlier work on coupled diffusional processes was also increasingly appreciated. Onsager had based his proof of force-flux reciprocal relations on the physical principle of detailed balance (as suggested by G. N. Lewis), which rules

out “circulating” reaction patterns in kinetic equilibria. By exploiting the expected connection to statistical fluctuations (“a spontaneous deviation from the equilibrium decays according to the same laws as one that has been produced artificially”), Onsager found a deep statistical mechanical justification for reciprocal relations that could be verified in great experimental detail [D. G. Miller. Thermodynamics of irreversible processes: the experimental verification of the Onsager reciprocal relations. *Chem. Revs* **60**, 15–37 (1960)]. This work was subsequently recognized by the award to Onsager of the 1968 Nobel Prize in Chemistry, which included mention of his work on electrolyte conductivity, the two-dimensional Ising problem, and quantized vortices in liquid helium as “milestones in the development of science,” but reserved special honor for discovery of the reciprocal relations, worthy to be recognized as a new thermodynamic law and “one of the great advances in science during this century.”

As Onsager’s fame grew, the jocular genius continued to generate a trail of anecdotes among bemused friends and colleagues. His regular Yale teaching assignments in statistical mechanics were popularly known as “Advanced Norwegian I, II.” Long-time colleague Joseph Hubbard, recalling the spectacle of Onsager crawling under the desk in his chaotic office to locate a misplaced 400-page thesis (and two-month old paycheck!), was moved to remark:

Here’s a fellow who scratches his left ear by reaching round the back of his head with his right hand. I wonder how he ties his shoes!

Onsager was also “an appalling correspondent,” said to take his letters out of the mailbox one by one, glance at them, and tear them up. A bewildered journal editor recalled the frustration of repeatedly prompting Onsager for an overdue referee’s report, only to finally receive the one-word reply: “Somehow.”

Onsager’s death in 1976 in Coral Gables, Florida (where he had accepted a distinguished professorship after involuntary retirement from Yale) brought a final chapter to his edgy rivalry with fellow Yale theorist J. G. Kirkwood. Onsager’s body was brought back to New Haven’s Grove Street Cemetery to be interred next to Kirkwood’s tombstone, which bore a long list of Kirkwood’s positions and honors, including the American Chemical Society Award in Pure Chemistry, the Richards Medal, and the Lewis Award. In contrast, Onsager’s tombstone inscription read simply “Nobel Laureate.” (The inscription was later altered by the family to include an asterisk and footnote “*etc.” in the lower-right corner.)

[C. Longuet-Higgins and M. E. Fisher. Lars Onsager 1903–1976. *Biographical Memoirs of the US National Academy of Sciences* (Washington DC, 1991), pp. 181–231.]

13.3 QUANTUM STATISTICAL THERMODYNAMIC ORIGINS OF CHEMICAL AND PHASE THERMODYNAMICS

An overarching goal of modern physical chemical studies is the *molecular*-level understanding of chemical and phase thermodynamics—a molecular theory of gases, liquids, and polymorphic solid phases of chemically reactive mixtures. Whereas thermodynamic sciences originated in an era when understanding of atomic and molecular phenomena

was limited by erroneous classical Newtonian conceptions, we now have a confident picture of the microscopic domain based on the rigorous foundations of quantum mechanics. Indeed, direct numerical solutions of Schrödinger's wave equation (of adequate chemical accuracy) are now recognized as valid alternatives to traditional experimental investigations of atomic and molecular properties. As a glance at modern research journals will show, *ab initio* (first-principles) theoretical methods increasingly permeate research investigation of valency, bonding, and aggregation phenomena in all areas of chemical, biochemical, and materials sciences. Accordingly, in this concluding section we wish to briefly survey some *microscopic* aspects of thermodynamic phenomena, focusing on molecular-level statistical mechanical and quantum statistical thermodynamical origins of thermodynamic geometry.

As Chapters 8–12 make clear, “chemical and phase thermodynamics” implies a richly detailed geometrical and dimensional structure ($\mathcal{M}_S$), dictated by the chemical components (c) and phases (p) of the system in question. Modern molecular understanding of thermodynamic geometry must therefore go considerably beyond accounting for simple convexity of a model free energy function, or verifying numerical stability of certain statistical descriptors of a model molecular dynamics trajectory. A noted statistical mechanician was fond of saying, “Thermodynamics tell you nothing,” but this statement reflects considerable underestimation of thermodynamics as well as the associated phenomena that statistical mechanical models should aspire to explain. Oversimplified models of Ising or lattice-gas type seem to offer little real prospect of understanding the chemical subtleties of observed thermodynamic behavior.

We now recognize—as scientists at the time of Clausius, Gibbs, and Boltzmann could not—that molecular-level phenomena are governed by *electronic* interactions of quantum mechanical nature. The electron mass is *thousands* of times smaller than total molecular masses, and the associated quantum effects (as reflected, for example, in the de Broglie wavelength or energy level spacings) are enormously larger than those for nuclei. The concept of the “classical limit” therefore has little real relevance for thermodynamic behavior that depends significantly on electronic interactions. This fact is well recognized for chemical reaction equilibria, but modern *ab initio* evidence suggests that many *phase* properties are dictated by electronic exchange-type interactions of chemical or “resonance-type” origin. Accordingly, we shall focus attention on a molecular-level description that incorporates realistic electronic structure effects and may (in principle or practice) be evaluated by modern electronic structure methods. Only in this quantum mechanical framework can we hope to achieve a realistic picture of chemical and phase thermodynamics that adequately describes the full (c, p) dependence and nonideality of actual chemical systems.

It is noteworthy that Gibbs himself was acutely aware of the qualitative failures of 19th-century molecular theory (as revealed, for example, by erroneous classical predictions of heat capacities; Sidebar 3.8). In the preface to his *Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics* (published in the last year of his life), Gibbs wrote:

In the present state of science, it seems hardly possible to frame a dynamic theory of molecular action which shall embrace the phenomena of thermodynamics, of radiation, and of the electrical manifestations which accompany the union of atoms . . . Even if we confine our attention to phenomena distinctively thermodynamic, we do not escape difficulties in as simple a matter as the number of degrees of freedom of a diatomic gas . . . Certainly, one is building on an insecure foundation, who rests his work on hypotheses concerning the constitution of matter.

Difficulties of this kind have deterred the author from attempting to explain the mysteries of nature, and have forced him to the more modest aim of deducing some of the more obvious propositions relating to the statistical branch of mechanics. Here, there can be no mistake in regard to the agreement of the hypotheses with the facts of nature, for nothing is assumed in that respect. The only error into which one can fall, is the want of agreement between the premises and the conclusions, and this, with care, one may hope, in the main, to avoid.

As indicated, Gibbs warily *averted* molecular dynamic assumptions to formulate an alternative *ensemble-based* reformulation of statistical mechanics that was able to seamlessly survive the revolutionary changes of 20th-century quantum theory, much to the approval of Einstein (see Sidebar 5.1) and others in the forefront of that revolution (see, e.g., Schrödinger’s statement, Sidebar 13.4).

In the following, we first describe (Section 13.3.1) a statistical mechanical formulation of Mayer and co-workers that anticipated certain features of thermodynamic geometry. We then outline (Section 13.3.2) the standard quantum statistical thermodynamic treatment of chemical equilibrium in the Gibbs canonical ensemble in order to trace the statistical origins of metric geometry in Boltzmann’s probabilistic assumptions. In the concluding two sections, we illustrate how modern *ab initio* molecular calculations can be enlisted to predict thermodynamic properties of chemical reaction (Sections 13.3.3) and cluster equilibrium mixtures (Section 13.3.4), thereby relating chemical and phase thermodynamics to a modern *ab initio* electronic structure picture of molecular and supramolecular interactions.

SIDEBAR 13.4: DYNAMICAL-VERSUS ENSEMBLE-BASED STATISTICAL THERMODYNAMICS

Although the branch of science known as statistical thermodynamics originally evolved from Boltzmann’s statistical analysis of the envisioned dynamics of atoms and molecules (then considered controversial; cf. Sidebar 13.7), Gibbs suggested a fruitful *ensemble-based* alternative that we adopt in the present work. Differences between these viewpoints, the meaning of “ensemble,” and apt reasons for adopting the ensemble-based approach are summarized in Schrödinger’s 1949 introduction to *Statistical Thermodynamics*, from which the following is quoted:

... there are two different attitudes as regards the physical application of the mathematical result [method of the most probable distribution]. We shall later, for obvious reasons, decidedly favour one of them; for the moment, we must explain them both.

The older and more naive application is to N actually existing physical systems in actual physical interaction with each other, e.g., gas molecules ... The N of them together represent the actual physical system under consideration. This original point of view is associated with the names of Maxwell, Boltzmann and others.

But it suffices only for dealing with a very restricted class of physical systems—virtually only with gases. It is not applicable to a system which does not consist of a great number of identical constituents with “private” energies ...

Hence a second point of view (or, rather, a different application of the same mathematical results, which we owe to Willard Gibbs, has been developed. It has a particular beauty of its own, is applicable quite generally to every physical system, and has some advantages to be mentioned forthwith. Here the N identical systems are mental copies of the one system under consideration—of the one macroscopic device that is actually erected on our laboratory table.

The “method of the most probable distribution” that underlies both Boltzmann-type and Gibbs-type formulations is outlined with admirable clarity in Chapter II of Schrödinger’s book.

[E. Schrödinger. *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1st edn 1946, 2nd edn 1952; Dover reprint of 2nd edn 1989), pp. 2–3].

13.3.1 Nonequilibrium Displacement Variables of Mayer and Co-workers

In 1965, Joseph E. Mayer (Sidebar 13.5) and co-workers published a paper [M. Baur, J. R. Jordan, P. C. Jordan, and J. E. Mayer. Towards a Theory of Linear Nonequilibrium Statistical Mechanics. *Ann. Phys. (NY)* **65**, 96–163 (1965)] in which the vectorial character of the thermodynamic formalism was suggested from a statistical mechanical origin. Although this paper attracted little attention at the time, its results suggest how thermodynamic geometry might be traced to the statistics of quantum mechanical phase-space distributions.

Mayer et al. begin with the fundamental equation in the Boltzmann-reduced entropy representation

$$s = -S/k = s(U, V, n_1, \dots, n_c) = s(\{X_i\}) \quad (13.50)$$

where k is Boltzmann’s constant. The key to their approach is a statistical mechanical expression for each extensive variable X_i (except V):

$$X_i = \mathfrak{S} W \chi_i \quad (13.51)$$

where $\mathfrak{S}$ is an integration operator that sums over particle numbers and integrates over all coordinates in phase space $p^{(N)}, q^{(N)}$, W is an ensemble probability distribution, and $\chi_i = \chi_i(p^{(N)}, q^{(N)})$ is a suitable function of the phase-space coordinates for an N -particle system. Here, “phase space” refers to the combined t -dependent position ($\vec{r}_i$) and momentum ($\vec{p}_i$) coordinates that specify the trajectories of each particle ($i = 1, 2, \dots, N$) of a many-particle system ($N \sim 10^{23}$).

The positive-definiteness of second derivatives of s ,

$$s_{ii} = (\partial^2 s)/(\partial X_i^2) \geq 0 \quad (13.52)$$

is used to write vector-type dot products of the form

$$\mathbf{i}_\alpha \cdot \mathbf{i}_\beta = s_{\alpha\beta} \quad (13.53)$$

where the $\mathbf{i}_\alpha$ are described as “basis vectors in a space of as many dimensions as there are values of α .” Many details of the association between the equilibrium equations (13.52) and

(13.53) are incompletely developed. Indeed, the motivation (and notation) for the vector-type equation (13.53) is more deeply tied to the statistical mechanical expression that is developed for relating phase-space dot products $F \cdot G$ to the corresponding thermodynamic ensemble averages $\langle F \cdot G \rangle$, namely,

$$\langle F \cdot G \rangle = N_A^{-1} \mathfrak{S} W_{\text{eq}} F \cdot G \quad (13.54)$$

Because $\mathfrak{S}$ is a linear operator and W_{eq} is a non-negative function, the expression on the right can be shown to have all the necessary properties of a scalar product, provided that W_{eq} and the phase-space functions F, G appropriate to each thermodynamic variable can be defined and the integral exists. Equation (13.53) is then treated as the equilibrium dot product “inherited” from the phase-space vectors and dot product in (13.54).

Mayer et al. interpret the vectors defined through (13.53) in terms of fluctuating displacements from the equilibrium state [see, e.g., H. B. Callen. *Thermodynamics* (Wiley, New York, 1960), Chap. 15], related to Onsager’s original treatment of the near-equilibrium limit. Displacements of the extensive variables from equilibrium are represented as coefficients of the vectors $\mathbf{i}_\alpha$, while those of the intensive variables are coefficients of the “reciprocal” (conjugate) set of vectors. (“Reciprocal” in this sense refers to the Fourier-transform relationship between $\vec{r}$ -space and $\vec{p}$ -space.) Attention is thereby drawn to the “orthonormal” variables (corresponding to the self-conjugate eigenmodes of Section 11.6), which cause the fluctuations to become uncoupled and the phenomenological Onsager flux-gradient matrix of their dynamical evolution to be diagonalized. Small displacements of these independent orthonormal variables from equilibrium are considered to decay independently with a single exponential factor $e^{-t/\tau}$.

In the treatment of Mayer et al., it is the independent displacement variables that come to the fore, equations such as (13.53) being employed primarily to establish the desirable properties of these variables for treating spontaneous decay of nonequilibrium displacements by statistical mechanics. The emphasis throughout is statistical mechanics rather than thermodynamics. [For example, the authors write (p. 113): “Were our interest limited to equilibrium thermodynamics the analogy between the intensive and extensive variables and the two reciprocal basis vector sets might be considered good clean mathematical fun, but of little scientific value.”] Given this emphasis, it is not surprising that the general thermodynamic significance of this “analogy”—its origin, scope, and consequences—was not pursued. However, their work clearly anticipates important aspects of the metric geometry of equilibrium thermodynamics ($\mathcal{M}_S$) as well as the associated diffusional geometry ($\mathcal{M}_D$) of the near-equilibrium limit, both of which share the same metric eigenvectors (the E_i of Sections 11.6 or 13.2).

The central formula (13.54) of Mayer et al. exhibits an explicit mapping relating the equilibrium geometry and dot products of $\mathcal{M}_S$ to those of an underlying phase-space description of the molecular dynamics. Their work also hints at how the formal mapping might be extended to quantum mechanics, based on the concept of orthonormal functions in phase space (rather than the usual Hilbert space of quantum mechanics). The phase-space orthonormality concept was considered by Mayer as a general means for linking statistical mechanics with quantum mechanics, and the entire logical sequence suggested by (13.54) seems to warrant further study and development in this direction.

SIDEBAR 13.5: JOSEPH E. MAYER (1904–83)



Joseph E. Mayer (although he graduated from Hollywood High School, not the movie mogul of the same name!) was born in New York City into the family of an Austrian immigrant bridge engineer. The family later moved to Canada and California, where Mayer completed undergraduate studies at Caltech (1924) and doctoral studies at the University of California at Berkeley (1927) under G. N. Lewis. After a year of postdoctoral study with Lewis, Mayer was awarded a Rockefeller Fellowship to James Franck's institute in Göttingen to work with Max Born on ionic crystals. There he also met and married a doctoral student in Born's group, Maria Goeppert, who was descended on her father's side from seven successive generations of university professors. Thus was formed one of the most powerful spousal teams in the history of science.

Mayer's academic appointments took the couple first to Johns Hopkins (1931–1939), then to Columbia (1939–1946). While Mayer's academic visibility grew steadily throughout this period (including appointment to editorship of the *Journal of Chemical Physics* in 1940), Maria Goeppert-Mayer was only able to obtain a succession of temporizing positions as research assistant or part-time lecturer, reflecting the anti-nepotism rules, academic sexism, and economic depression of the times. That she persisted "just for the fun of doing physics" was remarkable, but it led to unexpected dividends for chemical physics, for which she partially re-tooled herself under the influence of collaborations with her husband and other chemistry colleagues (including Alfred Sklar, with whom she co-authored an important methodological paper on the molecular orbital theory of benzene). This period also led to an influential collaborative monograph [J. E. Mayer and M. Goeppert-Mayer. *Statistical Mechanics* (Wiley, New York, 1940)] that expounded the powerful "Mayer cluster expansion" technique for virial description of real gases.

The Mayer cluster expansion approach is based on the recognition of the importance of physical "clustering," that is, aggregation into groups of molecules such that certain long-range interactions between molecules in different groups can be neglected. This in turn implies that Boltzmann-weighted potential energy functions factor into "cluster functions" and integrals over these functions into "cluster integrals," which can be further simplified by graphical techniques and gathered in virial-like fashion. Although cluster-based expansions are routinely used for dense gases, Mayer proved rigorously that identical methods are applicable to treating the properties of solutes in a solvent. This is surprising, for in the particular case of an ionic solute, the single-cluster integrals are *infinite*. Nevertheless, by properly summing terms of opposite sign prior to integration, the total sums that have thermodynamic significance were shown to be convergent. (Other extensions of cluster concepts to liquid properties are discussed in Section 13.3.3)

The couple's move to the University of Chicago (1946–1960) somewhat improved Maria's academic appointment [to "voluntary Associate Professor" of Physics at Chicago

and “Senior Physicist (half-time)” at Argonne National Laboratory] and greatly expanded her active involvement with nuclear physics. There she completed the work on the spin–orbit “magic number” nuclear shell model (with J. H. D. Jensen) that led to the Nobel Prize in Physics (1963). Joseph meanwhile held a professorial appointment in the Institute for Nuclear Studies while continuing to direct graduate physical chemistry research on statistical and quantum mechanical equations of state, also serving a term as president of the International Union of Pure and Applied Physics in this period.

