

Theoretical and Mathematical Physics

David A. Lavis

Equilibrium Statistical Mechanics of Lattice Models

 Springer

Equilibrium Statistical Mechanics of Lattice Models

Theoretical and Mathematical Physics

The series founded in 1975 and formerly (until 2005) entitled *Texts and Monographs in Physics* (TMP) publishes high-level monographs in theoretical and mathematical physics. The change of title to *Theoretical and Mathematical Physics* (TMP) signals that the series is a suitable publication platform for both the mathematical and the theoretical physicist. The wider scope of the series is reflected by the composition of the editorial board, comprising both physicists and mathematicians.

The books, written in a didactic style and containing a certain amount of elementary background material, bridge the gap between advanced textbooks and research monographs. They can thus serve as basis for advanced studies, not only for lectures and seminars at graduate level, but also for scientists entering a field of research.

Editorial Board

W. Beiglböck, Institute of Applied Mathematics, University of Heidelberg, Heidelberg, Germany

P. Chrusciel, Gravitational Physics, University of Vienna, Vienna, Austria

J.-P. Eckmann, Département de Physique Théorique, Université de Genève, Geneva, Switzerland

H. Grosse, Institute of Theoretical Physics, University of Vienna, Vienna, Austria

A. Kupiainen, Department of Mathematics, University of Helsinki, Helsinki, Finland

H. Löwen, Institute of Theoretical Physics, Heinrich-Heine-University of Düsseldorf, Düsseldorf, Germany

M. Loss, School of Mathematics, Georgia Institute of Technology, Atlanta, USA

N.A. Nekrasov, IHÉS, Bures-sur-Yvette, France

M. Ohya, Tokyo University of Science, Noda, Japan

M. Salmhofer, Institute of Theoretical Physics, University of Heidelberg, Heidelberg, Germany

S. Smirnov, Mathematics Section, University of Geneva, Geneva, Switzerland

L. Takhtajan, Department of Mathematics, Stony Brook University, Stony Brook, USA

J. Yngvason, Institute of Theoretical Physics, University of Vienna, Vienna, Austria

David A. Lavis

Equilibrium Statistical Mechanics of Lattice Models

David A. Lavis
Department of Mathematics
King's College London
London
UK

ISSN 1864-5879 ISSN 1864-5887 (electronic)
Theoretical and Mathematical Physics
ISBN 978-94-017-9429-9 ISBN 978-94-017-9430-5 (eBook)
DOI 10.1007/978-94-017-9430-5

Library of Congress Control Number: 2014955152

Springer Dordrecht Heidelberg New York London
© Springer Science+Business Media Dordrecht 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media B.V. Dordrecht is part of Springer Science+Business Media
(www.springer.com)

*To the memory of George MacDonald Bell
(1925–1993)*

Preface

This book was originally planned as an updated one-volume edition of the two works co-authored with the late George Bell (Lavis and Bell 1999a, b). However, as the project developed this became a substantially different book under my sole authorship and I am very grateful to Vanessa and Debbie Bell for permission to make use of some of the material from my work with their father.

This is not a book on the foundations of statistical mechanics; attention is restricted to equilibrium systems, with equilibrium defined simply in terms of the Gibbs distribution. The material is also restricted to deterministic systems on lattices. Within these constraints, most interesting and difficult problems arise from cooperative effects, that is from interactions between the constituent particles of the system. This leads frequently to the occurrence of phase transitions and much of the work of equilibrium statistical mechanics concentrates on the investigation of their properties, particularly their classification into universality classes. The book is structured into parts. Part I contains the basic elements of equilibrium thermodynamics and statistical mechanics, phenomenological scaling theory and a survey of a range of different models, many of which are studied in greater detail later in the book. A broad division of the rest of the material can be made between approximation methods and exact results. Parts II and IV are concerned with the former, II being devoted to classical, mean-field, methods and IV to series approximations and the renormalization group. Exact results and methods are presented in Part III. The last part of the book contains a collection of mathematical appendices on a range of relevant topics. For convenience, they have been divided into material related to graphs and lattices and (perhaps rather inexactlly) algebra and analysis. The intention is to minimise the need to refer to other mathematical sources.

Some parts of the material were originally written for an M.Sc. course at King's College London and I am grateful to the many students who, by their questions and comments, helped to reduce errors and obscurities. The exchange of ideas and information with friends and co-workers has contributed to this book. In particular I owe a great deal to the interaction with my two long-term collaborators, the late George Bell, to whom this work is dedicated, and Byron Southern of the University of Manitoba. I should also like to thank Reimer Kühn, who has always been

prepared to spend time in discussion of points of substance and presentation. As was the case with our previous books I acknowledge with gratitude the continued support and encouragement of Prof. Elliott Lieb. During the final stages of production, the book was read by three referees, for whose comments and suggestions I am very grateful. Finally, I wish to express my thanks to Prof. Wolf Beiglböck, Dr. Aldo Rampioni and the staff at Springer for their work in the production of this book.

London, June 2014

David A. Lavis

Contents

Part I Thermodynamics, Statistical Mechanical Models and Phase Transitions

1	Thermodynamics	5
1.1	Formulae and Variables	5
1.2	The Field-Density Representation	8
1.3	The Thermodynamic Limit	9
1.4	Particular Response Functions	10
1.4.1	Magnetic Systems	11
1.4.2	Fluid Systems	11
2	Statistical Mechanics	13
2.1	Distributions	13
2.1.1	Quantum Systems	14
2.1.2	The Connection to Thermodynamics	15
2.2	Variations of the Probability Function	16
2.3	Coupling Representations	17
2.3.1	The Case $n_f = 2$	19
2.4	Lattice Systems	21
2.4.1	Site-Variable Models	22
2.4.2	Edge-Variable Models	22
2.5	Correlation Functions and Symmetry Properties	23
2.5.1	A General Hamiltonian	23
2.5.2	Correlation Functions	24
2.5.3	Symmetry Properties	26
3	A Survey of Models	29
3.1	Upper and Lower Critical Dimensions	29
3.2	The Quantum Heisenberg Model	30
3.2.1	One-Dimensional Chains	31

3.3	Classical Vector Models	33
3.4	The Gaussian and Spherical Models	34
3.5	Ising Models	36
3.5.1	The Spin- $\frac{1}{2}$ Ising Model	36
3.5.2	The Spin-1 Ising Model.	44
3.6	State-Difference Models.	45
3.6.1	The Classical XY Model	45
3.6.2	The Ashkin–Teller Model	53
3.6.3	Potts Models	54
3.6.4	The Standard Potts Model	55
3.7	Chirality	57
3.7.1	Chiral Potts Models	58
3.7.2	An Extended 3-State Potts Model on the Triangular Lattice	59
3.8	Vertex Models	60
3.8.1	The Eight-Vertex Model	61
3.8.2	The Six-Vertex Model.	72
3.9	Dimer Models	80
3.9.1	The Modified KDP Model Equivalence	83
3.9.2	The Ising Model Equivalence.	85
4	Phase Transitions and Scaling Theory	89
4.1	The Geometry of Phase Transitions.	89
4.1.1	A Two-Dimensional Phase Space	90
4.1.2	A Three-Dimensional Phase Space	93
4.2	Universality, Fluctuations and Scaling	94
4.2.1	Universality	95
4.2.2	Scaling for the Ising Model	97
4.3	General Scaling Formulation	100
4.3.1	The Kadanoff Scaling Hypothesis	100
4.3.2	First-Order Transitions.	104
4.3.3	Effective Exponents	105
4.3.4	The Nightingale–’T Hooft Scaling Hypothesis	109
4.3.5	Constraints on Scaling.	110
4.3.6	Scaling Operators and Dimensions	115
4.3.7	Correlation Functions	118
4.3.8	Variable Scaling Exponents	119
4.3.9	Densities and Response Functions.	121
4.4	Critical Point and Coexistence Curve	122
4.4.1	Critical Exponents.	124
4.4.2	Exponent Inequalities	125
4.5	Scaling for a Critical Point.	125
4.5.1	Scaling Fields for the Critical Point.	126
4.5.2	Approaches to the Critical Point	128

4.5.3	Experimental Variables	129
4.5.4	The Density and Response Functions	130
4.5.5	Asymptotic Forms.	131
4.5.6	Critical Exponents and Scaling Laws.	132
4.5.7	Correlation Scaling at a Critical Point	134
4.6	Tricritical Point	136
4.7	Scaling for a Tricritical Point	141
4.7.1	Scaling Fields for the Tricritical Point	141
4.7.2	Tricritical Exponents and Scaling Laws	143
4.8	Corrections to Scaling	146
4.9	Scaling and Universality	148
4.10	Finite-Size Scaling	152
4.10.1	The Finite-Size Scaling Field	153
4.10.2	The Shift and Rounding Exponents.	155
4.10.3	Universality and Finite-Size Scaling	157
4.11	Conformal Invariance	159
4.11.1	From Scaling to the Conformal Group.	159
4.11.2	Correlation Functions for $d \geq 2$	159
4.11.3	Universal Amplitudes for $d = 2$	161
4.11.4	Schramm–Loewner Evolution.	163

Part II Classical Approximation Methods

5	Phenomenological Theory and Landau Expansions	169
5.1	Classical Methods.	169
5.1.1	A First-Order Transition	171
5.1.2	Metastability	174
5.2	The Van der Waals Equation	175
5.3	Landau Expansions with One Order Parameter.	176
5.3.1	The Spin- $\frac{1}{2}$ Ising Model.	182
5.4	Landau Expansions with Two Order Parameters	183
5.4.1	The Spin-1 Ising Model.	183
5.4.2	The 3-State Potts Model	184
5.5	Landau Theory for a Tricritical Point	188
5.5.1	Tricritical Exponents	190
5.6	Ginzburg–Landau Theory	193
5.6.1	A Critical Point	193
5.6.2	The Gaussian Approximation	194
5.6.3	Gaussian Critical Exponents.	197
6	Mean-Field Theory	205
6.1	The Ising Ferromagnet.	205
6.1.1	Mean-Field Fluctuations	209

6.2	A Model for Metamagnetism	212
6.2.1	The Paramagnetic State	215
6.2.2	The Antiferromagnetic State	216
6.2.3	A Neighbourhood of the Critical Curve	217
6.2.4	The First-Neighbour Antiferromagnet: $\lambda = 0$	221
6.2.5	The First-Order Transition	223
6.2.6	A Neighbourhood of the Tricritical Point	226
7	Cluster-Variation Methods	229
7.1	Improving Mean-Field Theory	229
7.2	The KHDeB Hierarchy of Approximations	231
7.2.1	Distribution Numbers	231
7.2.2	Extensive Quantities	233
7.2.3	The Hamiltonian and Free Energy	235
7.2.4	The Entropy	235
7.2.5	Minimization	237
7.2.6	Labelling Configurations	238
7.3	The Bethe-Pair Approximation for the Ising Model	239
7.4	Reduction to the Mean-Field Approximation	241
7.5	3-State Potts Model on a Triangular Lattice	243
7.6	A Lattice Model for Fluid Water	246

Part III Exact Results

8	Algebraic Methods	259
8.1	The Thermodynamic Limit	259
8.2	The Infinite-System Approach	261
8.3	Lower Bounds for Phase Transitions: The Peierls Method	263
8.3.1	The Simple Lattice Fluid	269
8.4	Grand Partition Function Zeros and Phase Transitions	271
8.4.1	Ruelle's Theorem	273
8.4.2	The Yang-Lee Circle Theorem	276
8.4.3	Systems with Pair Interactions	279
9	Transformation Methods	283
9.1	Related Systems	283
9.2	The Wegner Transformation	284
9.2.1	Duality for the ν -State Potts Model	286
9.2.2	Duality for the Spin- $\frac{1}{2}$ Ising Model	290
9.2.3	The Weak-Graph Transformation	291
9.3	The Regular Square-Lattice Eight-Vertex Model	293
9.3.1	Symmetry Properties and Transformations	294
9.3.2	The Case of Region I	297
9.3.3	Regions and Variables	301

9.4	The Star-Triangle Transformation	304
9.4.1	The ν -State Potts Model	306
9.4.2	The Spin- $\frac{1}{2}$ Ising Model	307
10	Edge-Decorated Ising Models	311
10.1	Primary and Secondary Sites	311
10.2	Super-Exchange or Bond-Dilution.	312
10.2.1	Critical Properties and Exponent Renormalization.	314
10.3	A Ferrimagnet	319
10.3.1	The Zero-Field Axis	320
10.3.2	Non-Zero Field.	322
10.4	A Competing-Interaction Magnetic Model	324
10.5	Decoration with Orientable Molecules	326
10.6	A Decorated Lattice Fluid	332
10.6.1	Case I: A Single Vapour/Liquid Transition.	335
10.6.2	Case II: A Water-Like Model.	337
10.6.3	Case III: Maxithermal, Critical Double and Cuspoidal Points.	339
11	Transfer Matrices: Incipient Phase Transitions	345
11.1	The Transfer Matrix Formulation	345
11.1.1	The Eigen Problem	346
11.1.2	The Partition Function.	347
11.1.3	Correlation Functions and Lengths	348
11.2	Incipient Phase Transitions.	353
11.3	Using Symmetry Properties	354
11.3.1	Block-Diagonalization of the Transfer Matrix.	355
11.3.2	Applications.	359
11.4	Analysis in the Complex Plane: The Wood Method	367
11.4.1	Evolution of Partition Function Zeros	367
11.4.2	Connection Curves and Cross-Block Curves.	369
11.4.3	The Spin- $\frac{1}{2}$ Square-Lattice Ising Model	372
11.4.4	Critical Points and Exponents.	376
12	Transfer Matrices: Exactly Solved Models	381
12.1	A General Eight-Vertex Model.	381
12.1.1	A Generalized Star-Triangle Transformation.	382
12.1.2	The Solution to the GST Transformation and the Elliptic Variable Formulation	383
12.1.3	Z-Invariance.	387
12.1.4	Edge Variables and Matrix Formulation.	391
12.1.5	Square-Lattice Models.	393

12.2	Square-Lattice Ising Models	394
12.2.1	The Modified Checkerboard Ising Model	399
12.2.2	Properties of the Transfer Matrices	403
12.2.3	The Reduction to Regular Ising Models	407
12.2.4	Transfer Matrix Eigenvectors	409
12.2.5	Notational Changes	410
12.2.6	Transfer Matrix Eigenvalues	412
12.2.7	The Standard Model	421
12.3	The Square-Lattice Eight-Vertex Model	431
12.3.1	The Low-Temperature Zone $\mathcal{R}_L(\text{I})$	431
12.3.2	The Low-Temperature Zone $\mathcal{R}_L(\text{III})$	433
12.3.3	The Transfer Matrix	434
12.3.4	Analysis in Terms of Pauli Matrices	437
12.3.5	Analysis of the Transfer Matrix	442
12.3.6	The VQ Equation	447
12.3.7	The Free Energy and Magnetization	475
12.3.8	Critical Behaviour	478
12.3.9	The Coupling Form and the Ising Model Limit	482
12.3.10	The Six-Vertex Model	485
12.3.11	The Eight-Vertex Model and Universality	491
13	Dimer Models	495
13.1	The Dimer Partition Function	495
13.2	Superposition Polynomials and Pfaffians	496
13.2.1	The Square-Lattice Case	500
13.2.2	The Honeycomb-Lattice Case	504
13.3	Vertex and Ising Model Equivalences	510
13.3.1	The Five-Vertex Model	510
13.3.2	The Honeycomb-Lattice Anisotropic Ising Model	512
13.4	K-Type and O-Type Transitions	514

Part IV Series and Renormalization Group Methods

14	Series Expansions	521
14.1	The Task and the Methods	521
14.2	Moment Expansions	524
14.2.1	At Low Temperatures	525
14.2.2	At High Temperatures	530
14.2.3	Duality for Graphs	534
14.3	Cumulant Expansions	536
14.3.1	The Low-Temperature Case	539
14.3.2	The High-Temperature Case	539
14.4	The Finite-Cluster Method	540

14.5	The Finite-Lattice Method	543
14.5.1	Block-Formation and Accuracy.	544
14.5.2	Constructing Block Partition Functions	548
14.5.3	Calculating the Series	551
14.6	The Analysis of Series: Second-Order Transitions.	555
14.6.1	Late-Term Analysis.	556
14.6.2	The Ratio Method.	557
14.6.3	Padé Approximants	559
14.6.4	Differential and Algebraic Approximants	562
14.7	The Analysis of Series: First-Order Transitions.	565
15	Real-Space Renormalization Group Theory	567
15.1	The Basic Elements of the Renormalization Group	567
15.2	RG Transformations and Weight Functions	570
15.3	Fixed Points and the Linear Renormalization Group	574
15.4	Free Energy and Densities	577
15.5	Decimation for the Ising Model	579
15.5.1	In One Dimension	579
15.5.2	In Two Dimensions.	585
15.6	The Kosterlitz–Thouless Transition	588
15.7	Upper-Bound and Lower-Bound Approximations	594
15.7.1	An Upper-Bound Method	595
15.7.2	A Lower-Bound Method	599
15.8	Finite-Lattice Approximations.	603
15.9	Variational Approximations	607
15.10	Phenomenological Renormalization.	609
15.10.1	The Square-Lattice Ising Model	611
15.10.2	Other Models	613
15.10.3	More Than One Coupling	614
15.11	Other Renormalization Group Methods	615

Part V Mathematical Appendices

16	Graphs and Lattices	619
16.1	Graphs	619
16.1.1	Introduction	619
16.1.2	The Cyclomatic Number	621
16.1.3	Triangulation of Graphs.	622
16.1.4	Oriented Graphs	622
16.1.5	The Dual Graph	623
16.2	Lattices	624
16.2.1	Types of Regular Lattices	624
16.2.2	Lattice Transformations	628
16.3	Rapidity Graphs and Lattices	633

16.4	Lattice Graphs	637
16.4.1	Augmented Graphs and the Whitney Polynomial	638
16.4.2	Hopping Matrices and the Canonical Flux Distribution	638
16.4.3	Embeddings and Topologies.	639
16.4.4	Lattice Constants	640
16.4.5	Partially-Ordered Sequences of Graphs and the T Matrix	644
16.4.6	Generating the Partially-Ordered Sequence.	647
16.4.7	Incorporating Sublattices	651
16.4.8	The Guggenheim–McGlashan Approach	654
16.4.9	Further Results	656
17	Algebra	659
17.1	Catastrophe Theory	659
17.1.1	Equivalence and Determinancy	659
17.1.2	Critical Points, Codimension and Unfoldings	662
17.1.3	Symmetry Considerations.	668
17.2	Matrix Algebra	670
17.2.1	Diagonalizability.	671
17.2.2	Commutativity	672
17.2.3	Reducibility	673
17.2.4	Theorems of Perron and Frobenius	673
17.2.5	Direct Products and Traces.	675
17.2.6	Defective Matrices	675
17.2.7	Groups of Matrices	676
17.3	Groups and Representations	676
17.3.1	Representations.	678
17.3.2	Permutation Representations and Equivalence Classes	682
17.3.3	Block Diagonalization Within an Equivalence Class. . .	684
17.3.4	Symmetry Groups.	687
17.4	The Conformal Group	691
17.5	Some Transformations in the Complex Plane	693
17.6	Algebraic Functions	695
17.7	Determinants of Cyclic Matrices.	700
18	Analysis	703
18.1	Fourier Transforms in d Dimensions	703
18.1.1	Discrete Finite Lattices	703
18.1.2	A Continuous Finite Volume	705
18.1.3	A Continuous Infinite Volume	707
18.1.4	Integrals Involving Bessel Functions	708
18.1.5	Lattice Green's Functions	710

18.2	Doubly-Periodic and Quasi-Periodic Functions	711
18.3	Elliptic Integrals and Functions.	714
18.3.1	Elliptic Integrals	714
18.3.2	Jacobi Theta Functions	717
18.3.3	Jacobi Elliptic Functions	720
18.3.4	Transformations in the Elliptic Modulus	724
18.3.5	The Modified Amplitude Function	726
18.3.6	Nome Series.	727
18.3.7	Special Results and Functions for Chap. 12	728
18.3.8	Baxter's Modified Theta Functions	731
18.4	The Potts Delta Function	737
18.4.1	The $\mu = 0$ Case	740
18.4.2	The $\mu \neq 0$ Case	742
18.5	Padé, Differential and Algebraic Approximants.	743
18.5.1	Padé Approximants	743
18.5.2	Dlog Padé Approximants	748
18.5.3	Differential Approximants	750
18.5.4	Algebraic Approximants	754
	References and Author Index.	757
	Index	783

Part I

Thermodynamics, Statistical Mechanical Models and Phase Transitions

Introduction

Thermodynamics is concerned with the laws governing the macroscopic properties of systems subject to thermal change. The only explicit recognition of the atomic nature of matter which occurs in the theory is the inclusion, for systems with a variable number of particles, of the chemical potential which is the energy a particle brings on entry to the system. Otherwise a thermodynamic system is regarded as a ‘black box’ with a set of variables related by differential relationships which apply when the system is in *equilibrium*. This is defined operationally by supposing that the system is left so that it “eventually reaches a state (the equilibrium state) in which no further change is perceptible, no matter how long one waits” (Pippard 1957, p. 6). So when the system is in equilibrium the variables will all assume time-independent values. The thermodynamic context of this book is that of a system in equilibrium, where changes of thermodynamic state are considered to be quasi-static (‘infinitely slow’) and reversible.¹

The principle aim of equilibrium statistical mechanics is to derive the macroscopic properties of a system from the laws governing the microscopic behaviour of its constituent particles. In particular, statistical mechanics gives rise to the equations of thermodynamics once the thermodynamic variables have been given an interpretation within the theory.² Since this is done in terms of the expectation

¹ An account of an extension of thermodynamics to non-equilibrium irreversible processes is given by Prigogine (1967) and Lieb and Yngvason (1998, 1999, 2013) (with a summary of the ideas of Lieb and Yngvason (1998, 1999) in Lieb and Yngvason (2000)) have developed a rigorous approach to thermodynamics which in Lieb and Yngvason (2013) expands the entropy concept to non-equilibrium states.

² A classic problem in statistical mechanics (Sklar 1993, Chap. 9, Callender 1999) is to understand in what sense thermodynamics can be said to *reduce*, to statistical mechanics. In particular the relationship between the meanings of equilibrium in thermodynamics and in statistical mechanics, together with the extension to non-equilibrium, are questions with no settled resolution. For a recent attempt to develop a solution to these problems see Lavis (2005, 2008, 2011).

values associated with a probability distribution, fluctuations of dependent variables are predicted for finite systems. Exact correspondence to a thermodynamic system at equilibrium occurs only in the thermodynamic limit (Sects. 1.3 and 8.1) of an infinite system, when, except close to a phase transition, the relative fluctuations of the dependent variables disappear.

Most physical systems can exist in a number of different *phases*, distinguished by their different types of molecular or atomic order. This order may be in the spatial configurations of one or more kinds of microsystems or it may be in the orientations or conformations of the microsystems themselves.³ In the case of the vapour (gas), liquid and solid phases of, for example, water or hydrogen, the order is spatial with no order in the vapour, a short-range clustering type of order in the liquid and long-range lattice order in the solid.⁴ The most well-known example of orientational order occurs in magnetic systems where ferromagnetism corresponds to the alignment of the unpaired spins of the microsystems (Sect. 3.5). Although a simple magnetic system may possess just one type of ferromagnetic phase, more complex ferrimagnetic systems can have a large number of such phases. In the case, for example, of cerium antimonide 14 different phases have been identified by neutron diffraction experiments (Fischer et al. 1978) and specific heat analysis (Rossat-Mignod et al. 1980). These phases exhibit different forms of anisotropic spatial ordering. The occurrence of different types of anisotropic ordering is also a feature which distinguishes between the nematic, smectic and columnar phases in liquid crystals (de Gennes and Prost 1993).

Chapter 3 contains a survey of lattice models with an indication of some of the relationships between them. Although the main emphasis of this book is on classical (as distinct from quantum) systems it is of interest to discuss the quantum Heisenberg model not least because of the relationships between d -dimensional quantum systems and $(d + 1)$ -dimensional classical systems (Polyakov 1987, Sect. 1.1). These are exemplified (Sect. 3.2) by the d -dimensional quantum Ising model and the $(d + 1)$ -dimensional classical Ising model (Suzuki 1976; Björberg and Grimmett 2009), the XXZ (Heisenberg–Ising) chain and the six-vertex model (Lieb 1967a, b, c, d) and the XYZ (general Heisenberg) chain and the eight-vertex model (Sutherland 1970).

The boundaries, or *transition regions*, between phases are characterized by discontinuous or singular behaviour in one or more of the thermodynamic variables associated with the system. A classification of phase transitions according to whether particular derivatives of the free-energy density have singularities or discontinuities across the transition region was proposed by Ehrenfest (1933) (See e.g. Pippard 1957). However, it is now known that many transitions do not fit into this classification. The attempt to place all phase transitions into classes according to these kinds of criteria has been abandoned and the more modest classification of Fisher (1967) has been largely adopted. In both Ehrenfest's and Fisher's terminologies regions of *first-order phase*

³ See Ziman (1979) for a discussion of the wide variety of possible types of order.

⁴ For water this is not the complete picture. There are at least nine different ice phases distinguished by their lattice structures and proton configurations (Eisenberg and Kauzmann 1969).

transitions are those across which densities are discontinuous. These can also be identified as regions of phase coexistence. In place of the more detailed proposals of Ehrenfest, Fisher suggested that all transition regions across which the densities are continuous should be called *continuous transitions*. These may, of course, involve discontinuities or singularities in second or higher derivatives of the free-energy density. In Chap. 4 we shall augment Fisher's scheme by referring to continuous transitions for which one or more of the thermodynamic response functions is singular as *second-order* and other continuous transitions as *higher-order*. The onset of ferromagnetism in iron is accompanied by a singularity in the susceptibility and is thus a second-order transition. The freezing of water is accompanied by a discontinuous reduction in density implying a first-order transition. In some of the discussion in Chap. 4 we shall need to make a distinction between different types of transition regions. The term *coexistence region* will be used for a region of first-order transitions and *critical region* for any region in which the transition is continuous.

Ehrenfest's classification of transitions is one attempt to produce a taxonomy of phase behaviour. With a rather different emphasis, the earlier formulation of the law of corresponding states is another. Scaling theory can be viewed in a similar light. It develops from an assumed form for the free-energy density, or equation of state, close to the transition region and the underlying idea is that the probability distribution of microstates is invariant under a uniform change of length scale. The asymptotic behaviour of thermodynamic functions as a transition region is approached can be characterized by a set of *critical exponents*. Using purely thermodynamic arguments, inequalities between these exponents can be derived (see, for example, Buckingham 1972). One of the salient features of scaling theory is that, subject to the validity of its assumptions, these inequalities become equalities, the so-called *scaling laws*. This means that a smaller set of variables, the *scaling exponents*, one for each independent field, are sufficient to characterize the leading asymptotic behaviour at the transition. Although scaling theory is unable to provide numerical values for scaling exponents, the generalization of the theory given by conformal invariance is able to do so in two dimensions. A conformal transformation can be understood as a generalization of a scaling transformation in which the rescaling factor varies continuously over the physical space of the system.

The idea of scaling was proposed independently by a number of authors (Widom 1965; Domb and Hunter 1965; Kadanoff 1975; Patashinskii and Pokrovskii 1966). The presentation used in Chap. 4 is based on the subsequent developments by Hankey and Stanley (1972), Hankey et al. (1972) and Nightingale and 't Hooft (1974). The application of conformal invariance to critical phenomena was first made by Polyakov (1970) and has been much developed in recent years (Cardy 1987; Christe and Henkel 1993). This approach is of particular importance in the case of two-dimensional systems where models are associated with specific values of c , the *central charge in the Virasoro algebra* of the generators of the conformal transformation. In the final section of Chap. 4 we give a brief introduction to *Schramm–Loewner evolution* indicating how under certain conditions it becomes a standard Brownian motion, for which the diffusion constant κ is related to the central charge of the conformal transformation.

Chapter 1

Thermodynamics

1.1 Formulae and Variables

Thermodynamic variables fall into two classes, intensive and extensive and intensive variables again are of two types, fields such as the temperature¹ $\tilde{T}$, the pressure $\tilde{P}$, the magnetic field $\tilde{\mathcal{H}}$ and the chemical potential μ , and densities which are defined below. For fixed values of the intensive variables, the extensive variables such as entropy $\tilde{S}$, volume $\tilde{V}$, magnetic moment $\tilde{\mathcal{M}}$ and number of molecules M are proportional to the size of the system. For temperature $\tilde{T}$, pressure $\tilde{P}$ and magnetic field $\tilde{\mathcal{H}}$ the fields are²

$$T := k_{\text{B}}\tilde{T}, \quad P := \tilde{P}v_0, \quad \mathcal{H} := \tilde{\mathcal{H}}m_0, \quad (1.1.1)$$

where k_{B} is Boltzmann's constant, v_0 is a standard volume and m_0 is the dipole moment of a single molecule so that each quantity is of the dimensions of energy.³ The extensive variables conjugate to fields are dimensionless, those corresponding to (1.1.1) being

$$S := \tilde{S}/k_{\text{B}}, \quad V := \tilde{V}/v_0, \quad \mathcal{M} := \tilde{\mathcal{M}}/m_0, \quad (1.1.2)$$

respectively. The *internal energy* U of a system is an extensive variable arising from the kinetic energies of the molecules and the energies of interaction between them.

¹ The reason for attaching the tilde symbol to these variables (apart from μ and M) is that it is convenient to define all fields so that they have the dimensions of energy and extensive variables so that they are dimensionless; μ and M are already of these types.

² To distinguish the use of the equals sign to signify a definition, as distinct from a derived equality, we use the symbol ':= ' to represent the former.

³ In this context it is interesting to note the work of Underwood, Sutton and de Podesta at the NPL (de Podesta 2013) which aims to set a new standardised temperature scale by measuring, through a determination of the speed of sound in a monotonic gas at low densities, the kinetic energy per particle and thus $k_{\text{B}}\tilde{T}$ for the gas.

For a system with $n_f + 1$ fields $T, \xi_i, i = 1, \dots, n_f$, with conjugate extensive variables $S, Q_i, i = 1, \dots, n_f$,

$$dU = TdS + \sum_{i=1}^{n_f} \xi_i dQ_i \quad (1.1.3)$$

applies to a quasi-static change between equilibrium states.⁴ It follows that⁵

$$T = \left(\frac{\partial U}{\partial S} \right)_{Q_i}, \quad \xi_i = \left(\frac{\partial U}{\partial Q_i} \right)_{S, Q_j}, \quad i = 1, \dots, n_f \quad (j \neq i). \quad (1.1.4)$$

Suppose that, with its internal constitution fixed, the size of the system is increased by a factor λ . Since all extensive variables will increase by the factor λ ,

$$U(\lambda Q_i, \lambda S) = \lambda U(Q_i, S). \quad (1.1.5)$$

Differentiation with respect to λ , and substitution from (1.1.4) gives

$$U = TS + \sum_{i=1}^{n_f} \xi_i Q_i, \quad (1.1.6)$$

when λ is put equal to 1, and

$$SdT + \sum_{i=1}^{n_f} Q_i d\xi_i = 0. \quad (1.1.7)$$

A subset, $Q_1, \dots, Q_\eta$, of the extensive variables can now be replaced as independent variables by their conjugate fields $\xi_1, \dots, \xi_\eta$ to give, using a Legendre transformation, the *enthalpy*

$$H_\eta := U - \sum_{i=1}^{\eta} \xi_i Q_i = TS + \sum_{i=\eta+1}^{n_f} \xi_i Q_i, \quad (1.1.8)$$

⁴ In this formula the contributions to the summation are assumed to be such that each field ξ_i causes an increase in the corresponding extensive variable Q_i . As we shall see in Sect. 1.4 this is the case for a magnetic system, when $\xi_i = \mathcal{H}$ and $Q_i = \mathcal{M}$. However, for a fluid system the pressure $\xi_i = P$ leads to a decrease in volume $Q_i = V$. Thus the contribution to (1.1.3) will in this case be $-PdV$.

⁵ Because of the various choices possible for the independent variables in any thermodynamic system, it is often necessary to specify, using subscripts, the variable or variables kept constant during a partial differentiation. In general this is not necessary for free energies which are defined to be functions of a specific set of variables.

with

$$dH_\eta = TdS - \sum_{i=1}^{\eta} Q_i d\xi_i + \sum_{i=\eta+1}^{n_f} \xi_i dQ_i. \quad (1.1.9)$$

The replacement of $Q_1, \dots, Q_\eta$ by their conjugate fields $\xi_1, \dots, \xi_\eta$ means that the system is open to energy interchange by way of these fields. The appropriate *free energy* is now⁶

$$F_\eta := H_\eta - TS = U - TS - \sum_{i=1}^{\eta} \xi_i Q_i = \sum_{i=\eta+1}^{n_f} \xi_i Q_i, \quad (1.1.10)$$

$$dF_\eta = -SdT - \sum_{i=1}^{\eta} Q_i d\xi_i + \sum_{i=\eta+1}^{n_f} \xi_i dQ_i. \quad (1.1.11)$$

From (1.1.10) and (1.1.11),

$$Q_i = -\frac{\partial F_\eta}{\partial \xi_i}, \quad i = 1, \dots, \eta, \quad S = -\frac{\partial F_\eta}{\partial T}, \quad (1.1.12)$$

$$\xi_i = \frac{\partial F_\eta}{\partial Q_i}, \quad i = \eta + 1, \dots, n_f, \quad H_\eta = \frac{\partial(F_\eta/T)}{\partial(1/T)}.$$

Finally we observe for later reference that

$$F_\eta(Q_{\eta+1}, \dots, Q_{n_f}, T, \xi, \dots, \xi_\eta) = F_{\eta-1}(Q_\eta, \dots, Q_{n_f}, T, \xi, \dots, \xi_{\eta-1}) - \xi_\eta Q_\eta, \quad (1.1.13)$$

where Q_η on the right-hand side of this equation is given, from (1.1.12), as a solution of

$$\frac{\partial F_{\eta-1}}{\partial Q_\eta} = \xi_\eta. \quad (1.1.14)$$

If we were to include Q_η explicitly on left-hand side of (1.1.13) by replacing $F_\eta(Q_{\eta+1}, \dots, Q_{n_f}, T, \xi, \dots, \xi_\eta)$ by $\tilde{F}_\eta(Q_\eta, Q_{\eta+1}, \dots, Q_{n_f}, T, \xi, \dots, \xi_\eta)$, then (1.1.14) is the condition for the required value of Q_η to be a stationary point of $\tilde{F}_\eta$ with respect to Q_η .

⁶ For enthalpy and free energies and later in the book for Hamiltonians and partition functions the subscript η (or its particular value) is used to indicate the number of external fields (excluding the temperature). In later chapters where there is no possibility of ambiguity this subscript is omitted.

1.2 The Field-Density Representation

The formulation of thermodynamics, given by (1.1.3)–(1.1.11), relates pairs of variables one of which is a field and the other an extensive variable; it is thus referred to as the *field-extensive variable representation*. We shall now introduce a transformation in which extensive variables are replaced by densities to give the *field-density representation*. Densities have the advantage that they remain finite in the thermodynamic limit of an infinite system (Sect. 1.3).