In 1960, the Mayers moved to the University of California at San Diego, where both could finally hold full professorships, his in chemistry and hers in physics. But shortly after arrival, Maria suffered a stroke and was troubled thereafter by poor health, succumbing finally (after a prolonged coma) in 1972. While continuing collaborative theoretical chemistry studies, Joseph Mayer helped to build the newly formed UCSD chemistry faculty to national prominence. He retired and remarried, then accepted a term of presidency of the American Physical Society (1973) and continued other professional activities until his death in 1983.

Mayer was remembered for the intensity of his classroom lectures on statistical mechanics, delivered

... with exemplary clarity and enthusiasm, chalk in one hand and lighted cigarette in the other, never confusing the two ...

Although his scientific contributions were naturally somewhat overshadowed by those of Maria, Joseph Mayer built a legacy of concepts and methods that influence many areas of modern physical chemistry research.

[W. H. Crouse et al. (eds.). *McGraw-Hill Modern Men of Science* (McGraw-Hill, New York, 1966), pp. 319–21; B. Zimm. *Bibliographic Memoirs of the US National Academy of Sciences* **65**, 210–21 (1994).]

13.3.2 Quantum Statistical Thermodynamics and the Statistical Origins of Metric Geometry

Although Boltzmann-style descriptions in terms of time-averaged dynamical trajectories continue to be employed (based on the “ergodic hypothesis” that such long-time dynamic averages coincide with the equilibrium ensemble average), the Gibbsian ensemble-based approach is the generally preferred method for treating chemical reaction equilibria (see Sidebar 13.4)



[Indeed, computer-implemented “molecular dynamics” simulations are usually based on model potential energy functions that cannot describe chemical bond-breaking processes, and are thus incapable of dealing realistically with (13.55).] In this section, we wish to briefly outline the formal quantum statistical thermodynamic theory of chemical reaction thermodynamics in the Gibbs canonical ensemble and examine the probabilistic origins of thermodynamic geometry. *Ab initio* numerical evaluations of thermodynamic partition functions for chemical reaction and cluster mixtures with modern *ab initio* electronic structure methods are described in the concluding sections.

In the canonical (T, V, N) ensemble, one envisions many replicas of the equilibrium system (13.55), each having the same fixed values of volume V and total particle number N and each coupled to a common thermal reservoir of temperature T . Each copy of the system has the same fixed quantum energy levels E_α , and the average energy E of each copy is fixed, but individual members of the ensemble exhibit a fluctuating *distribution* of energies among these allowed levels, with probability p_α that a given member of the ensemble is in quantum level E_α :

$$\sum_{\alpha} p_{\alpha} E_{\alpha} = E \quad (13.56)$$

Among the many possible energy distributions that can satisfy (13.56) subject to

$$\sum_{\alpha} p_{\alpha} = 1 \quad (13.57)$$

the *most probable distribution* (subject to given values of the control variables V, N, T) is that corresponding to the *Boltzmann probability* expression

$$p_{\alpha} = \frac{e^{-\beta E_{\alpha}}}{Q} \quad (13.58)$$

where Q is the canonical *partition function*,

$$Q = Q(T, V, N) = \sum_{\alpha} e^{-\beta E_{\alpha}} \quad (13.59)$$

and the exponential weighting factor is

$$\beta = 1/kT \quad (13.60)$$

(where k is Boltzmann's constant). Mathematical derivation of the Boltzmann (most probable) distribution (13.58) can be found in standard texts [see, e.g., Schrödinger (Sidebar 13.4) or T. L. Hill. *An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA, 1960; Dover reprint 1986)].

For an equilibrium ideal gas mixture containing n_A molecules of A, n_B molecules of B, and n_C molecules of C, with

$$n_A + n_B + n_C = N \quad (13.61)$$

the total partition function can be shown to factor into contributions Q_i from each reactant:

$$Q = \prod_i Q_i(n_i, V, T) = \left(\frac{q_A^{n_A}}{n_A!} \right) \left(\frac{q_B^{n_B}}{n_B!} \right) \left(\frac{q_C^{n_C}}{n_C!} \right) \quad (13.62)$$

Each molecular partition function q_i may in turn be factored into contributions from translational ($q_{i,\text{trans}}$), vibrational ($q_{i,\text{vib}}$), rotational ($q_{i,\text{vib}}$), and electronic ($q_{i,\text{elec}}$)

energy levels,

$$q_i = q_{i,\text{trans}} q_{i,\text{vib}} q_{i,\text{rot}} q_{i,\text{elec}} \quad (13.63)$$

neglecting vibration–rotation interactions and other small perturbations.

The molecular partition function allows evaluation of the associated chemical potential as

$$\mu_i = -kT \ln q_i \quad (13.64)$$

At equilibrium, the product of stoichiometric coefficient and chemical potential must be equalized for each species,

$$n_A \mu_A = n_B \mu_B = n_C \mu_C \quad (13.65)$$

subject to the reaction stoichiometry constraints, for example, for (13.55),

$$n_B = n_C \quad (13.66)$$

Equation (13.64) corresponds to the basic identity relating the canonical partition function to the Helmholtz free energy:

$$A = A(T, V, n_A, n_B, n_C) = -kT \ln Q \quad (13.67)$$

From this, one obtains the pressure P ,

$$P = kT \left(\frac{\partial \ln Q}{\partial V} \right)_{T, \underline{n}} = - \left(\frac{\partial A}{\partial V} \right)_{T, \underline{n}} \quad (13.68)$$

the chemical potential for each species, the Gibbs free energy $G = A + PV$, and so forth. Equations (13.61)–(13.68) allow the equilibrium populations n_A , n_B , n_C of each species to be obtained from known values of the molecular partition functions q_A , q_B , q_C , whose evaluation will be discussed in Section 13.3.3.

Of particular interest in the present context is the entropy $S = -(\partial A / \partial T)_{V, \underline{n}}$, which can be shown to satisfy the important identity (see Sidebar 13.6)

$$S = -k \sum_{\alpha} p_{\alpha} \ln p_{\alpha} \quad (13.69)$$

We note in passing that many alternative ensembles [with other control variables, such as the microcanonical (U , V , N) and grand canonical (μ , V , T) ensembles] might have been considered, all leading to the same predicted values of standard thermodynamic properties, but different fluctuation properties and other higher derivatives. Thus, there is no implied special role of the Helmholtz free energy, except as the natural thermodynamic potential for the control variables T , V , N of the chosen canonical ensemble. In particular, we are free to focus on the *entropy* expression (13.69) to investigate the statistical mechanical origin of the S -based metric geometry, as employed in preceding sections.

For this purpose, let us first rewrite (13.69) in terms of reduced *negentropy* contributions σ_α :

$$\sigma \equiv -S/k = \sum_{\alpha} p_{\alpha} \ln p_{\alpha} = \sum_{\alpha} \sigma_{\alpha} \quad (13.70a)$$

where

$$\sigma_{\alpha} = p_{\alpha} \ln p_{\alpha} \quad (13.70b)$$

The important aspect of (13.70b) is that each $p_{\alpha} = p_{\alpha}(U, V, N)$ has *maximal* (“most probable”) character with respect to the natural control variables of S . The constrained maximization procedure to find this optimal distribution by the method of Lagrange undetermined multipliers [see Schrödinger (1949), Sidebar 13.4, for further details] is very similar to that described in Section 5.2. In particular, the p_{α} must be *maximal* with respect to variations in each control variable, leading to the usual second-derivative curvature conditions such as

$$\left(\frac{\partial^2 p_{\alpha}}{\partial U^2} \right)_{V,N} < 0, \quad \left(\frac{\partial^2 p_{\alpha}}{\partial V^2} \right)_{U,N} < 0, \quad \left(\frac{\partial^2 p_{\alpha}}{\partial N^2} \right)_{U,V} < 0 \quad (13.71)$$

analogous to (5.21a–c). Requiring such maximal character for arbitrary linear combinations of the extensive control variables (i.e., envisioning alternative ensembles in which such linear combinations are the control variables) leads similarly to the symmetric and distributive properties of a Euclidean scalar product. Thus, we can associate with each allowed microstate α of the system a metric geometry $\mathcal{M}_{\alpha}$ with scalar products

$$\langle X_i | X_j \rangle_{\alpha} = - \left(\frac{\partial^2 p_{\alpha}}{\partial X_i \partial X_j} \right)_{\underline{X}} \quad (13.72)$$

At this deeper level, *Boltzmann probabilities are geometry!*

Let us now see how the metric geometrical properties of Boltzmann probabilities are “inherited” by the final thermodynamic properties. From (13.70a, b), we can write for σ and its first and second differentials

$$\sigma = \sum_{\alpha} p_{\alpha} \ln p_{\alpha} \quad (13.73a)$$

$$d\sigma = \sum_{\alpha} [dp_{\alpha} + \ln p_{\alpha} dp_{\alpha}] \quad (13.73b)$$

$$d^2\sigma = \sum_{\alpha} [d^2 p_{\alpha} + (dp_{\alpha})^2 + \ln p_{\alpha} d^2 p_{\alpha}] \quad (13.73c)$$

However, from the “probability conservation” constraints

$$\sum_{\alpha} p_{\alpha} = 1 \quad (13.74a)$$

$$\sum_{\alpha} dp_{\alpha} = 0 \quad (13.74b)$$

$$\sum_{\alpha} d^2 p_{\alpha} = 0 \quad (13.74c)$$

one can recognize that the first two terms on the right-hand side of (13.73c) must vanish, so that

$$d^2 \sigma = \sum_{\alpha} \ln p_{\alpha} d^2 p_{\alpha} \quad (13.75)$$

where

$$\ln p_{\alpha} \leq 0, \quad \text{all microstates } \alpha \quad (13.76)$$

From (13.71) and (13.76), we are therefore assured that

$$\left(\frac{\partial^2 \sigma}{\partial X_i^2} \right)_{\underline{X}} = \sum_{\alpha} \ln p_{\alpha} \left(\frac{\partial^2 p_{\alpha}}{\partial X_i^2} \right)_{\underline{X}} \geq 0 \quad (13.77)$$

which is the positive curvature that underlies the S -based metric geometry.

The generality of (13.77) rests on the association of probabilistic character with each allowed energy microstate α of the chemical system. Such probabilistic character is a deep feature of quantum mechanical description, but its introduction into classical mechanics by Ludwig Boltzmann's revolutionary statistical hypotheses marked a turning point in physical theory (Sidebar 13.7). The time-directionality of thermodynamic evolution introduced by Boltzmann's statistical assumptions [as captured in (13.77)] was recognized by Maxwell and others to be fundamentally inimical to classical dynamical theory, despite Boltzmann's many efforts to establish a consistent " H -theorem" within that framework. The deeply statistical nature of microscopic phenomena was of course fully exposed with the discovery of quantum mechanics, mooted the controversies surrounding Boltzmann's "atomistic" and "statistical" assumptions (neither of which appeared compatible with then-current understanding of physics).

In retrospect, one can see that Boltzmann's inspired conjecture (13.69) served, through (13.77), to anticipate an essential feature of the probabilistic quantum description that was to supplant classical determinism in the 20th century. Boltzmann and his followers specifically captured the key probabilistic feature (13.69) that could bring proper metric geometrical character (13.77) to macroscopic-level thermodynamic description, despite gross errors of then-current microscopic dynamical theory.

SIDEBAR 13.6: PROOF OF THE ENTROPY IDENTITY $S = -k \sum_{\alpha} p_{\alpha} \ln p_{\alpha}$

Problem Prove that

$$S = -k \sum_{\alpha} p_{\alpha} \ln p_{\alpha}$$

Solution From the basic definitions of Boltzmann's p_{α} and Q , we know that

$$Q p_{\alpha} = e^{-\beta E_{\alpha}}, \quad \text{all } \alpha \quad (\text{S13.6-1})$$

or, equivalently,

$$E_\alpha = -kT \ln(Qp_\alpha) \quad (\text{S13.6-2})$$

Starting from $A = -kT \ln Q$ and basic thermodynamic definitions, we obtain

$$S = -\left(\frac{\partial A}{\partial T}\right)_V = k\left(\frac{\partial(T \ln Q)}{\partial T}\right)_V = k\left[\ln Q + T\left(\frac{\partial \ln Q}{\partial T}\right)_V\right] \quad (\text{S13.6-3})$$

From the differential dQ of $Q = \sum_\alpha e^{-\beta E_\alpha}$, with help from (S13.6-1),

$$dQ = \sum_\alpha e^{-\beta E_\alpha}(-\beta dE_\alpha - E_\alpha d\beta) = \sum_\alpha Qp_\alpha(-\beta dE_\alpha - E_\alpha d\beta) \quad (\text{S13.6-4})$$

we find

$$\frac{dQ}{Q} = d \ln Q = -\beta \sum_\alpha p_\alpha dE_\alpha - \left(\sum_\alpha p_\alpha E_\alpha\right) d\beta \quad (\text{S13.6-5})$$

Noting that the energy levels E_α are independent of T , we obtain the desired logarithmic derivative

$$\left(\frac{\partial \ln Q}{\partial T}\right)_V = -\left(\frac{\partial \beta}{\partial T}\right)_V \sum_\alpha p_\alpha E_\alpha = \frac{1}{kT^2} \sum_\alpha p_\alpha E_\alpha \quad (\text{S13.6-6})$$

and, using (S13.6-2) and (S13.6-3), we obtain finally

$$\begin{aligned} S &= k \ln Q + kT \left(\frac{1}{kT^2} \sum_\alpha p_\alpha E_\alpha \right) \\ &= k \ln Q - k \sum_\alpha p_\alpha \ln(Qp_\alpha) \\ &= k \ln Q - k \ln Q \left(\sum_\alpha p_\alpha \right) - k \sum_\alpha p_\alpha \ln p_\alpha \\ &= -k \sum_\alpha p_\alpha \ln p_\alpha \end{aligned} \quad (\text{S13.6-7})$$

to establish the desired identity—*QED*.

SIDEBAR 13.7: LUDWIG BOLTZMANN (1844–1906)

Ludwig Boltzmann was born into the family of a Viennese tax collector and educated at his parent's home by private tutor until high school. After completing his PhD (on the kinetic theory of gases, with Josef Stefan) at Vienna, he became full Professor of Mathematical



Physics (at age 25!) at the University of Graz. His ongoing studies with Bunsen, Kirchhoff, Helmholtz, and others initially drew him away from Graz, for a time (1873–1876) as Professor of Mathematics at Vienna. However, he returned to Graz to marry his fiancée (who called Ludwig her “sweet fat darling”) and become Professor of Experimental Physics, where his students included Svante Arrhenius and Walther Nernst. His 14 years in Graz were the happiest and most productive of his career (despite being saddled briefly with university presidency). It was here that his statistical theories and support for atomic and molecular concepts were developed—both highly controversial.

In 1890, Boltzmann moved to the Chair of Theoretical Physics at Munich, and three years later to the corresponding chair at Vienna. There his advocacy of the atomic hypothesis brought him into fierce conflict with physicist/philosopher

Ernst Mach. Pugnacious and sharp-tongued, Boltzmann gave as well as he got in spirited disputes with Mach and other personal and professional colleagues. Boltzmann himself recognized his dangerous tendency toward mood swings from elation to severe depression, joking that he was born on the cusp between Mardi Gras and Ash Wednesday. His student Lise Meitner later chronicled his unsuccessful suicide attempts long before his ultimately successful effort in 1906.

Boltzmann also swung through a bewildering succession of academic switches between Vienna and Leipzig, until personal intervention by Emperor Franz Josef secured his final return to Vienna, where he was allowed to take over the philosophy course of his retired nemesis Mach. Boltzmann became a wildly popular teacher, with audiences for his natural philosophy lectures overflowing the largest halls available, culminating in a royal reception at the Emperor’s palace. Boltzmann’s scientific ideas were also gradually gaining the upper hand, although scientific and philosophical controversies (as well as personal health issues) continued to weigh on his mind as he struggled to re-work and refine his positions. Brilliant vindication of the atomic hypothesis lay just on the horizon, as Perrin, Einstein, and others were shortly to demonstrate. Whatever the causes, Boltzmann’s tragic suicide (at age 62, while vacationing with his wife and daughter in Italy) erased one of the pivotal figures of pre-quantum physics.

Boltzmann’s tombstone in Vienna bears the famous formula “ $S = k \log W$ ” ($W = \textit{Wahrscheinlichkeit}$ —probability) that was a signature of his audacious concepts. The alternative formula (13.69) (which reduces to Boltzmann’s in the limit of equal *a priori* probabilities p_α) was ultimately developed by Gibbs, Shannon, and others in a more general and productive way (see Sidebar 13.4), but the key step of employing probability to trump Newtonian determinism was his. Boltzmann was long identified with efforts to establish the “*H*-theorem” and “Boltzmann equation” within the context of classical mechanics, but each such effort to justify the second law (or existence of atoms) in the strict framework of Newtonian dynamics proved futile. Boltzmann’s deep intuition to elevate probability to a primary physical principle therefore played a key role in efforts to find improved foundation for atomic and molecular concepts in the pre-quantum era.

[For an engaging account of Boltzmann's scientific and personal life, see D. Lindley. *Boltzmann's Atom: The Great Debate that Launched a Revolution in Physics* (Free Press, New York, 2001).]

13.3.3 Evaluation of Molecular Partition Functions for Reactive Mixtures

The information needed to evaluate the molecular partition functions q_i , (13.63), may in principle be obtained from experimental spectroscopic measurements or theoretical calculations on each molecule i . Each "type" of energy contribution to q_i (translational, rotational, vibrational, electronic) in principle requires associated quantum energy levels ($\epsilon_{i\alpha,\text{type}}$):

$$q_{i,\text{type}} = \sum_{\alpha} \exp\left(\frac{-\epsilon_{i\alpha,\text{type}}}{kT}\right) \quad (13.78)$$

In practice, the energy spacings $\Delta\epsilon_{i,\text{type}}$ (if sufficiently small) and temperature T (if sufficiently high) may allow use of classical high- T /continuum approximations to the actual quantum level distribution, by replacing the discrete quantum summation (13.78) with a classical integral. [For further discussion, see D. A. McQuarrie. *Statistical Mechanics* (Harper & Row, New York, 1976).]

The translational partition function $q_{i,\text{trans}}$ is evaluated in the high- T continuum limit as

$$q_{i,\text{trans}} = \frac{V - V_{\text{excl}}}{\Lambda_i^3} \quad (13.79a)$$

where the thermal de Broglie wavelength Λ_i is given by

$$\Lambda_i \equiv \frac{h}{(2\pi m_i kT)^{1/2}} \quad (13.79b)$$

and where h is Planck's constant and m_i the mass of the molecular species. The numerator of (13.79a) is the available volume for free translational motion, obtained by subtracting from external V an "excluded" volume occupied by the molecular species (akin to the Van der Waals b parameter discussed in Section 2.4.2):

$$V_{\text{excl}} = \sum_{i=1}^c n_i V_i \quad (13.79c)$$

The molecular volume V_i of each species (considered impenetrable) may be estimated from empirical Van der Waals radii or other theoretical measures of molecular bulk, and all species are considered to have access to the same "free" volume in the numerator. However, V_{excl} is often neglected entirely for gaseous reactions at low density, and approximations to $q_{i,\text{trans}}$ are generally found to have little effect on overall ΔG_{rxn} or K_{eq} .

The rotational partition function $q_{i,\text{rot}}$ can be similarly evaluated in the continuum limit as

$$q_{i,\text{rot}} = \frac{1}{\sigma_{\text{rot}}} \left(\frac{\pi T^3}{\Theta_a \Theta_b \Theta_c} \right)^{1/2} \quad (13.80a)$$

where σ_{rot} is the rotational symmetry number (determined from the symmetry of the calculated equilibrium structure) and Θ_a , Θ_b , Θ_c are “rotational temperatures” with respect to the three principal rotation axes, determined from calculated moments of inertia I_a , I_b , I_c , with, for example,

$$\Theta_a \equiv \frac{h^2}{8\pi^2 k I_a} \quad (13.80b)$$

The classical continuum approximation (13.80) becomes questionable for light nuclei, where inertial moments are reduced and quantum rotational spacings proportionally increased, and in this case the quantum sum over angular momentum states may be substituted. Furthermore, the treatment assumes sufficient free volume for unhindered rotations, and is therefore only appropriate at the lower-density conditions of gaseous reactions.

For the vibrational term $q_{i,\text{vib}}$, a classical high- T continuum approximation is seldom valid, and evaluation of the discrete sum over states is therefore required over the quantum vibrational distribution. (As pointed out in Sidebar 5.13, accurate treatment of molecular vibrations is crucial for accurate assessment of entropic contributions to ΔG_{rxn} .) A simple quantum mechanical model of molecular vibrations is provided by the harmonic oscillator approximation for each of the $3N - 6$ normal modes of vibration of a nonlinear polyatomic molecule of N atoms (cf. Sidebar 3.8). In this case, the quantum partition function can be evaluated analytically as

$$q_{i,\text{vib}} = \prod_{j=1}^{3N-6} \left(1 - e^{-\theta_j/T} \right)^{-1} \quad (13.81a)$$

where θ_j is the “vibrational temperature” of mode j ,

$$\theta_j = \frac{h\nu_j}{k} \quad (13.81b)$$

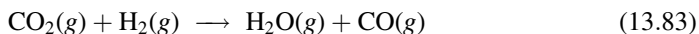
and ν_j is the associated harmonic vibrational frequency. The harmonic frequencies are determined from analytic second derivatives of the potential energy surface at the optimized equilibrium geometry of the species (cf. Sidebar 2.8). Although the harmonic approximation is adequate for higher-frequency modes at lower T ($h\nu_j \gg kT$), its accuracy deteriorates seriously for low-frequency modes (or higher T) where anharmonic corrections are required.

Finally, for the electronic contribution $q_{i,\text{elec}}$, it is generally only necessary to consider the first term (contribution from electronic ground-state energy level E_0) in the quantum sum over states, namely,

$$q_{i,\text{elec}} = \exp \left(-\frac{(E_0 - E_{\text{ref}})}{kT} \right) \quad (13.82)$$

because higher electronic states contribute negligibly for the usual T range of interest. In this equation, E_{ref} is a suitable reference energy (corresponding, for example, to separated electrons and nuclei at rest, the usual “zero of energy” for electronic structure calculations) that cancels in the overall chemical reaction. In principle, the ground-state electronic energy E_0 of species i should contain corrections due to the density-dependent average environment of other species present in the reaction mixture, but such higher-order corrections can be neglected in an ideal gas mixture of reactive gaseous species as will be considered here.

To give a specific example of the application of these equations, let us consider the “water gas shift reaction”



which is involved in the chemistry of catalytic converters. The evaluation of the reaction equilibrium constant K_{eq} from experimental data was previously described by T. L. Hill [*An Introduction to Statistical Thermodynamics* (Addison-Wesley, Reading, MA, 1960; Dover reprint 1986), pp. 187ff.]. We shall therefore focus on the corresponding evaluation by purely theoretical electronic structure methods. For this purpose, we employ a computational program [M. Frisch et al. *Gaussian 03* (Gaussian Inc., Wallingford, CT, 2004)] and theoretical method (labeled “B3LYP/6-311++G**” in the arcane nomenclature of computational quantum chemists) that are widely used throughout the modern chemical research literature. [For further background on quantum chemical methods and their applications to chemical phenomena, see J. B. Foresman and A. Frisch. *Exploring Chemistry with Electronic Structure Methods*, 2nd edn (Gaussian Inc., Pittsburgh, PA, 1993); F. Weinhold and C. R. Landis. *Valency and Bonding* (Cambridge University Press, Cambridge, 2005).]

Table 13.1 summarizes the computed geometrical, rotational, vibrational, and electronic energy properties (each specified in common units for the associated spectroscopy) pertaining to the water gas shift reaction, obtained from B3LYP/6-311++G** optimized geometries and analytic harmonic frequencies of each species. Table 13.2 displays the corresponding T -dependent Gibbs free energies G of each reaction species for $T = 298\text{K}$, 900K , 1200K . Table 13.3 shows the corresponding equilibrium constants and free energy values, comparing the purely theoretical (“B3LYP”) values with corresponding empirically based statistical thermodynamic (“Hill”) and directly measured (“Exp.”) values. As shown in the latter table, the gas-phase equilibrium constant is predicted to

TABLE 13.1 Calculated Molecular Properties (B3LYP/6-311++G** Level) for Reactants and Products of the Water Gas Shift Reaction, Showing Optimized Geometrical Bond Lengths and Angles, Rotational Moments of Inertia, Vibrational Frequencies, and Ground-State Energy E_0

Species	Geometry		Rotational Moment of Inertia (GHz)			Vibrational Frequency ν_j (cm ⁻¹)	Ground-State Energy E_0 (a.u.) ^a
	Bond Length (Å)	Bond Angle (°)					
			I_a	I_b	I_c		
H ₂ O	0.9621	105.05	823.8547	430.8613	282.5612	3922, 3817, 1603	−76.45853
CO	1.277		57.9638	57.9638	0.0000	2212	−113.34905
CO ₂	1.1608	180.00	11.7245	11.7245	0.0000	2420, 1373, 669, 699	−188.64691
H ₂	0.7440		1811.8181	1811.8181	0.0000	4420	−1.17957

^a 1 a.u. = 2626.50 kJ mol^{−1}.

TABLE 13.2 Calculated Gibbs Free Energy (B3LYP/6-311++G** Level) of Water Gas Shift Reaction Species [(13.83); cf. Table 13.1] for Various Temperatures (at $P = 1$ atm)

Species	Gibbs Free Energy (a.u.) ^a		
	298K	900K	1200K
H ₂ O	−76.45489	−76.50334	−76.53010
CO	−113.36313	−113.41284	−113.43977
CO ₂	−188.65592	−188.71127	−188.74228
H ₂	−1.18099	−1.21521	−1.23429

^a 1 a.u. = 2626.50 kJ mol^{−1}.**TABLE 13.3** T -dependent Equilibrium Constant (K_T), Gibbs Free Energy of Reaction (ΔG_T), and Overall Entropic Shift ($\Delta\Delta G \equiv \Delta G_{1200} - \Delta G_{900}$) for the Water Gas Shift Reaction (cf. Tables 13.1, 13.2, and Text), as Determined from Theoretically (“B3LYP”) or Empirically (“Hill”) Evaluated Statistical Thermodynamic Formulas Versus Experiment (“Exp.”)

	B3LYP	Hill	Exp.
K_{298}	2×10^{-9}	—	—
K_{900}	0.027	0.45	0.46
K_{1200}	0.172	1.41	1.37
ΔG_{900} (kJ mol ^{−1})	27.0	6.0	5.8
ΔG_{1200} (kJ mol ^{−1})	17.6	−3.4	−3.1
$\Delta\Delta G$ (kJ mol ^{−1})	−9.5	−9.4	−9.0

be quite unfavorable at room temperature, but to rise to appreciable values in the 900–1200K range. One can see from the tabulated comparisons that the B3LYP/6-311++G** calculation systematically underestimates the overall reaction exothermicity (by about 20 kJ mol^{−1}), but the overall T -dependent entropic shift ($\Delta\Delta G = \Delta G_{1200} - \Delta G_{900}$) is reproduced fairly well at this level. Many features of the electronic structure calculation could be readily improved (for example, by standard vibrational scaling, anharmonic corrections, complete basis set extrapolation, or more complete description of electron correlation). However, the tabulated results serve to illustrate the now-routine possibilities for predicting chemical reaction thermodynamics from a first-principles microscopic viewpoint in the framework of the statistical thermodynamic formulas (13.79)–(13.82).

13.3.4 Quantum Cluster Equilibrium Theory of Phase Thermodynamics

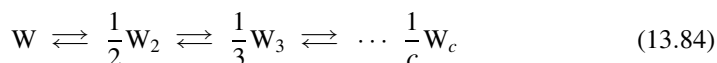
The statistical thermodynamics of phase equilibria involves more subtle aspects of molecular behavior than those underlying chemical reaction equilibria such as (13.55). In the case of a pure chemical component, the gas–liquid (or other) phase transition evidently involves changes of aggregation or coordination of constituent molecular units whose chemical identity remains *unaltered* in the transition. Accordingly, attention now focuses on weaker forms of *supramolecular* association (“clustering”) and rearrangement, involving a weaker range of interaction strengths than those involved in chemical bond breaking.

Nevertheless, the general principles of statistical thermodynamics are expected to extend to this weaker tier of clustering interactions. Molecular clustering is widely recognized to be

responsible for the nonideality of dense gases (Sidebar 13.5), and the essential thermodynamic *continuity* of gaseous and liquid states (as revealed by the discovery of gas–liquid critical points, Section 2.4.2) suggests that such clustering continues to be a significant structural feature of the liquid region. Of course, “structure” must now be understood in the context of the large-amplitude vibrational motions and facile rearrangements that are expected to characterize condensed-phase clustering.

An essential prerequisite for thermodynamic characterization of cluster equilibria is that the binding energies and rearrangement barriers be sufficiently robust (i.e., compared with kT) to allow survival of well-defined thermodynamic populations under the incessant disruptions of ambient thermal jostling. According to the best evidence of *ab initio* electronic structure theory, this criterion is rather comfortably satisfied by quantum mechanical interactions of hydrogen bonding or related “coordination” or “donor–acceptor” type bonding. In this concluding section, we wish to outline a simple extension of statistical thermodynamics to describe such cluster equilibria and the associated phase thermodynamics in hydrogen-bonded liquids such as water. This extension takes advantage of the fact that *ab initio* electronic structure methods can be routinely used to determine the rotational, vibrational, and electronic properties of hydrogen-bonded clusters, as well as their monomeric precursors (see Sidebar 13.8 for an overview of original literature references).

The starting point for “quantum cluster equilibrium” (QCE) theory is the envisioned aggregation of monomeric species (W) into supramolecular clusters (W_i , $i = 1, 2, \dots, c$) that are considered to be in simultaneous thermodynamic equilibria [cf. (13.55)]:



Each supramolecular cluster species W_i is associated with a partition function q_i , which may be factored into translational, rotational, vibrational, and electronic contributions in the usual manner [cf. (13.63)]. The equilibrium condition corresponding to (13.84) may now be expressed as [cf. (13.65)]

$$\mu_1 = \frac{1}{2}\mu_2 = \frac{1}{3}\mu_3 = \dots = \frac{1}{c}\mu_c \quad (13.85)$$

subject to [cf. (13.61)]

$$n_1 + 2n_2 + 3n_3 + \dots + cn_c = N_A \quad (13.86)$$

where n_i is the equilibrium population of cluster W_i , μ_i is its chemical potential, and N_A is Avogadro’s number. One can combine (13.85) with (13.64) to determine each population n_i as

$$n_i = q_i \left(\frac{n_r}{q_r} \right)^{i/r} \quad (13.87)$$

in terms of a chosen reference species (W_r , often taken as the monomer, $r = 1$), which becomes the single unknown in (13.86).

Although the ideal gas approximation contributes a relatively negligible error to description of chemical equilibrium [as discussed under (13.82)], the same cannot be true for the

cluster equilibria of (13.84). It is therefore essential to consider the nonideal “environmental” effects of interactions with other clusters. In effect, this environmental influence brings dependence on *all* the cluster populations $\{n_j\}$ into the expressions for each individual q_i , as described below. This converts the equilibration condition for cluster equilibria into a highly *nonlinear* problem that must be solved self-consistently for the final equilibrium populations, presenting a much more formidable mathematical challenge than the corresponding description of simple chemical equilibria (Section 13.3.3).

It may seem surprising that one can separate intracluster (“binding”) from intercluster (“environmental”) interactions in a non-arbitrary manner. This appears to be a characteristic consequence of quantum mechanical *cooperativity*, that is, the non-pairwise additivity of exchange-type forces that strongly distinguishes these “chemical” forces from pairwise-additive force laws, such as those of classical electrostatic type. In the case of hydrogen-bonding fluids, quantum cooperativity confers special stability on *proton-ordered* coordination patterns (“conjugated” chain-like or ring-like patterns, as illustrated in Fig. 13.3) that strongly distinguish whether a given hydrogen bond lies inside or outside a connected proton-ordered network, and hence appears “strong” (intracluster) or “weak” (intercluster) compared with the overall statistical average hydrogen-bond energy. As shown by Pauling’s evaluation of the zero-point entropy of ice (Sidebar 5.19), such net proton ordering is impossible in the 4-coordinate topology of crystalline ice, but it can play a dominant role in the fluxional coordination patterns that dominate gaseous and liquid phases. Of course, empirical potential models that fail to properly incorporate quantum cooperativity effects give little insight into the true “energy landscape” that dictates liquid-phase clustering patterns.

The simplest QCE model incorporates environmental effects of cluster–cluster interactions by (1) approximate evaluation of the excluded-volume effect on the translational partition function $q_{i,\text{trans}}$ (neglected in Section 13.3.3) and (2) explicit inclusion of a correction $\Delta E_0^{(\text{env})}$ for environmental interactions in the electronic partition function $q_{i,\text{elec}}$. Secondary environmental corrections on rotational and vibrational partition functions may also be considered, but are beyond the scope of the present treatment.

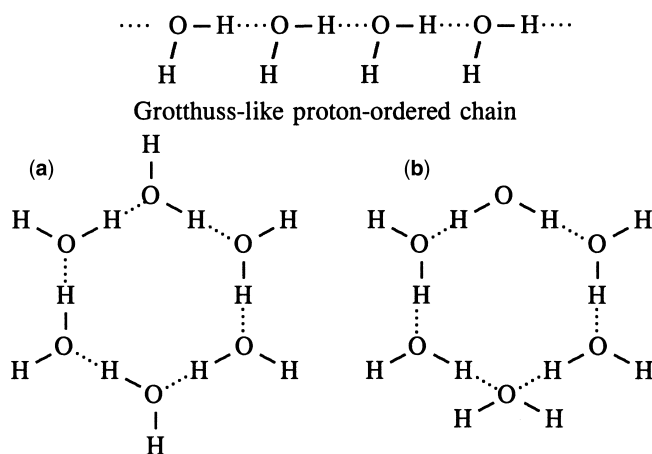


Figure 13.3 Proton ordering in hydrogen bonding, illustrating favorable (Grotthuss-like) ordering in a chain (above) or cyclic hexamer (a), contrasted with the unfavorable isomer (b), one of many similar that cannot survive thermodynamically. All clusters are “ice rule-compliant” (cf. Sidebar 5.18).