From (1.1.6) and (1.1.10) we see that, in order for the free energy $F_\eta(T, \xi_1, \dots, \xi_\eta, Q_{\eta+1}, \dots, Q_{n_f})$ not to be identically zero, we must have $\eta < n_f$; there must be at least one independent extensive variable Q_{n_f} . We take this as a fixed measure $Q := Q_{n_f}$ of the *size* of the system, with $dQ_{n_f} = 0$ in (1.1.3), (1.1.9) and (1.1.11). The remaining independent extensive variables are replaced as arguments of F_η by independent densities

$$\rho_i := Q_i/Q, \quad i = \eta + 1, \dots, n_f - 1, \quad (1.2.1)$$

with the free-energy density

$$f_\eta(Q, \rho_{\eta+1}, \dots, \rho_{n_f-1}, T, \xi_1, \dots, \xi_\eta) := \frac{F_\eta(Q, \rho_{\eta+1}, \dots, \rho_{n_f-1}, T, \xi_1, \dots, \xi_\eta)}{Q}. \quad (1.2.2)$$

From (1.1.6), (1.1.8) and (1.1.10),

$$u = sT + \xi_{n_f} + \sum_{i=1}^{n_f-1} \rho_i \xi_i, \quad (1.2.3)$$

$$h_\eta = u - \sum_{i=1}^{\eta} \rho_i \xi_i = sT + \xi_{n_f} + \sum_{i=\eta+1}^{n_f-1} \rho_i \xi_i, \quad (1.2.4)$$

$$f_\eta = h_\eta - sT = \xi_{n_f} + \sum_{i=\eta+1}^{n_f-1} \xi_i \rho_i, \quad (1.2.5)$$

and, from (1.1.3), (1.1.7), (1.1.9) and (1.1.11),

$$du = Tds + \sum_{i=1}^{n_f-1} \xi_i d\rho_i, \quad (1.2.6)$$

$$0 = sdT + d\xi_{n_f} + \sum_{i=1}^{n_f-1} \rho_i d\xi_i, \quad (1.2.7)$$

$$dh_\eta = du - \sum_{i=1}^{\eta} \rho_i d\xi_i - \sum_{i=1}^{\eta} \xi_i d\rho_i, \quad (1.2.8)$$

$$df_\eta = -s dT - \sum_{i=1}^{\eta} \rho_i d\xi_i + \sum_{i=\eta+1}^{n_f-1} \xi_i d\rho_i. \quad (1.2.9)$$

Another transformation, which is of importance for the development of scaling theory and the renormalization group, replaces the fields by dimensionless ratios of the fields with respect to T ; thus yielding the *coupling-extensive variable* and *coupling-density representations*, which are described in Sect. 2.3.

1.3 The Thermodynamic Limit

The system described above is *finite* with all the independent extensive variables Q_i , $i = \eta + 1, \dots, N_f$ having finite values. However, for statistical mechanics it is important to consider *infinite* systems since this is a necessary condition for the occurrence of a phase transition (see p. 253). There are two possible ways of dealing with this problem. The one adopted in this book is that in which an infinite system is obtained by taking the *thermodynamic limit* $Q \rightarrow \infty$ with the independent densities ρ_i , $i = \eta + 1, \dots, n_f - 1$ fixed. The conditions for the existence of the thermodynamic limit for a one-component lattice fluid are given in Sect. 8.1. An alternative to this is to consider an infinite system from the outset. For the sake of comparison a brief summary of this approach is given in Sect. 8.2.

Of course, the free-energy density, as defined by (1.2.2), and its field derivatives, which are the dependent densities s and ρ_i , $i = 1, \dots, \eta$, are all, in general, functions of Q . The thermodynamic limit can be achieved only if it leads to a finite value $f_\eta(\rho_{\eta+1}, \dots, \rho_{n_f-1}, T, \xi_1, \dots, \xi_\eta)$ for the free-energy density, and hence for the dependent densities.

In this book we shall make extensive use of the field-density representation defined in Sect. 1.2 and the coupling-density representation which will be defined in Sect. 2.3. In each case the free-energy density and the other dependent densities will be functions of the size parameter Q until the thermodynamic limit is applied. In many cases the form of the analysis is the same whether or not the limit has been applied and we shall not display the argument Q . In other cases it will be important to include a specific reference to whether the system is finite or infinite. For a system on a lattice $\mathcal{N}$ of $N := |\mathcal{N}|$ sites we often include $\mathcal{N}$ as an argument with $\mathcal{L}$ denoting the corresponding infinite lattice resulting from the limit $N \rightarrow \infty$.

1.4 Particular Response Functions

In Chap. 4 critical exponents are introduced. The most commonly used of these are associated with the limiting properties of densities and response functions, such as susceptibility, heat capacity and compressibility. These are given in the field-extensive variable representation by the second-order partial derivatives of the free energy F_η . They can in consequence be expressed easily in the field-density representation in terms of the second-order partial derivatives of the free-energy density f_η . Their forms in the coupling-density representation are more complicated as we shall see in Sect. 2.3. Here the discussion is restricted to the case $n_f = 2$, setting $\rho := \rho_1$ and $\xi := \xi_1$. From (1.2.5) and (1.2.9),

$$df_0 = -s dT + \xi d\rho, \quad (1.4.1)$$

$$df_1 = -s dT - \rho d\xi, \quad (1.4.2)$$

$$f_1 = \xi_2 = f_0 - \rho \xi. \quad (1.4.3)$$

There are four principle response functions

$$c_\xi := k_B T \left(\frac{\partial s}{\partial T} \right)_\xi, \quad (1.4.4)$$

$$c_\rho := k_B T \left(\frac{\partial s}{\partial T} \right)_\rho, \quad (1.4.5)$$

$$\varphi_T := T \left(\frac{\partial \rho}{\partial \xi} \right)_T, \quad (1.4.6)$$

$$\alpha_\xi := T \left(\frac{\partial \rho}{\partial T} \right)_\xi. \quad (1.4.7)$$

c_ξ and c_ρ are the *heat capacities* at constant field and density respectively, φ_T has the limiting properties of the *compressibility of a fluid* or the *susceptibility of a magnet* and α_ξ has the limiting properties of the *coefficient of thermal expansion*. The standard formula

$$c_\xi - c_\rho = k_B \frac{\alpha_\xi^2}{\varphi_T}, \quad (1.4.8)$$

relating the heat capacities at constant field and density can now be derived.

1.4.1 Magnetic Systems

Here the size parameter is $Q := M$, the number of magnetic dipoles, with $f_1 = \xi_2 := \mu$, the chemical potential per dipole. Then $\rho = m = \mathcal{M}/M$, the magnetization density and $\xi = \mathcal{H}$.⁷ $c_{\mathcal{H}}$ and c_m , given by (1.4.4) and (1.4.5) respectively, are the heat capacities per dipole at constant field and magnetization respectively. The *isothermal magnetic susceptibility* is

$$\chi_T := \frac{C\varphi_T}{T} = C \left(\frac{\partial m}{\partial \mathcal{H}} \right)_T, \quad \text{where} \quad C := \frac{m_0^2}{v_0 \mu_0}, \quad (1.4.9)$$

m_0 and v_0 being, as defined in Sect. 1.1 and μ_0 being the permeability of free space, $4\pi 10^{-7}$ Wb/(A m). The inclusion of the constant C , which is of the dimensions of energy, makes the susceptibility dimensionless.

1.4.2 Fluid Systems

Here one choice of size parameter is $Q := M$ the number of particles in the gas, with $f_1 = \xi_2 := \mu$, the chemical potential per particle. Then $v = V/M$ and $\xi = -P$.⁸ This is the *constant-pressure distribution* in statistical mechanics and c_P and c_v , given by (1.4.4) and (1.4.5) respectively, are the heat capacities per particle at constant pressure and volume respectively. The *isothermal compressibility* is

$$\mathcal{K}_T := -\frac{\varphi_T}{Tv} = -\frac{1}{v} \left(\frac{\partial v}{\partial P} \right)_T \quad (1.4.10)$$

and the *isobaric coefficient of thermal expansion* is

$$\mathcal{A}_P := -\frac{\alpha_P}{Tv} = -\frac{1}{v} \left(\frac{\partial v}{\partial T} \right)_P. \quad (1.4.11)$$

An alternative form of the field-density representation for fluids is achieved by scaling with respect to volume V , or, in the case of a lattice fluid (Sect. 3.5.2), the number of lattice sites N , rather than the number of particles. The independent field is now the chemical potential μ with conjugate density $\rho := M/V$ and $f_1 := -P$. The expression (1.4.10) for the isothermal compressibility is replaced by

$$\mathcal{K}_T := \frac{1}{\rho^2} \left(\frac{\partial \rho}{\partial \mu} \right)_T. \quad (1.4.12)$$

This choice gives the *grand-canonical distribution* in statistical mechanics.

⁷ When we consider lattice systems with a dipole on each lattice site $M = N$ the number of lattice sites.

⁸ That the field must correspond to *minus* the pressure has already been noted in Sect. 1.1.

Chapter 2

Statistical Mechanics

2.1 Distributions

The formalism described in Chap. 1 can be regarded as a macroscopic perspective on a system with a microscopic structure consisting of a collection of atoms or molecules of one or more types. The numbers of particles of each type, or component, will be among the set of extensive variables $Q_1, \dots, Q_{n_f}$.

A configuration of the particles of the system is called a *microstate* and is denoted by σ . For each microstate there is a value $E(\sigma)$ of the *statistical mechanical internal energy* E and values $\widehat{Q}_i(\sigma)$, $i = 1, \dots, \eta$, of the dependent statistical mechanical extensive variables.¹ The Hamiltonian²

$$\widehat{H}_\eta(\sigma) := E(\sigma) - \sum_{i=1}^{\eta} \xi_i \widehat{Q}_i(\sigma), \quad (2.1.1)$$

is a crucial quantity in the statistical mechanics of the system with independent variables $Q_{\eta+1}, \dots, Q_{n_f}, T, \xi_1, \dots, \xi_\eta$. It is the microscopic equivalent of the thermodynamic enthalpy defined by (1.1.8).

As indicated above we shall be concerned with the case where the system is in equilibrium and then the probability function for the *distribution* of σ is

$$p_\eta(T; \sigma) := \frac{\exp[-\widehat{H}_\eta(\sigma)/T]}{Z_\eta(T)}, \quad (2.1.2)$$

¹ In cases where we wish to use the same letter for a thermodynamic variable and its classical statistical mechanical counterpart we make the distinction by adding a ‘hat’ to the statistical mechanical variable. In particular this relationship applies to the enthalpy H_η and the Hamiltonian $\widehat{H}_\eta$ (Eq. 2.1.7).

² For a comment concerning the usage of the subscript η , specifying the number of independent fields (and hence the number of independent extensive variables, or densities) see the footnote on page 7. For the sake of brevity it is convenient in most cases to drop reference to the independent fields and extensive variables as arguments, retaining only T .

where

$$Z_\eta(T) := \sum_{\{\sigma\}} \exp[-\widehat{H}_\eta(\sigma)/T], \quad (2.1.3)$$

is the *partition function*.³ This is usually referred to as the *Gibbs distribution*. In the case when the system is a perfect gas it is the *Maxwell-Boltzmann distribution*. The classic account which justifies the role of (2.1.2) as the equilibrium distribution in statistical mechanics is that of Khinchin (1949), who established the result for the case $\eta = 0$, called the *canonical distribution*. This situation can be pictured as a system in a heat bath which controls the temperature. Other cases, where, for example, $\eta = 1$ and ξ_1 is the chemical potential (called the *grand-canonical distribution*, where in addition to the heat bath the system has permeable walls with the passage of particles controlled by the chemical potential), or $\eta = 1$ and ξ_1 is the pressure (called the *constant-pressure distribution*, where for a fluid system the volume is controlled by a piston used to vary the pressure), can be established in a similar way.⁴ Functions of the microstate of the system, like $\widehat{C}(\sigma)$, are random variables for which the expectation value

$$\langle \widehat{C}(\sigma) \rangle := \sum_{\{\sigma\}} p_\eta(T; \sigma) \widehat{C}(\sigma), \quad (2.1.4)$$

variance and covariance between pairs of such variables play a fundamental role in connecting the distribution to thermodynamics (Sect. 2.1.2).

2.1.1 Quantum Systems

This is a text on *classical* statistical mechanics. However, it is useful to give a brief indication of the corresponding development for quantum systems where the microstate σ will be some set of quantum operators. Then the Hamiltonian $\widehat{H}_\eta(\sigma)$ is an operator and $p_\eta(T; \sigma)$ is the *density operator*, which in terms of a representation of the system becomes the *density matrix*. Equations (2.1.3) and (2.1.4) are replaced by

$$Z_\eta(T) := \text{Trace}\{\exp[-\widehat{H}_\eta(\sigma)/T]\}, \quad (2.1.5)$$

³ In this and all subsequent formulae, the summation is over all microstates σ compatible with the given values of the independent extensive variables.

⁴ The different distributions, for different values of η are usually called *ensembles*. This term is associated with the (relative-frequency flavoured) picture of a collection (or ensemble) of systems whose phase points have the density (2.1.2) in the space of microstates (Gibbs 1902).

and

$$\langle \widehat{C}(\sigma) \rangle := \text{Trace}\{\rho_\eta(T; \sigma) \widehat{C}(\sigma)\}. \quad (2.1.6)$$

A *complete solution* to this problem, in both the quantum and statistical mechanical sense, would be a determination of all the eigenvalues and eigenvectors of the Hamiltonian. The complicating feature of a quantum system, as can be seen from the brief account of the Heisenberg model in Sect. 3.2, is that σ may contain non-commuting operators.

2.1.2 The Connection to Thermodynamics

This is achieved by the identifications

$$H_\eta := \langle \widehat{H}_\eta \rangle = -\frac{\partial \ln(Z_\eta)}{\partial (1/T)}, \quad (2.1.7)$$

$$Q_i := \langle \widehat{Q}_i \rangle = T \frac{\partial \ln(Z_\eta)}{\partial \xi_i}, \quad i = 1, \dots, \eta, \quad (2.1.8)$$

$$\xi_i := \left\langle \frac{\partial \widehat{H}_\eta}{\partial Q_i} \right\rangle = -T \frac{\partial \ln(Z_\eta)}{\partial Q_i}, \quad i = \eta + 1, \dots, n_f. \quad (2.1.9)$$

The ‘averaging’ relations (2.1.7)–(2.1.9) become identical to (1.1.12) if we set

$$F_\eta(T) := -T \ln[Z_\eta(T)]. \quad (2.1.10)$$

In the special case of the canonical distribution, $\eta = 0$, $\widehat{H}_0(\sigma)$ is the energy $E(\sigma)$, H_0 is the internal energy U and F_0 is the Helmholtz free energy, usually denoted by A . The second derivatives of F_η with respect to the independent variables $T, \xi_1, \dots, \xi_\eta$ are the *response functions*. It is not difficult to show that the response function with respect to the variables ξ_i, ξ_j is related to the covariance of the random variables $\widehat{Q}_i$ and $\widehat{Q}_j$ by

$$\frac{\partial^2 F_\eta}{\partial \xi_i \partial \xi_j} = -\frac{\partial Q_j}{\partial \xi_i} = -\frac{\partial Q_i}{\partial \xi_j} = -\frac{1}{T} \text{Cov}[\widehat{Q}_i, \widehat{Q}_j]. \quad (2.1.11)$$

This is a *fluctuation-response function relation*.

It follows from (2.1.3) and (2.1.10) that

$$F_\eta(T) = -T \ln \left\{ \sum_{\{Q_\eta\}} \exp[\{\xi_\eta Q_\eta - F_{\eta-1}(T)\}/T] \right\}. \quad (2.1.12)$$

Differentiating with respect to T and using the expression for S given by (1.1.10) gives

$$F_\eta(T) = \langle F_{\eta-1}(T) \rangle - \xi_\eta \langle \mathcal{Q}_\eta \rangle, \quad (2.1.13)$$

where the expectations are taken with respect to the η distribution. This formula is the statistical mechanical analogue of (1.1.13).

2.2 Variations of the Probability Function

For the Hamiltonian (2.1.1), let

$$\mathcal{F}_\eta(\Pi) := \sum_{\{\sigma\}} \Pi(\sigma) \hat{H}_\eta(\sigma) + T \sum_{\{\sigma\}} \Pi(\sigma) \ln[\Pi(\sigma)], \quad (2.2.1)$$

where $\Pi(\sigma) \geq 0$ satisfies the normalization condition

$$\sum_{\{\sigma\}} \Pi(\sigma) = 1. \quad (2.2.2)$$

It is easy to verify, from (2.1.2) and (2.1.10), that

$$\mathcal{F}(p_\eta) = F_\eta \quad (2.2.3)$$

and thus, from (1.1.10) and (2.1.7), that

$$S = - \sum_{\{\sigma\}} p_\eta(\sigma) \ln[p_\eta(\sigma)]. \quad (2.2.4)$$

Now consider variations of $\Pi(\sigma)$ in (2.2.1), subject to the normalization condition (2.2.2), for which we use the undetermined multiplier λ . Then

$$\begin{aligned} \delta \mathcal{F}_\eta(\Pi) = \sum_{\{\sigma\}} \left\{ [\hat{H}_\eta(\sigma) + T \ln \Pi(\sigma) + T + \lambda] \delta \Pi(\sigma) \right. \\ \left. + T \frac{(\delta \Pi(\sigma))^2}{2 \Pi(\sigma)} + O((\delta \Pi(\sigma))^3) \right\}, \end{aligned} \quad (2.2.5)$$

yielding a minimum given by

$$\hat{H}_\eta(\sigma) + T \ln \Pi(\sigma) + T + \lambda = 0, \quad (2.2.6)$$

which is the unique solution $\Pi(\sigma) = p_\eta(\sigma)$.

2.3 Coupling Representations

In general the statistical mechanical internal energy is a linear function of a set of energy parameters ε_ℓ of the form

$$E(\varepsilon_\ell; \sigma) := \sum_{\ell=1}^{n_e} E_\ell(\sigma) \varepsilon_\ell. \quad (2.3.1)$$

These parameters can be regarded as *internal fields* and it will be seen that the formulae of Sect. 2.1 can be simplified if the temperature is absorbed into the ratios

$$K_\ell := \varepsilon_\ell / T, \quad \ell = 1, \dots, n_e, \quad L_i := \xi_i / T, \quad i = 1, \dots, n_f, \quad (2.3.2)$$

called *internal* and *external couplings* respectively. This then gives

$$\hat{H}_\eta(K_\ell, L_i; \sigma) = \sum_{\ell=1}^{n_e} E_\ell(\sigma) K_\ell - \sum_{i=1}^{\eta} \hat{Q}_i(\sigma) L_i. \quad (2.3.3)$$

The discussion of scaling in Chap. 4 shows that the distinction between the internal couplings K_ℓ and the external couplings L_i is physical rather than mathematical.⁵ The latter arise from interactions between the system and external fields, like the magnetic field $\mathcal{H}$ or the chemical potential μ , whereas the former arise from the energies of interaction between the microsystems. The external couplings are, in general, open to change by the experimenter, either by individual variations of the fields ξ_i or together by variations of T . The internal couplings can be changed only by means of a change of T , so that

$$dK_\ell = -K_\ell dT/T, \quad \ell = 1, \dots, n_e. \quad (2.3.4)$$

From (2.1.3),⁶

$$Z_\eta(K_\ell, L_i) := \sum_{\{\sigma\}} \exp[-\hat{H}_\eta(K_\ell, L_i; \sigma)]. \quad (2.3.5)$$

Then, from (2.1.8)–(2.1.10),

$$\Phi_\eta(K_\ell, L_i) := F_\eta(T, \varepsilon_\ell, \xi_i) / T = -\ln[Z_\eta(K_\ell, L_i)] \quad (2.3.6)$$

⁵ Although we do need to be careful about the consequences of the difference of sign in the two terms on the right of (2.3.10).

⁶ Here it is clearer if we explicitly display representative members K_ℓ and L_i of the n_e internal couplings (all independent) and the η independent external couplings.

is the *dimensionless free energy*, with

$$Q_i = \langle \widehat{Q}_i(\boldsymbol{\sigma}) \rangle = -\frac{\partial \Phi_\eta}{\partial L_i}, \quad i = 1, \dots, \eta, \quad L_i = \frac{\partial \Phi_\eta}{\partial Q_i}, \quad i = \eta + 1, \dots, n_f, \quad (2.3.7)$$

$$\langle E_\ell(\boldsymbol{\sigma}) \rangle = \frac{\partial \Phi_\eta}{\partial K_\ell}, \quad \ell = 1, \dots, n_e. \quad (2.3.8)$$

Since $\Phi_\eta(K_\ell, L_i)$ is a function of T only through the couplings this must also be the case for the expectations $\langle \widehat{Q}_i(\boldsymbol{\sigma}) \rangle$ and $\langle E_\ell(\boldsymbol{\sigma}) \rangle$ and thus, from (2.3.1) and (2.3.3),

$$U(K_\ell) := \sum_{\ell=1}^{n_e} \langle E_\ell(\boldsymbol{\sigma}) \rangle K_\ell, \quad (2.3.9)$$

$$H_\eta(K_\ell, L_i) := \sum_{\ell=1}^{n_e} \langle E_\ell(\boldsymbol{\sigma}) \rangle K_\ell - \sum_{i=1}^{\eta} \langle \widehat{Q}_i(\boldsymbol{\sigma}) \rangle L_i. \quad (2.3.10)$$

The *coupling-density representation*, like the field-density representation, is obtained by forming densities by dividing the independent extensive variables by a size parameter which we again take to be $Q := Q_{n_f}$. Displaying a representative independent density ρ_i and extensive variable $Q_i := \rho_i Q$ explicitly, the *dimensionless free-energy density*⁷

$$\phi_\eta(\rho_i, K_\ell, L_i) := \frac{\Phi_\eta(Q, \rho_i Q, K_\ell, L_i)}{Q} \quad (2.3.11)$$

is a function of the independent couplings and densities, without explicit reference to the temperature, making them convenient tools for the formulation of scaling theory and renormalization group methods.

We can also, of course, defined densities $\hat{\rho}_i := \widehat{Q}_i/Q$, for the dependent extensive variables, $i = 1, 2, \dots, \eta$. These are random variables with $\rho_i := \langle \hat{\rho}_i \rangle$ defining thermodynamic densities and, from (2.1.11),

$$\text{Var}[\hat{\rho}_i] = -\frac{1}{Q} \frac{\partial^2 \phi_\eta}{\partial L_i^2} \quad i = 1, 2, \dots, \eta. \quad (2.3.12)$$

In the thermodynamic limit $Q \rightarrow \infty$, unless the response function $\partial^2 \phi_\eta / \partial L_i^2$ diverges faster than Q , fluctuations in $\hat{\rho}_i$ will tend to zero and the distinction between dependent and independent densities, and hence between the different choices of

⁷ In later work, when the context makes it unnecessary we shall omit the qualification ‘dimensionless’.

distribution (canonical, grand-canonical etc.), will disappear. We have a *thermodynamic equivalence of distributions* away from the exceptional points where the response functions diverge, which are associated with phase transitions (Sect. 4.1).

As in Sect. 1.3 we now omit explicit reference to Q unless we need to be specific about the system being finite. With

$$u_\ell := \langle E_\ell(\sigma) \rangle / Q, \quad v := u/T = \sum_{\ell=1}^{n_e} u_\ell K_\ell, \quad (2.3.13)$$

from (1.2.3)–(1.2.5),

$$v = s + L_{n_f} + \sum_{i=1}^{n_f-1} \rho_i L_i, \quad (2.3.14)$$

$$\phi_\eta = L_{n_f} + \sum_{i=\eta+1}^{n_f-1} \rho_i L_i \quad (2.3.15)$$

and, from (1.2.6)–(1.2.9) and (2.3.4),

$$ds = \sum_{\ell=1}^{n_e} K_\ell du_\ell - \sum_{i=1}^{n_f-1} L_i d\rho_i, \quad (2.3.16)$$

$$0 = \sum_{\ell=1}^{n_e} u_\ell dK_\ell - dL_{n_f} - \sum_{i=1}^{n_f-1} \rho_i dL_i, \quad (2.3.17)$$

$$d\phi_\eta = \sum_{\ell=1}^{n_e} u_\ell dK_\ell - \sum_{i=1}^{\eta} \rho_i dL_i + \sum_{i=\eta+1}^{n_f-1} L_i d\rho_i. \quad (2.3.18)$$

2.3.1 The Case $n_f = 2$

As in Sect. 1.4 we set $\rho := \rho_1$ with the one independent external coupling $L := L_1$. In place of the temperature we now have K_ℓ , $\ell = 1, \dots, n_e$ internal couplings. To reduce these to one variable which corresponds to thermal variation we set $\varepsilon := \varepsilon_1$, $K := \varepsilon/T$ with the parameters $\kappa_\ell := \varepsilon_\ell/\varepsilon$ kept constant and

$$v = u K, \quad u := \sum_{\ell=1}^{n_e} u_\ell K_\ell. \quad (2.3.19)$$

Now there are two free-energy densities: $\phi_0(K, \rho)$ and $\phi_1(K, L)$ with

$$d\phi_0 = u dK + L d\rho, \quad (2.3.20)$$

$$d\phi_1 = u dK - \rho dL. \quad (2.3.21)$$

They are related by the Legendre transformation

$$\phi_1 = \phi_0 - L\rho, \quad (2.3.22)$$

where $\rho = \rho(K, L)$ is the solution of

$$\frac{\partial \phi_0}{\partial \rho} = L. \quad (2.3.23)$$

Alternatively

$$\phi_0 = \phi_1 + \rho L, \quad (2.3.24)$$

where $L = L(K, \rho)$ is the solution of

$$\frac{\partial \phi_1}{\partial L} = -\rho. \quad (2.3.25)$$

In Sects. 1.4.1 and 1.4.2 we considered two examples of $n_f = 2$, a magnetic system where the two extensive variables are the magnetization $\mathcal{M}$ and the number of dipoles M and a fluid system where the two extensive variables are the volume V and the number of particles M . In the former case there is an obvious choice for scaling; taking the magnetization per dipole as the density, conjugate to the magnetic field. For the fluid the choice is less obvious: volume per particle, with conjugate field $-P$, or particles per unit volume, with conjugate field μ . It is useful to list the relationship between these two choices. Denoting one using a prime,

$$\begin{aligned} \rho' &= 1/\rho, & u' &= \rho' u, & \phi'(K, \rho') &= \rho' \phi_0(K, \rho), \\ L' &= \phi_1(K, L), & L &= \phi'(K, L'). \end{aligned} \quad (2.3.26)$$

It is now easy to check that Eqs. (2.3.20)–(2.3.25) take exactly the same form in the primed quantities. An example of the interchange between different choices of scaling variable is given for the van der Waals equation in Sect. 5.2.

We shall, in Chap. 4, need forms for the response functions, defined in Sect. 1.4, in terms of the coupling-density representation. From (2.3.20) and (2.3.21),

$$u = \frac{\partial \phi_0}{\partial K} = \frac{\partial \phi_1}{\partial K} \quad (2.3.27)$$

and, from (2.3.15),

$$s = -\phi_0 + uK = -\phi_1 + uK - L\rho. \quad (2.3.28)$$

Then from (1.4.5)–(1.4.7), (2.3.20)–(2.3.22) and (2.3.25),

$$c_\rho = -k_B K \left(\frac{\partial s}{\partial K} \right)_\rho = -k_B K^2 \left(\frac{\partial u}{\partial K} \right)_\rho = -k_B K^2 \frac{\partial^2 \phi_0}{\partial K^2}, \quad (2.3.29)$$

$$\begin{aligned} c_\xi &= -k_B \left\{ K \left(\frac{\partial s}{\partial K} \right)_L + L \left(\frac{\partial s}{\partial L} \right)_K \right\} \\ &= -k_B \left\{ K^2 \left(\frac{\partial u}{\partial K} \right)_L - 2KL \left(\frac{\partial \rho}{\partial K} \right)_K - L^2 \left(\frac{\partial \rho}{\partial L} \right)_K \right\} \\ &= -k_B \left(K^2 \frac{\partial^2 \phi_1}{\partial K^2} + 2KL \frac{\partial^2 \phi_1}{\partial K \partial L} + L^2 \frac{\partial^2 \phi_1}{\partial L^2} \right), \end{aligned} \quad (2.3.30)$$

$$\varphi_T = \left(\frac{\partial \rho}{\partial L} \right)_K = -\frac{\partial^2 \phi_1}{\partial L^2}, \quad (2.3.31)$$

$$\alpha_\xi = -K \left(\frac{\partial \rho}{\partial K} \right)_L - L \left(\frac{\partial \rho}{\partial L} \right)_K = L \frac{\partial^2 \phi_1}{\partial L^2} + K \frac{\partial^2 \phi_1}{\partial L \partial K}. \quad (2.3.32)$$

Alternatively c_ρ can be obtained from (2.3.30)–(2.3.32) using (1.4.8).

2.4 Lattice Systems

In Sect. 16.2.1 we define what is meant by a lattice, give diagrams for the regular lattices and consider the question of finite lattices and periodic boundary conditions. Here we need simply to assume that the lattice, denote by $\mathcal{N}$, is one of the regular types with periodic boundary conditions and $N := |\mathcal{N}|$ sites, which is taken to be the one independent extensive variable Q_{n_f} .⁸

Chapter 3 contains a survey of lattice models, some, but not all, of which are treated in this book. They are distinguished by the types of cooperative interaction which contribute to the Hamiltonian. These fall broadly into two classes, those for which the state variables are associated with lattice sites and those for which they are associated with lattice edges (connecting first-neighbour pairs of sites).

⁸ Of course, all the extensive quantities in the system, including the partition function and Hamiltonian are functions of N . However, it, or the lattice signifier $\mathcal{N}$, will be displayed explicitly only when it seems useful to do so. In the case of densities like ϕ it is also sometimes useful to indicate the underlying lattice by including the infinite lattice signifier $\mathcal{L}$.

2.4.1 Site-Variable Models

There is a microsystem at each site of the lattice and the state of the microsystem at site $\mathbf{r}$ is specified by the variable $\sigma(\mathbf{r})$. The symbol σ represents the collection or vector specifying a set of values of the states of all the microsystems on the lattice. The simplest contributions to the Hamiltonian are single-site terms with the extensive variables $\hat{Q}_i$ expressed in the form

$$\hat{Q}_i(\sigma) = \sum_{\{\mathbf{r}\}} q_i(\mathbf{r}; \sigma), \quad i = 1, \dots, n_f - 1, \quad (2.4.1)$$

where $q_i(\mathbf{r}; \sigma)$ is a function of the one component $\sigma(\mathbf{r})$ of σ at $\mathbf{r}$. Unless the states of the microsystems are very complicated there will be only a rather small number of independent extensive variables and different functions q_i .

The extensive internal energy variables E_ℓ in the Hamiltonian involve interactions between groups of more than one site. They can be expressed in the form

$$E_\ell(\sigma) = \sum_{\{\mathbf{r}\}} \epsilon_\ell(\mathbf{r}; \sigma), \quad \ell = 1, \dots, n_e, \quad (2.4.2)$$

where $\epsilon_\ell(\mathbf{r}; \sigma)$ is a function of $\sigma(\mathbf{r})$ and the state variables of some set of sites around $\mathbf{r}$. In most cases of interest the functions $\epsilon_\ell(\mathbf{r}; \sigma)$ are taken to be *local* in the sense that they involve only sites quite close to the site $\mathbf{r}$, although in Sects. 3.2 and 3.3 we consider cases where interactions, involving pairs of sites, are allowed to have an arbitrary range.

The simplest *pair-interaction models* are those where the terms are restricted to first-neighbour sites and this is the case for most site-variable models considered here. However, it is sometimes of interest to include second-neighbour interactions, particularly when there is ‘conflict’ between the types of ordering encouraged by the two different interactions (Sect. 3.5.1.1).

Another type of site-variable model which has attracted interest is the *face model* where the interaction term involves, in an indecomposable way, the variables $\sigma(\mathbf{r})$ for all sites $\mathbf{r}$ around a face of the lattice.⁹ Such a face model is considered in Sect. 3.7.2.

2.4.2 Edge-Variable Models

In this case a variable is attached to each edge of the lattice. The edge-decorated Ising models discussed in Chap. 10 are obvious examples of this, although in that case there are also state variables at the sites of the lattice and the model can equally

⁹ It is obvious that the Hamiltonian of a pair-interaction model can be represented (in a decomposable way) as a sum over lattice faces and this is often useful in cases where the Hamiltonian also includes an indecomposable face interaction.

well be thought of as a site-variable model on a lattice modified by the addition of a site on each lattice edge. The two types of true edge-variable models discussed in this book are the vertex models introduced in Sect. 3.8 and analyzed in Sect. 9.3 and Chap. 12, and the dimer model, introduced in Sect. 3.9 and analyzed in Chap. 13. In each model the edge variable has two states, which for the vertex model corresponds to the two directions of an arrow along the edge and for the dimer model indicates the presence or absence of a dimer on the edge.

Although vertex models can be formed on any regular two-dimensional lattice¹⁰ we shall restrict attention to the two two-dimensional regular lattices with coordination number four; that is the square and kagomé lattices (Sect. 16.2.1). There are then sixteen arrow arrangements around a lattice site. However, the cases which have been solved are the six-vertex model (Sect. 3.8.2) where the arrow configuration is constrained so that there is the same number of arrows pointing towards and pointing away from the lattice site and the eight-vertex model (Sect. 3.8.1) where there is an even number of arrows pointing towards the lattice site.

The dimer configuration is also constrained by a condition at each lattice site. In this case there can be at most (and in the case of a complete covering of the lattice with dimers, exactly) one dimer end at each lattice site (Sect. 3.9).

2.5 Correlation Functions and Symmetry Properties

2.5.1 A General Hamiltonian

The coupling-density representation gives a particularly elegant formulation of thermodynamics in the special case $\eta = n_f - 1$, when there are no independent densities. The function ϕ_{n_f-1} is dependent only on couplings and, from (2.3.15), is itself equal to the coupling L_{n_f} . We shall concentrate on this case, omitting subscripts indicating the number of independent external couplings.

The values of the two sets of quantities $q_i(\mathbf{r}; \sigma)$, $i = 1, \dots, n_f - 1$ and $e_\ell(\mathbf{r}; \sigma)$, $\ell = 1, \dots, n_e$ are determined by the microstate σ of the system. We shall therefore refer to them respectively as *external* and *internal state operators*. It is clear from the above discussion that the sets $\{Q_i, L_i, \hat{Q}_i(\sigma), q_i(\mathbf{r}; \sigma)\}$ and $\{U_i, K_i, E_i(\sigma), e_i(\mathbf{r}; \sigma)\}$ play the same mathematical role in the development of the theory.¹¹ We shall, therefore, use the generic set $\{C_j, \zeta_j, \hat{C}_j(\sigma), c_j(\mathbf{r}; \sigma)\}$ to stand for members of either the external or internal sets with

$$\hat{C}_j(\sigma) = \sum_{\{\mathbf{r}\}} c_j(\mathbf{r}; \sigma). \quad (2.5.1)$$

¹⁰ Or indeed on an arbitrary planar lattice (Baxter 1978) or a three-dimensional lattice.

¹¹ Although, of course, an internal coupling K_i is simply replaced by a ζ_j , whereas the coupling corresponding to an external coupling L_i is a $-\zeta_j$.

The state operators $c_j(\mathbf{r}; \sigma)$, $j = 1, \dots, n := n_f + n_e - 1$, with $\mathbf{r}$ ranging over all N sites of the lattice form a set of nN random variables distributed according to the probability function (2.1.2). In using (2.5.1) to express a general form for the Hamiltonian it is convenient, to include a configuration-independent term. Thus we set

$$\hat{H}(N, \zeta_j; \sigma) := N\zeta_0 + \sum_{j=1}^n \hat{C}_j(\sigma) \zeta_j. \quad (2.5.2)$$

This definition could, of course, be expressed in a more compact form by extending the summation to the range $j = 0, 1, \dots, n$ and setting $\hat{C}_0(\sigma) = N$. It will also be observed that

$$\phi(\zeta_0, \zeta_1, \dots, \zeta_n) = \zeta_0 + \phi(0, \zeta_1, \dots, \zeta_n), \quad (2.5.3)$$

and that the calculation of an expectation value (2.1.4) using the probability formula (2.1.2) is unaffected by the presence of ζ_0 . For this reason it is often called the *trivial coupling*, although, as we shall see in Sects. 4.3.6 and 15.4, it plays a crucial role in scaling and renormalization group theory.

It is convenient to generalize this scheme further by supposing that the coupling is different at each site of the lattice with the corresponding generalized Hamiltonian

$$\hat{H}^{(G)}(N, \zeta_j(\mathbf{r}); \sigma) = \sum_{j=0}^n \sum_{\{\mathbf{r}\}} c_j(\mathbf{r}; \sigma) \zeta_j(\mathbf{r}), \quad (2.5.4)$$

where $c_0(\mathbf{r}; \sigma) = 1$, for all $\mathbf{r}$. With a similar notation for other quantities,

$$\langle c_j(\mathbf{r}; \sigma) \rangle = -\frac{1}{Z^{(G)}} \frac{\partial Z^{(G)}}{\partial \zeta_j(\mathbf{r})} = N \frac{\partial \phi^{(G)}}{\partial \zeta_j(\mathbf{r})}. \quad (2.5.5)$$

2.5.2 Correlation Functions

Let $c_1(\mathbf{r}_1; \sigma), \dots, c_\tau(\mathbf{r}_\tau; \sigma)$ be some choice of a subset of these variables at τ sites of the lattice. In terms of these a number of different τ -point correlation functions can be defined:

- (i) The *total correlation function* $\langle c_1(\mathbf{r}_1; \sigma) \cdots c_\tau(\mathbf{r}_\tau; \sigma) \rangle$.
- (ii) The *fluctuation correlation function* $\langle \delta c_1(\mathbf{r}_1; \sigma) \cdots \delta c_\tau(\mathbf{r}_\tau; \sigma) \rangle$ where

$$\delta c_j(\mathbf{r}; \sigma) := c_j(\mathbf{r}; \sigma) - \langle c_j(\mathbf{r}; \sigma) \rangle, \quad j = 1, 2, \dots, \tau, \quad (2.5.6)$$

is a measure of the fluctuation of c_j at site $\mathbf{r}$ about its mean.

(iii) The *net* or *connected correlation function*, which is defined recursively by

$$\begin{aligned} \Gamma_\tau(c_1(\mathbf{r}_1) \cdots c_\tau(\mathbf{r}_\tau)) &:= \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \cdots c_\tau(\mathbf{r}_\tau; \boldsymbol{\sigma}) \rangle \\ &\quad - \sum \Gamma_{p_1}(c_{j_1}(\mathbf{r}_{j_1}), \dots, c_{p_1}(\mathbf{r}_{p_1})) \\ &\quad \cdots \Gamma_{p_k}(c_{j_k}(\mathbf{r}_{j_k}), \dots, c_{p_k}(\mathbf{r}_{p_k})), \end{aligned} \quad (2.5.7)$$

$$\Gamma_1(c_1(\mathbf{r}_1)) := \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle, \quad (2.5.8)$$

where the sum in (2.5.7) is over products of the connected correlation functions for all partitions of the τ lattice sites.

For $\tau = 1$ the total and connected correlation functions are the same and the fluctuation correlation function is zero. For $\tau = 2$,

$$\begin{aligned} \Gamma_2(c_1(\mathbf{r}_1), c_2(\mathbf{r}_2)) &= \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle - \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle \langle c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle, \\ &= \langle [c_1(\mathbf{r}_1; \boldsymbol{\sigma}) - \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle] [c_2(\mathbf{r}_2; \boldsymbol{\sigma}) - \langle c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle] \rangle \end{aligned} \quad (2.5.9)$$

and, for $\tau = 3$,

$$\begin{aligned} \Gamma_3(c_1(\mathbf{r}_1), c_2(\mathbf{r}_2), c_3(\mathbf{r}_3)) &= \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) c_2(\mathbf{r}_2; \boldsymbol{\sigma}) c_3(\mathbf{r}_3; \boldsymbol{\sigma}) \rangle \\ &\quad - \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle \Gamma_2(c_2(\mathbf{r}_2; \boldsymbol{\sigma}), c_3(\mathbf{r}_3; \boldsymbol{\sigma})) - \langle c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle \Gamma_2(c_3(\mathbf{r}_3; \boldsymbol{\sigma}), c_1(\mathbf{r}_1; \boldsymbol{\sigma})) \\ &\quad - \langle c_3(\mathbf{r}_3; \boldsymbol{\sigma}) \rangle \Gamma_2(c_1(\mathbf{r}_1; \boldsymbol{\sigma}), c_2(\mathbf{r}_2; \boldsymbol{\sigma})) - \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle \langle c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle \langle c_3(\mathbf{r}_3; \boldsymbol{\sigma}) \rangle \\ &= \langle [c_1(\mathbf{r}_1; \boldsymbol{\sigma}) - \langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \rangle] [c_2(\mathbf{r}_2; \boldsymbol{\sigma}) - \langle c_2(\mathbf{r}_2; \boldsymbol{\sigma}) \rangle] [c_3(\mathbf{r}_3; \boldsymbol{\sigma}) - \langle c_3(\mathbf{r}_3; \boldsymbol{\sigma}) \rangle] \rangle. \end{aligned} \quad (2.5.10)$$

The two- and three-point fluctuation and connected correlation functions are the same. That this identity does not persist for large values of τ can be seen if we expand the general τ -point fluctuation correlation function

$$\begin{aligned} \langle \delta c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \cdots \delta c_\tau(\mathbf{r}_\tau; \boldsymbol{\sigma}) \rangle &= \sum_{\{j_1, \dots, j_s\}} (-1)^{\tau-s} \langle c_{j_1}(\mathbf{r}_{j_1}; \boldsymbol{\sigma}) \cdots c_{j_s}(\mathbf{r}_{j_s}; \boldsymbol{\sigma}) \rangle \\ &\quad \times \langle c_{j_{s+1}}(\mathbf{r}_{j_{s+1}}; \boldsymbol{\sigma}) \rangle \cdots \langle c_{j_\tau}(\mathbf{r}_{j_\tau}; \boldsymbol{\sigma}) \rangle, \end{aligned} \quad (2.5.11)$$

where the sum is over all subsets $\{j_1, \dots, j_s\}$ of the indices $\{1, 2, \dots, \tau\}$. Each term in the expansion is of the form of an s -point total correlation function multiplied by $\tau - s$ one-point functions. As can be seen from (2.5.7), the expansion for $\Gamma_\tau(c_1(\mathbf{r}_1) \cdots c_\tau(\mathbf{r}_\tau))$ in terms of total correlation functions will contain terms which are products of p - and q -point total correlation functions where *both* p and q are greater than unity when $\tau > 3$.

Using the site-specific couplings defined above,

$$\langle c_1(\mathbf{r}_1; \boldsymbol{\sigma}) \cdots c_\tau(\mathbf{r}_\tau; \boldsymbol{\sigma}) \rangle = \frac{(-1)^\tau}{Z^{(G)}} \frac{\partial^\tau Z^{(G)}}{\partial \zeta_1(\mathbf{r}_1) \cdots \partial \zeta_\tau(\mathbf{r}_\tau)}. \quad (2.5.12)$$

It can also be shown by induction (Binney et al. 1993) that

$$\Gamma_\tau(c_1(\mathbf{r}_1), \dots, c_\tau(\mathbf{r}_\tau)) = N(-1)^{\tau+1} \frac{\partial^\tau \phi^{(G)}}{\partial \zeta_1(\mathbf{r}_1) \cdots \partial \zeta_\tau(\mathbf{r}_\tau)}, \quad (2.5.13)$$

for connected correlation functions.

Returning now to the situation where the couplings ζ_j are site independent

$$\begin{aligned} \frac{\partial^\tau \phi}{\partial \zeta_1 \cdots \partial \zeta_\tau} &= \sum_{\{\mathbf{r}_1, \dots, \mathbf{r}_\tau\}} \frac{\partial^\tau \phi^{(G)}}{\partial \zeta_1(\mathbf{r}_1) \cdots \partial \zeta_\tau(\mathbf{r}_\tau)} \\ &= \frac{(-1)^{\tau+1}}{N} \sum_{\{\mathbf{r}_1, \dots, \mathbf{r}_\tau\}} \Gamma_\tau(c_1(\mathbf{r}_1), \dots, c_\tau(\mathbf{r}_\tau)), \end{aligned} \quad (2.5.14)$$

where, in the second summation, the reduction to site-independent couplings is applied after differentiation. In the case of the connected two-point or *pair* correlation function this result can be derived directly without the use of (2.5.13). From (2.1.11), (2.5.1), and (2.5.9)

$$\frac{\partial^2 \phi}{\partial \zeta_1 \partial \zeta_2} = -\frac{1}{N} \text{Cov}[\widehat{C}_1, \widehat{C}_2] = -\frac{1}{N} \sum_{\{\mathbf{r}_1, \mathbf{r}_2\}} \Gamma_2(c_1(\mathbf{r}_1), c_2(\mathbf{r}_2)). \quad (2.5.15)$$

Like (2.1.11) this is a *fluctuation-response function relation*. Such relations play an important role in statistical mechanics, relating microscopic and macroscopic properties of a system. Of course, when the two variables are the same the covariance becomes the variance. Thus, for example, from (2.3.31), for the set $\{Q, L, \widehat{Q}(\sigma), q(\mathbf{r}; \sigma)\}$,

$$\varphi_\tau = -\frac{\partial^2 \phi}{\partial L^2} = \frac{1}{N} \text{Var}[\widehat{Q}] = \frac{1}{N} \sum_{\{\mathbf{r}_1, \mathbf{r}_2\}} \Gamma_2(q(\mathbf{r}_1), q(\mathbf{r}_2)). \quad (2.5.16)$$

2.5.3 Symmetry Properties

Suppose that the lattice is d -dimensional hypercubic with periodic boundary conditions and lattice vectors given by (18.1.1). The τ -point connected correlation function then has (Sect. 18.1.1) the Fourier transform

$$\begin{aligned} \Gamma_\tau^*(c_1, \dots, c_\tau; \mathbf{k}_1, \dots, \mathbf{k}_\tau) &= \sum_{\{\mathbf{r}_1, \dots, \mathbf{r}_\tau\}} \Gamma_2(c_1(\mathbf{r}_1), \dots, c_\tau(\mathbf{r}_\tau)) \\ &\quad \times \exp[-i(\mathbf{k}_1 \cdot \mathbf{r}_1 + \cdots + \mathbf{k}_\tau \cdot \mathbf{r}_\tau)], \end{aligned} \quad (2.5.17)$$

with

$$\begin{aligned} \Gamma_\tau(c_1(\mathbf{r}_1), \dots, c_\tau(\mathbf{r}_\tau)) &= \frac{1}{N} \sum_{\{\mathbf{k}_1, \dots, \mathbf{k}_\tau\}} \Gamma_\tau^*(c_1, \dots, c_\tau; \mathbf{k}^{(1)}, \dots, \mathbf{k}^{(\tau)}) \\ &\quad \times \exp[i(\mathbf{k}_1 \cdot \mathbf{r}_1 + \cdots + \mathbf{k}_\tau \cdot \mathbf{r}_\tau)], \end{aligned} \quad (2.5.18)$$

where the wave vectors are given by (18.1.6). It follows from (2.5.14) and (2.5.17) that

$$\frac{\partial^\tau \phi}{\partial \zeta_1 \cdots \partial \zeta_\tau} = \frac{(-1)^{\tau+1}}{N} \Gamma_\tau^*(\mathbf{c}_1, \dots, \mathbf{c}_\tau; \mathbf{0}, \dots, \mathbf{0}). \quad (2.5.19)$$

In many cases of interest the system has the property of invariance under *translations* of the form

$$\mathbf{r} \rightarrow \mathbf{r} + a[m^{(1)}\hat{\mathbf{r}}^{(1)} + m^{(2)}\hat{\mathbf{r}}^{(2)} + \cdots + m^{(d)}\hat{\mathbf{r}}^{(d)}], \quad (2.5.20)$$

for any integers $m^{(1)}, \dots, m^{(d)}$. The difference between the translation given by (18.1.2), (18.1.3) and (2.5.20), is that the former can be applied to each of the vectors $\mathbf{r}_1, \dots, \mathbf{r}_\tau$ of $\Gamma_\tau(\mathbf{c}_1(\mathbf{r}_1), \dots, \mathbf{c}_\tau(\mathbf{r}_\tau))$ independently, whereas (2.5.20) must be applied to them all simultaneously. One translation of the form (2.5.20) takes $\mathbf{r}_1$ to the origin to give

$$\Gamma_\tau(\mathbf{c}_1(\mathbf{r}_1), \dots, \mathbf{c}_\tau(\mathbf{r}_\tau)) = \Gamma_\tau(\mathbf{c}_1(\mathbf{0}), \mathbf{c}_2(\bar{\mathbf{r}}_1), \dots, \mathbf{c}_\tau(\bar{\mathbf{r}}_{\tau-1})), \quad (2.5.21)$$

where $\bar{\mathbf{r}}_j = \mathbf{r}_{j+1} - \mathbf{r}_1$, $j = 1, \dots, \tau - 1$ are the relative lattice vectors. The one-point connected correlation function, given from (2.5.8) by $\langle \mathbf{c}_1(\mathbf{r}_1) \rangle$ is independent of the site location $\mathbf{r}_1$. We also see, from (2.5.7) and (2.5.11), that the total and fluctuation correlation functions have a similar dependence on relative vectors. With translational invariance one of the summations in the final form of (2.5.14) can be performed to give

$$\frac{\partial^\tau \phi}{\partial \zeta_1 \cdots \partial \zeta_\tau} = (-1)^{\tau+1} \sum_{\{\bar{\mathbf{r}}_1, \dots, \bar{\mathbf{r}}_{\tau-1}\}} \Gamma_\tau(\mathbf{c}_1(\mathbf{0}), \mathbf{c}_2(\bar{\mathbf{r}}_1), \dots, \mathbf{c}_\tau(\bar{\mathbf{r}}_{\tau-1})), \quad (2.5.22)$$

and, if $\bar{\Gamma}_\tau^*(\mathbf{c}_1, \dots, \mathbf{c}_\tau; \bar{\mathbf{k}}_1, \dots, \bar{\mathbf{k}}_{\tau-1})$ is the Fourier transform defined with respect to the relative vectors,

$$\frac{\partial^\tau \phi}{\partial \zeta_1 \cdots \partial \zeta_\tau} = (-1)^{\tau+1} \bar{\Gamma}_\tau^*(\mathbf{c}_1, \dots, \mathbf{c}_\tau; \mathbf{0}, \dots, \mathbf{0}). \quad (2.5.23)$$

Thus, for example, from (2.5.16),

$$\varphi_T = \bar{\Gamma}_2^*(\mathbf{q}; \mathbf{0}). \quad (2.5.24)$$

Finally we note that if, in addition to translational invariance, the system has $N^{(1)} = N^{(2)} = \cdots = N^{(d)}$ and all its properties invariant under *rotation* between lattice directions the two-point correlation function is given by

$$\Gamma_2(\mathbf{c}_1(\mathbf{r}_1), \mathbf{c}_2(\mathbf{r}_2)) = \Gamma_2(\mathbf{c}_1(\mathbf{0}), \mathbf{c}_2(\mathbf{r}_2 - \mathbf{r}_1)) = \Gamma_2(\mathbf{c}_1, \mathbf{c}_2; |\mathbf{r}_2 - \mathbf{r}_1|). \quad (2.5.25)$$

Chapter 3

A Survey of Models

3.1 Upper and Lower Critical Dimensions

One of the main investigative interests of statistical mechanics is in the possibility of phase transitions and, as is seen in many places in this text, this depends on the lattice dimension d of the system in two ways:

- (i) It is one of the determining factors which specifies the universality class and hence the critical exponents of a phase transition (Sect. 4.2.1). As is seen in Sect. 5.6 these critical exponents assume dimensionally-independent values (their mean-field values) when the lattice is of dimension $d \geq d_{uc}$, the *upper critical¹ dimension*.
- (ii) It affects whether an ordered phase can exist. Such a phase, which corresponds to some kind of long-range spatial order, can, by fluctuations of the microstates, be broken up into domains with their own local order (Fig. 4.4). The free-energy cost of this will be ‘paid’ at the domain boundary with the boundary to bulk ratios increasing with lattice dimension. Thus it may be anticipated that the temperature range $[0, T_c]$ in which an ordered phase can exist will decrease monotonically with decreasing d , with a *lower critical dimension* d_{lc} such that no long-range ordered phase can occur at non-zero temperature when $d \leq d_{lc}$.² The Peierls method of Sect. 8.3 establishes the existence of a phase transition for the two-dimensional Ising model and for the one-dimensionally infinite Ising model there is no phase transition at non-zero temperature (Sect. 11.3.2), giving $d_{lc} = 1$, a result which applies more generally for models with discrete symmetry. Fluctuations have a greater effect for models with continuous symmetry. For the quantum Heisenberg model $d_{lc} = 2$ (Sect. 3.2) and this results carries over to

¹ These ‘critical’ dimensions are sometimes called ‘marginal’ or ‘borderline’ (see, for example, Fisher 1983, Sect. 4.6).

² Although there may, as in the case of the Kosterlitz–Thouless transition (Sect. 3.3), be a transition to a different type of phase.

classical models with continuous symmetry (Sect. 3.3). Simple arguments based on the free-energy cost of destabilising the ordered phase can be advanced for these conclusions (Cardy 1996, Chap. 6).

3.2 The Quantum Heisenberg Model

Ferromagnetism and other forms of permanent magnetism are due to exchange interaction between unpaired spins occurring either directly or through conduction electrons or non-magnetic component atoms, (Martin 1967; Sinha and Kumar 1980; Mattis 1981). The standard model for this is the *quantum Heisenberg model* for which, in its general anisotropic form, there are three exchange energies $J_x(\mathbf{r} - \mathbf{r}')$, $J_y(\mathbf{r} - \mathbf{r}')$ and $J_z(\mathbf{r} - \mathbf{r}')$ between a pair of sites $\mathbf{r}, \mathbf{r}' \in \mathcal{N}$ with the Hamiltonian taking the form³

$$\begin{aligned} \hat{H}(J_x, J_y, J_z, \mathcal{H}; \mathbf{S}) := & -\frac{1}{s^2} \sum_{\{\mathbf{r}, \mathbf{r}'\}} [J_x(\mathbf{r} - \mathbf{r}') S_x(\mathbf{r}) S_x(\mathbf{r}') \\ & + J_y(\mathbf{r} - \mathbf{r}') S_y(\mathbf{r}) S_y(\mathbf{r}') + J_z(\mathbf{r} - \mathbf{r}') S_z(\mathbf{r}) S_z(\mathbf{r}')] - \frac{\mathcal{H}}{s} \sum_{\{\mathbf{r}\}} S_z(\mathbf{r}). \end{aligned} \quad (3.2.1)$$

$S_x(\mathbf{r})$, $S_y(\mathbf{r})$ and $S_z(\mathbf{r})$ are the component operators associated with the spin on site $\mathbf{r}$ where s takes integer or integer + $\frac{1}{2}$ values and commutation rules

$$[S_a(\mathbf{r}), S_b(\mathbf{r})] = iS_c(\mathbf{r}), \quad (3.2.2)$$

where a, b, c is any of the three cyclic orderings of x, y, z , apply.⁴

From (3.2.2) it follows that each of the three component operators commutes with $\mathbf{S}^2(\mathbf{r}) = S_x^2(\mathbf{r}) + S_y^2(\mathbf{r}) + S_z^2(\mathbf{r})$. So a choice can be made of simultaneous eigenstates for $\mathbf{S}^2(\mathbf{r})$ and one of the three component operators. Conventionally this is taken to be $S_z(\mathbf{r})$ with the external field coupled to that operator. In most magnetic materials with localized spins there is a direction of *easy magnetization* (taken to be the z -direction) so that J_x and J_y are very much smaller than J_z . A field in the direction of easy magnetisation, as shown in (3.2.1) is called *longitudinal* and one orthogonal to that direction (replacing $S_z(\mathbf{r})$ in the final term in (3.2.1) by $S_x(\mathbf{r})$ or $S_y(\mathbf{r})$) is said to be *transverse*.

The question of the occurrence of ferromagnetic ordering (magnetism with $\mathcal{H} = 0$) or antiferromagnetic ordering (sublattice-order in the case of a bipartite lattice)

³ The factors of s^2 and s in the denominators of the first and second terms are included for expediting the limit $s \rightarrow \infty$ and the Landé factor and Bohr magnon have been absorbed into the field $\mathcal{H}$.

⁴ In most investigations of the model the exchange interactions are restricted to first-neighbour pairs, although Zener (1951) has suggested that a ferromagnetic exchange between more distant neighbours, mediated by the conduction electrons, can account for ferromagnetic behaviour in cases where the first-neighbour interaction alone would yield antiferromagnetism.

is important. This is symmetry breaking and the *Mermin–Wagner theorem* (Mermin and Wagner 1966) shows that, in the isotropic case ($J_x(\mathbf{r}) = J_y(\mathbf{r}) = J_z(\mathbf{r}) = J(\mathbf{r})$) this cannot occur for $d \leq 2$, when the second moment

$$M_2(J) := \sum_{\{\mathbf{r}\}} J(\mathbf{r})|\mathbf{r}|^2 \quad (3.2.3)$$

is finite.⁵ In particular, of course, this applies to the model with only first-neighbour interactions for which it was shown by Kennedy (1985) that, with any degree of anisotropy of the form $0 < J_x = J_y := \alpha J$, $0 < J_z := \delta J$, with $\alpha \neq \delta$, there is spontaneous order for $d \geq 2$ and sufficiently low temperatures.

A case of extreme isotropy is the Ising model with $\alpha = 0$, $\delta = 1$. The Hamiltonian is now diagonal in the representation in which the commuting operators $S_z(\mathbf{r})$ and $S^2(\mathbf{r})$ are diagonal, as long as the field term in (3.2.1) is longitudinal, and they can be regarded as classical scalars with the range of values $-s, -s+1, \dots, s-1, s$. This is the *general- s Ising model* which is discussed in Sect. 3.5. The quantum nature of the model (and the name ‘quantum Ising model’) is retained when a transverse field is added to the Hamiltonian (or replaces the longitudinal field). An important result for this model, shown for $s = \frac{1}{2}$ by Suzuki (1976) and for $s = 1$ by Björnberg and Grimmett (2009),⁶ is that the expression for the ground-state⁷ energy of the quantum Ising model, with both transverse and longitudinal fields, on a d -dimensional lattice, is equivalent to the free energy of a classical Ising model on a $(d+1)$ -dimensional lattice. This implies the existence of a phase transition in the transverse-field quantum Ising model for all d -dimensional hypercubic lattices *including* $d = 1$, a result which has been confirmed by high-temperature series expansions.

3.2.1 One-Dimensional Chains

Apart from the Ising models, exact results for cases of the Heisenberg model have been obtained only for the one-dimensional lattice (chain) and then mainly for the

⁵ More generally the Mermin–Wagner theorem states that, subject to a bound on the interactions like (3.2.3), there cannot be spontaneous breaking of a continuous symmetry in systems with $d \leq 2$. This applies to the isotropic Heisenberg model which has $O(3)$ symmetry; that is symmetry under arbitrary rotations in three dimensions. It should be noted that the Mermin–Wagner conditions are sufficient but not necessary to ensure the absence of a phase transition. This means, in general, for such systems $d_{LC} \geq 2$, although for the isotropic Heisenberg model $d_{LC} = 2$. A lower critical dimension of two has also been established for the superfluid and superconducting phases in Bose and Fermi liquids by Hohenberg (1967) and in quantum field theory in the absence of Goldstone modes by Coleman (1973).

⁶ But generally believed to be true for all s .

⁷ In this chapter the ‘ground state’ is understood in the informal sense of being the microstate which gives the smallest value for the Hamiltonian. A more formal definition of the concept is given in Sect. 8.3.

zero-field, $s = \frac{1}{2}$, case, when the Hamiltonian, in the representation in which $S_z(\mathbf{r})$ and $S^2(\mathbf{r})$ are diagonal, has the matrix form

$$\hat{H}(\{J_\zeta\}; \Sigma) := - \sum_{\{\zeta\}} \sum_{i=1}^N J_\zeta \Sigma_\zeta(i) \Sigma_\zeta(i+1), \quad (3.2.4)$$

where the outer summation is over $\zeta = x, y, z$,

$$\Sigma_\zeta(i) := \mathbf{I} \otimes \cdots \otimes \mathbf{I} \otimes \Sigma_\zeta \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I}, \quad (3.2.5)$$

$\mathbf{I}$ is the two-dimensional unit vector and Σ_ζ , which lies at the i th place in the product, is the appropriate Pauli spin matrix

$$\Sigma_x := \begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix}, \quad \Sigma_y := \begin{pmatrix} 0 & -i \\ +i & 0 \end{pmatrix}, \quad \Sigma_z := \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.2.6)$$

Bethe (1931) found the eigenvalues of the *isotropic*, $J_x = J_y = J_z := J$, model and determined the elementary excitation spectrum (in the form of m magnons for $m = 1, 2, \dots$) for the ferromagnetic ($J > 0$) case using a method now called the *Bethe ansatz*.⁸ For the antiferromagnetic ($J < 0$) case the ground state was determined by Hulthén (1938) and the elementary excitations by des Cloizeaux and Pearson (1962). In the ferromagnetic case, for general s , the ground-state and one-magnon and two-magnon excitations were obtained by Wortis (1963). A slightly different approach was used by Babujian (1982), who generalized the terms of the Hamiltonian (3.2.1) to polynomials in $S(\mathbf{r}) \cdot S(\mathbf{r}')$ and showed that, for a particular choice of polynomial coefficients the complete spectrum could be obtained.

It was shown by Yang and Yang (1966a) that, if one is interested only in the eigenvalues of the Hamiltonian then the signs of J_x , J_y and J_z , as long as they are of equal magnitude, are irrelevant. This is, of course, no longer the case for the more general *Heisenberg–Ising XXZ model* where $J_x = J_y := J$, $J_z := \delta J$, with $|\delta| \neq 1$. For this Lieb et al. (1961) obtained the ground-state and all the elementary excitations and Yang and Yang (1966b, c, d) showed rigorously that the Bethe ansatz is correct, deriving the minimum eigenvalue in the limit of an infinite chain.⁹ The parameter δ represents the anisotropy of the system and that and the sign of J are the quantities which characterize the system. With J positive/negative ferromagnetic/antiferromagnetic ordering predominate in the x – y plane with a similar effect in the z direction for $J\delta$ positive or negative if $|\delta| > 1$. When Lieb (1967a, b, c, d) solved the six-vertex model, he found that the Heisenberg–Ising Hamiltonian and the six-vertex transfer matrix commuted implying

⁸ For an historical account of the many uses of the Bethe ansatz, including in the solution of the six-vertex model, see Batchelor (2007).

⁹ For detailed accounts of the solution see Izyumov and Skryabin (1988, Chap.5) or Jimbo and Miwa (1995, Chap.1).

that they share the same eigenvectors (Sects. 12.3.4.2 and 12.3.10.2). He was, therefore, able to exploit many of the results of Yang and Yang.

For $|\delta| < 1$ the XXZ model is paramagnetic and the limit $\delta \rightarrow 0$ yields the *XX model* solved by Katsura (1962). This is the special case $\gamma = 0$ of the *XY model* for which $J_z := 0$, $J_x := \frac{1}{2}J(1 + \gamma)$, $J_y := \frac{1}{2}J(1 - \gamma)$. This was solved by Lieb et al. (1961) for a range of antiferromagnetic cases. With a magnetic field Jh , coupled to the $\Sigma_z(i)$, Kurmann et al. (1982) obtained the zero-temperature phase diagram in the γ - h plane. Within the unit circle the ground-state is of the classic Néel antiferromagnetic two-sublattice type and the line $h = 1$ separates ordered and disordered phases. For $\gamma = 1$ this transition is in the universality class of the quantum Ising model in a transverse field and otherwise it is in the universality class of the XX model.

In a result parallel to that for the XXZ and six-vertex models, Sutherland (1970) showed that the general XYZ Hamiltonian (3.2.4) and the eight-vertex transfer matrix commute (Sect. 12.3.4.2) implying that they share the same eigenvectors, and Baxter (1972a) (see also, Baxter 1982b, Sect. 10.14) showed that the Hamiltonian is effectively a logarithmic derivative of the eight-vertex transfer matrix, meaning that the ground-state energy can be found. A complete investigation of this case was carried out by Takhtadzhyan and Faddeev (1979).

For later reference we note that (3.2.4) can be re-expressed in the form

$$\begin{aligned} \hat{H}(J_x, J_y, J_z; \Sigma) = & -\frac{1}{4}(J_x + J_y) \sum_{i=1}^N [\{\Sigma_+(i)\Sigma_+(i+1) + \Sigma_-(i)\Sigma_-(i+1)\} \\ & + I_0\{\Sigma_+(i)\Sigma_-(i+1) + \Sigma_-(i)\Sigma_+(i+1)\} + 2I_z\Sigma_z(i)\Sigma_z(i+1)], \end{aligned} \quad (3.2.7)$$

where $\Sigma_{\pm}(i)$ are of the form (3.2.5) with

$$\Sigma_{\pm} := \Sigma_x \pm \Sigma_y, \quad (3.2.8)$$

and

$$I_0 := \frac{J_x - J_y}{J_x + J_y}, \quad I_z := \frac{2J_z}{J_x + J_y}. \quad (3.2.9)$$

3.3 Classical Vector Models

If, for the quantum Heisenberg model, $s_a(\mathbf{r}) := S_a(\mathbf{r})/s$, $a = x, y, z$, and the limit $s \rightarrow \infty$ is taken, it follows, from (3.2.2), that the three operators $s_x(\mathbf{r})$, $s_y(\mathbf{r})$ and $s_z(\mathbf{r})$ commute and become equivalent to the three components of the three-dimensional unit vector $\mathbf{s}(\mathbf{r}) := (s_x(\mathbf{r}), s_y(\mathbf{r}), s_z(\mathbf{r}))$. In the case $J_x(\mathbf{r}, \mathbf{r}') = J_y(\mathbf{r}, \mathbf{r}') = J_z(\mathbf{r}, \mathbf{r}') = J(\mathbf{r} - \mathbf{r}')$ the Hamiltonian (3.2.1) becomes

$$\widehat{H}(J(\mathbf{r} - \mathbf{r}'), \mathcal{H}; \mathbf{s}(\mathbf{r})) := - \sum_{\{\mathbf{r}, \mathbf{r}'\}} J(\mathbf{r} - \mathbf{r}') \mathbf{s}(\mathbf{r}) \cdot \mathbf{s}(\mathbf{r}') - \sum_{\{\mathbf{r}\}} \mathbf{s}(\mathbf{r}) \cdot \mathcal{H}, \quad (3.3.1)$$

where the ‘special status’ of the z -direction is now removed and $\mathcal{H}$ can be in any direction. This is the *classical Heisenberg model*, for which a generalization is the *n -vector model* proposed by Stanley (1968a)¹⁰ where, for all $\mathbf{r} \in \mathcal{N}$, $\mathbf{s}(\mathbf{r})$ is replaced by an n -dimensional unit vector and $\mathcal{H}$ is also, of course, n -dimensional.

The symmetry group when $\mathcal{H} = \mathbf{0}$ is $O(n)$, the group of $n \times n$ orthogonal matrices (Sect. 17.2.7). Apart from the classical Heisenberg model with symmetry group $O(3)$, the class of $O(n)$ models contains for $n = 2$ the classical XY model, for $n = 1$ the spin- $\frac{1}{2}$ Ising model and, in the limit $n \rightarrow 0$, the *self-avoiding walk* (de Gennes 1972; see also, Cardy 1996, Sect. 9.3). The Mermin–Wagner theorem holds when $n > 1$ and is independent of s applying therefore to the limit $s \rightarrow \infty$.

Of particular interest is the $O(2)$ model, where the spins rotate in a plane, with

$$\mathbf{s}(\mathbf{r}) = (\cos[\theta(\mathbf{r})], \sin[\theta(\mathbf{r})]), \quad (3.3.2)$$

$$\mathbf{s}(\mathbf{r}) \cdot \mathbf{s}(\mathbf{r}') = \cos(\theta(\mathbf{r}) - \theta(\mathbf{r}')). \quad (3.3.3)$$

This is the *classical XY model*, also called the *plane rotator model*. The spin-spin term (3.3.3) is of the state-difference form discussed in Sect. 3.6 and we reserve further discussion of this model to that section.

3.4 The Gaussian and Spherical Models

The spin- $\frac{1}{2}$ Ising model is, in the context of the last section, simply the one-vector model which escapes the application of the Mermin–Wagner theorem because the group $O(1)$ does not represent a continuous symmetry. There is, however, another interesting connection between the Ising and n -vector models.

Following Onsager’s solution of the two-dimensional Ising model in 1944, Berlin and Kac (1952) proposed two models, the *Gaussian model* and the *spherical model*, which were intended to capture the essential qualities of the Ising model while avoiding the complexities of its solution. The starting point for each of these models is to denote, as in the case of the spin- $\frac{1}{2}$ Ising model, the $n = 1$ version of the vector $\mathbf{s}(\mathbf{r})$ by $\sigma(\mathbf{r})$. Thus the Hamiltonian (3.3.1) becomes

$$\widehat{H}(J(\mathbf{r} - \mathbf{r}'), \mathcal{H}; \sigma) := - \sum_{\{\mathbf{r}, \mathbf{r}'\}} J(\mathbf{r} - \mathbf{r}') \sigma(\mathbf{r}) \sigma(\mathbf{r}') - \mathcal{H} \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}), \quad (3.4.1)$$

¹⁰ In fact he first used v in place of n and took the vectors to be of length $\sqrt{v}$. Subsequently (Stanley 1974) he replaced ‘ v ’ by ‘ D ’.

where $\sigma(\mathbf{r}) = \pm 1$ and, of course, $\sigma^2(\mathbf{r}) = 1$. Both the Gaussian and spherical models are predicated on a relaxation of this latter condition and an extension of the range of $\sigma(\mathbf{r})$ to all real values.

For the *Gaussian model* the weakened constraint on each of the spin variables is

$$\langle \sigma^2(\mathbf{r}) \rangle = 1, \quad \text{with a term} \quad \frac{1}{2} \sum_{\{\mathbf{r}\}} \sigma^2(\mathbf{r}) \quad (3.4.2)$$

added to the Hamiltonian, resulting in a Gaussian integral form for the partition function. Berlin and Kac (1952) showed that, for the first-neighbour case of this model and for hypercubic lattices of dimensions one, two and three, there is a critical temperature $T_c = 2Jd$. However, below the critical temperature there is no well-defined ordered phase since the free energy becomes complex.

For the *spherical model* the constraint condition is

$$\sum_{\{\mathbf{r}\}} \sigma^2(\mathbf{r}) = N \quad (3.4.3)$$

and the first-neighbour case was solved in one, two and three dimensions by Berlin and Kac (1952).¹¹ However, it can be solved for lattices of all dimensions (Joyce 1972; Baxter 1982b, Chap. 5). Although this model was introduced by reference to the Ising model ($n = 1$) it can be seen that the constraint (3.4.3), if applied to vectors of any dimension, would effectively combine the roles of N and n . In fact the properties of the model are more similar to those of the Heisenberg model than the Ising model and, for the first-neighbour case it has been shown (Stanley 1968b; Kac and Thompson 1971; Pearce and Thompson 1977) to be equivalent to the $n \rightarrow \infty$ limit of the n -vector model. Thus the Mermin–Wagner theorem implies that the lower critical dimension $d_{lc} = 2$. In this context, it is worth anticipating the discussion of Sect. 4.4.1 by quoting the critical exponents, which are (see, Baxter 1982b, Sect. 5.6)

$$\alpha = -\frac{4-d}{d-2}, \quad \beta = \frac{1}{2}, \quad \gamma = \frac{2}{d-2}, \quad \delta = \frac{d+2}{d-2}, \quad (3.4.4)$$

for $2 < d \leq 4$. At $d = 2$ the exponents α , γ and δ diverge and the critical temperature is zero. The critical exponents at $d = 4$ take their classical (Landau) values given by (5.3.30) and as in the Gaussian approximation to Ginzburg–Landau theory they retain these values for all $d > 4$ so that the upper critical dimensions $d_{uc} = 4$.