The *translational* environmental correction is crudely incorporated by estimating the volume V_i of each cluster [e.g., using the GEPOL algorithm: E. Silla, F. Villar, O. Nilsson, J. L. Pascual-Ahuir, and O. Tapia. *J. Mol. Graphics* **8**, 168–72 (1990)] and evaluating the population-weighted excluded volume that is inaccessible to free translational motions, namely,

$$V_{\text{excl}} = b_{\text{xv}} \sum_{i=1}^c n_i V_i^{(\text{GEPOL})} \quad (13.88)$$

As indicated, we incorporate a single overall “excluded-volume parameter” b_{xv} (of order unity) to compensate for the many errors of approximating V_{excl} as a simple sum of cluster volumes, neglecting, for example, cluster shape effects on interstitial packing vacancies. A nonzero b_{xv} value brings in the complex nonlinear dependence on *all* cluster populations,

$$q_{i,\text{trans}} = q_{i,\text{trans}}(b_{\text{xv}}, \{n_j\}) \quad (13.89)$$

which is a hallmark of QCE numerical solutions.

The electronic environmental correction can be crudely estimated by a mean-field (Van der Waals-type) monomer–monomer approximation of the form

$$\Delta E_0^{(\text{env})}[W_i] \propto \text{number density} \quad (13.90a)$$

such that doubling the environmental density doubles the strength of correction. The mean-field correction is proportional both to the number i of monomers in W_i and to the average number density in the given volume V :

$$\Delta E_0^{(\text{env})}[W_i] = a_{\text{mf}} \left(\frac{i}{V} \right) \quad (13.90b)$$

As indicated, a single “mean-field parameter” a_{mf} is included as the proportionality factor. It is noteworthy that the numerical value of a_{mf} is unimportant to the condensation phenomenon itself, so that even an infinitesimally small value (e.g., of order 10^{-6} a.u.), is sufficient to “reward” thermodynamic condensation and yield an alternative phase of greatly reduced V under appropriate conditions of temperature and pressure.

As in (13.89), a nonzero a_{mf} brings complex nonlinearity into the electronic partition function,

$$q_{i,\text{elec}} = q_{i,\text{elec}}(a_{\text{mf}}, \{n_j\}) \quad (13.91)$$

leading to an overall cluster partition function

$$q_i = q_i(a_{\text{mf}}, b_{\text{xv}}, \{n_j\}) \quad (13.92)$$

that depends on the chosen QCE parameters a_{mf} , b_{xv} as well as equilibrium populations of all other clusters. In practice, numerical values of a_{mf} , b_{xv} are chosen to match two known properties of the substance (e.g., normal boiling point and standard-state liquid density), and all remaining features of the phase diagram are predicted from self-consistent QCE solutions.

The numerical solutions of QCE equations are obtained by an iterative computer algorithm. Given the *ab initio* quantities to construct the q_i (initially for a guessed set of cluster populations) and the specific pressure P_0 and temperature T_0 of a desired state, the program finds the numerical volumes (and associated cluster population distributions) that satisfy (13.68) for the desired P_0 . This matching condition may be written in the form of a polynomial equation

$$P_0 = P(V^{(\alpha)}), \quad \alpha = 1, 2, \dots, n_{\text{roots}} \quad (13.93)$$

where the roots $V^{(\alpha)}$ (and associated cluster populations $n_i^{(\alpha)}$) satisfy the equilibrium conditions (13.85) and (13.68). Each “phase” α is associated [through (13.67)] with a Gibbs free energy $G^{(\alpha)}$,

$$G^{(\alpha)} = A^{(\alpha)} + P_0 V^{(\alpha)} \quad (13.94)$$

where the root $V^{(\alpha)}$ of lowest Gibbs free energy corresponds to the stable phase at T_0, P_0 . The cluster populations obtained [through (13.87)] for chosen $V^{(\alpha)}$ are then employed to construct new partition functions (13.92) and re-solve (13.93) in an iterative fashion, until fully self-consistent solutions are obtained. Further details of the iterative QCE search algorithms are described in the original papers (Sidebar 13.8).

Which clusters W_i are to be included in (13.84)? In principle, one should examine *all possible* cluster species for inclusion in the equilibrium cluster mixture. In practice, however, it is easy to use the QCE equations to “weed out” potential clusters that are negligibly populated, and thereby guide the search for more robust components of the equilibrium mixture. Rather surprisingly, it turns out that the search for robustly populated clusters quickly “saturates,” so that only *a remarkably small number of clusters with specific proton-ordered coordination patterns* are found to dominate the QCE equilibrium distributions of typical hydrogen-bonded fluids at near-standard conditions. This small number of characteristic “independent components” of equilibrium liquid structure provides an important simplifying feature in modeling many aspects of fluid-phase behavior. In principle, it is always possible to test the thermodynamic survival of proposed new cluster structures by freely adding them to the QCE cluster mixture and checking for significant changes in the equilibrium cluster distribution.

Why does the QCE cluster equilibrium sequence (13.84) truncate at small values of c ? The qualitative reason may be judged from simple enthalpic and entropic considerations. The key enthalpic descriptor is the average chemical potential per hydrogen bond, which effectively “plateaus” at hydrogen bonds centered in a chain (or ring) of hydrogen bonds with three or four proton-ordered hydrogen bonds extending in either direction. The enthalpic plateauing is seen most distinctly in proton-ordered *ring* patterns of near-linear (“unstrained”) O—H $\cdots$ O geometry, and such favored patterns are optimally achieved in cyclic W_5 or W_6 structures such as the cyclic hexamer shown in Fig. 13.3a. But even though the cyclic hexamer W_6 is slightly advantageous in this enthalpic respect, the cyclic pentamer W_5 gains considerable *entropic* advantage in competitive equilibria such as



because the right-hand side of (13.95) involves *six* “free” species (with free translations and rotations) instead of only five. Indeed, one can see that entropy will generally continue to favor *smaller* cluster species over larger ones, whenever their respective hydrogen bonds are

otherwise of equal average enthalpic strength. Moreover, clusters of larger size tend to be increasingly disfavored at higher T , even if they boast additional hydrogen bonds per cluster (which seems to confer an indisputable enthalpic advantage). Thus, it is not surprising that the equilibrium cluster distribution of near-ambient liquid water is found to be dominated by rather small cyclic clusters of proton-ordered topology, *far* less than the worst-case scenario (c of order 10^{23}) might have suggested. In fact, a two-component QCE mixture of monomers and cyclic pentamers provides a reasonable description of near-ambient thermodynamic liquid–gas properties, nearly equivalent to that of a much larger QCE mixture. Figure 13.4 illustrates a simple model QCE phase diagram for water that includes a

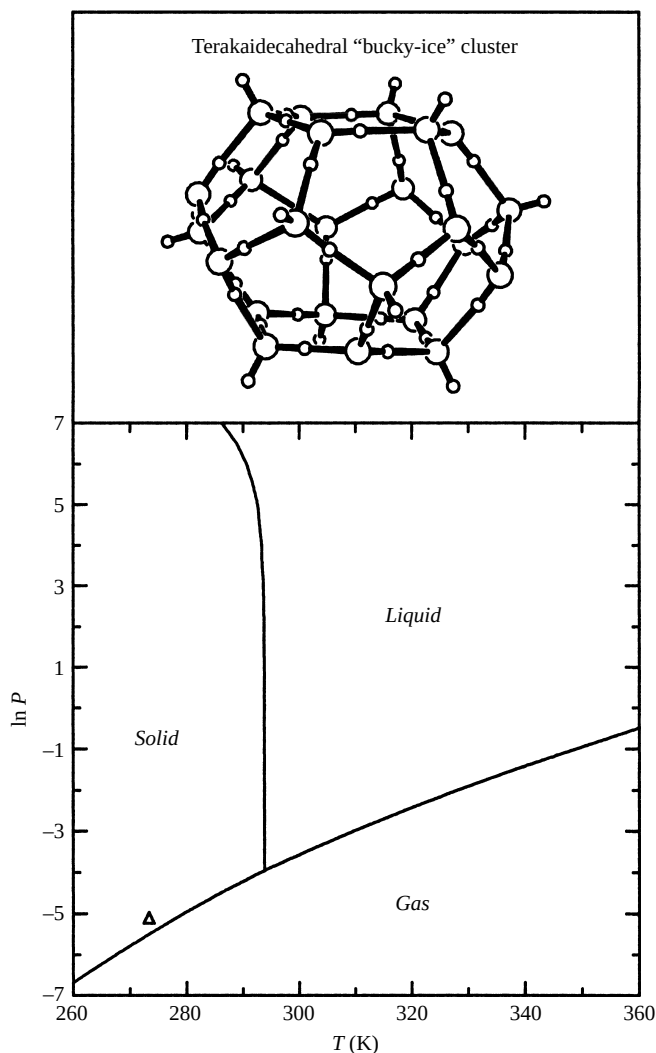


Figure 13.4 Low-level 18-cluster QCE model (RHF/3-21G level) of the water phase diagram, showing (above) the dominant W_{24} clathrate-type cluster of the “ice-like” solid phase, and (below) the overall phase diagram near the triple point (with a triangle marking the actual triple point). Note that numerous other clusters in the W_{20} – W_{26} range were included in the mixture, but only that shown (with optimal proton ordering) acquired a significant population.

triple point linking the vapor-pressure curve to a polymorphic solid phase dominated by clathrate-like W_{24} “buckyball” clusters. Sidebar 13.8 provides references to QCE studies of water and other hydrogen-bonded species (amides, amines, alcohols, acids) that indicate current possibilities and limitations of this approach.

The QCE model also allows numerical evaluation of the heat capacities, thermal coefficients, and compressibilities needed to construct the thermodynamic metric geometry. Unfortunately, the higher derivatives of Q that are needed to evaluate the QCE thermodynamic metric are subject to considerable errors, both from underlying theoretical approximations and from increasingly severe numerical errors in finite-difference evaluations. Significant improvements, including extension to multicomponent chemical mixtures and more accurate description of cluster–cluster interactions, are needed before QCE-like models can provide additional *ab initio* insights into the mysteries of nonideality in phase equilibria.

The surprising QCE prediction of proton-ordered chain or ring-type coordination patterns has received recent experimental support from novel X-ray spectroscopic measurements by Wernet et al. (see Sidebar 13.8). Current controversy still surrounds the coordination patterns and dynamics of liquid water, as well as the QCE suggestion that quantum mechanical “resonance-type” interactions are necessary to properly represent both the intermolecular (hydrogen-bonding) forces that underlie phase equilibria and the interatomic (chemical-bonding) forces that underlie chemical equilibria. Further progress in the thermodynamic sciences requires improved methods for interrogating clustering interactions by experimental means and more detailed electronic-level description of these interactions by theoretical means.

SIDEBAR 13.8: LEADING REFERENCES TO QCE THEORY AND APPLICATIONS

General Reviews and Perspectives

- F. Weinhold. Nature of H-bonding in clusters, liquids, and enzymes: an *ab initio*, natural bond orbital perspective. *J. Mol. Struct. (THEOCHEM)* [WATOC 1996 Symposium Issue] **398–399**, 181–97 (1997).
- F. Weinhold and C. R. Landis. Supramolecular bonding. In *Valency and Bonding: A Natural Bond Orbital Donor–Acceptor Perspective* (Cambridge University Press, Cambridge, 2005), Chap. 5.
- F. Weinhold. Resonance character of hydrogen-bonding interactions in water and other H-bonded species. *Adv. Protein Chem.* **72**, 121–55 (2006).

Basic Methodology and Illustrative Water Applications

- F. Weinhold. Quantum cluster equilibrium theory of liquids: general theory and computer implementation. *J. Chem. Phys.* **109**, 367–72 (1998).
- F. Weinhold. Quantum cluster equilibrium theory of liquids: illustrative application to water. *J. Chem. Phys.* **109**, 373–84 (1998).

Applications to Hydrogen-Bonded Fluids

- R. Ludwig, F. Weinhold, and T. C. Farrar. Experimental and theoretical studies of hydrogen bonding in neat, liquid formamide. *J. Chem. Phys.* **102**, 5118–25 (1995).

- R. Ludwig, F. Weinhold, and T. C. Farrar. Temperature dependence of hydrogen bonding in neat, liquid formamide. *J. Chem. Phys.* **103**, 3636–42 (1995).
- R. Ludwig, F. Weinhold, and T. C. Farrar. Experimental and theoretical determination of the temperature dependence of deuterium and oxygen quadrupole coupling constants of liquid water. *J. Chem. Phys.* **103**, 6941–50 (1995).
- R. Ludwig, F. Weinhold, and T. C. Farrar. Theoretical study of hydrogen bonding in liquid and gaseous *N*-methylformamide. *J. Chem. Phys.* **107**, 499–507 (1997).
- M. A. Wendt, J. Meiler, F. Weinhold, and T. C. Farrar. Solvent and concentration dependence of the hydroxyl chemical shift of methanol. *Mol. Phys.* **93**, 145–51 (1998).
- R. Ludwig, T. C. Farrar, and F. Weinhold. Quantum cluster equilibrium theory of liquids: molecular clusters and thermodynamics of liquid ammonia. *Ber. Bunsenges. Phys. Chem.* **102**, 197–204 (1998).
- R. Ludwig, T. C. Farrar, and F. Weinhold. Quantum cluster equilibrium theory of liquids: temperature dependent chemical shifts, quadrupole coupling constants, and vibrational frequencies of liquid ammonia. *Ber. Bunsenges. Phys. Chem.* **102**, 205–12 (1998).
- R. Ludwig, O. Reis, R. Winter, F. Weinhold, and T. C. Farrar. Quantum cluster equilibrium theory of liquids: temperature dependence of hydrogen bonding in liquid *N*-methylacetamide studied by IR spectra. *J. Phys. Chem. B* **102**, 9312–18 (1998).
- M. A. Wendt, T. C. Farrar, and F. Weinhold. Critical test of quantum cluster equilibrium theory: formic acid at B3LYP/6-31+G* hybrid density functional level. *J. Chem. Phys.* **109**, 5945–7 (1998).
- R. Ludwig, F. Weinhold, and T. C. Farrar. Quantum cluster equilibrium theory of liquids: molecular clusters and thermodynamics of liquid ethanol. *Mol. Phys.* **97**, 465–77 (1999).
- R. Ludwig, F. Weinhold, and T. C. Farrar. Quantum cluster equilibrium theory of liquids: temperature dependent chemical shifts, quadrupole coupling constants and vibrational frequencies in liquid ethanol. *Mol. Phys.* **97**, 479–86 (1999).
- R. Ludwig and F. Weinhold. Quantum cluster equilibrium theory of liquids: light and heavy QCE/3-21G water. *Phys. Chem. Chem. Phys.* **2**, 1613–19 (1999).
- R. Ludwig and F. Weinhold. Quantum cluster equilibrium theory of liquids: isotopically substituted QCE/3-21G model water. *Z. Phys. Chem.* **216**, 659–74 (2002).
- P. Borowski, J. Jaroniec, T. Janowski, and K. Wolinski. Quantum cluster equilibrium theory treatment of hydrogen-bonded liquids: water, methanol and ethanol. *Mol. Phys.* **101**, 1413–21 (2003). [This paper suggests the need for more accurate treatment of free cluster rotations that become hindered librations at normal liquid densities, thus reducing the populations of larger clusters.]

Other Applications (Ice, Sulfur Polymorphs)

- R. Ludwig and F. Weinhold. Quantum cluster equilibrium theory of liquids: freezing of QCE/3-21G water to tetrakaidecahedral “bucky-ice.” *J. Chem. Phys.* **110**, 508–15 (1999).

- R. Ludwig, J. Behler, B. Klink, and F. Weinhold. Molecular composition of liquid sulfur. *Angew. Chem.* **114**, 3331–5 (2002) [*Angew. Chem. Int. Ed. Engl.* **41**, 3199–202 (2002)].

Experimental Evidence for Proton-Ordered Ring/Chain Structures

- P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Naslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson. The structure of the first coordination shell in liquid water. *Science* **304**, 995–9 (2004).
-

Appendix: Units and Conversion Factors

Throughout this book, we follow a current tendency of the physical chemistry literature by frequently replacing “standard” SI units by common experimental alternatives [such as energy changes in thermochemical units (kcal mol^{-1}), structural parameters in crystallographic Angstrom units ($\text{\AA}$), vibrational frequencies in spectroscopic wavenumber units (cm^{-1}), and so forth], thereby facilitating communication between theoretical and experimental practitioners. The six tables below describe the SI system [*Système International d’Unités*, as originally established by the 11th General Conference on Weights and Measures (1960) and subsequently approved for “Quantities, Units and Symbols in Physical Chemistry” by the International Union of Pure and Applied Chemistry (IUPAC)] and its relationship to alternative “non-standard” units. Tables App. 1–2 summarize SI base names and symbols, and Tables App. 3–6 provide numerical values of fundamental constants and conversion factors that relate SI to atomic units (a.u.) and common thermochemical, crystallographic, and spectroscopic units. An abbreviated exponential notation is employed in which 6.02214(23) means 6.02214×10^{23} .

For many purposes, SI units appear theoretically and practically disadvantageous compared to other unit systems (such as the Gaussian electrostatic units preferred by many physicists or the atomic units preferred by many quantum chemists). The SI system presumes *seven* distinct “base units” (meter, kilogram, second, ampere, kelvin, candela, mole), rather than the three required on physical grounds (e.g., mass, length, time). These choices particularly affect the description of electromagnetic phenomena, where the SI units for current [ampere (A)] and charge [coulomb (C)] apparently alter the spatio-mechanical dimensionality of electric charge, and thus of all derived electrical and magnetic quantities. Ironically, adoption of the “coulomb” unnecessarily complicates the form of “Coulomb’s law” (the deepest expression of the spatial relationship between charge and mechanical force, apparently first discovered by Benjamin Franklin), requiring assignment of a mysterious “permittivity” property to empty space. For chemists, additional inconvenience is presented by the SI units for pressure [pascal (Pa)] or chemical concentration [moles per cubic meter (mol m^{-3})], both widely replaced in practice by atmosphere (atm) and molarity (mol L^{-1}), respectively. Sidebar 11.6 presents other evidence that SI units are intrinsically ill-suited for expressing fundamental geometrical relationships between thermodynamic properties, suggesting why (and how) IUPAC approval of SI units as an official pedagogical “standard” might be advantageously reconsidered.

TABLE APP. 1 Names and Symbols for Primary Physical Quantities in SI Units

Physical Quantity	SI Name	SI Symbol	Expression in SI Base Units
Electric capacitance	farad	F	$\text{C V}^{-1} = \text{kg}^{-1} \text{m}^{-2} \text{s}^4 \text{A}^2$
Electric charge	coulomb	C	A s
Electric conductance	siemens	S	$\text{A V}^{-1} = \text{kg}^{-1} \text{m}^{-2} \text{s}^3 \text{A}^2$
Electric current	ampere	A	A
Electric potential difference	volt	V	$\text{J C}^{-1} = \text{kg m}^2 \text{s}^{-3} \text{A}^{-1}$
Electric resistance	ohm	Ω	$\text{V A}^{-1} = \text{kg m}^2 \text{s}^{-3} \text{A}^{-2}$
Energy	joule	J	$\text{N m} = \text{kg m}^2 \text{s}^{-2}$
Force	newton	N	kg m s^{-2}
Frequency	hertz	Hz	s^{-1}
Magnetic flux	weber	Wb	$\text{V s} = \text{kg m}^2 \text{s}^{-2} \text{A}^{-1}$
Magnetic flux density	tesla	T	$\text{Wb m}^{-2} = \text{kg s}^{-2} \text{A}^{-1}$
Power = energy/time	watt	W	$\text{J s}^{-1} = \text{kg m}^2 \text{s}^{-3}$
Pressure = force/area	pascal	Pa	$\text{N m}^{-2} = \text{kg m}^{-1} \text{s}^{-2}$

TABLE APP. 2 Prefixes for Base-Multipliers in SI Units

Multiple	Prefix	Symbol	Submultiple	Prefix	Symbol
10	deca	da	10^{-1}	deci	d
10^2	hecto	h	10^{-2}	centi	c
10^3	kilo	k	10^{-3}	milli	m
10^6	mega	M	10^{-6}	micro	μ
10^9	giga	G	10^{-9}	nano	n
10^{12}	tera	T	10^{-12}	pico	p
10^{15}	peta	P	10^{-15}	femto	f
10^{18}	exa	E	10^{-18}	atto	a

TABLE APP. 3 Conversion Factors From Atomic to SI Units

Atomic Unit (Base Units)	SI Value	Name (Symbol)
Mass (m_e)	9.10939(−31) kg	Mass of the electron
Charge (e)	1.602188(−19) C	Electronic charge
Angular momentum ($\hbar$)	1.05457(−34) J s rad ^{−1}	Planck's constant/ 2π
Energy ($m_e c^4/\hbar^2$)	4.35975(−18) J	Hartree (H)
Length ($\hbar^2/m_e e^2$)	5.29177(−11) m	Bohr; Bohr radius (a_0)
Time ($\hbar^3/m_e e^4$)	2.41888(−17) s	Jiffy
Electric dipole moment ($\hbar^2/m_e e$)	8.47836(−30) C m	2.541765 Debye (D) units
Magnetic dipole moment ($eh/2m_e$)	9.27402(−24) J T ^{−1}	Bohr magneton (μ_B)

TABLE APP. 4 Energy Conversion Table for Non-SI Units

Unit	Value in Non-SI Units					
	a.u.	kcal/mol	eV	cm ⁻¹	Hz	K
a.u.	1	6.27510(2)	2.72114(1)	2.19475(5)	6.57968(15)	3.15773(5)
$\frac{\text{kcal}}{\text{mol}}$	1.59360(−3)	1	4.33641(−2)	3.49755(2)	1.04854(13)	5.03217(2)
eV	3.67493(−2)	2.30605(1)	1	8.06554(3)	2.41799(14)	1.16044(4)
cm ⁻¹	4.55634(−6)	2.85914(−3)	1.23984(−4)	1	2.99792(10)	1.43877
Hz	1.51983(−16)	9.53708 (−14)	4.13567(−15)	3.33564(−11)	1	4.79922(−11)
K	3.16683(−6)	1.98722(−3)	8.61739(−5)	6.95039(−1)	2.08367(10)	1

TABLE APP. 5 Fundamental Constants, in Atomic and SI Units

Physical Constant	Symbol	Value (a.u.)	Value (SI)
Rydberg constant	R_{∞}	2.29253(2)	1.09737(−23) m ⁻¹
Planck constant	h	6.28319 (=2 π)	6.62608(−34) J s
Speed of light	c	1.37036(2)	2.99792(8) m s ⁻¹
Proton mass	m_p	1.83615(3)	1.67262(−27) kg
Atomic mass unit	amu	1.82289(3)	1.66054(−27) kg
Fine structure constant	α	7.29735(−3)	7.29735(−3)

TABLE APP. 6 Other Constants and Conversion Factors

Quantity (Symbol)	SI Value or Equivalent
Avogadro's Number (N_A)	6.02214(23) mol ⁻¹
Kelvin (K)	°C + 273.15 (exactly)
Boltzmann constant (k)	1.38066(−23) J K ⁻¹
Faraday constant ($\mathcal{F}$)	9.64853(4) C mol ⁻¹
Kilocalorie (kcal)	4.184 kJ (exactly)
Atmosphere (atm)	1.01325(5) Pa (exactly)
Molarity (mol L ⁻¹)	0.001 mol m ⁻³ (exactly)

AUTHOR INDEX

- Abelson, P. H., 272
Adair, G. S., 259
Alder, B. J., 233
Amagat, E. H., 232
Aman, J. E., 422
Ampere, A. M., Appendix
Andresen, B., 422
Andrews, T., 37, 49, 50, 232
Archduke Ferdinand, 26
Arrhenius, S., 296, 451
Ashley, M. F., 190
Avogadro, A., 19, 30, 174
- Barnett, J. D., 233
Baur, M., 442
Beattie, J. A., 44
Behler, J., 463
Bejan, A., 423
Benedict, M. W., 44
Bengtsson, I., 422
Bent, H. A., 393
Bergmann, U., 463
Bernoulli, D., 30
Berry, R. S., 422, 427
Berthelot, M., 44, 110
Bird, R. B., 23, 45, 437
Black, J., 85, 117
Blander, M., 272
Block, S., 233
Boltzmann, L., 28, 36, 151, 174, 177, 441, 450, 446ff
Born, M., 105, 444
Borowski, P., 462
Boyle, R., 18, 19, 20, 45
Bridgeman, O. C., 44
Bridgman, P. W., 232, 393
Brody, D., 423
Brown, H. R., 145
Brown, S. C., 69
- Buckingham, A. D., 187
Bunting, E. N., 233
Butler, J. A. V., 416
- Cai, R.-G., 422
Callen, H. B., 443
Carnot, H., 118
Carnot, L., 118
Carnot, S., 118, 120, 123, 126, 135, 145
Carroll, C. W., 393
Carroll, B., 393
Casimir, H. B. G., 438
Cavalleri, M., 463
Celsius, A., 26
Charles, J. A. C., 19
Cho, J.-H., 422
Christian, C. S., 233
Clapeyron, E., 119, 120, 121
Clausius, R., 36, 44, 118, 120, 139, 145, 223
Coulomb, C. A., 375
Crawford, F. H., 348, 393
Crooks, G. E., 422, 427
Cross, P. C., 340
Crouse, W. H., 445
Curtiss, C. F., 23, 45
- Dalence, J., 26
Dalton, J., 19, 70
Davy, H., 67, 85
de Broglie, L., 452
De Donder, T. E., 285
De Vos, A., 423
Debye, P., 36, 184, 191, 438
Decius, J. C., 340
Defay, R., 416
Dennery, P., 330
Derbyshire, J., 429
Descartes, R., 58
DiCera, E., 423

- Dieterici, C., 44
 Dirac, P. A. M., 324
 Dirichlet, L., 428
 Donnan, F. G., 416
 Duhem, P., 201ff

 Eddington, A., frontispiece
 Ehrenfest, P., 227
 Einstein, A., 96, 151, 421, 427, 441, 451
 Elvius, P., 26
 Euler, L., 12, 14, 165, 202, 307, 308, 309

 Fahrenheit, D. G., 26, 223
 Faraday, M., 293, 302
 Farrar, T., 461, 462
 Feldmann, T., 423
 Fenn, J., 95
 Ferrara, S., 422
 Feynman, R. P., 59
 Fick, A. E., 433
 Fisher, M. E., 417, 439
 Flory, P. J., 271
 Foresman, J. E., 454
 Forman, R. A., 233
 Fowler, R. H., 145, 186, 378
 Franklin, B., Appendix
 Franzen, H. F., 416
 Frisch, A., 454
 Frisch, M., 454
 Frobenius, F. G., 336

 Gait, P. D., 422
 Galileo, 20, 26
 Gauss, K. F., 425, 428, 429
 Gay-Lussac, J. L., 19
 Gerstein, B. C., 416
 Giauque, W., 184, 190
 Gibbons, G. W., 422
 Gibbs, J. W., 6, 37, 121, 150, ...
 Glasstone, S., 416
 Goeppert-Mayer, M., 36, 444
 Goranson, R. W., 393
 Gram, J. P., 338, 346, 379ff
 Griffiths, R. B., 386
 Grotthuss, C. J. T., 296, 457
 Guggenheim, E. A., 55, 145, 186, 272, 378

 Haas, A., 416
 Haber, F., 105
 Hall, H. T., 233
 Heisenberg, W., 58
 Helmholtz, H. v., 68, 71, 150, 179

 Henderson, L. J., 122
 Henry, W., 238ff
 Hildebrand, J. H., 271
 Hill, T., 446, 454
 Hirsch, T. K., 463
 Hirschfelder, J. O., 23, 45
 Hooke, R., 20, 57, 72
 Horn, F., 423
 Hubbard, J., 439
 Hückel, E., 301
 Hudson, C. S., 253
 Huggins, M. L., 271
 Huygens, C., 58

 Ihrig, E., 422, 423
 Ising, E., 438, 440

 Jacobi, C. G. J., 9, 291, 428
 Janowski, T., 462
 Jaroniec, J., 462
 Jensen, J. H. D., 445
 Jordan, J. R., 442
 Jordan, P. C., 442
 Joule, J. P., 68, 70, 91

 Kallosh, R., 422
 Kamerlingh Onnes, H., 37, 44
 Kelvin (Wm. Thomson), 70, 118, 119, 121, 129, 145
 Kirchhoff, G. R., 107, 150
 Kirkwood, J. G., 439
 Klink, B., 463
 Konowalow, D., 416
 Kronecker, L., 321, 329
 Kryzwicki, A., 330
 Kwong, J. N. S., 44

 Landis, C. R., 454, 461
 Laplace, P.-S., 119
 Latimer, W. M., 294
 Lavoisier, A., 21, 68, 85
 Le Chatelier, H. L., 290
 Lebesgue, H., 107, 191
 Legendre, A.-M., 164, 309, 353, 428
 Lehrman, A., 393
 Leighton, R. B., 59
 Lennard-Jones, J. E., 35
 Lerman, F., 393
 Levelt-Sengers, J. M. H., 37
 Levine, R. D., 423
 Lewis, G. N., 180, 183, 186, 299, 438, 444
 Lightfoot, E. N., 437

- Lindley, D., 452
 Lippincott, R. R., 233
 Longuet-Higgins, C., 439
 Ludwig, R., 461, 462, 463

 Mach, E., 450
 Mamut, E., 423
 Manning, F. S., 393
 Manning, W. P., 393
 Margenau, H., 393
 Margules, M., 272
 Maxwell, J. C., 6, 30, 37, 146, 151, 313, 334, 449
 McQuarrie, D. A., 452
 Meiler, J., 462
 Miller, D. G., 439
 Mirsky, L., 336
 Misner, C. W., 425
 Mrugala, R., 423
 Murphy, G. W., 393

 Naslund, L. A., 463
 Nathanson, G., 430
 Newton, I., 21, 26
 Nillson, O., 458
 Nilsson, A., 463
 Noether, E., 58
 Nordlund, D., 463
 Nulton, J. D., 422, 423

 Odelius, M., 463
 Ogasawara, H., 463
 Ohm, G., 80
 Ojamäe, L., 463
 Ostwald, W., 151

 Pascal, B., Appendix
 Pascual-Ahuir, J. L., 458
 Paul, W., 233
 Pauli, W., 438
 Pauling, L., 113, 189, 291, 457
 Pelton, A. D., 272
 Pennini, F., 423
 Perrin, J., 450
 Pettersson, L. G. M., 463
 Pidokrajt, N., 422
 Piermarine, G. J., 233
 Pitzer, K. S., 55
 Planck, M., 96, 376, 452, frontispiece
 Plastino, A., 423
 Poisson, S.-D., 301
 Portesi, M., 423

 Prigogine, I., 416

 Randall, M., 180, 183, 186
 Raoult, F. M., 238
 Redlich, O., 44
 Regnault, H., 120
 Reis, O., 462
 Renaldi C., 26
 Rey, J., 26
 Riemann, B., 421, 424, 426, 428
 Rivier, N., 423
 Rømer, O., 26
 Roozeboom, H. W. B., 273
 Rowlinson, J. T., 362, 389
 Rubin, L. C., 44
 Rumford, Count (Benj. Thompson), 67, 68, 85, 117
 Ruppeiner, G., 423
 Rushbrooke, G. S., 386

 Salamon, P., 422, 423, 426, 427
 Sands, M., 59
 Sarkar, T., 422
 Schlögl, F., 423
 Schrödinger, E., 441, 446, 448
 Sengupta, G., 422
 Shannon, C. E., 176
 Shaw, A. N., 393
 Shen, J., 422
 Shilov, G. E., 380
 Shiner, J. S., 422
 Sieniutycz, S., 423
 Silla, A., 458
 Sinanoglu, O., 430
 Sklar, A., 444
 Stanley, H. E., 341, 385
 Stefan, J., 450
 Stewart, W. E., 437
 Stout, J. W., 190
 Strong, H. M., 233
 Su, R.-K., 422

 Tapia, O., 458
 Taylor, B., 15
 Thompson, Benj. (*see* Rumford)
 Thompson, J. B. J., 272
 Thomsen, J. S., 393
 Thomson, Jms., 119
 Thomson, Wm. (*see* Kelvin)
 Thorne, K., 425
 Tisza, L., 314, 393
 Tiwari, B. N., 422

Tobolsky, A., 393

Tsallis, C., 427

Uffink, J., 145

Van der Waals, J. D., 36, 37, 55

Van Valkenburg, A., 233

Van't Hoff, J. H., 151

Villar, F., 45

Vonnegut, K., 226

von Helmholtz, H., 68, 71,
150, 179

Wallis, J., 20

Wang, B., 422

Warschauer, D. M., 233

Watson, G. N., 437, 438

Watt, J., 48, 77, 122, 125

Webb, G. B., 44

Weinhold, F., 334, 454, 461, 462, 463

Weir, C. E., 233

Wendt, M. A., 462

Wentorf, R. H., 232, 233

Wernet, P., 463

Wheeler, J. A., 425

Wheeler, J. C., 386

Wheeler, L. P., 374

Whittaker, E. T., 437, 438

Wilkins, J., 20

Wilson, E. B., 340

Winter, R., 462

Wolinski, K., 462

Wren, C., 20

Wyman, J., 423

Young, T., 58

Zimm, B., 445

SUBJECT INDEX

- absolute zero
 - and “third law”, 183ff
 - and zero-point entropy, 192
 - in ideal gas scale, 27
 - in Kelvin scale, 131
- acceleration, 56
- acetone, 249, 366
- acquisitive convention, 74, 77, 123, 127
- activity, 181, 208, 260ff
 - absolute, 260
 - and chemical potential, 208, 260
 - and effective concentration, 208
 - and reaction equilibrium, 284
 - concept of, 182, 260
 - definition of, 182
 - determination from cell potentials, 296
 - for Henry’s law solutions, 263
 - mean ionic, 297
 - of ideal solutions, 241
 - practical system, 262
 - pure solids and liquids, 284
 - “rational system”, 262
 - relative, 261
- activity coefficient, 216ff
 - and concentration scale, 261
 - in Debye–Hückel theory, 304
 - mean ionic, 298, 300
 - of regular solution theory, 271
- adiabatic
 - container, 86
 - demagnetization, 184
 - equation of state for ideal gas, 98
 - expansion, 95, 97
 - (see Joule–Thomson experiment)
 - work, 86
- advancement of reaction coordinate, 281
- affinity (of a reaction), 285
- alloy (solid solutions), 266ff
 - and intermolecular forces, 269
 - in morphing phase diagrams, 268
 - shape memory, 272
- anode, 292
- atmosphere, ionic, 301ff
- atom, existence of, 39, 449, 451
- atomic units, 375
- attractive force (see intermolecular forces)
- Avogadro
 - constant, 30
 - hypothesis, 19, 174
- azeotrope, 248, 416
 - negative, 249
- Beattie–Bridgeman equation of state, 44
- Benedict–Webb–Rubin equation of state, 44
- benzene, 110, 243, 265, 444
 - geometric properties, 366ff
 - heat capacity, 96
- Berthelot
 - equation of state, 44
 - principle, 110
- binary solutions, 233ff
- binomial coefficient, 175
- biorthogonality, 348
 - in phase space, 443
 - vector construction, 349
- black hole thermodynamics, 421, 422
- boiling point
 - and activities, 262
 - diagram, 234, 247ff
 - relation to vapor–pressure diagram, 249
 - elevation, 254
- Boltzmann, L., 36, 151, 440
 - biographical sketch, 450
 - constant, 31, 174
 - distribution, 28, 446
 - and temperature, 28
 - entropy, 174
 - equation, 451