Joyce (1966) investigated the behaviour of a ferromagnetic spherical model for $d = 1, 2, 3$ with a long-range interaction of the form $J(\mathbf{r}) \sim 1/|\mathbf{r}|^{d+\varepsilon}$ ($\varepsilon > 0$). For such an interaction it is simple to show that the second moment $M_2(J)$ exists, implying from the Mermin–Wagner theorem the non-existence of a phase transition

¹¹ For an account of the origins of the spherical model see Kac (1964).

for $d = 1, 2$, when $\varepsilon > 2$. In fact Joyce showed that there is a phase transition for $d = 1, 2$ only when $0 < \varepsilon < d$. There is also, as expected, a transition for $d = 3$, and for $d = 1, 2, 3$ the exponents α , γ and δ are functions of ε . In particular

$$\gamma = \begin{cases} 1, & \text{for } 0 < \varepsilon < \frac{1}{2}d, \\ \frac{\varepsilon}{d - \varepsilon}, & \text{for } \frac{1}{2}d < \varepsilon < \min(2, d). \end{cases} \quad (3.4.5)$$

When $\varepsilon = \frac{1}{2}d$ the susceptibility has a logarithmic singularity and when $d = 3$, $\gamma = 2$, for $\varepsilon > 2$, coinciding with the value given in (3.4.4) for the first-neighbour model. The correspondence between the n -vector model and the spherical model, established by Stanley (1968b) in the limit $n \rightarrow \infty$ for the first-neighbour model, was extended to the case $n = 3$ by Joyce (1969) when the exchange interaction is $J(\mathbf{r})$. He showed that the isotropic classical Heisenberg model with such an exchange cannot be ferromagnetic or antiferromagnetic if the spherical model with exchange $|J(\mathbf{r})|$ does not exhibit a phase transition. This means that the isotropic classical Heisenberg model with $J(\mathbf{r}) \sim 1/|\mathbf{r}|^{d+\varepsilon}$ cannot be ferromagnetic or antiferromagnetic for $d < 3$ if $\varepsilon \geq d$.

3.5 Ising Models

As has been seen above the general spin- s Ising model arises as the special case $J_x(\mathbf{r}, \mathbf{r}') = J_y(\mathbf{r}, \mathbf{r}') = 0$ of the quantum Heisenberg model and the spin- $\frac{1}{2}$ Ising model as the special case $n = 1$ of the n -vector model. We shall initially concentrate on the latter case, with only a first-neighbour interaction.

3.5.1 The Spin- $\frac{1}{2}$ Ising Model

For $n = 1$ the sole component of the vector $\mathbf{s}(\mathbf{r})$ in (3.3.1) is denoted by $\sigma(\mathbf{r})$ with values ± 1 . Given that the field is in the direction of this component and the interaction energy is purely first neighbour the Hamiltonian becomes¹²

$$\hat{H}(K, L; \boldsymbol{\sigma}) := \hat{H}(J, \mathcal{H}; \boldsymbol{\sigma})/T = -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma(\mathbf{r})\sigma(\mathbf{r}') - L \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}), \quad (3.5.1)$$

where $K := J/T$ and $L := \mathcal{H}/T$, with $J > 0$ and $J < 0$ corresponding to ferro- and antiferromagnetism. When $\mathcal{H} = L = 0$, $\hat{H}(K, 0; \boldsymbol{\sigma})$ is invariant under the symmetry group $\mathcal{S}_2^{(m)} := \{I, Q^{(m)}\}$, where $Q^{(m)}$ is the spin-inversion operator

¹² Throughout this text the superscript ‘(k)’ indicates that a summation is confined to k th neighbour pairs.

$$Q^{(m)} : +\sigma(\mathbf{r}) \rightarrow -\sigma(\mathbf{r}), \quad \forall \mathbf{r} \in \mathcal{N}. \quad (3.5.2)$$

When $L \neq 0$ we must also consider the group $\mathcal{S}_2^{(L)} := \{I, Q^{(L)}\}$, where

$$Q^{(L)} : +L \rightarrow -L. \quad (3.5.3)$$

However, from the group $\mathcal{S}_2^{(m)} \times \mathcal{S}_2^{(L)} = \{I, Q^{(m)}, Q^{(L)}, Q^{(m)}Q^{(L)}\}$ the Hamiltonian is invariant only for the subgroup $\mathcal{S}_2^{(mL)} := \{I, Q^{(m)}Q^{(L)}\}$. This system is often referred to simply as ‘the Ising model’ since this was the case which was solved for $d = 1$ by Ising (1925).¹³ The extensive variable conjugate to $\mathcal{H}$ is the statistical mechanical magnetization

$$\hat{\mathcal{M}} := \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}), \quad (3.5.4)$$

with statistical mechanical and thermodynamic magnetization densities

$$\hat{m} := \hat{\mathcal{M}}/N, \quad m := \langle \hat{m} \rangle, \quad (3.5.5)$$

respectively and the partition function

$$Z(K, L) = \sum_{\{\sigma\}} \exp\{-\hat{H}(K, L; \sigma)\}. \quad (3.5.6)$$

Then from (2.3.7), (3.5.4) and (3.5.5)

$$m(K, L) = \langle \sigma(\mathbf{r}) \rangle = -\frac{\partial \phi(K, L)}{\partial L}. \quad (3.5.7)$$

Given that the lattice is one of the regular types described in Sect. 16.2.1, with coordination number z , there are $\frac{1}{2}Nz$ first-neighbour pairs and, from (2.3.27),

$$u(K, L) := u/J = -\frac{1}{2}z\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle^{(1)} = \frac{\partial \phi(K, L)}{\partial K}, \quad (3.5.8)$$

where $[\mathbf{r}, \mathbf{r}']$ is any first-neighbour pair of sites. Since, in general, we can consider the correlation $\langle \sigma(\mathbf{r}_1)\sigma(\mathbf{r}_2) \rangle$ between the spins on any pair of sites, the superscript ‘(1)’ is used to emphasis the fact that the correlation in this case is between first neighbours. It should also be noted that the relationship between the internal energy per site and the first-neighbour correlation is valid in the form (3.5.8) only when, as in the case of (3.5.1), the interaction term in the Hamiltonian consists

¹³ In fact this model was suggested to Ising by his research director Wilhelm Lenz and historians of science like Brush (1967) and Niss (2005) often call it the *Lenz-Ising model*. Interested readers are referred to these authors for accounts of historic developments related to the model.

Table 3.1 Critical parameters for the Ising model ferromagnet of dimension d and coordination number z

Source	Lattice or Approximation	d	z	$T_c/(zJ)$	$-2u_c/z$	$[\ln(2) - s_c]/\ln(2)$
(1)	Square exact	2	4	0.5673	0.7071	0.5579
(2)	Honeycomb exact	2	3	0.5062	0.7698	0.6181
(2)	Triangular exact	2	6	0.6068	0.6667	0.5235
(2)	Diamond series	3	4	0.6761	0.4370	0.2640
(3)	Simple cubic series	3	6	0.7518	0.3308	0.1951
(3)	B. C. cubic series	3	8	0.7942	0.2732	0.1603
(3)	F. C. cubic series	3	12	0.8163	0.2475	0.1485
(4)	Mean-field approx.	—	—	1.0	0.0	0.0
(5)	Bethe-pair approx.	—	3	0.6068	0.5000	0.2831
(5)	Bethe-pair approx.	—	4	0.7213	0.3333	0.1634
(5)	Bethe-pair approx.	—	6	0.8221	0.2000	0.0871
(5)	Bethe-pair approx.	—	8	0.8690	0.1429	0.0591
(5)	Bethe-pair approx.	—	12	0.9141	0.0909	0.0358

The sources of the data are (1) Sect. 9.2.2, (2), (3), (4) Sect. 6.1, (5) Sect. 7.3. The critical temperature and critical internal energy are scaled by the coordination number z in order that the table may include the results of classical approximations where the lattice type enters only in this way

only of first-neighbour terms with equal couplings. The generalization of (3.5.8) to the case of unequal couplings in different lattice directions or to include second- or further-neighbour interactions is obvious. It follows from the symmetry property $\hat{H}(J, \mathcal{H}; \sigma) = Q^{(m)} Q^{(L)} \hat{H}(J, \mathcal{H}; \sigma) = \hat{H}(J, -\mathcal{H}; -\sigma)$ that

$$m(K, -L) = -m(K, L), \quad u(K, -L) = u(K, L). \quad (3.5.9)$$

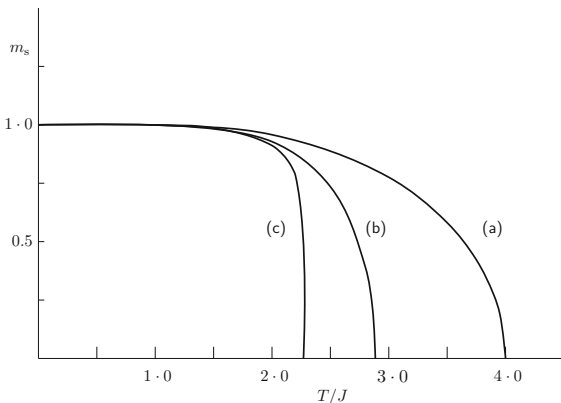
Thus, if the magnetization is single-valued on the zero-field axis, that value must be zero. However, as we shall see in detailed calculations throughout this book, the reduction of temperature in zero field leads to the onset of a spontaneous magnetism at a temperature T_c , called the *Curie temperature*. Consistency then requires that the magnetism will take one of two values $\pm m_s(K)$, ($m_s(K) > 0$). Critical data and curves of zero-field magnetization derived from calculations in different chapters of the book are given here in Table 3.1 and Fig. 3.1 respectively.

The simple (spin- $\frac{1}{2}$) Ising model can be generalized in various ways; second-neighbour pair interactions (as indicated above) or multi-spin interactions around each face can, for example, be included. However, if contributions are restricted to first-neighbour pairs and single-site terms, (3.5.1) is the most general Hamiltonian for the Ising model with $\sigma(\mathbf{r}) = \pm 1$. This still leaves the possibilities of interpreting the model as representing systems other than ferromagnets. Examples of this will be discussed in the following sections.

3.5.1.1 Antiferromagnetism

In the discussion of the ferromagnetic Ising model all lattice sites have been regarded as ‘equivalent’. This means that, as for all regular lattices listed in Sect. 16.2.1, the

Fig. 3.1 Spontaneous magnetization plotted against temperature for the ferromagnetic Ising model on the square lattice ($z = 4$) in (a) the mean-field approximation (Sect. 6.1), (b) the Bethe-pair approximation (Sect. 7.3), (c) the exact result (Sect. 12.3.9)



unoccupied lattice looks exactly the same when viewed from any site with each site having the same number z_p of p th neighbour sites for any p . The interaction of a microsystem with an external field is the same for all sites and the interaction energy of any two microsystems depends only on the relative position of the lattice sites they occupy.¹⁴ This last condition could easily be extended from two-body (pairwise) interactions to n -body interactions.

The equivalence of the lattice sites means that it may be possible to divide the lattice into two equivalent sublattices, labelled a and b . The sublattices are *interpenetrating* if all first-neighbour sites of any a -site are b -sites and vice-versa. A lattice divisible into two equivalent interpenetrating sublattices is called *loose-packed* (or *bipartite*), whereas a lattice for which a division of this kind is not possible is *close-packed*. The $d = 1$ lattice is loose-packed since it can be divided into interpenetrating sublattices simply by labelling alternate sites a and b . The a and b sublattices (identified respectively as the black and white sites) for the loose-packed square, honeycomb, simple cubic, body-centred cubic and diamond lattices are shown in Sect. 16.2.1.

The Hamiltonian of an antiferromagnet with first-neighbour interactions only on a loose-packed lattice is obtained by adapting (3.5.1) and is given by

$$\hat{H}(K, L; \sigma) = K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma^{(a)}(\mathbf{r}) \sigma^{(b)}(\mathbf{r}') - L \left[\sum_{\{\mathbf{r}\}} \sigma^{(a)}(\mathbf{r}) + \sum_{\{\mathbf{r}'\}} \sigma^{(b)}(\mathbf{r}') \right], \quad (3.5.10)$$

$\sigma^{(a)}(\mathbf{r})$ and $\sigma^{(b)}(\mathbf{r}')$ respectively denoting the spin variables for a site $\mathbf{r}$ of sublattice a and a site $\mathbf{r}'$ of sublattice b . The parameter K of (3.5.1) has been changed to $-K$ ($K := J/T > 0$) and the fact that all first-neighbour pairs consist of an a -site and a b -site has been used. For $\mathcal{H} = L = 0$ (3.5.10) can be written

¹⁴ ‘Relative position’ involves the direction as well as the length of the lattice vector connecting the two sites.

$$\hat{H}(K, 0; \sigma) = -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma^{(a)}(\mathbf{r}) [-\sigma^{(b)}(\mathbf{r}')]. \quad (3.5.11)$$

This is the Hamiltonian of a ferromagnet with spin variables $\sigma^{(a)}(\mathbf{r})$ and $-\sigma^{(b)}(\mathbf{r}')$. There is a one-to-one correspondence between microstates of equal energy in the zero-field ferromagnet and antiferromagnet. These microstates are connected by the operation of reversing the spins on one sublattice, leaving those on the other sublattice unchanged, while changing the sign of the interaction parameter. It follows that in zero field the partition function for the antiferromagnet is equal to that for the ferromagnet and the free energy, entropy, internal energy and heat capacity are the same. The critical temperature T_c , where there is a heat capacity singularity at $\mathcal{H} = 0$, is the Curie temperature for the ferromagnet and is termed the *Néel point* for the antiferromagnet. Experimentally, the phenomenon of antiferromagnetism is commoner than that of ferromagnetism (Martin 1967).

The type of long-range order in the antiferromagnet below the critical temperature is quite different from that in the ferromagnet. We define a relative magnetization for each sublattice

$$\begin{aligned} m_a &:= \langle \hat{m}_a \rangle, & \hat{m}_a &:= \frac{2}{N} \sum_{\{\mathbf{r}\}} \sigma^{(a)}(\mathbf{r}), \\ m_b &:= \langle \hat{m}_b \rangle, & \hat{m}_b &:= \frac{2}{N} \sum_{\{\mathbf{r}'\}} \sigma^{(b)}(\mathbf{r}'), \end{aligned} \quad (3.5.12)$$

where the sums are taken over all sites of sublattice a and sublattice b respectively. The overall relative magnetization m is then given by

$$m = \frac{1}{2}(m_a + m_b). \quad (3.5.13)$$

For the zero-field ferromagnet above the critical temperature T_c , $m_a = m_b = m = 0$ whereas below T_c , $m_a = m_b = \pm m_s$, ($m_s > 0$). Now, from (3.5.11), the zero-field equilibrium state of the antiferromagnet can be derived from that of the ferromagnet by reversing the magnetization of one of the sublattices. Hence, above the Néel point ($T > T_c$), the magnetizations of both sublattices are zero, but below it either

$$\left. \begin{aligned} m_a = m_s, & & m_b = -m_s, \\ & \text{or} & \\ m_a = -m_s, & & m_b = m_s, \end{aligned} \right\} \quad m = 0. \quad (3.5.14)$$

The magnetizations of the two sublattices are equal and opposite, giving zero overall magnetization. This is an example of *sublattice ordering*. The degeneracy displayed in (3.5.14) for $\mathcal{H} = 0$, $T < T_c$, parallels the degeneracy $m = \pm m_s$ of the ferromagnet (Sect. 6.1). As $T/J \rightarrow 0$, the Ising model antiferromagnet approaches the state $m_a = \pm 1$, $m_b = \mp 1$ where all first-neighbour spin pairs are antiparallel.

Since, from (3.5.14), m is a continuous function of $\mathcal{H}$ as the interval $(0, T_c)$ of the $\mathcal{H} = 0$ axis is crossed, this interval is no longer a first-order transition line, as it is in the ferromagnet (Sect. 6.1). The antiferromagnetic ground states are examples of commensurate phases.¹⁵

For $\mathcal{H} \neq 0$ there is no correspondence between the states of the antiferromagnet and those of the ferromagnet. From the equivalence of sites in the antiferromagnetic model it may be expected that, if either the magnetic field or thermal disordering is strong enough to offset the antiferromagnetic spin interaction, the equilibrium state will be one in which $m_a = m_b = m$. Such a state is called *paramagnetic* as opposed to the antiferromagnetic state in which the spins on sublattices a and b are oppositely oriented. The transition curve in the $\mathcal{H}$ – T plane on which the change of symmetry from the paramagnetic state to the antiferromagnetic state takes place for $d > 1$ has to be investigated by approximation or series methods even for $d = 2$, since no exact results are available for either the ferromagnet or antiferromagnet with $\mathcal{H} \neq 0$ and $T > 0$. However, the zero-field results given above indicate that the transition curve meets the $\mathcal{H} = 0$ axis at $T = T_c$ and is symmetric about $\mathcal{H} = 0$. Also, the points where it meets the $T = 0$ axis can be determined quite easily. It follows from (2.1.2) that as $T \rightarrow 0$ the probability distribution is dominated by the term or terms that give the ground-state Hamiltonian, which in this limit is equal to the enthalpy. In the paramagnetic ground state, all $\sigma^{(a)}(\mathbf{r}) = \sigma^{(b)}(\mathbf{r}') = \pm 1$, according as $\mathcal{H} \gtrless 0$ and, in the antiferromagnetic ground state, all $\sigma^{(a)}(\mathbf{r}) = 1$ and all $\sigma^{(b)}(\mathbf{r}') = -1$ (or vice-versa). If $\hat{H}_p$ and $\hat{H}_A$ denote the values of the Hamiltonian in the paramagnetic and antiferromagnetic states respectively, by (3.5.10),

$$\hat{H}_p = \frac{1}{2}zNJ - N|\mathcal{H}|, \quad \hat{H}_A = -\frac{1}{2}zNJ, \quad (3.5.15)$$

and $\hat{H}_A < \hat{H}_p$, making the antiferromagnetic state stable if

$$-\mathcal{H}_c < \mathcal{H} < \mathcal{H}_c := zJ, \quad (3.5.16)$$

where the last relation defines the *critical field* $\mathcal{H}_c$.

This simple antiferromagnetic model can be generalized by introducing a ferromagnetic interaction (i.e. tending to spin parallelism) between spins on the same sublattice. In a loose-packed lattice (Sect. 16.2.1) both sites of any second-neighbour pair lie on the same sublattice, and we introduce energy terms $-J_2\sigma^{(a)}(\mathbf{r})\sigma^{(a)}(\mathbf{r}')$ and $-J_2\sigma^{(b)}(\mathbf{r})\sigma^{(b)}(\mathbf{r}')$ for each such a – a and b – b pair. With $K_1 := J_1/T$ and $K_2 := J_2/T$, the Hamiltonian (3.5.10) is modified to

¹⁵ A phase is *commensurate* if its translational periodicity in each lattice direction is a rational (but not necessarily the same) multiple of the periodicity of the lattice. A homogeneous phase like the ferromagnetic ground state will be commensurate on all lattices. Of course, the construction of a commensurate antiferromagnetic phase is not confined to bipartite lattices. Thus, for example, on the triangular lattice the phase with a honeycomb of sites in one state and the centres of the honeycomb in the other state is both a commensurate and a lowest-energy state. However, it is one of a large number of arrangements with the same energy, leading to a macroscopic zero-point entropy, with no non-zero critical temperature to a long-range ordered phase (Sect. 9.4.2.2).

$$\begin{aligned} \widehat{H}(K_1, K_2, L; \sigma) = & K_1 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma^{(a)}(\mathbf{r}) \sigma^{(b)}(\mathbf{r}') - K_2 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(2)} \sigma^{(a)}(\mathbf{r}) \sigma^{(a)}(\mathbf{r}') \\ & - K_2 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(2)} \sigma^{(b)}(\mathbf{r}') \sigma^{(b)}(\mathbf{r}') - L \left(\sum_{\{\mathbf{r}\}} \sigma^{(a)}(\mathbf{r}') + \sum_{\{\mathbf{r}'\}} \sigma^{(b)}(\mathbf{r}') \right), \end{aligned} \quad (3.5.17)$$

where J_1 and J_2 are taken to be positive. At $\mathcal{H} = 0$ the correspondence with the ferromagnet is the same as for the case $J_2 = 0$ except that now the equivalent ferromagnetic model has both first- and second-neighbour interactions. The ground-state Hamiltonians for the paramagnetic and antiferromagnetic states are

$$\begin{aligned} \widehat{H}_p &= \frac{1}{2} z_1 N J_1 - \frac{1}{2} z_2 N J_2 - N |\mathcal{H}|, \\ \widehat{H}_a &= -\frac{1}{2} z_1 N J_1 - \frac{1}{2} z_2 N J_2 \end{aligned} \quad (3.5.18)$$

and the critical field is

$$\mathcal{H}_c = z_1 J_1. \quad (3.5.19)$$

The simple antiferromagnet is recovered by putting $J_2 = 0$, $J_1 = J$ and $z_1 = z$. It can be seen that the critical field is not altered by the introduction of the second-neighbour interaction. This is because the second-neighbour energy contribution is the same for the paramagnetic and antiferromagnetic ground states.

3.5.1.2 The Simple Lattice Fluid

A lattice fluid is simply a lattice model where a site is either occupied by a fluid particle or unoccupied (a ‘hole’). The lattice site variable is $\tau(\mathbf{r})$, which take the value one if the site is occupied by a particle and zero otherwise. Each particles has chemical potential μ and particles on first-neighbour lattice sites have an interaction energy $-\varepsilon$, which either encourages or discourages first-neighbour occupation according as $\varepsilon > 0$ or $\varepsilon < 0$. The Hamiltonian is

$$\widehat{H}(\varepsilon, \mu; \tau) = -\varepsilon \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \tau(\mathbf{r}) \tau(\mathbf{r}') - \mu \sum_{\{\mathbf{r}\}} \tau(\mathbf{r}). \quad (3.5.20)$$

By taking $\tau(\mathbf{r}) := \frac{1}{2}[1 + \sigma(\mathbf{r})]$, where $\sigma(\mathbf{r})$ is the Ising spin variable it can be shown that (3.5.20) is equivalent to (3.5.1) with an additional contribution

$$C := -\frac{1}{4}[\varepsilon + 4(\mu/z)], \quad (3.5.21)$$

per first-neighbour pair and

$$J := \frac{1}{4}\varepsilon, \quad \mathcal{H} := \frac{1}{4}(z\varepsilon + 2\mu). \quad (3.5.22)$$

The fluid density is

$$\rho := \langle \tau(\mathbf{r}) \rangle = \frac{1}{2}[1 + \langle \sigma(\mathbf{r}) \rangle]. \quad (3.5.23)$$

It follows from the symmetry properties of the Ising model that, on the line $\mu = -\frac{1}{2}z\varepsilon$ in the μ - T plane, the density is one-half at high temperatures and, if in the corresponding Ising model spontaneous magnetization $\pm m_s(T)$ occurs for $T < T_c$, this line becomes a line of coexistence between phases (call them ‘liquid’ and ‘vapour’) with densities

$$\rho_\ell := \frac{1}{2}[1 + m_s(T)], \quad \rho_v := \frac{1}{2}[1 - m_s(T)]. \quad (3.5.24)$$

In general for any fluid system in a state with coexistent vapour and liquid phases the *diameter of the coexistence region* is

$$\rho_d(T) := \frac{1}{2}[\rho_\ell(T) + \rho_v(T)] \quad (3.5.25)$$

and the *width of the coexistence region* is

$$\rho_w(T) = \rho_\ell(T) - \rho_v(T). \quad (3.5.26)$$

In the simple lattice fluid $\rho_d = \frac{1}{2}$. This is a special case of the *rectilinear diameter law* which states that $\rho_d(T)$ is an asymptotically linear function of temperature in a neighbourhood of the critical point. It is now known that this law is not generally true. There are a number of models (Hemmer and Stell 1970; Widom and Rowlinson 1970; Mermin 1971a, b; Wheeler 1977) in which the coexistence diameter is singular at the critical point. One type of such model is treated in Sect. 10.6. In fact, it may be shown (Mermin and Rehr 1971), on the basis of some reasonable assumptions, that, unless the model has some special symmetry, as is the case for the Ising model, $d\rho_d/dT$ is at least as singular as the constant-volume heat capacity c_ρ .

This equivalence between a lattice fluid and the spin- $\frac{1}{2}$ Ising model, which was first shown by Lee and Yang (1952), holds between any one-component lattice fluid of particles of chemical potential μ and a spin- $\frac{1}{2}$ system on the same lattice in a magnetic field $\mathcal{H}$.

3.5.1.3 A Binary Mixture

We consider a mixture of two types of atoms **A** and **B** on a lattice. The chemical potentials are μ_A and μ_B and there are first-neighbour interaction energies ε_{AA} , ε_{AB} and ε_{BB} between **A**-**A**, **A**-**B** and **B**-**B** pairs. By identifying $\sigma(\mathbf{r}) = +1$ and $\sigma(\mathbf{r}) = -1$ with types **A** and **B** respectively it is easy to show that the system is equivalent to the spin- $\frac{1}{2}$ Ising model with Hamiltonian (3.5.1) with an additional constant contribution

$$C := \frac{1}{4}(2\varepsilon_{AB} + \varepsilon_{AA} + \varepsilon_{BB}) - \frac{1}{z}(\mu_A + \mu_B) \quad (3.5.27)$$

to each pair of sites and

$$J := \frac{1}{4}(2\varepsilon_{AB} - \varepsilon_{AA} - \varepsilon_{BB}), \quad \mathcal{H} := \frac{1}{4}[2(\mu_A - \mu_B) - z(\varepsilon_{AA} - \varepsilon_{BB})]. \quad (3.5.28)$$

Unlike the lattice fluid, which has phase separation and is equivalent to a ferromagnetic Ising model, the binary mixture can be equivalent to either ferromagnetic or antiferromagnetic models. In the former case phase separation will occur to A- and B-rich phases. The latter case yields sublattice ordering with one lattice preferentially occupied by As and the other by Bs. Such a phenomenon occurs in, for example, FeAl on a body-centered cubic lattice (Bragg and Williams 1934) and AuCu on a face-centred cubic lattice (Shockley 1938).

3.5.2 The Spin-1 Ising Model

For the case where $\sigma(\mathbf{r})$ at any lattice site $\mathbf{r}$ can take the three values ± 1 and 0 the Hamiltonian with only first-neighbour and single-site terms has additional contributions and takes the most general form

$$\begin{aligned} \hat{H}(J, \varepsilon, \varepsilon', \mathcal{H}, \mu; \sigma) = & -J \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma(\mathbf{r})\sigma(\mathbf{r}') - \varepsilon \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma^2(\mathbf{r})\sigma^2(\mathbf{r}') \\ & - \varepsilon' \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} [\sigma^2(\mathbf{r})\sigma(\mathbf{r}') + \sigma(\mathbf{r})\sigma^2(\mathbf{r}')] - \mathcal{H} \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}) - \mu \sum_{\{\mathbf{r}\}} \sigma^2(\mathbf{r}). \end{aligned} \quad (3.5.29)$$

There are now two densities m , conjugate to $\mathcal{H}$ and

$$\rho := \langle \hat{\rho} \rangle, \quad \hat{\rho} := \frac{1}{N} \sum_{\{\mathbf{r}\}} \sigma^2(\mathbf{r}), \quad (3.5.30)$$

conjugate to the chemical potential μ . This is the *spin-1 Ising model*, which is also called the *Blume–Emery–Griffiths model* in reference to the work of Blume et al. (1971) (see also, Ellis et al. 2005), who used the model to investigate the λ -transition and phase separation in ^3He – ^4He mixtures.¹⁶

The spin-1 Ising model Hamiltonian (3.5.29), with $J = \varepsilon' = \mathcal{H} = 0$ and the replacement $\sigma^2(\mathbf{r}) \rightarrow \tau(\mathbf{r})$, also gives (3.5.20) and is thus equivalent to a lattice

¹⁶ They solved the model in the mean-field approximation, with $\mathcal{H} = \varepsilon' = 0$. In that case the Hamiltonian, like (3.5.1), with $\mathcal{H} = 0$, is invariant under spin inversion on all the sites of the lattice.

fluid.¹⁷ In general it is used to model a dilute magnetic system, where the dilution state $\sigma(\mathbf{r}) = 0$ is either a vacancy, giving a lattice fluid of magnetic particles, or a second non-magnetic component, in which case the conflicting effects of magnetic and sublattice ordering can be studied (Thompson and Lavis 1967). It can also be used to investigate ternary mixtures (Mukamel and Blume 1974).

3.6 State-Difference Models

A very general class of site-variable models has Hamiltonians of the form

$$\hat{H}(\boldsymbol{\varepsilon}) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \epsilon(\varepsilon(\mathbf{r}) - \varepsilon(\mathbf{r}')), \quad (3.6.1)$$

where $\varepsilon(\mathbf{r})$ is a state variable at $\mathbf{r} \in \mathcal{N}$ and ϵ is an energy function. In this section we shall suppose that $\epsilon(\cdot)$ is an even function so that the ordering of the site vectors $\mathbf{r}$ and $\mathbf{r}'$ in (3.6.1) is immaterial. In Sect. 3.7 we introduce the idea of *chirality* for which this condition is relaxed.

Here the important distinction is between models where $\varepsilon(\mathbf{r})$ takes a continuum of values over some interval $[\varepsilon_0, \varepsilon_1)$ and those where the values of ε are integers $n_0, n_0 + 1, \dots, n_0 + \nu - 1$. The zero-field spin- $\frac{1}{2}$ Ising model is an example of the latter with

$$\varepsilon(\mathbf{r}) := \frac{1}{2}[1 - \sigma(\mathbf{r})], \quad \epsilon(\varepsilon) := K \cos(\pi \varepsilon), \quad (3.6.2)$$

when the Hamiltonian (3.5.1) take the form (3.6.1). We now consider a model with continuously-valued state variables, before turning to discrete-valued cases in Sects. 3.6.2–3.7.1.

3.6.1 The Classical XY Model

It can be seen from (3.3.3) and (3.6.2) that the Hamiltonian

$$\hat{H}(K, \boldsymbol{\theta}) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} K \cos[\theta(\mathbf{r}) - \theta(\mathbf{r}')], \quad (3.6.3)$$

could be taken to represent either the Ising model or the classical XY model, the difference being that in the Ising case $\theta(\mathbf{r}) = \pm \frac{1}{2}\pi$ and in the XY case $\theta(\mathbf{r}) \in [0, 2\pi)$.

¹⁷ Except that now the particles have a state degeneracy of two (the states ± 1) which will contribute a trivial term to the entropy.

Here we explore the second of these. For low temperatures we expect $|\theta(\mathbf{r}) - \theta(\mathbf{r}')|$ to be small with

$$\widehat{H}(K, \boldsymbol{\theta}) = \frac{1}{4}K \sum_{\{\mathbf{r}, \mathbf{r}'\}} \Delta(\mathbf{r} - \mathbf{r}') \{[\theta(\mathbf{r}) - \theta(\mathbf{r}')]^2 + O([\theta(\mathbf{r}) - \theta(\mathbf{r}')]^4\}, \quad (3.6.4)$$

where the summation is over all $\mathbf{r}$ and $\mathbf{r}' \in \mathcal{N}$ with $\Delta(\mathbf{r} - \mathbf{r}') := 1$ or 0 , according as $[\mathbf{r}, \mathbf{r}']$ is or is not a first-neighbour pair, and a constant term has been neglected. For simplicity we suppose that $\mathcal{N}$ is a d -dimensional hypercubic lattice of lattice spacing a , with $z = 2d$ and linear dimension $L = aN^{1/d}$. Let $x_1, x_2, \dots, x_d$ be cartesian coordinates in the lattice space with unit lattice vectors $\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_d$. Then, in the continuum limit $N \rightarrow \infty$, $a \rightarrow 0$ and L fixed,

$$\theta(\mathbf{r} \pm a\mathbf{i}_j) - \theta(\mathbf{r}) \simeq \pm a \frac{\partial \theta(\mathbf{r})}{\partial x_j}; \quad (3.6.5)$$

giving

$$\sum_{\{\mathbf{r}'\}} \Delta(\mathbf{r} - \mathbf{r}') \{[\theta(\mathbf{r}) - \theta(\mathbf{r}')]^2\} = 2a^2 [\nabla \theta(\mathbf{r})]^2. \quad (3.6.6)$$

Truncating (3.6.4) at the quadratic (harmonic) term gives

$$\widehat{H}(K, \boldsymbol{\theta}) = \frac{K}{2a^{d-2}} \int_{\mathcal{V}} d\mathcal{V} [\nabla \theta(\mathbf{r})]^2, \quad (3.6.7)$$

where $\mathcal{V}$ is the hypercube of edge length L which has replaced the lattice. With periodic boundary conditions applied to $\mathcal{V}$, $\theta(\mathbf{r})$ is related to its Fourier transform by formulae like (18.1.14) and (18.1.15). That is

$$\theta(\mathbf{r}) = \sum_{\{\mathbf{k}\}} \theta^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad \theta^*(\mathbf{k}) := \frac{1}{L^d} \int_{\mathcal{V}} d\mathcal{V} \theta(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}), \quad (3.6.8)$$

where $\overline{\theta^*(\mathbf{k})} = \theta^*(-\mathbf{k})$ and the values of $\mathbf{k}$ are given by (18.1.16) with $L^{(1)} = \dots = L^{(d)} = L$. Substituting into (3.6.7) gives

$$\widehat{H}(K, \boldsymbol{\theta}^*) = \frac{L^d K}{a^{d-2}} \sum_{\{\mathbf{k}\}} |\theta^*(\mathbf{k})|^2 k^2. \quad (3.6.9)$$

Then

$$\langle \cos(\theta(\mathbf{r})) \rangle = \langle \cos(\theta(\mathbf{0})) \rangle = \Re \left\{ \frac{\int_{\Theta^*} d\Theta^* \exp(-\widehat{H}'(K, \boldsymbol{\theta}^*))}{\int_{\Theta^*} d\Theta^* \exp(-\widehat{H}(K, \boldsymbol{\theta}^*))} \right\}, \quad (3.6.10)$$

where Θ^* is the space of the variables θ^* and

$$\widehat{H}'(K, \theta^*) := \widehat{H}(K, \theta^*) - i\theta(\mathbf{0}). \quad (3.6.11)$$

Let $\theta^*(\mathbf{k}) := \theta_r^*(\mathbf{k}) + i\theta_i^*(\mathbf{k})$, where $\theta_r^*(\mathbf{k})$ and $\theta_i^*(\mathbf{k})$ are, respectively, even and odd real functions of $\mathbf{k}$. Then

$$\begin{aligned} \widehat{H}'(K, \theta^*) = \frac{L^d K}{a^{d-2}} \sum_{\{\mathbf{k}\}} \left\{ \theta_r^*(\mathbf{k}) - \frac{ia^{d-2}}{2L^d K k^2} \right\}^2 k^2 \\ + \frac{a^{d-2}}{4KL^d} \sum_{\{\mathbf{k}\}} k^{-2} + \frac{L^d K}{a^{d-2}} \sum_{\{\mathbf{k}\}} \{\theta_i^*(\mathbf{k})\}^2 k^2. \end{aligned} \quad (3.6.12)$$

The integrands are Gaussian which are steep for large L allowing the ranges of integration to be extended to $(-\infty, \infty)$. It follows that

$$\langle \cos(\theta(\mathbf{r})) \rangle = \exp[\eta(d; K)I_1(d)], \quad (3.6.13)$$

where

$$\eta(d; K) := (2\pi a^{2-d} K)^{-1} = \frac{Ta^{d-2}}{2\pi J}, \quad (3.6.14)$$

$$I_1(d) := \frac{\pi}{2L^d} \sum_{\{\mathbf{k}\}} k^{-2}. \quad (3.6.15)$$

At this point it is useful to introduce two scale parameter $L_0 \sim L$ for the size of the system and $a_0 \sim a$, the minimum resolution length. Then

$$\ell_0 := 2\pi/L_0, \quad k_0 := 2\pi/a_0, \quad (3.6.16)$$

become natural low- and high-frequency boundaries with the summation (3.6.15) approximated by an integral over the hyperspherical shell $\ell_0 < k < k_0$. From (18.1.28) and (18.1.35)

$$\begin{aligned} I_1(d) &= \frac{(2\sqrt{\pi})^{2-d}}{4\Gamma(\frac{d}{2})} \int_{\ell_0}^{k_0} dk k^{d-3}, \\ &= \begin{cases} \frac{(2\sqrt{\pi})^{2-d}}{4(d-2)\Gamma(\frac{d}{2})} \left[1 - \left(\frac{\ell_0}{k_0} \right)^{d-2} \right], & d \neq 2, \\ \frac{1}{4} \ln(k_0/\ell_0), & d = 2; \end{cases} \end{aligned} \quad (3.6.17)$$

which, when $k_0 \gg \ell_0$, gives

$$\langle \mathbf{s}(\mathbf{r}) \rangle = \begin{cases} \exp\left(-\frac{\eta(d; K)(k_0/2\sqrt{\pi})^{d-2}}{4 \Gamma\left(\frac{d}{2}\right)(d-2)}\right), & d > 2, \\ \left(\frac{\ell_0}{k_0}\right)^{\frac{1}{4}\eta(2; K)}, & d = 2, \\ \exp\left(-\frac{\eta(1; K)}{2\ell_0}\right), & d = 1. \end{cases} \quad (3.6.18)$$

In a similar way

$$\begin{aligned} \langle \mathbf{s}(\mathbf{r} + \mathbf{r}') \cdot \mathbf{s}(\mathbf{r}') \rangle &= \langle \cos[\theta(\mathbf{r}) - \theta(\mathbf{0})] \rangle \\ &= \Re \left\{ \frac{\int_{\Theta^*} d\Theta^* \exp\{-\widehat{H}''(K, \boldsymbol{\theta}^*)\}}{\int_{\Theta^*} d\Theta^* \exp(-\widehat{H}(K, \boldsymbol{\theta}^*))} \right\}, \end{aligned} \quad (3.6.19)$$

where

$$\begin{aligned} \widehat{H}''(K, \boldsymbol{\theta}^*) &:= \widehat{H}(K, \boldsymbol{\theta}^*) - i[\theta(\mathbf{r}) - \theta(\mathbf{0})] \\ &= -Kd\left(\frac{L}{a}\right)^d + \frac{L^d K}{a^{d-2}} \sum_{\{\mathbf{k}\}} \left\{ \theta_{\mathbf{r}}^*(\mathbf{k}) - \frac{i a^{d-2} [\cos(\mathbf{r} \cdot \mathbf{k}) - 1]}{2L^d K k^2} \right\}^2 k^2 \\ &\quad + \frac{L^d K}{a^{d-2}} \sum_{\{\mathbf{k}\}} \left\{ \theta_{\mathbf{i}}^*(\mathbf{k}) + \frac{i a^{d-2} \sin(\mathbf{r} \cdot \mathbf{k})}{2L^d K k^2} \right\}^2 k^2 \\ &\quad + \frac{a^{d-2}}{2KL^d} \sum_{\{\mathbf{k}\}} k^{-2} [1 - \exp(i\mathbf{r} \cdot \mathbf{k})]. \end{aligned} \quad (3.6.20)$$

Thus

$$\langle \cos[\theta(\mathbf{r}) - \theta(\mathbf{0})] \rangle = \exp[\eta(d; K) I_2(d; \mathbf{r})], \quad (3.6.21)$$

where

$$I_2(d; \mathbf{r}) := \frac{\pi}{L^d} \sum_{\{\mathbf{k}\}} k^{-2} [\exp(i\mathbf{r} \cdot \mathbf{k}) - 1]. \quad (3.6.22)$$

From (18.1.28), (18.1.35) and (18.1.36),

$$\begin{aligned} I_2(k; \mathbf{r}) &= \frac{(2\sqrt{\pi})^{2-d}}{2\pi \Gamma\left(\frac{d}{2}\right)} \int_{\ell_0}^{k_0} dk k^{d-3} \int_0^{2\pi} d\psi [\exp(ikr \cos \psi) - 1] \\ &= \frac{(2\sqrt{\pi})^{2-d}}{\Gamma\left(\frac{d}{2}\right)} \int_{\ell_0}^{k_0} dk k^{d-3} [J_0(rk) - 1] \end{aligned}$$

$$= -\frac{(2r\sqrt{\pi})^{2-d}}{\Gamma\left(\frac{d}{2}\right)} \mathfrak{F}(r\ell_0, rk_0; d-3), \quad (3.6.23)$$

where

$$\mathfrak{F}(\ell_0, k_0; x) := \int_{\ell_0}^{k_0} k^x \{1 - J_0(k)\} dk. \quad (3.6.24)$$

Interest in this integral can be restricted to $x = -2, -1, 0, 1, \dots$ and in Sect. 18.1.4 we show, by dividing the range of integration in parts that the limit $\ell_0 \rightarrow 0$ can be taken in all cases. Substituting from (18.1.49)–(18.1.54), for $|\mathbf{r} - \mathbf{r}'| := r \rightarrow \infty$, with $k_0 \gg 0$,

$$\langle \mathbf{s}(\mathbf{r}) \cdot \mathbf{s}(\mathbf{r}') \rangle \simeq \begin{cases} \exp\left(-\frac{\eta(d; K)(k_0/2\sqrt{\pi})^{d-2}}{\Gamma\left(\frac{d}{2}\right)(d-2)}\right), & d \geq 3, \\ (rk_0)^{-\eta(2; K)}, & d = 2, \\ \exp[-2\eta(1; K)r], & d = 1. \end{cases} \quad (3.6.25)$$

For $d > 2$ the spin-spin correlation function is non-zero in the limit $r \rightarrow \infty$, indicating the presence of a long-range ordered phase. This conclusion is supported in three dimensions where high-temperature series analysis yields a second-order transition in the universality class of the λ -transition in liquid ^4He (Butera et al. 1993). For $d = 1$ the correlation function decays exponentially relative to a correlation length $1/2\eta(1; K) = \pi aJ/T$ which diverges as $T \rightarrow 0$, and $\langle \mathbf{s}(\mathbf{r}) \rangle \rightarrow 0$ as $L \rightarrow \infty$, for all $T > 0$, indicating that for non-zero temperature there is no long-range order.

However, it is the two-dimensional case which makes this model of particular interest. In the theoretical sense this arises from its archetypal status as a representative of the universality class of $O(2)$ models which exhibit a *Kosterlitz–Thouless transition*. The experimental importance of this transition, and hence of the two-dimensional XY model, is that it models the superfluid transition in ^4He films (Bishop and Reppy 1978; Minnhagen 1987; Chaikin and Lubensky 1995, pp. 551–555) and the transition in weakly-disordered two-dimensional superconductors (Resnick et al. 1981; Erez and Meir 2010).

3.6.1.1 The Two-Dimensional XY Model: Kosterlitz–Thouless Transition

Moore (1969) concluded from high-temperature series calculations that the evidence for a phase transition in the two-dimensional XY model “is as strong in two dimensions as it is in three”. However, the Mermin–Wagner theorem excludes the existence of a long-range ordered phase, a conclusion supported by (3.6.18) which yields $\langle \mathbf{s}(\mathbf{r}) \rangle = 0$ for all temperatures in the limit $L_0 \rightarrow \infty$ of an infinite system. On the other hand the correlation function is algebraic, like the critical-point correlation

in standard scaling theory (Eq.(4.5.55)), with a temperature-dependent exponent $\eta(2; K)$ indicative of a line of fixed points.¹⁸ It is not to be expected that the correlation function can persist in this form as $T \rightarrow \infty$; there one expects exponential decay on the scale of the correlation length (Eq.(5.6.37)) and this is borne out by the renormalization group analysis of Sect. 15.6. Thus some kind of phase transition must be anticipated, but from a phase which is not excluded by the conditions of the Mermin–Wagner theorem. This puzzle was resolved by the work of Berezinskii (1971, 1972) and Kosterlitz and Thouless (1972, 1973). They showed that there is indeed a phase transition (now called the *Kosterlitz–Thouless transition*) to a form of *quasi-long range order* very different from that for the three-dimensional model.

To understand the nature of this phase we must re-examine the assumptions underlying the analysis given above. The most obvious of these is that $|\theta(\mathbf{r}) - \theta(\mathbf{r}')|$, for a pair of first-neighbour sites, is small. This assumption, valid at low temperatures, is the basis for the replacement of the Hamiltonian (3.6.3) first by (3.6.4) and then in the continuum limit by (3.6.7), where the set of periodic angle variables $\theta(\mathbf{r})$ for all $\mathbf{r} \in \mathcal{N}$ becomes a *single* scalar field $\theta(\mathbf{r})$ defined at all $\mathbf{r} \in \mathcal{V}$.

Whilst we may choose to accept the harmonic assumption, what must be focused on is the inherited property arising from the periodicity, and, therefore, non-single-valued nature, of the angle variables. Suppose $\mathbf{r}^*$ is a point at the centre of one of the squares of $\mathcal{N}$. That is to say, at a site of the dual lattice $\mathcal{N}^*$ (Sect. 16.2.2.1). And suppose that all the angle variables are aligned in the same (clockwise or anticlockwise) direction along tangents to circles centre $\mathbf{r}^*$. In the continuum limit this becomes a scalar field for which we define the annulus

$$\mathcal{R}(\mathbf{r}^*) := \{\mathbf{r} : a \leq |\mathbf{r} - \mathbf{r}^*| \leq R\} \quad (3.6.26)$$

and

$$\theta^{(\omega)}(\mathbf{r}^*; \mathbf{r}) := \begin{cases} \omega \arccos \left[\frac{(\mathbf{r} - \mathbf{r}^*) \cdot \mathbf{r}^*}{|\mathbf{r}^*| |\mathbf{r} - \mathbf{r}^*|} \right], & \mathbf{r} \in \mathcal{R}(\mathbf{r}^*), \\ 0, & \mathbf{r} \notin \mathcal{R}(\mathbf{r}^*), \end{cases} \quad (3.6.27)$$

where $\omega = \pm 1$ according as the orientation about $\mathbf{r}^*$ is anticlockwise or clockwise. This field satisfies

$$\oint_{\mathcal{C}} \nabla \theta^{(\omega)}(\mathbf{r}^*; \mathbf{r}) \cdot d\mathbf{r} = 2\pi\omega, \quad (3.6.28)$$

where $\mathcal{C}$ is any simple curve, in the anti-clockwise direction, enclosing $\mathbf{r}^*$ and lying in $\mathcal{R}(\mathbf{r}^*)$. $\theta^{(\omega)}(\mathbf{r}^*; \mathbf{r})$ is the field of a *vortex* v centre $\mathbf{r}^*$ with *winding number* ω , which measures the number of anticlockwise rotations of the field around a circle centre $\mathbf{r}^*$ (so it can, henceforth, be taken to be any positive or negative integer). We have introduced via the annulus $\mathcal{R}(\mathbf{r}^*)$ a core of radius the lattice distance a and

¹⁸ The coupling J/T is marginal like the four-spin coupling in the spin-representation of the eight-vertex model (Sect. 12.3.9).

an outer radius R . Ignoring these constrictions, $\theta^{(v)}(\mathbf{r}^*; \mathbf{r})$, given by the first line in (3.6.27), has a branch-point at $\mathbf{r} = \mathbf{r}^*$ and satisfies

$$\nabla^2 \theta^{(v)}(\mathbf{r}^*; \mathbf{r}) = 2\pi\omega \delta^D(\mathbf{r} - \mathbf{r}^*). \quad (3.6.29)$$

From (3.6.7) the energy of v is

$$\hat{H}^{(v)}(K) = \frac{1}{2}K \int_{\mathcal{R}(\mathbf{r}^*)} dV [\nabla \theta^{(v)}(\mathbf{r}^*; \mathbf{r})]^2. \quad (3.6.30)$$

With polar variables (q, ψ) with respect $\mathbf{r}^*$ and the unit axis vector $\hat{\mathbf{i}}$,

$$\theta^{(v)}(\mathbf{r}^*; \mathbf{r}) = \omega \psi, \quad \nabla \theta^{(v)}(\mathbf{r}^*; \mathbf{r}) = \frac{\omega}{q} \hat{\boldsymbol{\psi}}. \quad (3.6.31)$$

Substituting into (3.6.30) gives

$$\hat{H}^{(v)}(K) = \pi\omega^2 K \int_a^R \frac{dq}{q} = -\pi\omega^2 K \ln\left(\frac{a}{R}\right). \quad (3.6.32)$$

Consider now the case of a second vortex of winding number $\tilde{\omega}$ located at $\mathbf{r}^* + a\hat{\mathbf{i}}$. It is not difficult to show that

$$\tilde{\theta}^{(v)}(\mathbf{r}^* + a\hat{\mathbf{i}}; \mathbf{r}) = \tilde{\omega} \left\{ \psi + \frac{|\sin \psi|}{q} a - \frac{|\sin \psi| \cos \psi}{2q^2} a^2 + O(a^3) \right\}, \quad (3.6.33)$$

giving for the vortex pair $v + \tilde{v}$

$$\hat{H}^{(v)}(K) = \begin{cases} 2\pi K[(\omega + \tilde{\omega})^2 + O(a)] \ln\left(\frac{R}{a}\right), & \tilde{\omega} \neq -\omega, \\ \pi K \left[1 - \left(\frac{a}{R}\right)^2 \right] + O(a^3), & \tilde{\omega} = \omega. \end{cases} \quad (3.6.34)$$

Whilst the energy of a single vortex is $O(K \ln[R/a])$, that of a bound pair with opposite winding numbers on neighbouring sites is $O(K)$. However, of course, the thermodynamic quantity of interest is the free energy and one expects bound pairs of vortices to dissociate if there is a free-energy advantage in doing so. A free vortex has entropy of the order of $\ln[(R/a)^2]$. So the free energy released when a single vortex of winding number ω becomes unbound is of the order of $[\pi\omega^2 - 2K^{-1}] \ln(R/a)$. This decreases as T increases becoming negative for $T > \pi\omega^2 J/2$. For

$$T < T_{\text{KT}} := \pi J/2 \simeq 1.5708J, \quad (3.6.35)$$

all free vortices are suppressed and as temperature increases unbinding of vortices of higher winding number becomes advantageous. The *Kosterlitz–Thouless transition* is

this unbinding transition. As will be seen in the real-space renormalization treatment of Sect. 15.6 it occurs at a temperature on a critical line controlled by a fixed point at the *Kosterlitz–Thouless temperature* T_{KT} .

Consider now the case of a number of vortices with¹⁹

$$\theta(\mathbf{r}) = \sum_{j=1}^{N_v} \theta_j^{(\vee)}(\mathbf{r}_j^*; \mathbf{r}), \quad (3.6.36)$$

where $\theta_j^{(\vee)}(\mathbf{r}_j^*; \mathbf{r})$ is the field of a vortex v_j , centre $\mathbf{r}_j^* \in \mathcal{N}^*$ and winding number ω_j . From (3.6.29),

$$\nabla^2 \theta(\mathbf{r}) = 2\pi \rho(\mathbf{r}), \quad (3.6.37)$$

where

$$\rho(\mathbf{r}) := \sum_{j=1}^{N_v} \omega_j \delta^{\text{D}}(\mathbf{r} - \mathbf{r}_j^*) \quad (3.6.38)$$

is the *vortex distribution function*. Imposing the condition that the total vorticity of the system is zero gives

$$\int_{\mathcal{V}} \rho(\mathbf{r}) d\mathcal{V} = \sum_{j=1}^{N_v} \omega_j = 0. \quad (3.6.39)$$

Using (18.1.19) to take the Fourier transform of (3.6.37) gives²⁰

$$k^2 \theta^*(\mathbf{k}) = -\frac{2\pi}{L^2} \sum_{j=1}^{N_v} \omega_j \exp(-i\mathbf{k} \cdot \mathbf{r}_j^*). \quad (3.6.40)$$

Then using the *lattice Green's function* $G(2; r)$ defined in Sect. 18.1.5, from (3.6.9),

¹⁹ In fact the total field of the system will contain a conservative single-valued term $\theta^{(\text{s})}(\mathbf{r})$, which thus satisfies (3.6.28) with $\omega = 0$ (Kosterlitz and Thouless 1973). It makes an additively separate contribution to the Hamiltonian and will, therefore, for the sake of simplicity be ignored, although its perturbative effect is the cause of the absence of long-range below the critical point.

²⁰ The asterisk ‘*’ for the Fourier transform should not be confused with the star ‘★’ for the dual lattice.

$$\begin{aligned}
\hat{H}(K) &= \frac{4\pi^2 K}{L^2} \sum_{j=1}^{N_v} \sum_{j'=1}^{N_v} \omega_j \omega_{j'} \sum_{\{k\}} \frac{\exp(-ik \cdot [\mathbf{r}_{j'}^* - \mathbf{r}_j^*])}{k^2} \\
&= 4\pi^2 K \sum_{j=1}^{N_v} \sum_{j'=1}^{N_v} \omega_j \omega_{j'} G(2; |\mathbf{r}_j^* - \mathbf{r}_{j'}^*|) \\
&= -\pi K \sum_{\{j \neq j'\}} \omega_j \omega_{j'} \ln \left| \frac{\mathbf{r}_j^* - \mathbf{r}_{j'}^*}{a} \right| - K\mu \sum_{j=1}^{N_v} \omega_j^2, \tag{3.6.41}
\end{aligned}$$

where in the first summation each pair is counted only once and the last term is the sum of the energies of isolated vortices. The last line of (3.6.32) is of the form of the Hamiltonian of a *Coulomb gas* on the dual lattice $\mathcal{N}^*$ (Minnhagen 1987; Nienhuis 1987) but slightly restricted since the second summation, which corresponds to the fugacity of the gas is related via the vortex winding numbers to the charge energies given by the coefficients in the first summation.

Other models related to the XY model by duality transformations are the *sold-on-solid model* (Knops 1977; see also Cardy 1996, Sect. 6.3) and the *Villain model* (Villain 1975; José et al. 1977; see also Chaikin and Lubensky 1995, Appendix 9B). Further models exhibiting a Kosterlitz–Thouless are discussed by Nelson (1983). A real-space renormalization group treatment of the Kosterlitz–Thouless transition is given in Sect. 15.6.

3.6.2 The Ashkin–Teller Model

Ashkin and Teller (1943) proposed a model with four types of microsystems, A, B, C and D, on the sites of the lattice $\mathcal{N}$ with first-neighbour interaction energies w_0 for AA, BB, CC and DD pairs, w_1 for AB and CD pairs, w_2 for AC and BD pairs and w_3 for AD and BC pairs. In the special case $w_3 = w_1$, if the states A, B, C and D at $\mathbf{r} \in \mathcal{N}$ are identified respectively with $\varepsilon(\mathbf{r}) := \sigma(\mathbf{r}) = 0, 1, 2, 3$, the Ashkin–Teller model has the Hamiltonian (3.6.1) with $\epsilon(0) = w_0$, $\epsilon(\pm 1) = \epsilon(\pm 3) = w_1$, $\epsilon(\pm 2) = w_2$.

Fan (1972) formulated the model in terms of two spin- $\frac{1}{2}$ Ising variables $\tilde{\sigma}(\mathbf{r})$ and $\tilde{\tau}(\mathbf{r})$ at each site $\mathbf{r} \in \mathcal{N}$ with a site occupation identified by $(\tilde{\sigma}, \tilde{\tau})$ with A := (+1, +1), B := (+1, -1), C := (-1, +1) and D := (-1, -1). The Hamiltonian then takes the form

$$\begin{aligned}
\hat{H}(K_0, K_1, K_2, K_3; \tilde{\sigma}, \tilde{\tau}) &:= -K_0 - \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} [K_1 \tilde{\sigma}(\mathbf{r}) \tilde{\sigma}(\mathbf{r}') + K_2 \tilde{\tau}(\mathbf{r}) \tilde{\tau}(\mathbf{r}')] \\
&\quad + K_3 \tilde{\sigma}(\mathbf{r}) \tilde{\tau}(\mathbf{r}) \tilde{\sigma}(\mathbf{r}') \tilde{\tau}(\mathbf{r}'), \tag{3.6.42}
\end{aligned}$$

where $K_i := J_i/T$, $i = 0, \dots, 3$, and

$$\begin{aligned}
J_0 &:= -\frac{1}{4}(w_0 + w_1 + w_2 + w_3), & J_1 &:= -\frac{1}{4}(w_0 + w_1 - w_2 - w_3), \\
J_2 &:= -\frac{1}{4}(w_0 - w_1 + w_2 - w_3), & J_3 &:= -\frac{1}{4}(w_0 - w_1 - w_2 + w_3).
\end{aligned}
\tag{3.6.43}$$

This result applies to any lattice. However, if a square lattice $\mathcal{N}$ is chosen it is clear that (3.6.43) represents two first-neighbour Ising models linked by a four-site interaction. Using this result and the spin formulation of the eight-vertex model developed by Kadanoff and Wegner (1971) and Wu (1971) (Sect. 3.8.1.3), Wegner (1972) showed that the Ashkin–Teller and eight-vertex models are related by a duality relation.

3.6.3 Potts Models

The problem suggested to Renfrey Potts by his research supervisor Cyril Domb was to generalize the Ising condition (3.6.2) to one in which the vector states can point in ν directions, expressed by ‘recycling’ the Ising variable $\sigma(\mathbf{r})$ to give

$$\theta(\mathbf{r}) = 2\pi\sigma(\mathbf{r})/\nu, \quad \sigma(\mathbf{r}) = 0, 1, \dots, \nu - 1. \tag{3.6.44}$$

With $L = 0$ the Hamiltonian now takes the form

$$\widehat{H}(K; \sigma) = -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \cos\left(\frac{2\pi}{\nu}[\sigma(\mathbf{r}) - \sigma(\mathbf{r}')] \right). \tag{3.6.45}$$

Domb (1974a) called this the *planar Potts model* and it is also called the *vector Potts model* or *clock model*. Potts (1952) obtained the critical point on the square lattice for $\nu = 2, 3, 4$.

An obvious generalization to (3.6.45) is to adopt the Hamiltonian (3.6.1) with $\varepsilon(\mathbf{r}) := \sigma(\mathbf{r}) = 0, 1, \dots, \nu - 1$ and

$$\mathfrak{e}(\sigma + n\nu) = \mathfrak{e}(\sigma), \quad n = 0, \pm 1, \pm 2, \dots \tag{3.6.46}$$

which is invariant under the cyclic group $\mathcal{C}_\nu$, where $\mathcal{C}_\nu : \sigma \rightarrow \sigma + 1 \pmod{\nu}$. An additional property, satisfied by some but not all of the Potts models (Sect. 3.7) is that $\mathfrak{e}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))$ is invariant under the interchange of states on $\mathbf{r}$ and $\mathbf{r}'$. This implies that

$$\mathfrak{e}(\nu - \sigma) = \mathfrak{e}(\sigma). \tag{3.6.47}$$

The Ashkin–Teller model with $w_1 = w_3$ satisfies both (3.6.46) and (3.6.47) with $\nu = 4$. From (3.6.46) and (18.1.9),

$$\epsilon(\sigma) = \frac{1}{\nu} \sum_{k=0}^{\nu-1} \kappa_k \exp\left(i \frac{2\pi k \sigma}{\nu}\right). \quad (3.6.48)$$

with (3.6.47) giving $\kappa_{\nu-k} = \kappa_k$. For the special case of the clock-model Hamiltonian (3.6.45) $\kappa_1 = \kappa_{\nu-1} = \frac{1}{2} K \nu$ and $\kappa_k = 0$ otherwise.

3.6.4 The Standard Potts Model

One possibility suggested by Potts (1952) is when the interaction between a first-neighbour pair has one value when the states are the same and another value when they are different, irrespective of what that difference may be. This model, called by Domb (1974a) the *standard Potts model*, is the one most usually associated with the name of Potts, to the extent that it is commonly (and henceforth in this text) referred to simply as *the Potts model* (Wu 1982; Martin 1991),

Without loss of generality let the energy of a like pair be $-R$ and of an unlike pair be zero. Then in (3.6.1), with $\varepsilon(\mathbf{r}) := \sigma(\mathbf{r}) = 0, 1, \dots, \nu - 1$, $\epsilon(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')) := K \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))$ and $K := R/T$,

$$\widehat{H}(K; \sigma) = \widehat{H}(R; \sigma)/T := -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')). \quad (3.6.49)$$

From (18.1.11), (3.6.48) applies with $\kappa_k = 1$, $k = 0, 1, \dots, \nu - 1$. The Hamiltonian (3.6.49) is invariant under the symmetric group $\mathbb{S}_\nu$, which permutes the ν microstates among themselves. As in the case of the Ising model, it is said to be *ferromagnetic* when $R > 0$, encouraging the members of a first-neighbour pair to be in the same state and *antiferromagnetic* when $R < 0$, encouraging the members of a first-neighbour pair to be in different states. The Ashkin–Teller model with $w_1 = w_2 = w_3$ is equivalent to the (standard) four-state Potts model and Temperley and Lieb (1971) showed the equivalence between this model on a square lattice with different interactions in the two lattice direction and a staggered six-vertex model (Sect. 3.8.2.1).

Another form for the partition function can be obtained using

$$\exp[K \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))] = 1 + v \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')), \quad (3.6.50)$$

where

$$v := \exp(K) - 1. \quad (3.6.51)$$

This gives

$$\begin{aligned}
 Z(\mathcal{N}, v) &= \sum_{\{\sigma\}} \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} [1 + v \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))] \\
 &= \sum_{\{\mathbf{g}\}}^{(\text{AG})} v^{c(\mathbf{g})} v^{e(\mathbf{g})} = W(v, v),
 \end{aligned} \tag{3.6.52}$$

where, as defined Sect. 16.4.1, the summation is over augmented graphs and $W(v, v)$ is the Whitney polynomial defined in (16.4.2). This is the *cluster form* for the Potts partition function, the components of $\mathbf{g}$ being *Fortuin–Kasteleyn clusters* (Kasteleyn and Fortuin 1969).²¹

Although (3.6.49) represents the ‘pure’ Potts model it is often useful to regard it as a kind of lattice fluid with $\sigma = 0$ and $\sigma \neq 0$ corresponding respectively to a hole and a particle. Then the Hamiltonian is modified to

$$\begin{aligned}
 \hat{H}(K, L; \sigma) &:= \hat{H}(R, \mu; \sigma)/T := -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')) \\
 &\quad - L \sum_{\{\mathbf{r}\}} [1 - \delta^{\text{Kr}}(\sigma(\mathbf{r}))],
 \end{aligned} \tag{3.6.53}$$

where $L := \mu/T$ and μ is the chemical potential of a particle. Of course, the Hamiltonian (3.6.53) no longer has $\mathbb{S}_v$ symmetry. In the next sections we consider equivalences with Ising models.

3.6.4.1 $v = 2$: The Spin- $\frac{1}{2}$ Ising Model

The spin- $\frac{1}{2}$ Ising variable which we now denote by $\tilde{\sigma}(\mathbf{r})$, is related to the Potts variable by²²

$$\tilde{\sigma}(\mathbf{r}) := 2\sigma(\mathbf{r}) - 1, \tag{3.6.54}$$

with

$$\begin{aligned}
 \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')) &= \frac{1}{2} [1 + \tilde{\sigma}(\mathbf{r}) \tilde{\sigma}(\mathbf{r}')], \\
 1 - \delta^{\text{Kr}}(\sigma(\mathbf{r})) &= \frac{1}{2} [1 + \tilde{\sigma}(\mathbf{r})].
 \end{aligned} \tag{3.6.55}$$

²¹ It should be noted that these clusters are not identical to the set of graphs obtained by joining sites with the same microstate. The difference is that they also include graphs derived from the latter set by omitting one or more edges. A small calculation for a lattice consisting of a triangle of sites illustrates this point.

²² Here and in the rest of the chapter we shall need some ‘temporary’ notation to express relations between different models.

Substituting from (3.6.55) into (3.6.49) and comparing with (3.5.1) we see that the Hamiltonians and partition functions for the two-state Potts model (PM2) and spin- $\frac{1}{2}$ Ising model (IMH) are related by

$$\widehat{H}^{(\text{PM2})}(R, \mu; \sigma) = -\frac{1}{4}N(zR + 2\mu) + \widehat{H}^{(\text{IMH})}\left(\frac{1}{2}R, \frac{1}{2}\mu; \tilde{\sigma}\right), \quad (3.6.56)$$

$$Z^{(\text{PM2})}(N, K, L) = \exp\left(\frac{1}{4}N[zK + 2L]\right) Z^{(\text{IMH})}\left(N, \frac{1}{2}K, \frac{1}{2}L\right), \quad (3.6.57)$$

where $K := R/T$ and $L := \mu/T$.

3.6.4.2 $\nu = 3$: The Spin-1 Ising Model

The spin-1 Ising variable, which we again denote by $\tilde{\sigma}(\mathbf{r})$, is related to the Potts variable by

$$\tilde{\sigma}(\mathbf{r}) := \sigma(\mathbf{r}) - 1, \quad (3.6.58)$$

with

$$\begin{aligned} \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')) &= 1 + \frac{1}{2}\tilde{\sigma}(\mathbf{r})\tilde{\sigma}(\mathbf{r}') + \frac{3}{2}\tilde{\sigma}^2(\mathbf{r})\tilde{\sigma}^2(\mathbf{r}') \\ &\quad - \tilde{\sigma}^2(\mathbf{r}) - \tilde{\sigma}^2(\mathbf{r}'), \end{aligned} \quad (3.6.59)$$

$$1 - \delta^{\text{Kr}}(\sigma(\mathbf{r})) = \tilde{\sigma}^2(\mathbf{r}).$$

Substituting from (3.6.59) into (3.6.49) and comparing with (3.5.29) we see that the Hamiltonians and partition functions for the three-state Potts model (PM3) and spin-1 Ising model (IMO) are related by

$$\widehat{H}^{(\text{PM3})}(R, \mu; \sigma) = -\frac{1}{2}NzR + \widehat{H}^{(\text{IMO})}\left(\frac{1}{2}R, \frac{3}{2}R, 0, 0, \mu - zR; \tilde{\sigma}\right), \quad (3.6.60)$$

$$Z^{(\text{PM3})}(N, K, L) = \exp\left(\frac{1}{2}zKN\right) Z^{(\text{IMO})}\left(N, \frac{1}{2}K, \frac{3}{2}K, 0, 0, L - zK\right). \quad (3.6.61)$$

This result can also be derived by imposing $\mathcal{S}_3$ symmetry on (3.5.29).

3.7 Chirality

The term ‘chirality’ was introduced by Lord Kelvin (William Thomson) (Thomson 1904) who chose to “call any geometrical figure, or group of points, chiral, and say it has chirality, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself” (ibid, p. 619). This idea, suitably interpreted (see, for example, Le Guennec 1998), now plays an important role in physics, chemistry and mathematics.

In giving a definition of chirality tailored to our present needs it should first be noted that Kelvin's definition is couched in terms of the absence of something (call it 'achirality'). In statistical mechanics, an achiral system is one which, expressed in rather an imprecise way, has a Hamiltonian which is invariant under orientation reversal. To make this idea more precise we concentrate on two-dimensional lattices with the orientation in question being the ordering (clockwise or anticlockwise) of sites around a lattice face with respect to a chosen direction of the axis orthogonal to the plane. As we saw in Sect. 17.3 the lattice symmetry group for operations in the plane consists of translations and a point group $\mathcal{C}_{nv}$ for some n . Of these group operations, translations and rotations are direct isometries (preserving orientation) and the reflections are indirect isometries (reversing orientation). Then

Definition 3.7.1 The system is *achiral* if the value of the Hamiltonian is unchanged under the operation of all the indirect isometries in the lattice symmetry group.

Thus a system is chiral if there is at least one indirect isometry in the lattice symmetry group which alters the value of the Hamiltonian.

3.7.1 Chiral Potts Models

From the general Potts interaction $\epsilon(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))$ defined in Sect. 3.6.3, the invariant condition under interchange of the states on sites $\mathbf{r}$ and $\mathbf{r}'$ is removed. This is most simply represented by giving a vector-direction $\mathbf{r} - \mathbf{r}'$ to the lattice edge and replacing $\epsilon(\sigma)$ with $\epsilon(\sigma; \mathbf{r} - \mathbf{r}')$, so that (3.6.47) is replaced by

$$\epsilon(v - \sigma; \mathbf{r}) = \epsilon(\sigma; -\mathbf{r}) \quad (3.7.1)$$

and (3.6.48) becomes

$$\epsilon(\sigma; \mathbf{r}) = \frac{1}{v} \sum_{k=0}^{v-1} \kappa_k(\mathbf{r}) \exp\left(i \frac{2\pi k \sigma}{v}\right), \quad (3.7.2)$$

with $\kappa_{v-k}(\mathbf{r}) = \kappa_k(-\mathbf{r})$.

$$\hat{H}(K; \sigma) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \epsilon(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'); \mathbf{r} - \mathbf{r}') \quad (3.7.3)$$

is the most general form for the Hamiltonian of the *chiral Potts model*. A special case is the *chiral clock model*

$$\hat{H}(K, \Delta; \sigma) := -K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \cos\left(\frac{2\pi}{v} [\sigma(\mathbf{r}) - \sigma(\mathbf{r}') - \Delta \cdot (\mathbf{r} - \mathbf{r}')] \right), \quad (3.7.4)$$

for which $\kappa_1(\mathbf{r}) = \frac{1}{2}K\nu \exp(-2\pi i \mathbf{\Delta} \cdot \mathbf{r}/\nu)$, $\kappa_{\nu-1}(\mathbf{r}) = \frac{1}{2}K\nu \exp(2\pi i \mathbf{\Delta} \cdot \mathbf{r}/\nu)$ and $\kappa_k(\mathbf{r}) = 0$ otherwise. This model was first proposed by Ostlund (1981) and Huse (1981) to investigate a commensurate-incommensurate phase transition. For the square lattice, with $\mathbf{r} = a(n_x \hat{\mathbf{x}} + n_y \hat{\mathbf{y}})$ and $\mathbf{\Delta} = \Delta \hat{\mathbf{x}}$ the ground states will consist of lines of sites in the y direction in the same state. Corresponding to the $\sigma = 0$ ground state chosen from the ν ferromagnetic ground states of the non-chiral ($\mathbf{\Delta} = \mathbf{0}$) clock model we now choose the line $\mathbf{r}' = an_y \hat{\mathbf{y}}$ of sites to be in state $\sigma = 0$ and the line of site $\mathbf{r} = a(n_x \hat{\mathbf{x}} + n_y \hat{\mathbf{y}})$ will be in the state $\sigma = s \pmod{\nu}$ for which $s - \Delta n_x a$ has the least value for all integer values of s . These ground states identify long-range ordered commensurate phases from which there are phase transitions to incommensurate and disordered phases (Ostlund 1981; Huse 1981; Duxbury et al. 1984; Houlrik and Knak Jensen 1986).

The search for a form for the chiral Potts interaction $\epsilon(\sigma; \mathbf{r})$, which yields a solvable model has attracted a great deal of interest. Such a case (now often simply referred to as ‘the chiral Potts model’) was obtained by Baxter et al. (1988). Over the next two years the free energy of the model was derived (Baxter 1988, 1990; Baxter et al. 1990)²³ and in 2005 Baxter obtained the order parameter (Baxter 2005a, b, 2006).

3.7.2 An Extended 3-State Potts Model on the Triangular Lattice

A different type of chirality involving a three-site term for the 3-state Potts model on a triangular lattice arose as a ‘biprodut’ of the development of a lattice model for water by Bell and Lavis (1970).

Suppose that the triangular lattice $\mathcal{N}_T$ is divided into three equivalent sublattices $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$, which can be identified respectively with the black, grey and white sites in Fig. 16.4. It will be noted that the triangles of $\mathcal{N}_T$ fall into two sets: those for which the sublattice order $\mathfrak{a} \rightarrow \mathfrak{b} \rightarrow \mathfrak{c} \rightarrow \mathfrak{a} \rightarrow \dots$ is clockwise around the triangle and those for which the ordering is anticlockwise; we refer to these two types as *right-handed* and *left-handed* respectively. The symmetry group of the lattice is C_{3v} , where the axis of rotation is orthogonal to the lattice plane and through the centre of an arbitrarily chosen triangle and the three planes of reflection $\Pi_{\mathfrak{x}}$ contain the axis and the vertex of the triangle $\mathfrak{x}$, for $\mathfrak{x} = \mathfrak{a}, \mathfrak{b}, \mathfrak{c}$. It will be seen that the handedness of the triangles are invariant under C_3 but are interchanged by the three reflections.

The full Bell–Lavis water model, which can be represented as a spin-1 Ising model with additional terms, has attracted considerable attention.²⁴ However, the interest

²³ For a review of this work together with related developments see Au-Yang and Perk (1995).