- Boltzmann, L. (*Continued*)
 H-theorem, 177, 451
 statistical hypothesis, 449
 Boltzmann geometry, 448
 Boltzmann probability, 446ff
 metric origins of, 441, 448
 bond
 dissociation energy, 115
 energy, 113
 enthalpy, 114
 Born–Haber cycle, 105
 boundary, definition of, 60
 Boyle, R., 18
 biographical sketch, 20
 Boyle’s “law”, 19
 Boyle temperature, 45
 bra vector (see Dirac notation)
 Bridgman, P. W., 393
 biographical sketch, 232
 bubble-point surface, 236
- caloric
 and theory of heat, 67, 85
 calorie, 85
 calorimeter, 85, 90
 calorimetry, 85, 90
 capacitance, 81
 capacitor (condenser), 81
 carbon
 graphite-diamond conversion, 232
 phase diagram, 230
 standard elemental form, 105
 carbon dioxide
 Boyle temperature, 46
 compressibility factor, 32, 43
 critical point, 48
 in water gas shift reaction, 454
 phase diagram, 48
 2nd virial coefficient, 46
 carbon monoxide, 188
 in water gas shift reaction, 454
 carbon tetrachloride, 55, 366ff
 Carnot, S.
 and formulation of 2nd law, 118
 biographical sketch, 118
 influence on Clapeyron and Kelvin, 120
 pre-Gibbs role, 145
 steam engine analysis, 123
 Carnot
 efficiency, 126
 principle of, 126
 theorem, 135
 Carnot cycle (engine), 124ff
 ideal gas in, 131
 in TS coordinates, 137
 reverse (heat pump), 126
 refrigerator efficiency, 127
 Cartesian coordinates, 315, 366
 cathode, 292
 catlinite (pipestone), 232
 cell (see electrochemical cell)
 cell potential
 and Gibbs energy, 293
 measurement of, 293, 295
 temperature dependence of, 191
 cell reaction, 295
 Celsius (centigrade) scale, 28
 chain rule, 4, 331, 332
 and distributive metric property, 334
 change of state
 adiabatic, 86
 and Riemann geometry, 424ff
 at constant pressure (isobaric), 22
 at constant temperature (isotherm), 22
 at constant volume (isochoric), 22
 vs. change in size, 215
 metric aspects, 417ff
 progress variable, 426
 reversible vs. irreversible, 76
 Charles’ “law”, 19
 (also called “Gay–Lussac’s law”)
 chemical component
 at maxima/minima of phase diagram, 245, 247, 265
 counting, 213
 definition of, 211
 distribution in phases, 211
 in phase rule, 211, 308
 independent, 211
 ion as, 296
 single, 216ff, 231
 variable number of, 231
 chemical equation, 102, 281
 chemical equilibrium, 281ff
 approach to, 285
 criterion of, 282
 effect of unreactive species on, 292
 in electrochemical cells, 292ff
 in ideal gas mixtures, 207
 in real gas mixtures, 195ff, 204, 206
 non-ideality effects, 300
 chemical potential, 156, 308
 and activity, 260ff

- and mass migration, 205
- and osmotic phenomena, 256
- and partition function, 447
- and phase coexistence, 219
- and slope of phase boundary, 220
- as “escaping tendency”, 206
- as “chemical pressure”, 206
- convention for electrically charged species, 296ff
- in dilute solutions, 240
- in ideal solutions, 204
- in QCE theory, 456
- of electrolytes, 296
- of ideal gas mixture, 206
- of solute in binary solution, 204
- of mixtures, 197
- physical nature of, 204
- chemical reaction, 281ff
 - “active”, 211, 214, 231, 308
 - entropy changes in, 186, 288
 - equation form, 102, 281
 - heat of (see thermochemistry)
 - statistical thermodynamics of, 445
 - temperature dependence, 107
- Clapeyron equation, 219, 221, 413
 - and colligative properties
 - and solid–liquid equilibria, 263
 - melting transition, 221
 - slopes of phase boundaries, 220
 - vaporization transition, 222
 - (see Clausius–Clapeyron equation)
- Clapeyron matching condition, 391
- Clapeyron vector, 387
 - in co-Clapeyron orientation, 390
- Clausius, R., 36, 118, 136, 145, 440
 - and virial of force, 44
 - biographical sketch, 120
 - equation of state, 44
 - inequality of, 139
 - principle of, 130
- Clausius–Clapeyron equation, 223
 - and relative humidity, 224
 - integrated form, 223
- clusters
 - and virial expansions, 45, 444
 - as constituents of liquids, 36, 444
 - buckyball, 460
 - in phase thermodynamics, 455
- co-Clapeyron orientation, 390
- coefficient of performance (see efficiency)
- coefficient of thermal expansion
 - (see thermal expansion coefficient)
- coexistence
 - co-Clapeyron vectors, 390
 - coefficients, 410, 411
 - conditions, 219, 390, 410
 - critical slope, 392
 - double-tangent construction for, 270
 - (see phase transition)
- cohesive energy
 - of ionic crystals, 105
- colligative properties, 253ff
 - boiling-point elevation, 254
 - freezing-point depression, 255
 - osmotic pressure, 255ff
 - vapor–pressure lowering, 254
- collisions
 - conservation laws for, 58
 - elastic, 58
 - pressure due to, 30
 - (see kinetic molecular theory)
- common ion effect, 277
- components (see chemical components)
- composition (see concentration units)
- compressibility
 - adiabatic, 363
 - behavior near absolute zero, 379
 - divergence at critical point, 384ff
 - in thermodynamic metric, 353ff
 - isothermal, 23
- compressibility factor, 31ff
 - of Van der Waals gas, 41
- concentration unit (composition)
 - for binary solutions, 234
 - for chemical reaction, 284
 - molality, 261, 296, 297
 - molarity, 261, 297
 - mole fraction, 261
- condensation
 - gas–liquid, 47
- condenser (see capacitor)
- congruent melting, 265
- conjugate coordinate
 - and metric biorthogonality, 348
 - and conjugate vectors, 348
 - intensive–extensive, 84, 138, 204, 306
- conservation of energy, law of, 58ff, 87ff, 307
 - and Noether’s theorem, 58
 - Feynman’s parable, 59
 - (see first law of thermodynamics)
- conservation of mass
 - and chemical equations, 102, 105, 281
- conservative force, 57
- consolute temperature (see critical point)

- constants
 - fundamental, 376, Appendix (see Boltzmann, Faraday, ...)
- constant heat summation, law of (see Hess' law)
- continuity
 - gas–liquid, 37, 50
 - of equations of state, 22
- convention
 - for activity, 260ff
 - for current flow, 80
 - for electrode potentials, 294
 - for ion formation enthalpy, 112
 - for standard form of element, 105
 - for standard state, 104, 262
- corresponding state
 - experimental deviations, 55
 - principle of, 54
 - theoretical limits on accuracy of, 55
- cooperativity, 457
- Coulomb's law, 299, 375
- critical constants, 50
- critical eigenangle, 383, 384
- critical point, 218
 - and degrees of freedom, 219
 - amplitudes, 385
 - coexistence line slope, 392
 - consolute points, 250ff, 276
 - divergences, 49, 251, 384
 - exponents, 384, 385
 - (in)equalities, 382, 383, 386
 - inflection properties, 52
 - mechanical, 66
 - metric eigenvalues, 383, 392
 - of helium, 226, 228
 - opalescence, 49, 251
 - plait point, 275
- critical solution temperature (plait point), 275
- critical state, 47, 49
 - instability limit, 379ff
- cubic equation, 51
- cycle,
 - Born–Haber, 105
 - definition of, 88
 - first law, 103
 - for Kirchhoff equation, 108
 - reversible, 140
- cyclic (Jacobi) identity, 9
- cyclic integral, 88
- Dalton's law of partial pressure, 19, 207, 237, 239, 244, 260
- de Broglie wavelength, 452
- Debye–Hückel theory, 300, 301ff, 438
- Debye screening length, 302
- De Donder's affinity, 285
- degrees of freedom, 18, 211, 379
 - change of, 379
 - number of, 65, 211 (see Gibbs phase rule)
 - of a function, 3
 - of compressibility factor, 32
 - spatio-geometric, 428 (see space, dimension)
 - thermodynamic, 65, 332
 - vibrational, 96
- density
 - and slope of phase boundary, 222
 - as non-intensive variable, 64, 210
- derivative
 - condition for equilibrium, 156
 - definition of, 4, 5
 - experimental determination, 7
 - formulas, 5
 - graphical, 5
 - low-order for thermodynamics, 27, 332
 - third and higher, 332, 417
- desalination, 260
- dew-point surface, 236
- dialysis, 259
- diameter, molecular, 40
- diamond (see carbon)
 - anvil, 233
- diathermal (non-adiabatic)
 - container, 86
 - work, 86
- diatomic molecule, 96
- dielectric
 - constant, 82, 301
 - medium, 81
 - polarization in, 81
- Dieterici equation of state, 44
- diethyl ether, 366
- differential, 3
 - exact (perfect), tests of, 11
 - exact and inexact, 10
 - finite magnitude, 4, 7
 - for thermodynamic potentials, 164, 168
 - for changes of state, 417
 - Riemannian line element, 426
- diffusion, 429ff
 - coefficient, 433
 - of mass, 205
- diffusional geometry, 429ff, 436, 443

- dimensionality (see degrees of freedom)
 - change of, 379
- dilution, 109
- dipole moment, 83
 - induced, 83
 - permanent, 83
- Dirac notation, 323ff, 345
 - bra–ket scalars, 324
 - ket–bra (dyadic) operators, 324
- dissipative (see irreversible)
- distillation, 247ff, 416
 - fractional, 247
 - of azeotropic mixtures, 248
- distribution
 - Boltzmann, 28, 446
 - probability, 176
- domain structure, 209
- donor–acceptor interaction, 456
- dot (scalar) product, 56
- droplet
 - and new phase appearance, 212, 236
 - formation of solution from, 110, 198
- Ehrenfest classification of phase
 - transitions, 227
- efficiency (coefficient of performance)
 - of heat engine, 122
 - of heat pump (refrigerator), 122
- eigenvector, 321
 - degeneracy, 325
 - orthonormality, 325, 326
- eigenvalue, 321
 - spectrum of, 322
- Einstein, A., 429, 441, 451
 - gravitation theory, 421
 - index convention, 427
 - statistical mechanics, 151
- electric field, 81
- electric potential, 81
- electrical work (see work, electrical)
- electrochemical cell, 80, 292ff
 - diagram, 292, 295
 - redox reactions of, 293
- electrochemical potential, 292
- electrode, 292
- electrode potential, 295
- electrolyte
 - activity of, 297
 - chemical potential of, 296
 - dissociation of, 296
 - solubility of, 111
- electromotive force (emf), 80
- electron affinity, 106
- electronic structure
 - (see quantum chemistry)
- element, standard state of, 105
- emf, 80
- endoergic process, 285
- endothermic process, 103
- energy
 - average per molecule, 96
 - bond, 113, 114, 115
 - concept of, 58ff
 - conservation of, 58, 70, 84
 - electrical (see work, electrical)
 - equipartition of, 31
 - interaction, 34, 267
 - internal, 87
 - ionization, 106
 - kinds of, 58
 - kinetic, 58
 - magnetic, 83
 - mechanical, 57, 58
 - nuclear, 58
 - potential, 34, 58
 - rotational, 96
 - thermal, 31
 - translational, 31, 96
 - vibrational, 96
 - (see Gibbs energy, Helmholtz energy, enthalpy, work, first law)
- energy level, 192
- ensemble, 63, 441, 445, 447
- enthalpy, 90
 - bond, 113
 - criterion of equilibrium, 163
 - excess, 271
 - of ideal gas mixtures, 207
 - of formation, 104
 - partial molar, 199
 - role in proton-ordered clustering, 460
 - standard, 104
 - temperature dependence of, 107
- entropy, 137ff, 307
 - and information theory, 176
 - and probability, 173, 174, 447, 449
 - and vibrational flexibility, 173, 177
 - Boltzmann, 174, 192
 - definition of, 137
 - excess, 271
 - maximal character (in isolation), 149
 - analytic formulation, 153
 - vs. energy minimization, 160
 - nuclear spin, 60, 189

- entropy (*Continued*)
 - of mixing, 173, 20, 271
 - of the universe, 144
 - partial molar, 199
 - physical nature of, 177
 - role in clustering, 459
 - statistical definition of, 174
 - vectors of common laboratory fluids, 371, 374
- entropy-based metric geometry, 337, 430, 447, 449
- entropy change
 - in chemical reactions, 288
 - in ideal gas isothermal mixing, 207
 - spontaneous (in isolation), 141
- entropy-energy duality, 160ff, 307, 430
- entropy production, 433, 435
- equation of state, 18ff
 - empirical, 43
 - of gas mixture, 204
 - thermodynamic, 204, 306
 - (see ideal, Van der Waals, etc.)
- equilibration
 - “minus-first” law of, 145
 - reverse, 431
 - time-evolution, 65
- equilibrium
 - and activity, 208
 - between phases, 209ff
 - conditions for, 149ff
 - gas–solid, 171
 - heterogeneous, 209
 - homogeneous, 195ff
 - in electrochemical cells, 293
 - liquid–gas, 218ff, 234ff
 - liquid–liquid, 250ff
 - metastable, 229
 - multi-component, 279
 - phase (see phase equilibrium)
 - solid–liquid, 172, 278
- equilibrium constant, 284
 - and Gibbs energy, 284
 - enthalpic and entropic contributions, 287
 - pressure, 284
 - pressure dependence of, 289
 - temperature dependence of, 288
- equipartition of energy, 93, 371
- equivalent weight, 293
- escaping tendency, 206
- ethanol, 116, 199, 245, 366
- Euler theorem for homogeneous functions, 202, 308
- Euler criterion for exactness
 - general, 14
 - in Maxwell relations, 165, 307, 309
 - two degrees of freedom, 12
- eutectic, 264ff
 - diagram, 264
 - in morphing phase diagram, 268
- exact differential, 10ff, 307
 - energy as, 88
 - entropy as, 136
 - properties of, 10
- exoergic process, 285
- exothermic process, 74
- expansion work (see work, pressure–volume)
- extensive variable
 - biorthogonality to intensive variables, 348ff
 - conjugate intensity of, 84, 204
 - definition of, 63
 - enthalpy as, 104
 - “excess” (redundant), 410
 - in Gibbs theory, 305
 - in heat, 85
 - in work forms, 84
 - in multiphase systems, 210
 - non-extensive thermodynamics, 427
- extent of reaction (see advancement of reaction)
- $f(f+1)$ rule, 167, 348
- Fahrenheit, D. G., 26
 - temperature, 223
- Faraday constant, 293, 402
- Feynman’s parable, 59
- Fick’s law of diffusion, 433
- first law of thermodynamics, 67ff
 - general statements, 87ff
 - alternative form, 131
 - and symmetric property of thermodynamic metric, 334
- Flory–Huggins theory, 271
- flow, 433, 435
- fluctuations, 436
 - different, in different ensembles, 447
 - in critical point phenomena, 49, 417
 - in Onsager theory, 439
 - in theory of Mayer et al., 443
 - omission from thermodynamic-level description, 62
 - T -dependent, 173
 - thermal contribution to molecular vibrational entropy, 177
- flux, 433

- force, 56
 - and momentum transfer, 30
 - conservative, 57
 - and potential energy, 34
 - thermodynamic, 435
- force constant, 36, 72, 340
- formation reaction, 105
 - standard enthalpy, 105
 - for ions, 112
- Fourier, J., 119
 - heat theorem, 433
 - series, 428
- fractional distillation (see distillation)
- free energy (see Gibbs energy, Helmholtz energy)
- freezing-point depression, 255
 - curves, 264
- friction, 57, 84
 - incorporation in “thermo”-dynamics, 57, 138
- Frobenius’ theorem, 336
- fugacity, 181ff, 208
 - coefficient, 182
- function, 3
 - single-variable, 3
 - multi-variable, 4
 - physical vs. mathematical conception, 8
- fundamental constants, 376, Appendix
- “fundamental equations” of thermodynamics, 204, 306
 - exclusion of Legendre transforms, 353
- fusion, heat of, 221
- gas constant, 19, 27
- gas imperfection (see compressibility factor)
- gases
 - continuity with liquids, 37
 - heat capacity of, 92
 - ideal, 18ff, 95
 - kinetic molecular theory of, 30
 - liquefaction of, 47, 50, 95
 - real vs. ideal, 47
- geometry
 - definition of, 313
 - metric, 314
 - Euclidean, 315, 428
 - criteria for, 328, 424
 - Dirac notation for, 325, 329
 - Riemann, 421ff, 424ff
 - and measure, 424
 - geodesics of, 426
 - Minkowski, 427
 - (see thermodynamic geometry)
 - (see diffusional geometry)
 - (see Boltzmann geometry)
- Gibbs, J. W., 6, 37, 121, 150, 181
 - biographical sketch, 150
 - criterion for units, 27, 374
 - change of perspective, 145, 149ff, 429ff
 - criteria of equilibrium, 150, 159
 - alternative statements, 157ff
 - model, 6, 161, 313
 - phase rule, 211
 - advocacy of graphical methods, 313
 - P – T slope of univariant systems, 414
 - stationary points on phase boundaries, 416
 - statistical mechanics, 440, 442
- Gibbs’ thermodynamic formulation, synopsis of, 305
- Gibbs–Duhem equation, 201ff, 203, 386, 413
 - and scaling, 201
 - special form for constant T and P , 203
 - in Gibbs phase rule, 212, 308
 - in osmotic equilibria, 257
 - in chemical equilibria, 282
 - and metric dimensionality, 337
 - and null metric eigenvectors, 339, 347, 352, 386
 - and generalized homogeneity, 341
 - vector form, 337ff, 413
- Gibbs–Duhem vector, 386, 409
- Gibbs free energy, 162ff
 - and cell potential, 293
 - and non- PV work, 173
 - as function of reaction advancement, 282
 - composition dependence of, 207
 - convexity, 159
 - criterion of equilibrium, 163
 - definition of, 163
 - double-tangent construction, 270
 - electrical contribution to, 292
 - for roots of QCE polynomial, 459
 - “free” aspect, 172
 - of a mixture, 267
 - of cell reaction, 293
 - of chemical reaction, 281
 - of electrolyte solution, 293, 296ff
 - of formation, 286, 287
 - of gas–solid transition, 171
 - of ideal gas, 173, 179
 - of liquid–solid transition, 172
 - of mixing, 267
 - of water gas shift reaction, 454
 - partial molar, 199

- Gibbs free energy (*Continued*)
 physical nature of, 172
 properties of, 170ff
 standard state for, 283, 286
 T , P dependence of, 178
- Gibbs–Helmholtz equation, 179
- Gibbs–Konowalow equation, 414ff
- Gibbs phase rule, 211, 307, 332
 and intensive properties, 210
 proof of, 212
 and “active” chemical reactions, 214
 for 1-component systems, 216
 for 2-component systems, 223
 for 3-component systems, 273
 and null eigenvectors of metric, 339, 343
 intensive vs. extensive vector asymmetry, 353
 volume contribution, 427
 (see space, dimension)
- Gibbs (U , S , V) space, 6, 137, 313, 426
- Gibbs units, desiderata for, 27, 376
- gradients in diffusional processes, 432
- Gram matrix, 338, 346, 379, 380
 Gramian determinant, 381
 and dimensional collapse at critical limit, 381ff
- gravitation
 work, 72, 78
 weight vs. mass, 56
 effect on phase domain boundaries, 209
 general relativistic theory of, 421
- Grotthuss proton ordering, 296, 457
- ground state, 192
- half-cell, 293, 295
- half-reaction (oxidation/reduction), 293, 295
- harmonic oscillator, 453
 partition function for, 453
- heat, 85ff
 and molecular motions, 67, 85
 constant-pressure (see enthalpy)
 constant-volume (see internal energy)
 definition of, 86
 exclusion from Newtonian mechanics, 57
 high- T vs. low- T , 123, 131
 mechanical definition, 86
 mechanical equivalent of, 67
 nature of, 85
 of dilution, 109
 of formation, 105
 of fusion, 221
 of phase transition, 221
 of pressure variation, 354
 of reaction, 101ff
 at constant pressure, 104
 dependence on temperature, 107
 measurement of, 85, 90, 105
 of solution, 108, 200
 of solution, differential, 109, 200
 of sublimation, 221
 of vaporization, 222
- heat capacity, 85ff, 89ff
 at constant pressure, 89, 90
 at constant volume, 89, 90
 behavior in state changes, 420
 critical point divergence, 384ff
 Debye contribution, 191
 difference $C_P - C_V$, 91, 166, 355, 391
 in thermodynamic metric, 353ff
 low-temperature behavior, 378
 of ideal gases, 92
 monatomic, 96
 polyatomic, 96
 ratio C_P/C_V , 98, 354, 370
 saturation, 362, 389
 vibrational, 96
- heat conduction (see diffusional geometry)
- heat engine, 122ff
 efficiency of, 125
 temperature dependence of, 133
- heat pump (refrigerator)
 as reverse Carnot cycle, 126
- heat reservoir, 61, 123
- Heisenberg uncertainty principle, 58
- helium
 zero-point entropy, 188
 lambda-line, 227
 phase diagram, 226
- Helmholtz, H. v., 68, 150
 biographical sketch, 71
- Helmholtz free energy, 162ff
 definition of, 163
 criterion of equilibrium, 163
 and canonical ensemble, 447
- Henry’s law, 238
 and solubility of gases, 239
 and osmotic equilibria, 257
 activity convention, 262
- Hermitian (see operator, Hermitian)
- Hess’ law, 104, 113
- heterogeneous equilibrium, 209
 (see phase equilibrium)

- homogeneous equilibrium
 - definition of, 209
 - metric of, 353
- homogeneous function, 202
 - generalized, 341ff
- Hooke's law, 57, 72
- humidity
 - and vapor pressure, 218, 224
- hydrogen bonding
 - and proton-ordered cluster patterns, 457ff
 - and zero-point entropy of ice, 189
 - in ice polymorphs, 225
 - effects on equilibria, 287
 - effects on non-ideality, 199
- hydrogen chloride (HCl)
 - 3rd-law entropy of, 191
 - non-ideality of, 247, 249, 300, 305
- hydrogen electrode, standard, 294
- ice
 - anomalous volume-change of
 - melting, 222
 - zero-point entropy, 189
 - polymorphs, 224
 - rule, 189, 457
- ideal solution, 237
 - chemical potentials in, 204
 - standard states for, 241
 - liquid and vapor composition, 244
 - dilute limit for osmotic equilibria, 257
- ideal gas, 95ff
 - as working fluid of Carnot engine, 131
 - chemical potential of pure, 205
 - entropy of isothermal mixing, 173
 - heat capacity, 92
 - in Carnot cycle, 131ff
 - in Clausius–Clapeyron equation, 222
 - kinetic molecular theory of, 30
 - properties of 21ff, 167
 - T -dependence, 92, 167
 - thermodynamic geometry of, 356, 370
 - thermodynamics of, 95
 - work of isothermal expansion, 77
- ideal gas “law”, 21
 - deviations from, 31, 47
 - low-pressure limit, 27, 33
- ideal gas mixture, 204ff
 - partial molar quantities in, 207
 - entropy of, 173
- impossibility (impotence) axioms, 88, 129, 130, 186
- incongruent melting point, 266
- indicator diagram, 48, 77
 - (see Watt's indicator)
- inductive laws of thermodynamics, 18
 - (see first law, second law, ...)
- inductive theory, 17
- inflection point, 52, 66
- integrals
 - cyclic, 11
 - line (path), 10, 13, 57, 87, 426
- integrating factor, 14, 136
- intensive variables
 - and definition of state, 65
 - as state descriptors, 215
 - as thermodynamic vector, 332
 - definition, 63
 - in Gibbs theory, 306
 - in heat, 85
 - in multiphase systems, 210
 - in work forms, 84
 - independent, number of (see Gibbs phase rule)
 - reference vectors, 358
- intensive vectors
 - “axis”, 409
 - basis transformations, 335
 - “excess” (redundant), 409
 - linear dependence, 337, 339
 - (see thermodynamic vectors)
- interaction energy, 34
 - (see bond energy)
- intermolecular forces, 34
 - attractive, 33, 35, 40
 - repulsive, 33, 35, 39
 - and phase behavior, 267ff
- internal energy (see energy, internal)
- International System (SI) of units, 374, Appendix
- ion product of water, 284
- ionic atmosphere (see Debye–Hückel theory)
- ionic crystals
 - cohesive energy of, 105
 - solubility of, 111
- ionic equations, net, 112
- ionic solutions (see electrolyte solutions)
- ionic strength, 299
- ionization potential, 106
- irreversible thermodynamics, 117, 140
 - comparison with reversible, 76
 - heat flow, 142
 - volume flow, 143
 - gas expansion, 180

- irreversible changes of state, 71, 75, 140
 - and diffusional geometry, 429ff
- Ising problem, 438, 440
- isobar, 22
- isochore (isometric), 22
- isotherm, 22
- isotope
 - and zero-point entropy, 189
 - quantum effects in helium, 226
- Jacobi identity, 9, 291
- Jacobian determinant, 336, 403
 - evaluation of thermodynamic derivatives, 393, 401
- Joule, J. P., 68
 - biographical sketch, 70
 - experiment, 91
- Joule heat, 68, 81
- Joule–Thomson expansion, 93
 - as cryogenic technique, 183
 - coefficient, 93, 166, 400
 - T -dependence of, 94
 - inversion temperature, 95, 183
- Kelvin, Lord (Wm. Thomson), 70
 - biographical sketch, 119
 - and 2nd law formulation, 118
 - principle of, 129
- Kelvin temperature scale, 130ff
 - equivalence to ideal scale, 131
- kinetic energy
 - average, 31
 - of a molecule, 62
 - of random motion, 30
 - and *vis viva*, 58
- kinetic molecular theory (KMT), 30
 - of heat, 85
- Kirchhoff equation, 107
- Kronecker delta function, 321, 329
- Lagrange undetermined multipliers
 - and constrained entropy maximization, 155
 - and constrained probability maximization, 448
 - illustration, 154
 - method of, 153
- latent heat, 107, 191, 221, 222
- laws of thermodynamics, 18
 - (see first law, second law, ...)
- Lebesgue integral, 107
- Le Chatelier's principle, 290ff
 - in phase diagrams, 225
 - in chemical equilibrium, 283, 289
 - mathematical formulation and proof, 291
- Legendre transformation, 164, 309, 353
- length
 - thermodynamic, 345
 - vector, 315
 - unit of, 375
- Lennard–Jones potential, 35
- lever rule, 241
 - proof of, 242
 - in solid–liquid diagrams, 265
 - in ternary (triangle) diagrams, 275
- Lewis, G. N., 180, 444
 - and third law, 183, 186
 - and ionic strength, 299
 - and detailed balance, 438
- line (path) integral, 10
 - and energy conservation, 87
 - and work definition, 57, 77
 - for change of state, 426
 - illustrative example, 13
- liquefaction, 47, 50, 95
- liquids
 - continuity with gases, 37, 50, 456
 - liquidus curve, 264
 - liquidus surface, 236
- macroscopic
 - definition of, 63ff
- magic square, 168ff
- magnetic field, 84
- magnetic moment, 84
- Margules expansion, 272
- mass
 - relation to weight, 56
 - migration, driving force for, 205
 - unit of, 375
- mathematical background
 - affine spaces, 313, 424
 - eigenvalue equation, 321, 325
 - Euclidean spaces, 313ff, 329
 - exact and inexact differentials, 10ff
 - field, 430, 432
 - integrating factor, 14
 - matrix algebra, 315ff
 - method of Lagrange undetermined multipliers, 153
 - metric spaces, 328ff
 - vector algebra, 315ff

- matrix
 - adjoint, 317, 320
 - conjugate, 352
 - eigenvalue equation, 321, 325, 367
 - eigenvalue, 321, 326, 367, 380
 - eigenvector, 321, 380
 - orthonormality (for self-adjoint), 321, 326
 - completeness, 322, 327
 - Gram, 338, 346, 379, 380
 - Hessian, 338, 340, 431
 - inverse, 32
 - normal, 320
 - partitioned (supermatrix), 323, 419
 - positive-definite, 376
 - rank, 339
 - representation of an operator, 325
 - representation of geometry, 319
 - self-adjoint (Hermitian), 320
 - transpose, 317, 320
 - unitary, 320
- matrix algebra, 315ff, 319ff
 - addition, 316
 - advocacy by Gibbs, 151, 315
 - conformability, 317
 - equality, 326
 - exponential, 319
 - functions, 323
 - inverse, 319
 - multiplication
 - by a scalar, 316
 - by a matrix, 317
 - powers, 319
 - positivity, 376
 - resolution of the identity, 327, 394
 - spectral theorem, 322, 327
- maximum work, 74ff, 173
- Maxwell, J. C., 6, 30, 37, 146, 151, 313, 334, 449
- Maxwell relations, 164ff
 - magic square for, 168
 - and metric symmetry, 347
- Mayer, J. E.
 - biographical sketch, 444
 - cluster theory, 36
 - geometry of non-equilibrium
 - displacement variables, 442ff
- Mayer, J. R., 67
 - biographical sketch, 68
- mechanical
 - definition of heat, 86
 - equivalent of heat, 67
 - stability, 66
 - temperature scale, 25
- melting curve, 217
- mercury, 366
- metastable
 - state, 66
 - phase, 225, 229ff, 270
- metric, 329ff
 - and changes of state, 329
 - conjugate, 352
 - diffusional, 431
 - eigenproperties of common laboratory
 - liquids, 373
 - correlation with critical point, 373
 - internal (non-singular), 347
 - matrix, 329, 346
 - null eigenvectors of, 339, 408
 - of a homogeneous fluid, 353
 - space, 328
 - thermodynamic, 334
 - positivity and the "third law", 376ff
 - eigenmodes, 403
- microstate, 175, 193
- "minus first" law of thermodynamics, 145
- miscibility, partial, 250ff
 - in liquids, 250
 - in solids, 266, 267
 - propensity for, 245, 246
 - in ternary phase diagrams, 276
- mixing
 - entropy of, 173, 199
 - Gibbs energy of, 199, 267
 - heat of, 108ff
 - volume of, 198ff
- mixtures, 195ff
 - ideal gas, 207
 - chemical potential of, 195
 - (see composition variables)
- modes of kinetic and potential energy, 96
- molality
 - definition of, 261, 296
 - mean ionic, 298
 - comparison with molarity, 297
- molar mass (molecular weight)
 - determination of, 29, 258
- molar polarization, 83
- molarity
 - definition of, 261, 297
 - comparison with molality, 297
 - and activity in dilute ideal solutions, 284

- mole fraction, 201, 234, 235
- molecule
 - existence of, 36
 - properties affecting solubility, 111
 - properties affecting phase coexistence, 269ff
- molecular beam scattering, 34, 95
- molecular diameter, 40
- molecular partition functions
 - electronic, 453
 - rotation, 453
 - translation, 452
 - vibration, 453
- molecular weight (see molar mass)
- moment of inertia, 453
- momentum, 30, 56
 - conservation of, 30
- Morse potential, 36
- motion
 - rotational, 96
 - translational, 30, 96
 - vibrational, 96
- natural (spontaneous) change, 117ff
 - driving forces for, 110, 162, 164
 - entropy and, 141, 162
- Nernst, W., 151, 451
 - and "third law", 183ff
- Nernst equation, 294
- Nernst heat theorem, 181, 185, 378
- Newtonian mechanics, 56ff
 - in Boltzmann theory, 175
 - failures, 440, 449, 451
- normal modes
 - as metric eigenvectors, 364
 - Cartesian analogs, 364, 366
 - of molecular vibrations, 340, 408
 - of *S*-based metric, 431ff
 - of thermodynamic metric, 341, 363ff
 - self-conjugacy of, 364
- nuclear spin, 189, 226
- Ohm's law, 80
- Onsager, L., 435, 443
 - biographical sketch, 437
 - coefficients, 436
 - eigenvalue equation, 436
 - matrix, 436
 - reciprocal relations, 436
- operator, 324
 - adjoint, 320
 - gradient, 34, 433
 - Hermitian (self-adjoint), 320, 325
 - Laplacian, 301
 - matrix representation of, 314, 325
 - normal, 320
 - unitary, 320
- orthogonality (perpendicularity), 321
- orthonormal set, 321, 322
- ortho-hydrogen, 189
- osmosis, 255ff
 - reverse, 259
- osmotic pressure, 256
 - measurement of, 255
 - Van't Hoff equation for, 257
- para-hydrogen, 189
- partial molar quantities, 197ff
 - of ideal gas mixtures, 207
- partial pressure
 - concept, 218
 - Dalton's law, 19
 - equilibrium constant for, 284
- partition function
 - and chemical potential, 447
 - and equilibrium constant, 454
 - contributions to, 447
 - canonical, 446
 - electronic, 453, 458
 - for reactive mixtures, 452
 - rotational, 453
 - translational, 452, 458
 - vibrational, 453
- path integral (see line integral)
- Pauling, L., 113, 189, 291, 457
- peritectic point, 266
- permanent gas, 50
- permittivity
 - dielectric medium, 82
 - vacuum, 82, 301, 375
- perpetual motion
 - limits on, 122
 - of first kind, 88
 - of second kind, 129
- phase
 - coexistence, 219
 - condensed, 47ff
 - definition of, 209
 - diagram, 48
 - interface, 209
 - role in Gibbs theory, 308
 - stability of, 48
 - transition, 49