²⁴ Its water-like properties have been investigated by Lavis (1973, 1975) and Lavis and Christou (1977) using cluster-variation methods and Lavis (1976) using a transfer matrix method. More recently exact calculations on the Husimi cactus have been used by Barbosa and Henriques (2008)

here is the special case which arises when, as in Sect. 3.6.4.2, the spin-1 formulation reduces to a 3-state Potts model. This takes the form

$$\begin{aligned} \widehat{H}^{(\text{EP})}(K_1, K_2, K_3; \sigma) &:= -K_1 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \delta^{K_1}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}')) \\ &- \sum_{\{\mathbf{r}, \mathbf{r}', \mathbf{r}''\}}^{\Delta} \{K_2 [\mathfrak{E}_c(\sigma(\mathbf{r}), \sigma(\mathbf{r}'), \sigma(\mathbf{r}''))]^2 + K_3 \mathfrak{E}_c(\sigma(\mathbf{r}), \sigma(\mathbf{r}'), \sigma(\mathbf{r}''))\}, \end{aligned} \quad (3.7.5)$$

where the ordering of sites $\mathbf{r} \rightarrow \mathbf{r}' \rightarrow \mathbf{r}''$ in the second summation is clockwise on right-handed triangles and anticlockwise on left-handed triangles, with

$$\mathfrak{E}_c(\tilde{\sigma}, \tilde{\sigma}', \tilde{\sigma}'') := \frac{1}{2}[\tilde{\sigma} - \tilde{\sigma}'][\tilde{\sigma}' - \tilde{\sigma}''][\tilde{\sigma}'' - \tilde{\sigma}]. \quad (3.7.6)$$

The presence of this term means that this is a face model in the sense defined in Sect. 2.4.1. It is also chiral in that $\mathfrak{E}_c(\tilde{\sigma}, \tilde{\sigma}', \tilde{\sigma}'')$ is invariant under the rotational elements of the lattice symmetry group but reversed in sign by the reflections. This model has been investigated using a finite-lattice RSRG method by Young and Lavis (1979) and using Monte-Carlo simulations by Wang et al. (2003). The K_2 term in the Hamiltonian, which can be conveniently represented as the square of the chiral term is simply a pure Potts antiferromagnetic three-site interaction. Its inclusion is necessary for the renormalization calculations and means that the results obtained by Young and Lavis (1979) became a generalization of those of Schick and Griffiths (1977).

3.8 Vertex Models

A vertex model is one where the microstate of the system is represented by placing an arrow on each edge of the lattice $\mathcal{N}$. If no restrictions are applied, the number of arrow arrangements around a lattice site is 2^z , where z is the coordination number. Vertex models on three-dimensional lattices have been investigated (Thibaudier and Villain 1972; Hu and Hou 1996) and also on the honeycomb (Wu 1974) and triangular (Sacco and Wu 1975) lattices. However, most interest has been directed to the square lattice, with the extension, by Baxter (1978), to an arbitrary plane lattice consisting of straight lines intersecting at points with coordination number four. Here the most general case, the *sixteen-vertex model*, which can be shown to be equivalent to an Ising model with two-, three- and four-site interactions and with an external field (Suzuki and Fisher 1971), is unsolved.²⁵

(Footnote 24 continued)

to compare the properties of the Bell–Lavis model to those of the antiferromagnetic Blume–Emery–Griffiths model and a Monte-Carlo study has been carried out by Fiore et al. (2009).

²⁵ As may be expected, since the Ising model with just a two-site interaction and an external field is unsolved.

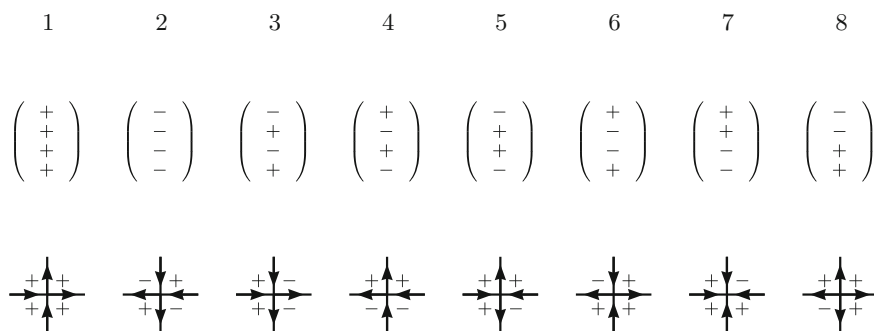


Fig. 3.2 The eight-vertex configurations showing, in row two, the edge variables clockwise in a column starting at the *top* with the southern vertical edge and in row three the *arrow* directions assigned according to the direction rule of Baxter (1972b)

The first square-lattice vertex model to be solved was the *six-vertex model* (Lieb 1967a, b, c, d). Here the vertices are subject to the *ice-rule*²⁶ which restricts the arrow arrangements around a vertex to those (labelled 1–6 in Fig. 3.2) where there are two inward-pointing and two outward-pointing arrows. This model is explored in more detail in Sect. 3.8.2. If the ice rule is relaxed to one in which there is an even number of arrows pointing in and out, two additional vertex-types are added to give the *eight-vertex model*.

3.8.1 The Eight-Vortex Model

The square-lattice zero-field partition function for this model was derived by Baxter (1972b). In this approach the eight allowed configurations of arrows at a vertex were labelled using a prescribed orientation of the lattice with lattice lines horizontal and vertical (Baxter 1982b, Chap. 10) and toroidal boundary conditions.²⁷ The other four-coordinated regular lattice is the kagomé lattice and the square-lattice results can be

²⁶ A water molecule can be regarded as having two negative and two positive ‘arms’ (Eisenberg and Kauzmann 1969, Chap. 1) and two molecules are hydrogen-bonded by the junction of a positive arm from one and a negative arm from the other. In a fully hydrogen-bonded structure like Ice I (ibid, Chap. 3), the positive arms of each molecule, that is those corresponding to the positions of the two protons, can be labelled with a single arrow on the segment between the centres of the bonded pair. A completely hydrogen-bonded network is thus represented by an oriented graph (Sect. 16.1.4) with the centres of the molecules at the vertices with inward arrows on two of the four edges meeting at each vertex and outward arrows on the other two. As well as water, certain other substances, such as KH_2PO_4 (potassium dihydrogen phosphate), crystallize as four-coordinated hydrogen-bonded networks. KH_2PO_4 is a *ferroelectric* with a spontaneous electrical polarization in zero field below a critical temperature.

²⁷ Toroidal boundary conditions strain the notion of horizontal/vertical, top/bottom, right/left a little but the problem can be resolved by choosing directions for one site and applying them in a consistent way to the other sites.

extended to this case (Baxter 1982b, Chap. 11), but now the situation is complicated by the fact that a vertex can have three different orientations (Fig. 16.6).

Let $\mathcal{N}^{\otimes}$ consist of straight lines intersecting at vertices with coordination number four.²⁸ A base line is used to give a direction to all the lines of $\mathcal{N}^{\otimes}$. For the square lattice the base line is chosen so that the directions of the two sets of lines are vertically upwards and to the right, that is in the direction of the unit vectors of (16.2.2). For the kagomé lattice a suitable choice of base line gives the directions of the three sets of lattice lines in the directions of the unit vectors of (16.2.4).²⁹ For the square lattice all vertices are equivalent, and for the kagomé lattice they can be designated as lying on sublattice $\mathcal{N}_{13}^{\otimes}$, $\mathcal{N}_{32}^{\otimes}$ or $\mathcal{N}_{13}^{\otimes}$ according to the directions of the unit vectors of the lines for which they are the intersection.

A property of $\mathcal{N}^{\otimes}$ (Sect. 16.3) is that some of its faces can be shaded so the every edge is a boundary between a shaded face and an unshaded face. Suppose that a vertex is oriented so that the directed lines passing through it are pointing north and east as shown in Fig. 16.16 and label the edges clockwise starting with the south-pointing edge. Then the sites are divided into two sublattices: $\mathcal{N}_E^{\otimes}$ consisting of sites for which the first (south-pointing) edge has shading in the clockwise direction and $\mathcal{N}_O^{\otimes}$ consisting of sites for which that edge has shading in the anticlockwise direction.³⁰ For the square lattice shading is a convenient mechanism for the division into two sublattices and the possible introduction of different Boltzmann weights. For the kagomé lattice one choice is to shade the triangular faces giving $\mathcal{N}_E^{\otimes} = \mathcal{N}_{13}^{\otimes} + \mathcal{N}_{32}^{\otimes}$, $\mathcal{N}_O^{\otimes} = \mathcal{N}_{12}^{\otimes}$, with the alternative, shading of the honeycomb faces, interchanging $\mathcal{N}_E^{\otimes}$ and $\mathcal{N}_O^{\otimes}$. Although, as in the square lattice, shading could be used here to introduce different Boltzmann weights, it relies on a rather arbitrary combination of two of the three sublattice. A proper treatment of this problem Baxter (1982b, Chap. 11) involves different weights on all three sublattices.

Arrow arrangements giving the eight vertex types are shown in the third line of Fig. 3.2 and, differentiating between sublattices, in the third and fourth lines of Fig. 3.3. Edge variables ς_i , $i = 1, 2, 3, 4$ are assigned around a vertex.

Definition 3.8.1 The *direction rule*, used by Baxter for the square lattice (see, Baxter 1982b, Chap. 10) and the kagomé lattice (see, Baxter 1982b, Chap. 11), is equivalent to setting $\varsigma_i = +1$ or -1 according as the arrow points in the direction the same as, or opposite to, the line direction. These edge-variable assignments are shown as a vertical column (with index increasing downwards) in the second line of Fig. 3.2.

Definition 3.8.2 The *shading rule*, used by Baxter (1978), sets $\varsigma_i = +1$ or -1 according as the arrow is pointing in the direction with shading on the left or right respectively. With arrows reversed between $\mathcal{N}_E^{\otimes}$ and $\mathcal{N}_O^{\otimes}$, the vertical column of edge-variable assignments shown in the second line of Fig. 3.3 apply to both sublattices.

²⁸ The use of this symbol for the ‘arrow-bearing’ lattice is for conformity with Sect. 16.3, since it is a *rapidity lattice*.

²⁹ That is ‘generally to the right’ as needed in Baxter (1982b, Chap. 11).

³⁰ This notation, in which $_E$ and $_O$ stands respectively for ‘even’ and ‘odd’, is in keeping with the labelling of sublattices in Chap. 12 (Fig. 12.3).

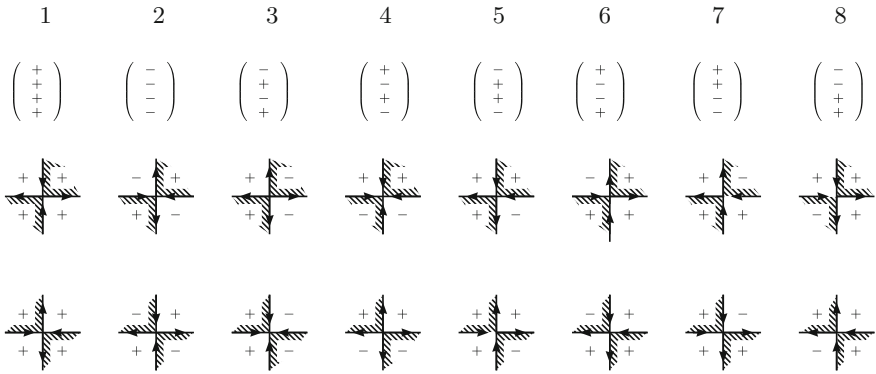


Fig. 3.3 The eight-vertex configurations with the edge variables in row two as in Fig. 3.2. The arrow directions which are different for sublattices $\mathcal{N}_E^{\oplus}$ and $\mathcal{N}_O^{\oplus}$ (and different from those given by the direction rule) are assigned using the shading rule of Baxter (1978) and shown in lines three and four, respectively, for the two sublattices

In terms of edge variables, assigned by either rule, the eight-vertex constraint condition is

$$\zeta_1 \zeta_2 \zeta_3 \zeta_4 = 1. \quad (3.8.1)$$

Sites can be placed on the shaded faces of $\mathcal{N}^{\oplus}$ to form a lattice $\mathcal{N}$ and on the unshaded faces to form its dual $\mathcal{N}^*$. Spins are now assigned to the sites of $\mathcal{N} + \mathcal{N}^*$ with the spins around a vertex labelled $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ in the clockwise direction starting with the spin on the face immediately following the first-labelled edge. It is a consequence of the eight-vertex rule (3.8.1) that spin-variable values can be related to edge-variable values, assigned by either rule, to satisfy

$$\begin{aligned} \zeta_i = \sigma_{i-1} \sigma_i, \implies \zeta_i \zeta_{i+1} = \sigma_{i-1} \sigma_{i+1} \quad \text{and} \quad \zeta_i \zeta_{i+2} = \sigma_1 \sigma_2 \sigma_3 \sigma_4, \\ i = 1, \dots, 4 \pmod{4}. \end{aligned} \quad (3.8.2)$$

To see that this is the case, suppose that, for any arrow configuration, the configuration graph $\mathfrak{g}$ is formed on $\mathcal{N}^{\oplus}$ by placing a graph edge on all lattice edges with $\zeta_i = -1$. From (3.8.1) every vertex of $\mathfrak{g}$ is even and every edge is a boundary between regions of $\mathcal{N} + \mathcal{N}^*$ with different homogeneous spin assignments. It is clear that the spin assignment for a particular assignment of edge-variable values is unique to within an overall reversal of spins on the lattice. The spin assignments shown in line three of Fig. 3.2 and lines three and four of Fig. 3.3 are one possibility.

Both the direction rule and shading rule for vertex assignments will be useful for our discussion of the eight-vertex model and the problem is to retain and compare them with a minimum of duplication. There are two possibilities:

- (i) To give the same ordering of configurations with respect to arrow directions. This has the advantage of retaining the labelling of ferroelectric/antiferroelectric ground states, but loses the correspondence for spins between ferromagnetic

Table 3.2 Notation comparison between that adopted here and in Baxter (1978)

	Direction Rule	Shading Rule	Baxter (1978)
(1,2)	a	a	c
(3,4)	b	b	d
(5,6)	c	c	a
(7,8)	d	d	b
	Δ	Δ	$-\Delta$
	Γ	Γ	$-\Gamma$
I	FEN&F/FMU&S	AFEL/FMU&S	FEN&F/AFML
II	FEV&H/FMV&H	AFER/FMV&H	FEV&H/AFMR
III	AFER/AFMR	FEN&F/AFMR	AFER/FMU&S
IV	AFEL/AFML	FEV&H/AFML	AFEL/FMV&H

/antiferromagnetic ground states. It also means that the Boltzmann weights for the two rules have different forms in terms of spins and edge variables.

- (ii) To give the same ordering of configurations in terms of edge variables and spins. The advantages and disadvantages in terms of electric and magnetic ground states are reversed, but the eight Boltzmann weights in terms of spins or edge variables now have the same form for each rule.

Baxter (1978)³¹ uses (i), but we use (ii). Table 3.2 provides a comparison of notation.

Let the Boltzmann weight of the vertex $\mathbf{r} \in \mathcal{N}^{\otimes 4}$ be $\mathfrak{X}(\mathbf{r}; \varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4)$, in terms of edge variables, and $\mathfrak{Y}(\mathbf{r}; \sigma_1, \sigma_2, \sigma_3, \sigma_4)$, in terms of spins. Then for the eight vertex configurations

$$\begin{aligned}
 \mathfrak{X}_1(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; +1, +1, +1, +1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \pm 1, \pm 1, \pm 1), \\
 \mathfrak{X}_2(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; -1, -1, -1, -1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \mp 1, \pm 1, \mp 1), \\
 \mathfrak{X}_3(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; +1, -1, +1, -1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \pm 1, \mp 1, \mp 1), \\
 \mathfrak{X}_4(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; -1, +1, -1, +1) = \mathfrak{Y}(\mathbf{r}; \mp 1, \pm 1, \pm 1, \mp 1), \\
 \mathfrak{X}_5(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; -1, +1, +1, -1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \pm 1, \pm 1, \mp 1), \\
 \mathfrak{X}_6(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; +1, -1, -1, +1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \mp 1, \pm 1, \pm 1), \\
 \mathfrak{X}_7(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; -1, -1, +1, +1) = \mathfrak{Y}(\mathbf{r}; \pm 1, \pm 1, \mp 1, \pm 1), \\
 \mathfrak{X}_8(\mathbf{r}) &:= \mathfrak{X}(\mathbf{r}; +1, +1, -1, -1) = \mathfrak{Y}(\mathbf{r}; \mp 1, \pm 1, \pm 1, \pm 1).
 \end{aligned} \tag{3.8.3}$$

³¹ Except for an unimportant interchange of (3,4) and (5,6).

3.8.1.1 A Staggered Eight-Vertex Model

Let us step back for a moment from the details so far presented for the analysis of this model and ask more directly the question: “Given that there are eight arrangements of arrows at a vertex, how do we distinguish between them?” In the absence of an orientation to the lattice, directions to the lattice lines or shading, and using the listing of vertex types in Fig. 3.2, the groups $\mathbb{T}_1 := \{1, 2, 3, 4\}$, $\mathbb{T}_2 := \{5, 6\}$ and $\mathbb{T}_3 := \{7, 8\}$ are of indistinguishable vertex types. The initial approach, which was used by Baxter, and which distinguished all eight types, was to give a physical orientation to the lattice (square or kagomé). Our proposed direction rule, giving directions to the lattice lines, is effectively equivalent to that and if *both* directions to the lattice lines *and* shading are included the effect is to distinguish sixteen different vertex types as shown in lines three and four of Fig. 3.3. Here we address the question posed by the abandonment of physical orientation (and line direction) and the reliance simply on shading. All vertex types equivalent under rotation must be counted together and the only distinguishing feature is the orientation of arrows with respect to the two shaded and unshaded faces. This will group together the arrow configurations p_E and p_O , $p = 1, 2, \dots, 8$, of lines three and four, respectively, of Fig. 3.3 as follows:

- $\mathbb{S}_1 := \{5_E, 6_E, 7_O, 8_O\}$: arrows on the edges of the unshaded regions follow each other round one clockwise and the other anticlockwise.
- $\mathbb{S}_2 := \{5_O, 6_O, 7_E, 8_E\}$: arrows on the edges of the shaded regions follow each other round one clockwise and the other anticlockwise.
- $\mathbb{S}_3 := \{2_E, 2_O\}$: arrows on the edges of both shaded and unshaded regions follow each other round clockwise on the shaded regions and anticlockwise on the unshaded regions.
- $\mathbb{S}_4 := \{1_E, 1_O\}$: arrows on the edges of both shaded and unshaded regions follow each other round clockwise on the unshaded regions and anticlockwise on the shaded regions.
- $\mathbb{S}_5 := \{3_O, 4_E\}$: arrows on the edges of both shaded and unshaded regions point in.
- $\mathbb{S}_6 := \{3_E, 4_O\}$: arrows on the edges of both shaded and unshaded regions point out.

This is a *staggered eight-vertex model* with the Boltzmann weights on the two sublattices interchanged between 3 and 4 and pairs (5,6) and (7,8). The effect of shading is to divide $\mathbb{T}_1$ into $\mathbb{S}_1$ and $\mathbb{S}_2$, $\mathbb{T}_2$ in $\mathbb{S}_3$ and $\mathbb{S}_4$, and $\mathbb{T}_3$ in $\mathbb{S}_5$ and $\mathbb{S}_6$.

3.8.1.2 The Regular Square-Lattice Model

The Boltzmann weights are now site and sublattice independent and denoted by $\mathfrak{X}_p$. The assumption of sublattice independence is, as can be seen by comparing lines 3

and 4 of Fig. 3.3, the assumption that the vertex weights are unchanged under arrow reversal.³² From this it then follows that

$$\begin{aligned} \mathfrak{x}_1 = \mathfrak{x}_2 &=: a, & \mathfrak{x}_3 = \mathfrak{x}_4 &=: b, \\ \mathfrak{x}_5 = \mathfrak{x}_6 &=: c, & \mathfrak{x}_7 = \mathfrak{x}_8 &=: d. \end{aligned} \tag{3.8.4}$$

The partition function for the model is

$$Z^{(8\text{vm})}(\mathcal{N}^{\otimes}, a, b, c, d) := \sum^{[8]} a^{n_{12}} b^{n_{34}} c^{n_{56}} d^{n_{78}}, \tag{3.8.5}$$

where n_{pq} is the numbers of p and q vertices and the sum is over all allowed arrow configurations.

The number of arrow configurations is the *entropy of the eight-vertex ice*. This is obtained by setting all the vertex energies in (3.8.5) to zero. However, this is the same as counting the number of spin configurations. Taking into account that one arrow configuration corresponds to two spin configurations, the eight-vertex ice entropy is

$$Z^{(8\text{vm})}(\mathcal{N}^{\otimes}, 1, 1, 1, 1) = 2^{N^{\otimes}-1}. \tag{3.8.6}$$

For large $N^{\otimes}$, in the absence of interactions, the reduced entropy per site s , defined by Eqs. (1.1.2) and (1.2.1) (with $Q := N^{\otimes}$), is thus given by

$$s = \frac{1}{N^{\otimes}} \ln Z^{(8\text{vm})}(N^{\otimes}, 1, 1, 1, 1) = \ln(2) = \frac{1}{N^{\otimes}} \ln[\Omega(N^{\otimes})], \tag{3.8.7}$$

where

$$\Omega(N^{\otimes}) = 2^{N^{\otimes}} = \left(\frac{8}{4}\right)^{N^{\otimes}} \tag{3.8.8}$$

is the number of possible configurations with the eight-vertex types shown in Fig. 3.2. The numerator of this expression $8^{N^{\otimes}}$ would be the value of $\Omega(N^{\otimes})$ if the eight vertex types could be freely distributed over the $N^{\otimes}$ lattice sites. The restriction that any first-neighbour pair of vertices must consist of types with compatible arrows along the connecting edge therefore has the effect of reducing the number of vertex configurations by a factor $4^{N^{\otimes}}$.

³² Vertices 5–8 are polarized with directions of polarization given by the vector sum of the arrows, so an electric field could be introduced which would make different contributions to the energies of vertices. For these vertices the assumption is that the electric field is zero.

3.8.1.3 Ising Model Equivalences

Following Wu (1971) and Kadanoff and Wegner (1971), diagonal second-neighbour Ising interactions with energies $-J^{(\text{f})}(\mathbf{r})$ from bottom-left to top-right and $-J^{(\text{b})}(\mathbf{r})$ from bottom-right to top-left are introduced between these spins³³ with a four-spin interaction $-J^{(4)}(\mathbf{r})$ connecting all four spins around a site of $\mathcal{N}^{(*)}$. Denoting couplings in the usual way and including a constant coupling $K^{(0)}$, the Boltzmann weights in terms of spins are

$$\mathfrak{Y}(\mathbf{r}; \sigma_1, \sigma_2, \sigma_3, \sigma_4) = \exp[K^{(0)} + K^{(\text{f})}(\mathbf{r})\sigma_1\sigma_3 + K^{(\text{b})}(\mathbf{r})\sigma_2\sigma_4 + K^{(4)}(\mathbf{r})\sigma_1\sigma_2\sigma_3\sigma_4] \quad (3.8.9)$$

and, in terms of the edge variables, are

$$\mathfrak{X}(\mathbf{r}; \varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) = \exp[K^{(0)} + K^{(\text{f})}(\mathbf{r})\varsigma_2\varsigma_3 + K^{(\text{b})}(\mathbf{r})\varsigma_3\varsigma_4 + K^{(4)}(\mathbf{r})\varsigma_1\varsigma_3]. \quad (3.8.10)$$

This yields a site-dependent and lattice-independent version of (3.8.4) with

$$\begin{aligned} \mathbf{a}(\mathbf{r}) &= \mathfrak{X}(\mathbf{r}; \pm 1, \pm 1, \pm 1, \pm 1) = \exp[K^{(0)} + K^{(\text{f})}(\mathbf{r}) + K^{(\text{b})}(\mathbf{r}) + K^{(4)}(\mathbf{r})], \\ \mathbf{b}(\mathbf{r}) &= \mathfrak{X}(\mathbf{r}; \pm 1, \mp 1, \pm 1, \mp 1) = \exp[K^{(0)} - K^{(\text{f})}(\mathbf{r}) - K^{(\text{b})}(\mathbf{r}) + K^{(4)}(\mathbf{r})], \\ \mathbf{c}(\mathbf{r}) &= \mathfrak{X}(\mathbf{r}; \pm 1, \mp 1, \mp 1, \pm 1) = \exp[K^{(0)} + K^{(\text{f})}(\mathbf{r}) - K^{(\text{b})}(\mathbf{r}) - K^{(4)}(\mathbf{r})], \\ \mathbf{d}(\mathbf{r}) &= \mathfrak{X}(\mathbf{r}; \pm 1, \pm 1, \mp 1, \mp 1) = \exp[K^{(0)} - K^{(\text{f})}(\mathbf{r}) + K^{(\text{b})}(\mathbf{r}) - K^{(4)}(\mathbf{r})]. \end{aligned} \quad (3.8.11)$$

For later reference we define

$$\begin{aligned} \Delta(\mathbf{r}) &:= \sinh[2K^{(\text{f})}(\mathbf{r})] \sinh[2K^{(\text{b})}(\mathbf{r})] \\ &\quad + \tanh[2K^{(4)}(\mathbf{r})] \cosh[2K^{(\text{f})}(\mathbf{r})] \cosh[2K^{(\text{b})}(\mathbf{r})] \\ &= \frac{\mathbf{a}^2(\mathbf{r}) + \mathbf{b}^2(\mathbf{r}) - \mathbf{c}^2(\mathbf{r}) - \mathbf{d}^2(\mathbf{r})}{2[\mathbf{a}(\mathbf{r})\mathbf{b}(\mathbf{r}) + \mathbf{c}(\mathbf{r})\mathbf{d}(\mathbf{r})]}, \end{aligned} \quad (3.8.12)$$

and

$$\Gamma(\mathbf{r}) := \tanh[2K^{(4)}(\mathbf{r})] = \frac{\mathbf{a}(\mathbf{r})\mathbf{b}(\mathbf{r}) - \mathbf{c}(\mathbf{r})\mathbf{d}(\mathbf{r})}{\mathbf{a}(\mathbf{r})\mathbf{b}(\mathbf{r}) + \mathbf{c}(\mathbf{r})\mathbf{d}(\mathbf{r})}. \quad (3.8.13)$$

For the regular model the partition functions of the eight-vertex and Ising models satisfy

$$Z^{(8\text{vM})}(\mathcal{N}^{(*)}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = \frac{1}{2} \exp(N^{(*)} K^{(0)}) Z^{(\text{IM})}(\mathcal{N} + \mathcal{N}^*, K^{(\text{f})}, K^{(\text{b})}, K^{(4)}). \quad (3.8.14)$$

³³ The notation refers to the directions of the interactions being respectively those of the forward and backward slash punctuation marks.

If we impose the condition $K^{(4)} = 0$ then, from (3.8.11),

$$ab = cd \quad (3.8.15)$$

and the model decouples into first-neighbour Ising models on $\mathcal{N}$ and $\mathcal{N}^*$, each with unequal couplings $K^{(f)}$ and $K^{(b)}$ and (3.8.14) reduces to

$$\begin{aligned} Z^{(8vM)}(\mathcal{N}, a, b, c, d) &= \frac{1}{2} \exp(N^{\otimes} K^{(0)}) \\ &\times Z^{(IM)}(\mathcal{N}, K^{(f)}, K^{(b)}) Z^{(IM)}(\mathcal{N}^*, K^{(f)}, K^{(b)}). \end{aligned} \quad (3.8.16)$$

(3.8.14)–(3.8.16) readily generalize to the case of site-dependent couplings.

If the couplings are site independent except that $K^{(f)}(\mathbf{r}) = K^{(b)}(\mathbf{r}') = K$, $K^{(b)}(\mathbf{r}) = K^{(f)}(\mathbf{r}') = K^*$, for $\mathbf{r} \in \mathcal{N}_E^{\otimes}$ and $\mathbf{r}' \in \mathcal{N}_O^{\otimes}$, then $a(\mathbf{r}) = a(\mathbf{r}') =: a$, $b(\mathbf{r}) = b(\mathbf{r}') =: b$, $c(\mathbf{r}) = d(\mathbf{r}') =: c$, $d(\mathbf{r}) = c(\mathbf{r}') =: d$. The model is staggered and the partition functions of the eight-vertex and Ising models satisfy

$$Z^{(8vM)}(\mathcal{N}^{\otimes}, a, b, c, d) = \frac{1}{2} \exp(N^{\otimes} K^{(0)}) Z^{(IM)}(\mathcal{N} + \mathcal{N}^*, K, K^*, K^{(4)}). \quad (3.8.17)$$

If $K^{(4)} = 0$, then the model decomposes into isotropic Ising models on $\mathcal{N}$ and $\mathcal{N}^*$ with first-neighbour couplings K and K^* respectively.

$$Z^{(8vM)}(\mathcal{N}^{\otimes}, a, b, c, d) = \frac{1}{2} \exp(N^{\otimes} K^{(0)}) Z^{(IM)}(\mathcal{N}, K) Z^{(IM)}(\mathcal{N}^*, K^*). \quad (3.8.18)$$

3.8.1.4 Ground States of the Regular Square-Lattice Model

The essence of our approach is to take the edge variables, or equivalently the spin variables, as the definers of the vertex weights. The arrows are then simply representations, with the arrow configurations, for the direction and shading rules related by *a reversal of arrows on all edges where the line direction gives shading on the right*. Having said that, the ground states of the system can be described either in terms of a perfect ordering of arrow directions, which can be spoken of in ferroelectric language, or a perfect ordering of spins, which can be spoken of in ferromagnetic terms. The matching between these two will of course differ according to the rule used for arrow assignment.

For the discussion of patterns of arrow assignment it is convenient to translate to a configuration graph $\mathbf{g}$, where a graph edge is drawn for every edge carrying an arrow in the opposite direction to the line direction (that is downwards or to the left). Since this is the procedure used above for the graph constructed for the direction rule it is convenient to use, for reference purposes, the numbering of configurations in Fig. 3.2. The eight-vertex rule implies that the valency of every vertex of $\mathbf{g}$ is even and the possibilities are:

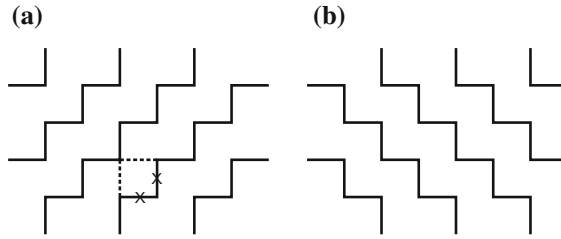


Fig. 3.4 For the square lattice the **a** BL-TR and **b** BR-TL staircase graphs. A perturbation of BL-TR is given by removing the *crossed lines* and replacing them with the *broken lines*

FEN: The null graph; all vertices are of type 1.

FEF: The full lattice graph; all vertices are of type 2.

FEV: The graph consisting of vertical lines; all vertices are of type 3.

FEH: The graph consisting of horizontal lines; all vertices are of type 4.

AFER: A bottom-left-top-right (BL-TR) graph of staircases; vertices 5 and 6 alternate on each staircase.

AFEL: A bottom-right-top-left (BR-TR) graph of staircases; vertices 7 and 8 alternate on each staircase.

FEN, FEF, FEV and FEH can be understood as ferroelectric states. Since vertex energies are unchanged under arrow reversal, they are energetically degenerate in pairs which we denote as FEN&F and FEV&H. AFER and AFEL, which are sublattice-ordered antiferroelectric states, are shown as graphs Fig. 3.4. The effect of arrow reversal is to translate the staircases by one edge.

For spin-ordering the vertex numbering is the same for both rules and the ordered states are:

FMU: Spins aligned uniformly; all vertices are of type 1.

FMS: Spins aligned in different directions on the sublattices $\mathcal{N}$ and $\mathcal{N}^*$; all vertices are of type 2.

FMV: Spins aligned in vertical rows with antiparallel alignment between neighbouring rows; all vertices are of type 3.

FMH: Spins aligned in horizontal rows with antiparallel alignment between neighbouring rows; all vertices are of type 4.

AFMR: Spins aligned in bottom-left-top-right double rows of spins with antiparallel alignment between double rows; vertices 5 and 6 alternate along BL-TR diagonal lines.

AFML: Spins aligned in bottom-right-top-left double rows of spins with antiparallel alignment between double rows; vertices 7 and 8 alternate along BR-TL diagonal lines.

FMU and FMS are degenerate, and together denoted as FMU&S, since, in the absence of first-neighbour couplings round a square (as in Sect. 3.8.1.5 and as shown in Fig. 12.5) the directions of ordered spin on $\mathcal{N}$ and $\mathcal{N}^*$ are independent. FMV and FMH are degenerate, and together denoted as FMV&H.

Here and in Sect. 9.3.2 it is convenient to expose the temperature-dependence of the Boltzmann weights by using vertex energies $\varepsilon_q > 0$, with

$$q =: \exp(-\varepsilon_q/T), \quad q = a, b, c, d. \quad (3.8.19)$$

The phase space of (a, b, c, d) is divided into four parts associated with ground states, which can be identified with the different spin-ordered states and, depending on the rule of assignment used, with arrow-ordered states. The regions are:

$$\text{I:} \quad \{\varepsilon_a < \varepsilon_b, \varepsilon_c, \varepsilon_d\} \Rightarrow \{a > b, c, d\},$$

$$\text{II:} \quad \{\varepsilon_b < \varepsilon_c, \varepsilon_d, \varepsilon_a\} \Rightarrow \{b > c, d, a\},$$

$$\text{III:} \quad \{\varepsilon_c < \varepsilon_d, \varepsilon_a, \varepsilon_b\} \Rightarrow \{c > d, a, b\},$$

$$\text{IV:} \quad \{\varepsilon_d < \varepsilon_a, \varepsilon_b, \varepsilon_c\} \Rightarrow \{d > a, b, c\},$$

and the associations with ordered states are given in Table 3.2.

3.8.1.5 The Free-Fermion Model

In this section we consider a regular model with the same vertex weights at every site, but do not assume that the lattice is necessarily square. An alternative Ising-spin representation to that given above, in which the four-spin coupling is replaced by interactions with a spin σ_0 on the vertex site itself was developed by Jüngling and Obermair (1974) and Jüngling (1975).

Let the forward couplings to σ_0 below and above the site be $K_B^{(f)}$ and $K_A^{(f)}$ with a similar notation for the backward couplings.³⁴ Also included are horizontal couplings $\frac{1}{2}K_B^{(h)}$ and $\frac{1}{2}K_A^{(h)}$ for the pairs of sites below and above σ_0 . The vertical couplings between the pairs to the left and right of σ_0 are denoted by $K_L^{(v)}$ and $K_R^{(v)}$. (This means that in the lattice as a whole the horizontal and vertical couplings between pairs of sites on $\mathcal{N} + \mathcal{N}^*$ are $\frac{1}{2}[K_B^{(h)} + K_A^{(h)}]$ and $\frac{1}{2}[K_L^{(v)} + K_R^{(v)}]$, respectively.) Summing over the states of σ_0 the Boltzmann weight is

³⁴ The use of the designations ‘below’, ‘above’, ‘horizontal’ and ‘vertical’ refer to the orientation in which edge number 1 (the south-pointing edge) is vertically downwards and, as for the square lattice the east-west line is horizontal. This is just topological convenience and this description of the free-fermion model is not restricted to that the square lattice. ‘B’ and ‘A’ (signifying *below* and *above* the site of $\mathcal{N}^{(*)}$) is a usage in conformity with the transfer matrix labelling in Sect. 12.2.

$$\begin{aligned}
\mathfrak{V}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = & 2 \exp\{K^{(0)} + K^{(f)}\sigma_1\sigma_3 + K^{(b)}\sigma_2\sigma_4 \\
& + \frac{1}{2}[K_B^{(h)}\sigma_1\sigma_4 + K_L^{(v)}\sigma_1\sigma_2 + K_A^{(h)}\sigma_2\sigma_3 + K_R^{(v)}\sigma_3\sigma_4]\} \\
& \times \cosh(K_B^{(f)}\sigma_1 + K_A^{(b)}\sigma_2 + K_A^{(f)}\sigma_3 + K_B^{(b)}\sigma_4). \quad (3.8.20)
\end{aligned}$$

With $\mathfrak{V}_p$, $p = 1, 2, \dots, 8$ given using (3.8.3),³⁵ the primary interest in this section is in models which satisfy

$$\mathfrak{V}_1\mathfrak{V}_2 + \mathfrak{V}_3\mathfrak{V}_4 = \mathfrak{V}_5\mathfrak{V}_6 + \mathfrak{V}_7\mathfrak{V}_8. \quad (3.8.21)$$

This is the *free-fermion condition* (Fan and Wu 1970; Jüngling 1975) The free-fermion model on a square lattice is solvable using Pfaffians (Fan and Wu, *ibid*) or directly by diagonalization of the transfer matrix (Felderhof 1973a), even in presence of a field (Felderhof 1973b).