- phase diagram, 48, 216, 234
 - carbon dioxide, 49
 - pure substance, 224ff
 - sulfur, 229
 - water, 217, 224
 - water-acetic acid-vinyl acetate, 275
 - water-nicotine, 253
 - helium, 226
 - carbon, 230
 - binary solutions, 233ff
 - water- NH_4Cl - $(\text{NH}_4)_2\text{SO}_4$, 277
 - of QCE model, 460
- phase rule (see Gibbs phase rule)
- phase space
 - configurations, 65
 - Mayer displacement variables in, 442
- phase transition
 - binary systems, 233ff
 - boundary (meniscus), 252
 - coexistence coefficients, 410
 - density and boundary slope, 222
 - Ehrenfest classification, 227
 - gas-liquid, 217, 218
 - gas-solid, 171, 217
 - kinetic aspects of, 230
 - latent heat, 107
 - liquid-solid, 217, 263ff
 - metastable, 230
 - slopes of boundaries, 222
 - (see saturation properties)
- plait point, 275
 - (see critical point)
- Planck's constant, 376, 452
- Poisson's equation, 301
- Poisson-Boltzmann equation, 302
- polarizability
 - orientational, 83
 - distortion (induction), 83
- polarization
 - electrical, 81, 83
 - magnetic, 83
- potential
 - electrode, 292
 - half-cell, 294
- potential energy, 34
- precipitation (see solid-liquid equilibria)
 - in ternary systems, 278
- pressure, 30, 77
 - and partition function, 447
 - high, 232
 - kinetic molecular theory of, 30
 - polynomial of QCE theory, 459
 - saturation vector, 391
 - vectors of common laboratory liquids, 371
- pressure-volume work
 - (see work, pressure-volume)
- probability
 - and entropy, 174, 176, 447, 449ff
- progress coordinate, 281
- property, 61
 - (see thermodynamic property)
- proton order
 - in ice, 189
 - in water, 457, 459
 - experimental evidence for, 461
- pseudoelasticity, 272
- QCE (see quantum cluster equilibrium)
- quadruple point, 219, 225
- quantum chemistry, 34
 - role in phase phenomena, 440
 - evaluation of partition functions, 454
- quantum cluster equilibrium (QCE) theory, 455ff
 - bibliography, 461
- quantum mechanics
 - limits on energy conservation, 58
 - quantum energy levels, 192
 - in Hilbert space, 324
 - matrix-mechanical representation, 324
 - wave-mechanical representation, 324
 - and statistical mechanics, 449
- quantum statistical thermodynamics, 439ff
 - statistical origins of metric geometry, 445
- radioactive decay, 62
- randomness
 - kinetic energy of, 30
 - and entropy, 173, 446ff
- Raoult's law, 238
 - activity convention, 262
 - and non-ideal solution, 238ff
 - and solution composition, 244
- rational activity, 262
- reaction Gibbs energy, 282
- reaction quotient, 283
- real gases, 30ff
 - Gibbs energy of, 197ff
 - isotherms of, 33
- real processes (finite time), 76, 429ff
- Redlich-Kwong equation of state, 44
- reduced variables, 54
- refrigerator (see heat pump)
- relative humidity, 224

- reservoir (work, heat, mass), 60
- resistance, 80
- response functions, 307
- reversibility
 - of electrochemical cell, 294
 - microscopic, 438
- reversible change of state, 71
 - path, definition of, 74
 - comparisons with irreversible, 76
 - isothermal expansion (ideal gas), 97
 - adiabatic expansion (ideal gas), 97
- reversible engine, 124ff
- Riemann, B., 421, 426
 - biographical sketch, 428
 - geometry, 329, 424ff
 - line element, 426
- rotational partition function, 453
- Royal Society of London, 20
- rubber elasticity, 79
- Rumford, Count (Benjamin Thompson), 67, 85
 - biographical sketch, 68
 - reverse-Rumford process, 117
- saturation properties, 360ff, 389ff, 397ff
 - definition of, 360
 - identities for, 362, 397, 398
- scalar product, 56, 318
 - complex-valued, 330
 - Dirac notation, 324
 - Euclidean, criteria for, 328, 329
 - in metric spaces, 328
- scale factor, 408, 410
 - matrix, 410, 412
- scaling
 - and generalized homogeneity, 341
 - and Gibbs–Duhem equation, 201
 - in multiphase systems, 212
 - of extensive properties, 305
- Schrödinger equation, 34, 324, 440
- screened Coulomb potential, 303
- second law of thermodynamics, 117ff
 - alternative forms, 130, 131, 164
 - and positivity of thermodynamic metric, 334
 - Clausius statement of, 130, 141, 144
 - in Gibbsian form, 149ff
 - Kelvin statement of, 129
 - misconceptions, 144
- semi-permeable membrane, 255, 259
- series expansion (see Taylor series)
- shape memory alloy, 272
- SI (Système Internationale) units
 - unsuitability, 367, 374ff
- size
 - as irrelevant state descriptor, 215
 - (see macroscopic; scale factor)
- SHE (standard hydrogen electrode), 294
- solid
 - metastable phases, 229
 - polymorphs, 224, 229
 - shape memory alloys, 272
 - sluggish equilibration, 229
 - solid–liquid equilibria, 263
 - solid–solid phase transitions, 272, 266ff
- solid solution (see alloy)
- solidus curve, 264
- solubility
 - common-ion effect on, 277
 - curves, 264
 - ideal law of, 269, 271
 - in ternary systems, 277
 - of gases and Henry’s law, 239
 - of salts, 110, 111
- solubility product, 284
- solute
 - activity convention for, 262
 - chemical potential measurement, 204
 - non-volatile, 253
- solution
 - binary, 233ff
 - chemical potential measurement, 204
 - heat of, 108, 200
 - Henry’s law, 240
 - ideal, 207, 237
 - equivalence of definitions, 240
 - phenomenological models, 270
 - regular, 271ff
 - solid, 266
- solvent
 - activity definition for, 262
- space, 315
 - affine, 317, 333
 - basis vectors, 332
 - Euclidean, 315, 324, 329
 - criteria for, 328
 - Hilbert, 324, 443
 - linear manifold, 424
 - mathematical requirements, 332
 - metric, 328ff
 - Minkowski, 427
 - non-Euclidean, 329
 - of thermodynamic responses, 331ff
 - Riemann, 329, 424ff

- specific heat, 86
- spectroscopy, 34
 - microwave, 28
 - nuclear magnetic resonance, 189
- spontaneity
 - conditions for, 139ff
 - of heat flow, 141
 - of volume flow, 143
 - in laboratory conditions, 170
 - vaporization of a solid, 171
 - of mass flow, 205
- stability conditions
 - for chemical equilibrium, 283
 - for mechanical equilibrium, 66
 - for phases, 48, 210
 - in Gibbs equilibration, 157, 307
 - in metric geometry, 356
 - with respect to self-contact, 64
- standard electrode potential, 294
 - determination of, 295
 - equilibrium constants from, 294
- standard hydrogen electrode, 294
- standard reaction Gibbs energy, 283
- standard state, 104, 261ff
 - and concentration units, 262
 - arbitrariness of, 262
 - for Gibbs energy, 181, 283
 - for ideal gas, 181
 - for ideal solutions, 241
 - for practical activity, 262
 - for pressure, 181, 206
 - for rational activity, 262
 - form of elements, 105
- state
 - definition of, 64
 - equation of, 204
 - reversible change of, 71
 - time variation of, 65
 - space of, 306
 - distinguished role of intensities, 353
- state property
 - definition of, 64
 - energy as, 88
 - enthalpy as, 90, 103
 - entropy as, 136
 - of other Legendre transforms, 164
 - of cell potentials, 294
- statistical definition of
 - entropy, 174
- statistical mechanics
 - Boltzmann, 174, 441
 - Gibbsian, 151, 440ff
 - comparison of Boltzmann and Gibbsian formulations, 441
- steam engine, 48, 77, 122ff
 - (see Watt's indicator)
- stoichiometry, 102, 281
- sublimation, 49
 - heat of, 106
 - Gibbs energy of, 171
 - curve, 217
- sulfur
 - phase diagram for, 229
 - liquid, 230, 463
- supercooling, 229
- supercritical fluid, 50, 218, 226
- superfluidity, 226
- superheating, 230
- surface energy, 64
 - thermodynamic treatment, 79
 - irrelevance in bulk thermodynamics, 64
- surface tension (see work, surface tension)
- surroundings
 - definition of, 60
- symmetry breaking, 49, 379
- symmetry number, 453
- system
 - adiabatic, 86
 - closed, 61
 - definition of, 60
 - diathermal, 86
 - isolated, 61
 - open, 61
 - properties of, 61
 - state of, 64
- Taylor series expansion, 15
 - double, for osmotic equilibrium, 256
 - in Poisson–Boltzmann equation, 302
 - in virial equation of state, 45, 46
 - invalidity for Gibbs energy, 417
 - for entropy change, 158
 - for fugacity, 183
 - for heat capacity, 108, 420
 - for metric, 417
 - Margules expansion, 272
- temperature
 - absolute zero of (see absolute zero)
 - Boyle, 45
 - concept of, 24ff
 - critical (see critical point)
 - gas scale, 27
 - Joule–Thomson inversion, 95
 - “negative”, 29

- temperature (*Continued*)
 - reservoirs, 123
 - saturation vector, 391
 - “true” scale of, 25ff
 - vectors of common laboratory liquids, 370
- temperature scale, 24ff
 - absolute, 27
 - current definition of, 27
 - ideal gas, 27
 - Kelvin, 28
 - thermodynamic, 167, 196
 - Celsius, 28
 - Fahrenheit, 223
 - mechanical, 25
 - Boltzmann, 28
- ternary system, 273ff
- theoretical plate, 247, 267
- thermal contact, 24
- thermal equilibrium, 24
 - condition for, 150, 159, 164
 - “minus first” law of, 145
 - transitive character of, 25
- thermal expansion coefficient, 23
 - divergence at critical point, 384ff
 - in thermodynamic metric, 353ff
 - saturation value, 397
- thermal stability (see thermal equilibrium)
- thermochemistry, 101
 - chemical reactions, 101ff
 - first-law consequences, 89
- thermodynamic derivatives, 393ff
 - for phase boundaries, 408, 411
 - higher-dimensional, 405
 - higher-order, 417
 - Jacobian method, 401
 - of metric, 419
 - vector evaluation of, 395, 396, 398, 399, 401, 404, 407
- thermodynamic equation of state
 - (see equation of state)
- thermodynamic function
 - dependence on composition, 195ff
 - partition function and, 447
- thermodynamic geometry, 331ff
 - lack of “tensorial” character, 427
 - basis vectors, 333
 - bibliography, 422
 - conjugacy relationships, 359
 - dimension, 332, 337
 - general transformation theory, 357ff
 - isomorphism to Gibbsian thermodynamics, 335
 - ket vector (intensive, extensive), 332
 - metric, 333ff
 - of a monatomic ideal gas, 356
 - scalar product, 333, 334
- thermodynamic property, 60, 427
 - characteristics of, 62
 - extensive, 63
 - ignorable, 62
 - intensive, 63
 - force (affinity) 285, 433
 - geometrical characterization, 331
 - macroscopic, 63
 - saturation, 360
- thermodynamic vector, 345ff
 - biorthogonality, 348
 - components, 347
 - length, 345
 - linear dependence, 346
 - physical significance, 346
 - units of, 346, 362
- “third law” of thermodynamics
 - alternative forms, 131
 - as post-Gibbsian development, 181
 - as second-law stability consequence, 378
 - Buckingham statement, 187
 - common textbook statement, 186
 - entropies (derived from), 187, 190ff
 - experimental exceptions to, 188ff
 - in standard entropy of formation, 288
 - Lewis–Randall statement, 186
 - statistical rationalization, 192
 - superfluous character of, 187, 376ff
 - tautological character of, 187
 - zero-point entropy, 188ff
- tie-line (see lever rule)
- time
 - as pseudo-metric component, 427
 - arrow of, 118
 - de-equilibration, 430
 - evolution toward equilibrium, 65
 - direction of natural (spontaneous) process, 117, 449
 - in dynamic phase transitions, 233
 - in shape memory alloys, 270
 - process in finite, 75
 - role in pre-Gibbs formulations, 145
 - unit of, 375
- translational energy, 58, 96
- transport
 - diffusive, 429ff
 - general near-equilibrium equation for, 433
 - properties, 437

- triple point
 - of ice polymorphs, 225
 - of water, 27, 218
 - (see eutectic, peritectic)
- turnover rule (Hermitian operators), 325
- units
 - choice of, 374
 - of energy, 58
 - SI and other, 374, Appendix
- universe, 60
 - energy of, 87
 - entropy of, 144
- Van der Waals, J. D.
 - and corresponding states principle, 55
 - biographical sketch, 37
- Van der Waals equation, 36ff
 - constants of, 42
 - compressibility factor of, 41
 - critical properties of, 50
 - without derivatives, 53
 - isotherms, 48, 51
 - mean-field aspect, 458
 - reduced form, 54
 - second-law violation in, 51
- Van't Hoff equation for osmosis, 257
 - molecular-weight determination, 258
- Van't Hoff equation for T -dependence, 288
- vapor pressure
 - curve, 217
 - diagram, 234ff, 249
 - lowering, 254
 - molecular interactions underlying, 245, 246
 - of binary solutions, 233ff
 - positive and negative deviations, 243
 - saturation, 218
- vaporization, heat of, 222
- variable(s)
 - active, 4
 - as geometrical vectors, 331ff
 - complement, 394
 - composition, 195, 261, 297
 - conjugate, 394
 - conjugate complement, 394
 - control, 4
 - dependent, 3
 - independent, 213
 - natural, 138, 152
 - of 1st/2nd/3rd laws, 138
 - of metric state space, 337
 - of state, definition of, 60
 - phase-coexistence, 221, 335
 - reduced, 54
 - saturation, 360
- vector algebra, 315ff
 - adjoint vector, 317
 - and thermodynamic geometry, 345ff
 - angle, 316
 - basis vectors, 327, 332
 - conjugate vector, 348
 - dyadic product, 318
 - equality, 326
 - length, 315
 - linear combination, 332
 - null vector, 339, 388
 - orthonormality, 321, 325
 - representation of geometry, 315, 323
 - scalar product, 318, 328
 - Schwarz inequality, 328, 329, 346
 - triangle inequality, 328, 346
- velocity, 56
 - components of, 30
 - average, 30
- vibration-rotation coupling, 447
- virial equation of state, 44ff
- virial coefficients, 45ff
 - for fugacity expansion, 183
 - for osmotic pressure, 258
 - of Van der Waals gas, 47
 - theoretical expression for $B(T)$, 45
- voltage (see electromotive force)
- volume
 - excluded, 39, 452, 458
 - of mixing/solution, 199
 - partial molar, 198
 - experimental determination, 200
 - vectors of common laboratory liquids, 372
- von Helmholtz, H., 68, 150
 - biographical sketch, 71
- wall (see boundary)
- water
 - anomalous density, 222
 - geometric properties, 366
 - ion product, 284
 - phase diagram, 217, 224
 - (see ice)
- water gas shift reaction, 454ff
- Watt's indicator, 125
- work, 56, 76ff, 84
 - elastic, 79
 - electric polarization, 81
 - molecular view, 83

work (*Continued*)

- electrical, 80, 292
- expansion, 75
- general forms, 84
- gravitation, 57
- magnetic polarization, 83
- maximum and minimum, 75
- non- PV and Gibbs energy, 173
- of “charging up” an ion, 304
- of isothermal ideal gas expansion, 77
- pressure–volume, 76, 78

- reversible and irreversible, 71

- spring-extension, 57

- surface tension, 78, 79

- thermodynamic, definition of, 84

working fluid

- role in heat engines, 123

zero-point entropy (see third law)

zeroth law of thermodynamics,

- 25, 145

zone melting, 267

SIMPLIFIED, INNOVATIVE ALTERNATIVE TO STANDARD CALCULUS-BASED THERMODYNAMICS

This textbook proves how equilibrium thermodynamics, a traditionally difficult subject, can be accurately expressed using basic high school geometry concepts. Specifically, the text deals with classical equilibrium thermodynamics and its modern reformulation in metric geometric terms. The author emphasizes applications to chemical and phase equilibria in complex chemical systems, statistical mechanical origins, and extensions to near-equilibrium transport properties.

The text is divided into three parts:

- *Part One: Inductive Foundations of Classical Thermodynamics*
- *Part Two: Gibbsian Thermodynamics of Chemical and Phase Equilibria*
- *Part Three: Metric Geometry of Equilibrium Thermodynamics*

In addition to a rigorous introduction to the principles of chemical and phase thermodynamics, this text takes a novel, innovative approach, reformulating and extending these principles in the framework of Euclidean and Riemannian geometry. The resulting "thermodynamic geometry" provides a powerful alternative to the traditional calculus-based formulation of equilibrium thermodynamics. Students will find that this approach simplifies many mathematical derivations and clarifies thermodynamic logic. Moreover, this approach suggests novel extensions to equilibrium and nonequilibrium systems of arbitrary chemical and phase complexity.

This text also provides a general introduction to vector- and matrix-algebraic methods that are highly effective for thermodynamics as well as applications in quantum mechanics and other advanced topics in physical chemistry.

With its unique simplified approach to thermodynamics coupled with its discussions of the historical and conceptual development of the field, this text is ideal for both undergraduate and graduate students in physical chemistry, chemical engineering, thermal physics, and materials sciences.

FRANK WEINHOLD

received his PhD from Harvard and held postdoctoral and teaching appointments at Oxford, Berkeley, and Stanford before moving to the University of Wisconsin-Madison in 1976, where he served as Professor of Chemistry and a member of the Theoretical Chemistry Institute until achieving emeritus status in 2007. He is the author of more than 170 technical publications and software packages, including the Natural Bond Orbital (NBO) analysis program, widely incorporated in modern ab initio quantum chemistry packages. In addition to the metric geometric theory of equilibrium thermodynamics, Dr. Weinhold's research examines upper and lower bounds for quantum-mechanical properties, complex-coordinate rotation methods for autoionizing resonances, and quantum cluster equilibrium descriptions of water and other H-bonded fluids.

Subscribe to our free Chemistry eNewsletter at
wiley.com/enewsletters

Visit wiley.com/chemistry

 **WILEY**
wiley.com

ISBN 978-0-470-40236-8



9 780470 402368