In terms of the spin couplings it will be observed that the values of the horizontal and vertical coupling do not affect this, since they cancel in (3.8.21). However, this condition imposes constraints on the remaining couplings. Two cases of note are:

- If $K^{(f)} = K^{(b)} = 0$ the free-fermion condition (3.8.21) is satisfied by the vertex weights given by (3.8.3) and (3.8.20). When $\mathcal{N}^{(0)}$ is the square lattice, this is a *checkerboard Ising model* on $\mathcal{N} + \mathcal{N}^*$ (Baxter 1986, and Sect. 12.2). We have thus shown that the checkerboard Ising model is equivalent to the square-lattice free-fermion model. Allowing complex couplings, a special case of some interest (Eq. (12.2.60)) is when

$$K_A^{(f)} = -K_B^{(b)}, \quad K_A^{(b)} = K_B^{(f)} + \frac{1}{2}i\pi, \quad (3.8.22)$$

giving

$$\mathfrak{V}_5 = \mathfrak{V}_7 = 0, \quad \mathfrak{V}_1 = -\mathfrak{V}_3, \quad \mathfrak{V}_2 = \mathfrak{V}_4. \quad (3.8.23)$$

- When $K_A^{(f)} = K_B^{(b)} = K_B^{(f)} = K_A^{(b)} = 0$ the Boltzmann weights are of the form $\mathfrak{V}_p = \exp(-\varepsilon_p/T)$ in terms of vertex energies ε_p , $p = 1, 2, \dots, 8$ (not necessarily satisfying (3.8.4)) and the free-fermion condition can be satisfied at all temperatures only if

$$\begin{aligned}
\varepsilon_1 + \varepsilon_2 = \varepsilon_5 + \varepsilon_6 \quad & \text{and} \quad \varepsilon_3 + \varepsilon_4 = \varepsilon_7 + \varepsilon_8 \\
& \text{or} \\
\varepsilon_1 + \varepsilon_2 = \varepsilon_7 + \varepsilon_8 \quad & \text{and} \quad \varepsilon_3 + \varepsilon_4 = \varepsilon_5 + \varepsilon_6.
\end{aligned} \quad (3.8.24)$$

With $K_B^{(h)} = K_A^{(h)}$ and $K_L^{(v)} = K_R^{(v)}$ these conditions become respectively $K^{(b)} = 0$ and $K^{(f)} = 0$ which in each case, when $\mathcal{N}^{(0)}$ is a square lattice, is a *triangular Ising model* with different couplings in the three lattice directions.

³⁵ $\mathfrak{V}_p$ rather than $\mathfrak{X}_p$ is used here since the interest is in the spin formulation.

Although the validity of the free-fermion condition is not affected by the values of the horizontal and vertical couplings, a case of particular interest is when $K^{(h)} := K_B^{(h)} = -K_A^{(h)}$ and $K^{(v)} := K_L^{(v)} = -K_R^{(v)}$ (Sect. 12.2) since these terms cancel in the partition function of the whole lattice. Then it is not difficult to show that:

(i) if $K^{(h)}$ and $K^{(v)}$ are given by³⁶

$$\tanh(2K^{(h)}) = \frac{\sinh(2K_A^{(f)}) \sinh(2K_A^{(b)}) - \sinh(2K_B^{(f)}) \sinh(2K_B^{(b)})}{\cosh(2K_A^{(f)}) \cosh(2K_A^{(b)}) + \cosh(2K_B^{(f)}) \cosh(2K_B^{(b)})}, \quad (3.8.25)$$

$$\tanh(2K^{(v)}) = \frac{\sinh(2K_A^{(f)}) \sinh(2K_B^{(b)}) - \sinh(2K_B^{(f)}) \sinh(2K_A^{(b)})}{\cosh(2K_A^{(f)}) \cosh(2K_B^{(b)}) + \cosh(2K_B^{(f)}) \cosh(2K_A^{(b)})}, \quad (3.8.26)$$

$$\mathfrak{Y}_5 = \mathfrak{Y}_6, \quad \mathfrak{Y}_7 = \mathfrak{Y}_8. \quad (3.8.27)$$

Relationships (3.8.25) and (3.8.26) satisfy the transformations:

$$\begin{aligned} \text{(a)} \quad (K_B^{(f)}, K_B^{(b)}) &\longleftrightarrow (K_A^{(f)}, K_A^{(b)}) \\ &\implies (K^{(h)}, K^{(v)}) \rightarrow (-K^{(h)}, -K^{(v)}), \\ \text{(b)} \quad (K_B^{(f)}, K_A^{(f)}) &\longleftrightarrow (K_B^{(b)}, K_A^{(b)}) \\ &\implies (K^{(h)}, K^{(v)}) \rightarrow (K^{(h)}, -K^{(v)}). \end{aligned} \quad (3.8.28)$$

(ii) if $\{K_A^{(f)} = K_B^{(f)} \text{ and } K_A^{(b)} = -K_B^{(b)}\}$ or $\{K_A^{(f)} = -K_B^{(f)} \text{ and } K_A^{(b)} = K_B^{(b)}\}$

$$\mathfrak{Y}_1 = \mathfrak{Y}_2, \quad \mathfrak{Y}_3 = \mathfrak{Y}_4. \quad (3.8.29)$$

The conditions (3.8.27) and (3.8.29) are independent, but if they are both satisfied the model is equivalent to the regular eight-vertex model with vertex weights a, b, c and d and free-fermion condition

$$a^2 + b^2 = c^2 + d^2. \quad (3.8.30)$$

3.8.2 The Six-Vertex Model

Here the arrow assignments are made using the direction rule; that is as shown in Fig. 3.2. The regular six-vertex model on the square lattice corresponds to the

³⁶ To compare with the formulation of Baxter (1986) it should be noted that his M and P correspond respectively to $K^{(v)}$ and $-K^{(h)}$.

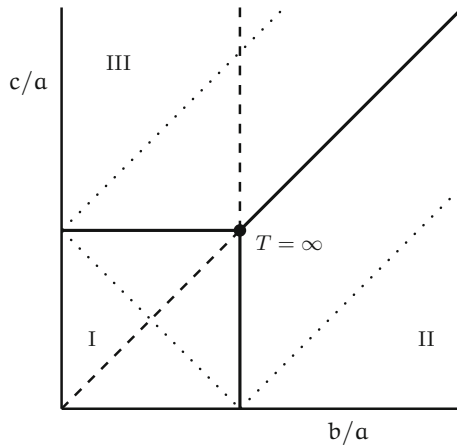


Fig. 3.5 Parameter space for the six-vertex model, showing the ferroelectric regions I and II and the antiferroelectric region III. The *broken lines* $a > b = c$ and $c > a = b$ correspond respectively to the KDP and F models. The *dotted lines* represent the transitions discussed in Sect. 9.3.2.2

limiting case $\varepsilon_d \rightarrow \infty$, ($d \rightarrow 0$) of the corresponding eight-vertex model. The ‘ice-rule’ applies and vertices 7 and 8 are eliminated. The partition function is

$$Z^{(6\text{vm})}(\mathcal{N}, a, b, c) := Z^{(8\text{vm})}(\mathcal{N}, a, b, c, 0) = \sum^{[6]} a^{n_{12}} b^{n_{34}} c^{n_{56}}, \quad (3.8.31)$$

where the sum is now over all arrow configurations. If these are pictured as configuration graphs $\mathfrak{g}$, drawn as described in the last section, top-right and bottom-left corners which make up the BR-TL staircase shown in Fig. 3.4b are excluded. As we saw above, the calculation for the entropy of ‘eight-vertex ice’ (which does not obey the ice-rule) was made quite simple by the mapping from arrows to spins. For the six-vertex model, where six-vertex (square) ice does obey the ice rule, the calculation is much more difficult (Lieb 1967a, b and Sect. 12.3.10.4).

The space of the variables (a, b, c) is divided into regions I, II and III of the set for the eight-vertex model (Sect. 3.8.1.4). These are shown in Fig. 3.5 in the plane of the variables b/a and c/a .

The special cases of the Slater KDP model of a ferroelectric,³⁷ for which $\varepsilon_a < \varepsilon_b = \varepsilon_c$ ($a > c = b$), and the Rys F model for an antiferroelectric,³⁸ for which $\varepsilon_a = \varepsilon_b > \varepsilon_c$ ($a = b < c$) were solved by Lieb (1967d) and Lieb (1967c), respectively. The continuation of the line $\varepsilon_b = \varepsilon_c$ with $\varepsilon_b < \varepsilon_a$ ($a < c$) along the boundary between regions II and III gives the inverse KDP (IKDP) model and

³⁷ Slater (1941) introduced a model for a hydrogen-bonded ferroelectric called the *KDP model* since it was intended to represent KH_2PO_4 and similar substances (Nagle 1966b).

³⁸ The *F model* was originally proposed by Rys (1963) as an interesting statistical mechanical model. However, it is now viewed (Nagle 1966b) as a reasonable model for a hydrogen-bonded antiferroelectric.

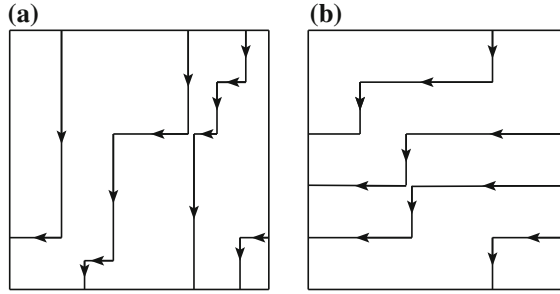


Fig. 3.6 Configuration graphs for the six-vertex model, with the lattice on a vertical cylinder

the continuation of the line $\varepsilon_a = \varepsilon_b$ with $\varepsilon_a < \varepsilon_b$ ($a < c$) along the boundary between regions I and II gives the inverse F (IF) model. The more general model with $\varepsilon_a = 2\varepsilon - \delta$, $\varepsilon_b = 2\varepsilon + \delta$, $\varepsilon_c = \varepsilon$ ($\varepsilon \geq 0$), which includes the ice model ($\varepsilon = \delta = 0$), the KDP model ($\varepsilon = \delta$, with the interchange $(1, 2) \leftrightarrow (3, 4)$ ³⁹) and the F model ($\delta = 0$) was solved by Sutherland (1967).

To consider configuration graphs g for the six-vertex model it is convenient first just to impose periodic boundary conditions in the horizontal direction so that the lattice forms a vertical cylinder. The only boundary sites are now those on the top and bottom rows. Also we adopt the convention that the ‘cross’ pattern of vertex 2 is to be regarded as a pair of ‘corners’ like those at vertices 5 and 6. A typical graph is shown on Fig. 3.6a. It consists of three paths which can be regarded as followed in the sense of the arrows, which is the same as that of the original arrows on the lines making up the path.⁴⁰ From the permitted vertex patterns, no sequence of lines can terminate except at a boundary site. Only steps downward or to the left are allowed so that a path leaving the top row of sites must finish at the bottom. So any path passes through a given row of vertical edges exactly once. This has an important consequence. If n_d denotes the number of downward arrows on a row of vertical edges in the original arrow configuration, then in the configuration graph the number of paths passing through this row is n_d , meaning that n_d is the same for all rows of vertical edges. In Fig. 3.6a, for instance, $n_d = 3$ and Fig. 3.6b shows a configuration graph consisting of a single path spiralling round the cylinder so that $n_d = 1$. This preservation property for n_d means that, with an electric field $\mathcal{E}$, an electrostatic field term $\mathcal{E}(N - 2n_d)$ can be included in the energy (Sutherland 1967; Lieb 1969). Any closed path must encircle the cylinder. Since no upward steps are allowed, no downward steps can occur either, and the path is confined to a single ring of horizontal edges, with $n_d = 0$ and all vertices of type 4. If one ring graph exists all the sites of the lattice must be parts of ring graphs or of type 1.

The ferroelectric ordered states, are the same as those for the eight-vertex model described in Sect. 3.8.1.4. However, there is an important difference between the two

³⁹ Allowed by the six-vertex version of the symmetry transformation $\mathcal{T}_0$ of (9.3.2).

⁴⁰ The apparently isolated piece of path in the bottom right-hand corner is a continuation on the cylinder of the left-hand path.

models. The smallest perturbation of any of the ordered ferroelectric states, FEN, FEF, FEV or FEH, takes the form of a square of sites occupied by vertices 5, 6, 7 and 8. Since vertices 7 and 8 are excluded from the six-vertex model this has the effect of ‘freezing in’ the ferroelectric order. This contrasts with the Ising ferromagnetic ordered state where the complete spin alignment can be perturbed by overturning one spin. The antiferroelectric ordered state has configuration graph shown in Fig. 3.4a where perturbations like that shown by the broken lines can still occur.

3.8.2.1 The Staggered Six-Vortex Model and the Potts Model

The equivalence, on the square lattice, between a staggered six-vertex model with different vertex energies on the two equivalent sublattices and the ν -state Potts model was first demonstrated by Temperley and Lieb (1971). This relationship was generalized to any (not necessarily regular) four-coordinated planar lattice by Baxter et al. (1976). We shall adopt the graphical approach of this latter work. However, for simplicity, we restrict attention to the cases where the Potts model resides on a square, triangular or honeycomb lattice $\mathcal{N}$,⁴¹ showing the equivalence to a six-vertex model on the medial (rapidity) lattice $\mathcal{N}^\circledast$, which (Sect. 16.2.2.2) is for the square lattice a square lattice and for triangular and honeycomb lattices a kagomé lattice.

The formula (3.6.52) can be generalized to include different couplings in different lattice direction. For the square lattice there are two such directions (horizontal and vertical) and for both the triangular and honeycomb lattices there are three. Thus

$$Z^{(\text{PM})}(\mathcal{N}, \{v^{(k)}\}) = \sum_{\{\mathbf{g}\}}^{\text{(AG)}} \nu^{c(\mathbf{g})} \prod_{k=1}^{\eta} (v^{(k)})^{e^{(k)}(\mathbf{g})} = W(\nu, \{v^{(k)}\}), \quad (3.8.32)$$

where $W(\nu, \{v^{(k)}\})$ is the obvious generalization of the Whitney polynomial (Sect. 16.4.1) and $\eta = 2$ for the square lattice and $\eta = 3$ for the triangular and honeycomb lattices.

Consider one of the graphs $\mathbf{g}$ on $\mathcal{N}$, augmented so all disconnected points of $\mathcal{N}$ are counted as single-site components. From (16.1.1), the cyclomatic number of $\mathbf{g}$ is

$$\mathfrak{C}(\mathbf{g}) = c(\mathbf{g}) + \sum_{k=1}^{\eta} e^{(k)}(\mathbf{g}) - N. \quad (3.8.33)$$

A site of the medial lattice $\mathcal{N}^\circledast$ lies either on an edge of $\mathcal{N}$ which is also an edge of $\mathbf{g}$ or on edge of $\mathcal{N}$ which is not an edge of $\mathbf{g}$. $\mathcal{N}^\circledast$ is transformed into a set of polygons by cutting the edges at the vertices and reconnecting them to form polygons, so that each connected component of $\mathbf{g}$ is bounded on the outside by a polygon and each component consisting of a cycle is also bounded on the inside by a polygon. Each

⁴¹ No periodic boundary conditions are assume.

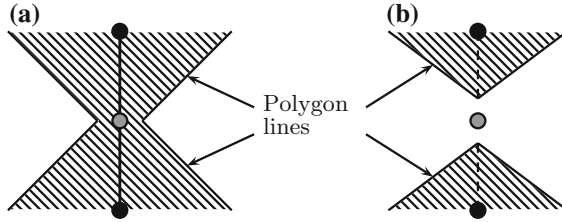


Fig. 3.7 Polygon construction on $\mathcal{N}$ (black sites) from the medial lattice $\mathcal{N}^{\otimes}$ (grey sites), **a** corresponds to the cases where a (heavy black) line of $\mathbf{g}$ passes through the site of $\mathcal{N}^{\otimes}$, making it an interior site and **b** where it does not, making the $\mathcal{N}^{\otimes}$ site an exterior site

region bounded by a polygon and containing one component of $\mathbf{g}$ is shaded, as is shown in Fig. 3.7 for the square lattice, and the number of polygons is

$$\mathfrak{N}(\mathbf{g}) = c(\mathbf{g}) + \mathfrak{C}(\mathbf{g}). \quad (3.8.34)$$

It is clear that any augmented graph $\mathbf{g} \subset \mathcal{N}$ is in one-to-one correspondence with a polygon decomposition ϖ of $\mathcal{N}^{\otimes}$, for which sites can be classified as *interior* if they lie on an edge of $\mathbf{g}$ (and hence in the shaded area of the plane) or *exterior* if they do not lie on an edge of $\mathbf{g}$. Both $\mathfrak{e}^{(k)}(\mathbf{g})$ and $\mathfrak{N}(\mathbf{g})$ can be unambiguously relabelled as functions of ϖ and the summation in (3.8.32) taken as that over a polygon decomposition ϖ . Then, from (3.8.33) and (3.8.34),

$$Z^{(\text{PM})}(\mathcal{N}, \{u^{(k)}\}) = v^{\frac{1}{2}N} \sum_{\{\varpi\}} v^{\frac{1}{2}\mathfrak{N}(\varpi)} \prod_{k=1}^{\eta} (u^{(k)})^{\mathfrak{e}^{(k)}(\varpi)}, \quad (3.8.35)$$

where

$$u^{(k)} := v^{(k)} / \sqrt{v} = [\exp(K^{(k)}) - 1] / \sqrt{v}. \quad (3.8.36)$$

To re-express this sum in the form of a staggered six-vertex model on $\mathcal{N}^{\otimes}$ with η different sets of vertex energies (depending on the orientations of the corresponding edges of $\mathcal{N}$), the summation must be over terms local to the sites of $\mathcal{N}^{\otimes}$. This is already true for the product term and the problem is to ‘localize’ $v^{\frac{1}{2}\mathfrak{N}(\varpi)}$.

We first choose a direction round each of the polygons. One possibility for the case of the square lattice is shown by the arrows in Fig. 3.8. At each vertex of $\mathcal{N}^{\otimes}$ the directions along the polygons gives a configuration of arrows and these, together with the shading, can be used (when the polygons are deformed back onto the edges) to identify the groups of vertex types $\mathbb{S}_1, \dots, \mathbb{S}_6$ listed in Sect. 3.8.1.1. However, there are two differences that must be noted. The first is that, because of the polygon construction, arrows at a vertex follow each other round the shaded or unshaded regions (or both). This means that the six-vertex ice-rule is satisfied and $\mathbb{T}_3 = \mathbb{S}_5 \cup \mathbb{S}_6$ is excluded. The second caveat is that it is necessary to distinguish between the cases

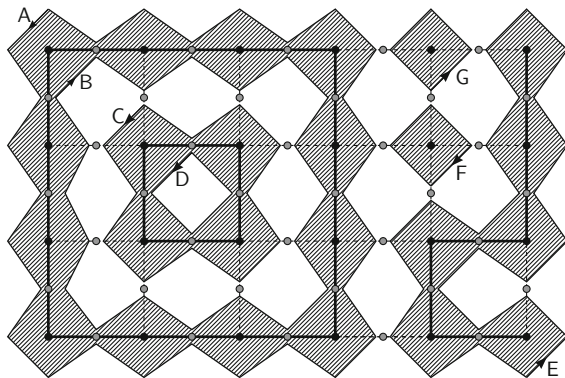


Fig. 3.8 Polygons formed on the square lattice

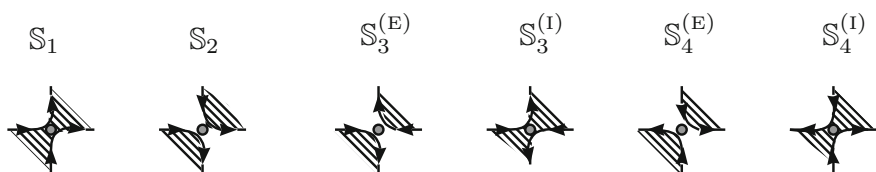


Fig. 3.9 The six vertex types for an interior and an exterior site of $\mathcal{N}^{\otimes}$

of shaded regions close to an interior site and to an exterior site. Since the arrows must follow each other around the shaded region, vertices of type S_1 can occur only at an interior site, and vertices of type S_2 can occur only at an exterior site; S_3 and S_4 type vertices can occur in both ways, giving the six possibilities shown in Fig. 3.9, where exterior and interior cases are distinguished by an appropriate superscript. As an example, the polygons in Fig. 3.9, labelled A–F, are assigned arrow directions. Clearly the interior sites within polygon E must be $S_3^{(i)}$ or $S_4^{(i)}$ irrespective of the arrow direction and in a similar way the exterior site between parts of E must be $S_3^{(e)}$ or $S_4^{(e)}$. All other sites have the possibility of being one of two types. With the arrow directions shown, the interior C/D sites are S_1 and the interior A/B sites are $S_3^{(i)}$ or $S_4^{(i)}$; the exterior B/C, A/E, A/G and E/G sites are $S_3^{(e)}$ or $S_4^{(e)}$, and the exterior A/F, E/F and G/F sites are S_2 .

Consider a particular single polygon with a particular direction of traverse (arrow direction). As an exterior site of $\mathcal{N}^{\otimes}$ is passed the polygon makes a clockwise or anticlockwise turn through an angle $\pi - \theta$ and as it passes an interior site it makes a clockwise or anticlockwise turn through θ , where θ is just the angle subtended at a site by the shade region; that is $\theta = \frac{1}{2}\pi$ when $\mathcal{N}$ is the square lattice, $\theta = \frac{1}{3}\pi$ when $\mathcal{N}$ is the honeycomb lattice and $\theta = \frac{2}{3}\pi$ when $\mathcal{N}$ is the triangular lattice. Given that, in the passage around the polygon, m sites are passed, let α_j , $j = 1, 2, \dots, m$

be the angles turned at the successive sites with plus and minus values attributed to clockwise and anticlockwise turns, respectively. Then replace the factor $\sqrt{\nu}$ in (3.8.35) by the product of factors q^{α_j} , for some as yet undefined q and all the sites visited by the passage around the polygon. Since $\alpha_1 + \alpha_2 + \dots + \alpha_m = \pm 2\pi$, with the plus and minus signs corresponding to clockwise and anticlockwise passages, respectively and since the summation must be over both directions of passage it follows that q is given by⁴²

$$q^{2\pi} + q^{-2\pi} = \sqrt{\nu}. \quad (3.8.37)$$

The index $k = 1, 2, \dots, \eta$ was attached to the edges of $\mathcal{N}$ to distinguish orientations and Potts interactions. It applied equally well to the sites of $\mathcal{N}^{\otimes}$ dividing them into sublattices $\mathcal{N}_k^{\otimes}$. The Boltzmann weights for the vertices of $\mathcal{N}_k^{\otimes}$ are denoted by $\mathfrak{z}_j^{(k)}$, $j = 1, 2, 3, 4, k = 1, 2, \dots, \eta$, for the four types of vertices shown in Fig. 3.9. These arise from contributions from the two polygons at the site, together with, the edge contribution $\{u^{(k)}\}$, for an interior site. Thus

$$\begin{aligned} \mathfrak{z}_1^{(k)} &= u^{(k)}, & \mathfrak{z}_2^{(k)} &= 1, \\ \mathfrak{z}_3^{(k)} &= q^{2\theta} u^{(k)} + q^{2\theta-2\pi}, & \mathfrak{z}_4^{(k)} &= q^{-2\theta} u^{(k)} + q^{2\pi-2\theta}, \end{aligned} \quad (3.8.38)$$

giving

$$Z^{(\text{PM})}(\mathcal{N}, \{u^{(k)}\}) = \nu^{\frac{1}{2}N} Z^{(6\text{VM})}(\mathcal{N}^{\otimes}, \{\mathfrak{z}_j^{(k)}\}), \quad (3.8.39)$$

where

$$Z^{(6\text{VM})}(\mathcal{N}^{\otimes}, \{\mathfrak{z}_j^{(k)}\}) := \sum_{j=1}^{[6]} \prod_{k=1}^4 \prod_{j=1}^{\eta} (\mathfrak{z}_j^{(k)})^{n_j^{(k)}}, \quad (3.8.40)$$

with the summation over all arrow configurations satisfy the six-vertex ice rule. Since $|\mathcal{N}^{\otimes}| = 2N$

$$\phi^{(\text{PM})}(\mathcal{N}, \{u^{(k)}\}) = -\frac{1}{2} \ln(\nu) + 2\phi^{(6\text{VM})}(\mathcal{N}^{\otimes}, \{\mathfrak{z}_j^{(k)}\}). \quad (3.8.41)$$

We pause to note that this result is, with reference to the first six arrow configurations of Fig. 3.2, for a six-vertex model with the Boltzmann weights of 1 and 2 (equivalent to $\mathbb{S}_1$), and 3 and 4 (equivalent to $\mathbb{S}_2$) set equal but not 5 and 6 (which here are equivalent to $\mathbb{S}_3$ and $\mathbb{S}_4$, respectively) and also with different weights on different sublattices.

⁴² As pointed out by Baxter et al. (1976) this formula can be made to apply to the case where the lattice lies on a vertical cylinder, by choosing a vertical line on the cylinder and adding/subtracting a factor of 2π to the sum $\alpha_1 + \alpha_2 + \dots$ whenever the polygon crosses the line from the left/right.

It is clear that the argument leading to (3.8.39) could equally well be carried out for the dual lattice $\mathcal{N}^*$. The angle θ is replaced by $\theta^* = \pi - \theta$ and the parameters $\{u^{(k)}\}$ by $\{u^{*(k)}\}$. This gives

$$Z^{(\text{PM})}(\mathcal{N}^*, \{u^{*(k)}\}) = \nu^{\frac{1}{2}N^*} Z^{(6\text{VM})}(\mathcal{N}^{\otimes}, \{\mathfrak{Z}_j^{*(k)}\}), \quad (3.8.42)$$

where

$$\begin{aligned} \mathfrak{Z}_1^{*(k)} &= u^{*(k)}, & \mathfrak{Z}_2^{*(k)} &= 1, \\ \mathfrak{Z}_3^{*(k)} &= q^{2\theta^*} u^{*(k)} + q^{2\theta^*-2\pi}, & \mathfrak{Z}_4^{*(k)} &= q^{-2\theta^*} u^{*(k)} + q^{2\pi-2\theta^*}. \end{aligned} \quad (3.8.43)$$

The aim, of course, is to have the same six-vertex partition function appearing in (3.8.39) and (3.8.42). This is achieved by taking $u^{*(k)} = 1/u^{(k)}$, or equivalently from (3.8.36),

$$[\exp(K^{(k)}) - 1][\exp(K^{*(k)}) - 1] = \nu \quad (3.8.44)$$

and interchanging $\mathbb{S}_1$ with $\mathbb{S}_2$. Extracting a factor of $u^{*(k)}$ from each vertex weight $\mathfrak{Z}_j^{(k)}$ in the six-vertex partition function in (3.8.42) and reverting to the use of couplings in the Potts model yields

$$Z^{(\text{PM})}(\mathcal{N}^*, \{K^{*(k)}\}) = \left\{ \prod_{k=1}^{\eta} \frac{[\exp(K^{*(k)}) - 1]^{N/\eta}}{[\exp(K^{(k)}) - 1]^{N^*/\eta}} \right\} Z^{(\text{PM})}(\mathcal{N}, \{K^{(k)}\}). \quad (3.8.45)$$

Except in two special cases the staggered square-lattice ($\eta = 2$, $\theta = \frac{1}{2}\pi$) six-vertex model with weights (3.8.38) is unsolved. These two cases correspond to

$$\xi^{(1)}\xi^{(2)} = \pm 1, \quad \text{where} \quad \xi^{(k)} := \frac{u^{(k)}q^{2\pi} + 1}{u^{(k)} + q^{2\pi}} \quad (3.8.46)$$

and we consider them separately.

- The case $\xi^{(1)}\xi^{(2)} = 1$ was solved by Baxter (1973d). It corresponds to $u^{(1)}u^{(2)} = 1$ which is equivalent to

$$[\exp(K^{(1)}) - 1][\exp(K^{(2)}) - 1] = \nu \quad (3.8.47)$$

and the vertex weights satisfy the conditions

$$\frac{\mathfrak{Z}_1^{(1)}}{\mathfrak{Z}_2^{(2)}} = \frac{\mathfrak{Z}_2^{(1)}}{\mathfrak{Z}_1^{(2)}} = \frac{\mathfrak{Z}_3^{(1)}}{\mathfrak{Z}_4^{(2)}} = \frac{\mathfrak{Z}_4^{(1)}}{\mathfrak{Z}_3^{(2)}} = u^{(1)}. \quad (3.8.48)$$

This represents a uniform proportionality between vertex weights which, when $u^{(1)}$ is positive, makes the model effectively equivalent to the regular (non-staggered) six-vertex model. Comparing (3.8.47) and (3.8.44) shows that the Potts model is self-dual. When $K^{(1)} = K^{(2)} = R/T > 0$, (3.8.47) gives the critical temperature

$$T_c = R / \ln(\sqrt{\nu} + 1), \quad (3.8.49)$$

of the ferromagnetic Potts model (Sect. 9.2.1) at which, for the corresponding six-vertex model, from (3.8.37) and (3.8.38),

$$a = b = 1, \quad c = q^\pi + q^{-\pi} = (2 + \sqrt{\nu})^{\frac{1}{2}}. \quad (3.8.50)$$

The staggered property is now absent and it is a case of the F model.

- The case $\xi^{(1)}\xi^{(2)} = -1$ was solved by Baxter (1982a). This constraint reduces to

$$(u^{(1)}\sqrt{\nu} + 2)(u^{(2)}\sqrt{\nu} + 2) = 4 - \nu, \quad (3.8.51)$$

which fixes the temperature of the anisotropic Potts model with the condition

$$4 - \nu = [\exp(K^{(1)}) + 1][\exp(K^{(2)}) + 1]. \quad (3.8.52)$$

Although it is of interest to consider non-physical situations (Baxter 1982a), (3.8.52) has a physical solution only when $\nu \leq 3$ and then only when at least one of $K^{(1)}$ and $K^{(2)}$ (that is of $R^{(1)}$ and $R^{(2)}$) is negative. When $R^{(1)} = R^{(2)} = R < 0$ (3.8.52) gives the critical temperature

$$T_c = R / \ln(\sqrt{4 - \nu} - 1), \quad R < 0 \quad (3.8.53)$$

of the antiferromagnetic square-lattice Potts model. For $\nu = 3$ the critical temperature of the antiferromagnetic 3-state Potts model is $T_c = 0$. Unlike the ferromagnetic case the staggered nature of the underlying six-vertex model does not disappear at the critical temperature and evaluation of the free energy in this case is rather more complicated (Baxter 1982a).

3.9 Dimer Models

Most of the work in this book is on models where each microsystem occupies one site and where two microsystems can occupy any pair of sites. In such systems there can be no provision for size and shape effects although, as in the water models of Bell and Lavis (1970) and Bell (1972) (Sect. 7.6) directional bonding can be included. In the lattice fluid picture (Sect. 3.5.1.2), one possible modification is to introduce an infinite repulsion between microsystems extending as far as a certain range of neighbours;

such models are called *exclusion models*. The hard-square and hard-hexagon models (Sect. 8.4.3) are of this type.

A mathematically similar approach is to suppose that the microsystem itself is a connected subgraph of the lattice $\mathcal{N}$ (Sect. 16.1.1). An *arrangement* of microsystems on the lattice then consists of a subgraph each of whose components is a microsystem. A *covering* is an arrangement which is a *spanning graph* (Sect. 16.4), that is one in which every lattice site is a vertex of a microsystem. In this context a microsystem with a single vertex and no lines is called a *monomer* and a microsystem consisting of a single line and two vertices of valency one is a *dimer*. In this section and in Chap. 13 we are concerned with *pure dimer assemblies*, for which each arrangement or *configuration* is a covering of the lattice consisting entirely of dimers. For a lattice $\mathcal{N}$, it is clear that a dimer covering is possible only when $N := |\mathcal{N}|$ is even, a situation henceforth assumed, and then the number of dimers in the covering is $\frac{1}{2}N$. Figure 3.10 shows a pure dimer configuration on the square lattice with configurations on the honeycomb lattice shown in Figs. 3.11 and 3.12a.

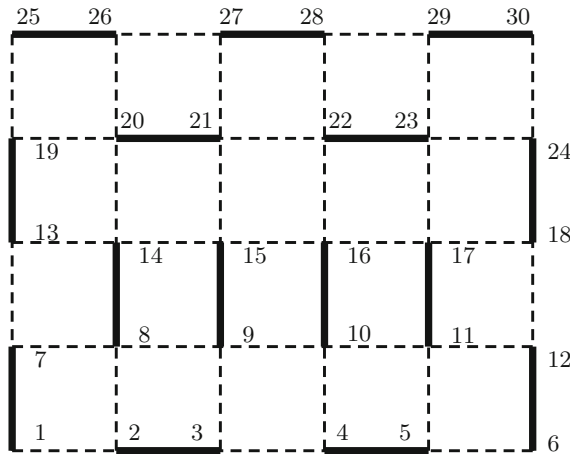


Fig. 3.10 A pure dimer configuration on the square lattice with a numbering of the lattice sites

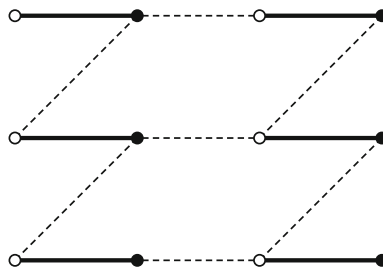


Fig. 3.11 A deformation of the honeycomb lattice with a pure dimer configuration

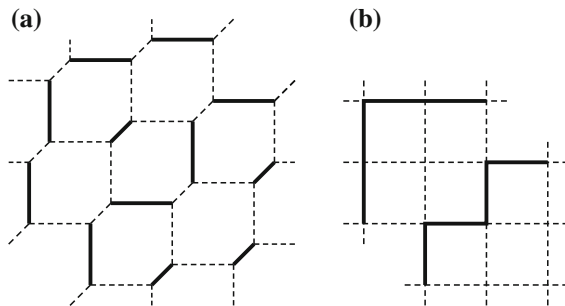


Fig. 3.12 **a** A pure dimer configuration on the honeycomb lattice. **b** The square lattice achieved by contracting the diagonal lines of (a) to zero

For the regular plane lattices the edges can be divided into η types according to spatial orientation. For the square lattice $\eta = 2$, with horizontal and vertical edge orientations, for the honeycomb, triangular and kagomé lattices $\eta = 3$. The Boltzmann weight for a dimer in orientation k is $\mathfrak{z}_k$ and we suppose that there are n_k dimers in this orientation. Then, if $\Omega(\mathcal{N}; \{n_k\})$ is the number of dimer configurations on the lattice for given values of $\{n_k\}$, the *dimer partition function* is

$$Z(\mathcal{N}, \{\mathfrak{z}_k\}) = \sum_{\{\sum_{k=1}^{\eta} n_k = \frac{1}{2}N\}} \Omega(\mathcal{N}; \{n_k\}) \prod_{k=1}^{\eta} \mathfrak{z}_k^{n_k}. \quad (3.9.1)$$

The Boltzmann weights for dimers, like those for lattice fluids are usually called *fugacities*; they can have various physical interpretations.

Let $\Pi_0 := (1, 2, \dots, N)$ be some labelling of the sites of $\mathcal{N}$ so that the partition into pairs $(1, 2), (3, 4), \dots, (N-1, N)$, with (i, j) being a first-neighbour pair occupied by a dimer, represents a possible dimer covering of $\mathcal{N}$. In the case of the square lattice with an even number of sites in each horizontal row, Π_0 could be a labelling from left to right along successive rows, as shown in Fig. 3.10, with the corresponding covering having all the dimers horizontal. A similar horizontal arrangement of dimers can be effected on the triangular lattice and on the honeycomb lattice if it is deformed in the manner shown in Fig. 3.11. Let $\Pi := (\pi_1, \pi_2, \dots, \pi_N)$ be some permutation of Π_0 with the properties

$$\begin{aligned} \pi_1 < \pi_2, \quad \pi_3 < \pi_4, \dots, \quad \pi_{N-1} < \pi_N, \\ \pi_1 < \pi_3 < \pi_5 < \dots < \pi_{N-1} \end{aligned} \quad (3.9.2)$$

and Λ be an $N \times N$ matrix with elements $\lambda(i, j)$ satisfying the conditions

$$\lambda(i, j) := \begin{cases} \lambda(j, i) > 0, & i \text{ and } j \text{ a first-neighbour pair,} \\ 0, & i \text{ and } j \text{ not a first-neighbour pair.} \end{cases} \quad (3.9.3)$$

Then $(\pi_1, \pi_2), (\pi_3, \pi_4), \dots, (\pi_{N-1}, \pi_N)$ denotes a dimer covering of $\mathcal{N}$ if and only if

$$\lambda(\pi_1, \pi_2)\lambda(\pi_3, \pi_4) \cdots \lambda(\pi_{N-1}, \pi_N) \neq 0. \quad (3.9.4)$$

As an example, in the case of the square lattice with the dimer configuration shown in Fig. 3.10, it can be seen that the corresponding partition

$$\Pi = \{1, 7, 2, 3, 4, 5, 6, 12, 8, 14, 9, 15, 10, 16, 11, 17, 13, 19, \\ 18, 24, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30\} \quad (3.9.5)$$

satisfies (3.9.2). For the matrix $\mathbf{A}$, let

$$Z(\mathcal{N}; \mathbf{A}) := \sum_{\{\Pi\}} \lambda(\pi_1, \pi_2)\lambda(\pi_3, \pi_4) \cdots \lambda(\pi_{N-1}, \pi_N). \quad (3.9.6)$$

In view of (3.9.4) a term in the summation in (3.9.6) not corresponding to a dimer covering is zero and the summation can therefore be restricted to dimer coverings. Then (3.9.6) is an expression for the dimer partition function (3.9.1) when $\lambda(i, j)$ is equal to the same fugacity value for all parallel dimers in any particular principal lattice direction.

3.9.1 The Modified KDP Model Equivalence

Figure 3.12a shows a pure dimer distribution on the honeycomb lattice. Suppose that the lattice is contracted by reducing the lengths of the diagonal edges to zero and letting the two vertices at the ends of each such edge coalesce into one. The resulting square lattice is shown in Fig. 3.12b and it can be seen that the possible vertex line patterns deriving from the horizontal and vertical dimers in Fig. 3.12a are those in Fig. 3.2, labelled 1, 3, 4, 5 and 6. The vertex line patterns, 2, 7 and 8 are absent because they would correspond to overlapping dimers on the original honeycomb lattice. With suitable relations between dimer fugacities and vertex energies the honeycomb lattice dimer model is thus equivalent to a square-lattice *five-vertex model*. This model can be regarded as a special case of the six-vertex model, with vertex 2 absent and different energies for vertices 3 and 4. Of course, this in turn is a special case of the eight-vertex model. So, with $\mathfrak{X}_p := \exp(-\varepsilon_p/T)$, $p = 1, 2, \dots, 8$,⁴³

⁴³ The properties of this model are identical to those of the model where ε_2 is finite and $\varepsilon_1 = \infty$. However, for that situation the derivation of the equivalence between the dimer model and the five-vertex model is more complicated and involves an intermediate lattice. This is a consequence of the convention for vertex line configurations used in vertex models.

$$\begin{aligned} \varepsilon_1 = \varepsilon_a, \quad \varepsilon_3 = \varepsilon_b, \quad \varepsilon_4 = \varepsilon_{b'}, \quad \varepsilon_5 = \varepsilon_6 = \varepsilon_c, \quad \varepsilon_2 = \varepsilon_7 = \varepsilon_8 = \infty, \\ \mathfrak{X}_1 = a, \quad \mathfrak{X}_3 = b, \quad \mathfrak{X}_4 = b', \quad \mathfrak{X}_5 = \mathfrak{X}_6 = c, \quad \mathfrak{X}_2 = \mathfrak{X}_7 = \mathfrak{X}_8 = 0. \end{aligned} \quad (3.9.7)$$

In the honeycomb dimer model shown in Fig. 3.12a, let the fugacities $\mathfrak{Z}_1$, $\mathfrak{Z}_2$ and $\mathfrak{Z}_3$ be associated respectively with diagonal, vertical and horizontal edges. Since the contraction of a dimer-occupied diagonal edge results in a type 1 vertex on the square lattice we must set

$$\mathfrak{X}_a = \mathfrak{Z}_1. \quad (3.9.8)$$

For each vertical dimer the fugacity $\mathfrak{Z}_2$ can be decomposed into two factors $\sqrt{\mathfrak{Z}_2}$ and one of these assigned to each of the two sites occupied by the dimer segments. Treating the fugacity $\mathfrak{Z}_3$ in a similar way and comparing Figs. 3.12a, b and 3.2, we have

$$\mathfrak{X}_b = \mathfrak{Z}_2, \quad \mathfrak{X}_{b'} = \mathfrak{Z}_3, \quad \mathfrak{X}_c = \sqrt{\mathfrak{Z}_2 \mathfrak{Z}_3}, \quad (3.9.9)$$

from which it follows that a condition for equivalence between the dimer and five-vertex models is that

$$\varepsilon_b + \varepsilon_{b'} = 2\varepsilon_c. \quad (3.9.10)$$

As was seen in Sect. 3.8.2, the ferroelectric six-vertex model is one where the ground-state vertices are of type 1, 2, 3 or 4 and the antiferroelectric model is one where the ground-state vertices are of type 5 or 6. For the latter to be the case it must be that $\varepsilon_c < \varepsilon_b$ and $\varepsilon_c < \varepsilon_{b'}$, a situation which is excluded by (3.9.10). When (3.9.10) holds the partition function of the five-vertex model on the square lattice $\mathcal{N}_s$ is related to that of a dimer model on the honeycomb lattice $\mathcal{N}_h$ by

$$Z^{(5\text{VM})}(\mathcal{N}_s, \mathfrak{X}_a, \mathfrak{X}_b, \mathfrak{X}_{b'}, \mathfrak{X}_c) = Z^{(\text{DM})}(\mathcal{N}_h, \mathfrak{Z}_1, \mathfrak{Z}_2, \mathfrak{Z}_3), \quad (3.9.11)$$

where $|\mathcal{N}_s| = \frac{1}{2}N$, $|\mathcal{N}_h| = N$.

The case of the six-vertex model for which the only non-zero vertex energies are those of type 1 and 2, which are equal and negative, is the *KDP model* (Sect. 3.8.2). The case of the present model, where

$$\varepsilon_a < 0, \quad \varepsilon_b = \varepsilon_{b'} = \varepsilon_c = 0, \quad (3.9.12)$$

differs from the KDP model in that the energy of a type 2 vertex is infinity rather than ε_a . It is called the *modified KDP model*.

3.9.2 The Ising Model Equivalence

Using the relationship

$$\exp[K\sigma(\mathbf{r})\sigma(\mathbf{r}')] = \cosh(K)(1 + v\sigma(\mathbf{r})\sigma(\mathbf{r}')), \quad (3.9.13)$$

where⁴⁴

$$v := \tanh(K), \quad (3.9.14)$$

the partition function (3.5.6) with zero field becomes

$$\begin{aligned} Z^{(\text{im})}(\mathcal{N}, v) &= [\cosh(K)]^{\frac{1}{2}zN} \sum_{\{\sigma\}} [1 + \sigma(\mathbf{r})\sigma(\mathbf{r}') v] \\ &= [\cosh(K)]^{\frac{1}{2}zN} \sum_{\{\sigma\}} \sum_{\{\mathbf{g}\}} v^{e(\mathbf{g})} \prod_{\mathbf{g}} \sigma(\mathbf{r})\sigma(\mathbf{r}'), \end{aligned} \quad (3.9.15)$$

where the inner sum is over all graphs $\mathbf{g}$ on $\mathcal{N}$ (Sect. 16.1) and the product is over all spin pairs on the edges of $\mathbf{g}$. It is clear that on summation over $\{\sigma(\mathbf{r})\}$ every term containing any spin variable to an odd power will cancel, giving

$$Z^{(\text{im})}(\mathcal{N}, v) = 2^N [\cosh(K)]^{\frac{1}{2}zN} \sum_{\{\mathbf{g}\}}^{(\text{ZF})} v^{e(\mathbf{g})}, \quad (3.9.16)$$

where the graphs are restricted to those which have only even vertices. For reasons explained in Sect. 14.2.2 these are called *zero-field graphs*. Given any such graph $\mathbf{g}$ placed in a particular way on the lattice, the set of $\frac{1}{2}zN - e(\mathbf{g})$ edges of the lattice which do not belong to $\mathbf{g}$ is called the *complement* of $\mathbf{g}$ and denoted by $\mathbf{g}'$. With a sum over these graphs (3.9.16) can be re-expressed in the form

$$Z^{(\text{im})}(\mathcal{N}, v) = 2^N [\sinh(K)]^{\frac{1}{2}zN} \sum_{\{\mathbf{g}'\}}^{(\text{CZF})} (v^{-1})^{e(\mathbf{g}')}. \quad (3.9.17)$$

This result can be generalized to an anisotropic Ising model. Suppose the lattice is divided into η classes of edges with s_k edges of class k , where $s_1 + s_2 + \cdots + s_\eta = \frac{1}{2}zN$. Let the first-neighbour coupling on the edges of class k be K_k with $v_k := \tanh(K_k)$. Then in place of (3.9.17) we have

⁴⁴ This variable features in the duality relationship (9.2.22) ($\mathfrak{X}^* = 1/v$, with $v = 2$) and as the appropriate variable for the high-temperature series expansion in Sect. 14.2.2 (again with $v = 2$).

$$Z^{(\text{IM})}(\mathcal{N}, \{v_k\}) = 2^N \left\{ \prod_{k=1}^{\eta} [\sinh(K_k)]^{s_k} \right\} \sum_{\{\mathbf{g}'\}}^{(\text{CZF})} \left[\prod_{k=1}^{\eta} (v_k^{-1})^{e_k(\mathbf{g}')} \right], \quad (3.9.18)$$

where $e_k(\mathbf{g}')$ is the number of edges of class k in the graph $\mathbf{g}'$. The equivalence between pure dimer and Ising models is established by finding a relationship between complementary graphs and dimer graphs, formed, as in Figs. 3.10 and 3.12a, by placing a line on every edge occupied by a dimer. However, for dimer graphs, exactly one line is incident at each vertex, while for zero-field graphs, or their complements, several lines can be incident. The way to deal with this is indicated by the procedure of Sect. 3.9.1, where the five-vertex graph of Fig. 3.12b is obtained from the dimer graph of Fig. 3.12a by deleting edges and coalescing sites. The reverse process would be the expansion of the square lattice of Fig. 3.12b by bifurcating each site into two sites connected by a new edge. Fisher (1966) gave a general prescription for splitting sites and inserting edges so as to yield a one-to-one relationship between complementary graphs on the original lattice and dimer graphs on the new lattice. The simplest application of Fisher's method is to the honeycomb lattice Ising model and an outline of this procedure will now be given.

Since all vertices of zero-field graphs have even valency, those on the honeycomb lattice must have either no lines or two lines incident at each vertex. A complementary graph must, therefore, have either one or three lines incident at each vertex. Suppose that each honeycomb site is split into three sites connected by a triangle of new edges, each $\bullet$ site in Fig. 3.11 being replaced by a downward pointing triangle and each $\circ$ site by an upward pointing triangle. This new lattice has two types of elementary polygons, the new triangles and dodecagons with six sides originating from the honeycomb hexagons and six sides from the new triangles. Hence the term *3-12 lattice* used by Nagle et al. (1989). The six sites A, B, C, D, E and F shown in Fig. 3.13 can be regarded as a *unit cell* in the new lattice, the sites A', B', E' and F' being first-neighbour sites in adjacent unit cells. Now consider a dimer graph on this lattice. If none of the new edges AB, BC and CA is occupied by a line of the dimer graph then the three edges E'A, F'B and CD originating from the honeycomb lattice will all be so occupied. Otherwise not more than one of the three edges AB, BC and CA can be occupied by a dimer line. If, say, there is a dimer line on AB then there will also be one on CD, but not on E'A or F'B. Similar arguments apply for BC and

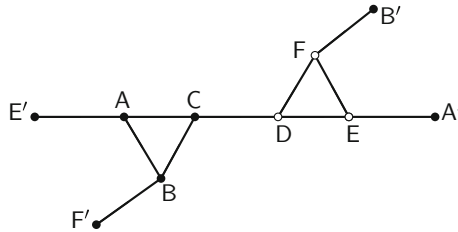


Fig. 3.13 The 3-12 lattice obtained by splitting the sites of a honeycomb lattice

CA. Hence, if the honeycomb lattice is recovered by replacing each triangle by a single site, then either three lines or one line are incident at this site. It follows that there is a one-to-one correspondence between dimer graphs on the new 3–12 lattice denoted by $\mathcal{N}_{\text{TT}}$, where $|\mathcal{N}_{\text{TT}}| = 3N$, and complementary graphs on the honeycomb lattice $\mathcal{N}_{\text{H}}$ of N sites.

For the dimer assembly on $\mathcal{N}_{\text{TT}}$ let the fugacities on the edges of the triangles be unity and those on the honeycomb edges of which **CD**, **EA'** and **FB'** are representative be $\mathfrak{Z}_1$, $\mathfrak{Z}_2$ and $\mathfrak{Z}_3$ respectively. Because of the one-to-one correspondence between dimer graphs on the $\mathcal{N}_{\text{TT}}$ and complementary graphs on the $\mathcal{N}_{\text{H}}$, the dimer partition function on $\mathcal{N}_{\text{TT}}$ is given by

$$Z^{(\text{DM})}(\mathcal{N}_{\text{TT}}, \mathfrak{Z}_1, \mathfrak{Z}_2, \mathfrak{Z}_3) = \sum_{\{\mathfrak{g}'\}}^{(\text{CZF})} \mathfrak{Z}_1^{e_1(\mathfrak{g}')} \mathfrak{Z}_2^{e_2(\mathfrak{g}')} \mathfrak{Z}_3^{e_3(\mathfrak{g}')}, \quad (3.9.19)$$

where the summation is over all complementary graphs on $\mathcal{N}_{\text{H}}$. In fact this summation is just the $\eta = 3$ case of the summation in (3.9.18) with $\mathfrak{Z}_k := v_k^{-1}$, $k = 1, 2, 3$. For the honeycomb lattice $s_1 = s_2 = s_3 = \frac{1}{2}N$ and, from (3.9.18) and (3.9.19),

$$Z^{(\text{IM})}(\mathcal{N}_{\text{H}}, \{v_k\}) = \left[4 \prod_{k=1}^3 \sinh(K_k) \right]^{N/2} Z^{(\text{DM})}(\mathcal{N}_{\text{TT}}, v_1^{-1}, v_2^{-1}, v_3^{-1}). \quad (3.9.20)$$

Chapter 4

Phase Transitions and Scaling Theory

4.1 The Geometry of Phase Transitions

We consider a system in the coupling-density representation in which, as in Sect. 2.5, there are no independent densities ($\eta = n_f - 1$).¹ It will also be supposed, until the discussion of finite-size scaling in Sect. 4.10, that the thermodynamic limit has been taken with respect to the one independent extensive variable. From (2.3.15) and (2.3.18),

$$\phi = L_{n_f} = f/T, \quad (4.1.1)$$

$$\phi = -s + \sum_{\ell=1}^{n_e} u_{\ell} K_{\ell} - \sum_{i=1}^{n_f-1} L_i \rho_i, \quad (4.1.2)$$

$$d\phi = \sum_{\ell=1}^{n_e} u_{\ell} dK_{\ell} - \sum_{i=1}^{n_f-1} \rho_i dL_i. \quad (4.1.3)$$

The system has $n = n_f + n_e - 1$ independent couplings and the free-energy density ϕ can be plotted as an n -dimensional hypersurface in the $(n + 1)$ -dimensional space of the variables $\phi, K_1, \dots, K_{n_e}, L_1, \dots, L_{n_f-1}$. This space and hypersurface are unaffected by the choice of Q_{n_f} in Sect. 2.1.² From (4.1.3) it can be seen that components of the gradient of the surface at any point correspond to the densities and the elements of the curvature matrix are thermodynamic response functions like the one appearing on the left-hand side of equation (2.5.15).

¹ In such general cases the subscript will be omitted from ϕ_{n_f-1} (and f_{n_f-1}), it being retained only when n_f has some small specified value (e.g. $n_f = 2$ for a critical point and $n_f = 3$ for a tricritical point).

² Since $\phi = L_{n_f}$ a different choice of an independent extensive variable would correspond simply to a relabelling of the axes.

The free-energy density $\phi(K_1, \dots, K_{n_e}, L_1, \dots, L_{n_f-1})$ is, for most values of its arguments, a smooth function.³ The exceptional points are associated with some kind of critical or phase-coexistence behaviour. Phase boundaries correspond to $(n - 1)$ -dimensional regions on the hypersurface of ϕ in which some kind of non-smooth behaviour occurs. These are the transition regions of highest dimension. In general a transition region may have any dimension between $n - 1$ and zero (a point). The projection of all these transition regions into the n -dimensional space of couplings yields the phase diagram of the system. The simplest form of a phase boundary is when some or all of the densities are discontinuous across the boundary and we have a *first-order transition or region of two-phase coexistence*. In a similar way a region of p -phase coexistence is in general of dimension $n + 1 - p$. Since this region must have a non-negative dimension we might conclude that largest number of phases which can coexist for n independent couplings is $n + 1$. This restriction is the *Gibbs phase rule*.

There are exceptions, which arise when the phase diagram has some special symmetry. One of these is the triangular ferrimagnetic Ising model for which the phase diagram has reflectional symmetry about the zero-field axis. A compensation point⁴ occurs on this axis where four first-order transitions meet. However, mean-field (Bell 1974a) and renormalization group (Lavis and Quinn 1983, 1987) calculations indicate that, at the compensation point, the discontinuities in density are zero, with the phases forming two pairs of the same ordering structure. Such a compensation point, at the coincidence of four first-order transitions, also occurs in the decorated lattice model of Sect. 10.3 and a further example of the ‘violation’ of the Gibbs phase rule is provided by the metamagnet with a staggered field used as an example in Sect. 4.1.2 (see also Sect. 6.2). Here the phase diagram is three-dimensional with reflectional symmetry about the plane on which the staggered field is zero. Mean-field calculations indicate (Kincaid and Cohen 1975) that, for a certain parameter range, this plane contains a line of four-phase coexistence (Fig. 4.3a, b); one more phase than would be allowed by the Gibbs phase rule.

4.1.1 A Two-Dimensional Phase Space

We here consider the case $n_f = 2, n_e = 1, n = 2$, with free-energy densities $\phi_0(K, \rho)$ and $\phi_1(K, L)$ satisfying the differential forms (2.3.20) and (2.3.21), respectively, and related by the Eqs. (2.3.22)–(2.3.25).

For this case phases can be separated by lines of transitions. A *critical point* occurs at a second-order transition terminating a line of first-order transitions (Fig. 4.1a).⁵

³ See Definition 17.1.1.

⁴ For a discussion of ferrimagnetism and compensation points see Martin (1967) pp. 38–44.

⁵ In this and subsequent figures first-order transitions are indicated by broken lines and second-order transitions by continuous lines.

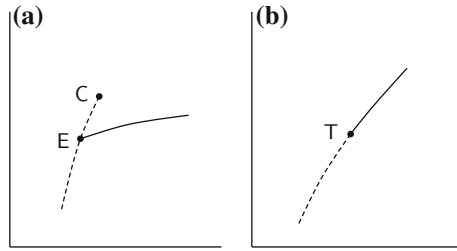


Fig. 4.1 **a** A critical point **C** terminating a line of first-order transitions and a critical end-point **E** where a line of second-order transitions terminates on a line of first-order transitions. **b** A tricritical point **T** where a line of first-order transitions meets a line of second-order transitions

Examples of this are the Curie point in a magnet and the termination of the vapour-liquid transition in a fluid system.

A line of first-order transitions shows as a discontinuity of gradient in the surface of ϕ_1 plotted as a function of K and L (Fig. 4.2a). This line can also be regarded as representing the coexistence of two phases with different densities given, from (2.3.25), by the negatives of the gradients of ϕ_1 with respect to L on each side of the line. In the corresponding picture (Fig. 4.2b) for ϕ_0 plotted as a function of ρ the points **A** and **B** represent the two sides of the phase transition. These points are joined by a double tangent with slope L^* and, from (2.3.24),

$$\phi_0^{(+)} - \phi_0^{(-)} = L^*[\rho^{(+)} - \rho^{(-)}]. \quad (4.1.4)$$

As K is varied **A** and **B** describe two lines on the ϕ_0 surface bordering a flat region of phase coexistence. Typically the two lines will meet at a critical point (Sect. 4.4). If (2.3.23) is used to plot L as a function of ρ along an isotherm, a first-order transition appears as a flat part of the curve (Fig. 4.2c). A *triple point* occurs at the confluence of three lines of first-order transitions; an example of this is the point where the first-order transitions between ice, liquid water and steam meet.⁶ Other types of behaviour, which are grouped under the general heading of *multicritical phenomena* can also occur. When a line of first-order transitions is continued by a line of second-order transitions the two lines are said to meet at a *tricritical point* (Fig. 4.1b). When a line of second-order transitions terminates on a line of first-order transitions at an intermediate point this meeting point is called a *critical end-point* (Fig. 4.1a). Tricritical points and critical end-points occur for different parameter ranges when mean-field theory is applied to a metamagnetic model (Sect. 6.2). In that model spins are arranged in parallel planes with a ferromagnetic interaction within planes and an antiferromagnetic interaction between neighbouring planes. FeCl_2 is an experimental example of this kind of system. In this case the

⁶ Both the Kelvin and Celsius temperature scales are defined by the triple point of water: 273.16 K and 0.01 °C respectively. (But see the footnote on p. 5 for an alternative way to define the temperature scale.)

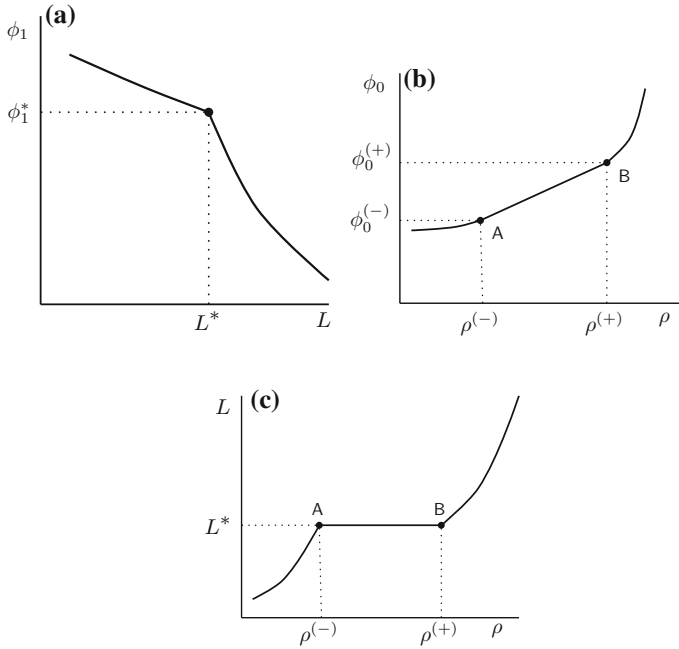


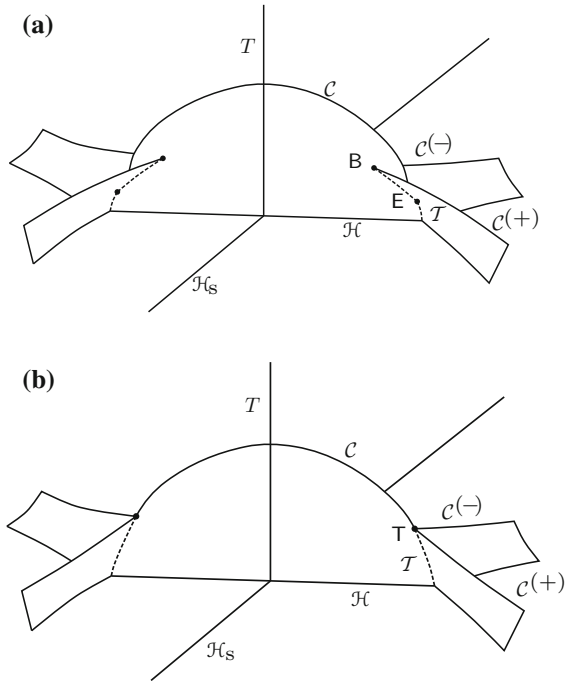
Fig. 4.2 **a** A first-order transition showing as a discontinuity of slope in an isothermal section of the ϕ_1 surface. **b** A first-order transition showing as the double tangent AB in an isothermal section of the ϕ_0 surface. **c** A first-order transition showing as a horizontal part of an isotherm of L plotted against ρ

second-order transition line terminates at the *Néel point* on the zero-field axis and meets a first-order transition line at a point in the plane, with this latter terminating on the zero-temperature axis. The meeting of the two transition curves can be a tricritical point or a critical end-point⁷ according to the relative strengths of the ferromagnetic and antiferromagnetic interactions. As can be seen from Fig. 4.3a (see also Sect. 6.2.5) the bicritical end-point in the metamagnetic model, occurs *inside* the antiferromagnetic phase. This differs from the behaviour in the *Blume-Emery-Griffiths model* (Blume et al. 1971) where the phase topology is similar to that shown in Fig. 4.1a.

A further type of multicritical behaviour which can occur in a two-dimensional phase space is when the critical lines of two phase transitions meet at a point, which is also the end of a coexistence line of first-order phase transitions separating the two ordered phases. The meeting point is called a *bicritical point*. Bicritical points have been observed most clearly in weakly anisotropic antiferromagnets such as

⁷ The name ‘tricritical point’ in the context seems slightly odd since we are apparently considering a situation with only two phases present. A full understanding of tricritical points needs a phase space of more dimensions (Sect. 4.1.2). The same applies to the designation of the critical end-point as a *bicritical end-point*.

Fig. 4.3 a Phase diagram for a metamagnetic model in the space of temperature T , magnetic field $\mathcal{H}$ and staggered field $\mathcal{H}_s$. The critical curve $\mathcal{C}$ meets the triple line $\mathcal{T}$ at the critical end-point E . The critical curves $\mathcal{C}^{(+)}$ and $\mathcal{C}^{(-)}$ meet on the plane $\mathcal{H}_s = 0$ at the bicritical end-point B . (Reprinted from Kincaid and Cohen (1975), p. 114, by permission of the publisher Elsevier Science.) **b** As for (a) but now the critical end-point and bicritical end-point have coalesced to form a triple-point T . (Reprinted from Kincaid and Cohen (1975), p. 110, by permission of the publisher Elsevier Science.)



GdAlO_3 and MnF_2 in carefully oriented magnetic fields (Rohrer 1975). A theoretical realization of a bicritical point occurs in the temperature–electric field plane of a six-vertex ferroelectric (Lieb and Wu 1972, Fig. 30).

4.1.2 A Three-Dimensional Phase Space

When the phase space acquires a third dimension, so that $n = 3$, the boundaries between pairs of phases become surfaces. A first-order surface may be bounded by a line of critical points or *critical line*. Three first-order surfaces, separating three distinct phases can meet at a *triple line*. With these basic components phase diagrams can exhibit a rich diversity of structure. Many of these appear when a *staggered field*, alternating in sign between layers, is added to the metamagnetic model referred to above. Figure 4.3a reproduces a phase diagram for this model obtained by Kincaid and Cohen (1975) using mean-field methods (Sect. 6.2). This form of phase diagram occurs for a particular range of the ratio between the ferromagnetic interaction within layers and the antiferromagnetic interaction between layers (Sect. 4.6). Three phase boundaries meet along a triple line $\mathcal{T}$, which lies between the zero-temperature plane and the critical end-point E , where it meets the critical curve $\mathcal{C}$ and a line of *four-phase coexistence* BE . The point B is a *bicritical end-point*, where the critical curves

$\mathcal{C}^{(+)}$ and $\mathcal{C}^{(-)}$ meet on the plane $\mathcal{H}_s = 0$. By varying the parameter ratio so that **B** and **E** coalesce, a tricritical point can be achieved and the phase diagram takes the form shown in Fig. 4.3b. The tricritical point is the meeting point of the critical curve and the triple line clarifying the reason why Griffiths (1970) chose the name ‘tricritical’ for a point of coincidence of a first-order and critical line in a plane. It will be clear from these brief remarks that even within the constraints of a phase space of at most three dimensions not all the possible types of phase behaviour have been covered.⁸

4.2 Universality, Fluctuations and Scaling

As indicated above, one aim of the theory of critical phenomena is to formulate a classification of phase transitions. Criteria are sought in terms of which it can be said that certain phase transitions are ‘of the same type’. A early attempt of this kind arose from the observation by van der Waals (1873) that his equation could be expressed in terms of variables scaled with respect to their critical values (Sect. 5.2). A *law of corresponding states* was formulated suggesting that a wide class of physical systems would, in terms of appropriately scaled variables, obey this same equation of state. The experimental fact that the *critical ratio*

$$K_c := \frac{\tilde{P}_c \tilde{V}_c}{M_A \tilde{T}_c} = k_B \frac{P_c V_c}{M_A T_c}, \quad (4.2.1)$$

where M_A is Avogadro’s number, is not the same for all gases caused some concern and was the first sign that the law of corresponding states has only limited applicability. However, as has been pointed out by Levelt Sengers (1974):

Although this law has limited validity experimentally, and is not obeyed by most other analytic equations of state, its role in the development of the physical, chemical and engineering sciences dealing with fluids can hardly be exaggerated.

It led to many attempts to construct more general analytical equations of state and ultimately to the theory of Landau (1937) described in Chap. 5. The first steps towards the development of this approach were taken by van der Waals (1893). In this article on the theory of capillarity he investigated the asymptotic critical behaviour of his equation by expanding about the critical point in terms of the volume and temperature deviations from their critical values. This yielded the first calculation of the critical exponent, now known as β , for which he obtained the value $\frac{1}{2}$. This exponent measures the curvature of the coexistence curve in the vicinity of the critical point and, towards the end of the nineteenth century, experimental results began to appear indicating disagreement with the theoretical value $\beta = \frac{1}{2}$. The initial reaction was to suppose that the experimental measurements were not taken sufficiently close to

⁸ For a review of theoretical and experimental work on multicritical phenomena see Pynn and Skjeltorp (1984) and for tricritical phenomena, with further examples, the review by Lawrie and Sarbach (1984).

criticality. In 1900 Verschaaffelt was, however, able to show by analyzing the results of Young (1892) for isopentane, that $\beta = 0.3434$, with no evidence of the value tending towards $\frac{1}{2}$ as the critical point was approached. This was the beginning of the realization that critical exponents (both β and the other critical exponents associated with the asymptotic properties of thermodynamic functions, see Sect. 4.4) were crucial for the classification of phase transitions. The development of generalized forms of the equation of state incorporating fractional exponents took another 65 years. By this time there was ample evidence that experimental systems yield critical exponents with values different from those predicted by classical (Landau) theory. Theoretical impetus was provided by the Onsager (1944) solution of the two-dimensional Ising model which demonstrated non-analytic behaviour at the critical point. We shall not, with the benefit of hindsight, attempt to trace the details of the early development of generalized equations of state leading to scaling theory. Rather we shall concentrate on trying to distinguish those properties of a system which affect its critical behaviour.

4.2.1 Universality

At points near a transition region in phase space a system will be in different stable phases. This entails the existence of fluctuations, which can be regarded as the short-time and local appearance of fragments of one of the phases in another. An example of this is the critical opalescence present in a fluid near the critical point of the liquid-vapour transition. Thus fluctuations are the appearance of locally-ordered regions. These persist for a certain mean length of time t called the *correlation time* and extend over a mean distance r called the *correlation length*. Roughly speaking, dynamic and static critical behaviours are governed by t and r respectively. Since we are concerned exclusively with static critical behaviour it is the correlation length, measuring the mean extension of correlations, which will be of interest to us. Given that, in a neighbourhood of a transition region, correlations extend over distances large with respect to the detailed nature (short-range interactions, lattice structure etc.) of the system, it is intuitively easy to understand that this microstructure will not affect the characteristics of the system. This led to the formulation of a *universality hypothesis* which can be expressed in the following way (Kadanoff 1976b):

All critical problems can be divided into *universality classes* differentiated by:

- (a) the (physical) dimension d of the system,
- (b) the symmetry group of the *order parameter*,
- (c) perhaps other criteria.

Within each class, the critical properties, and in particular the critical exponents, are supposed to be identical or, at worst, to be continuous functions of very few parameters. Although, in theoretical models, systems of any dimension can be easily realized it might be supposed that experiment is relevant only to the case $d = 3$. However, polymers (linear macromolecules) and solids made up of chains

of weakly-coupled molecules provide examples of one-dimensional, or quasi-one-dimensional systems and two-dimensional experimental systems can be realized in films, adsorbed phases, weakly-coupled planes and surface phenomena.

In a rather general way the order parameter of a phase can be defined as the expectation value of a quantity which measures the extent to which the microscopic configuration mirrors a state of perfect ordering characteristic of this phase. However, the order parameter is not always easy to define. In a fluid system undergoing a liquid-gas transition the order parameter is a density difference and of dimension one. In a magnetic transition it is the magnetization which may be of dimension one, two or three according to the constraints imposed on the system. The change of the nature of a phase transition according to the symmetry group of the order parameter is illustrated in the context of Landau theory in Chap. 5. The inclusion of item (c) by Kadanoff could be viewed as a means of providing for exceptional cases which will almost inevitably arise in any kind of classification of this type. An example of this is the bond-dilute Ising model where the order parameter is a scalar quantity, so one would, on the basis of (a) and (b) alone, expect the exponents to be those of the pure Ising model, for the corresponding value of d . It can, however, be shown (Sect. 10.2.1) that instead of the Ising exponents α and γ for the heat capacity and susceptibility respectively the actual exponents α_r and γ_r are given in terms of their Ising values by $\alpha_r = -\alpha/(1-\alpha)$ and $\gamma_r = \gamma/(1-\alpha)$. This change of exponent, whenever $\alpha \neq 0$, ($\alpha = 0$ for $d = 2$) is called *exponent renormalization* (Fisher 1968). It arises whenever the fraction of secondary sites occupied by solute atoms is treated as an independent variable. In the six-vertex model in an electric field the order parameter is the scalar polarization. However, it can be shown (Lieb and Wu 1972, Sect. 5E) that, at the transition to the polarized state, the heat capacity diverges with an exponent $\alpha = \frac{1}{2}$ compared to the zero value of the corresponding Ising model. These two different types of transition are also exhibited by dimer models. They are discussed in Sect. 13.4 and by Nagle et al. (1989) who called them K-type (for Kasteleyn) and O-type (for Onsager). Finally the eight-vertex model should be mentioned. This yields continuously varying critical exponents (Sect. 12.3.8). The way in which such exponents can arise within the context of scaling theory is investigated in Sect. 4.3.7 and the particular case of the eight-vertex model is considered in Sect. 12.3.11. It should be emphasized that the concept of a universality class is associated with a phase transition and not with a physical or theoretical system as a whole. Systems with sufficient complexity may exhibit different phase transitions, in different parts of their phase spaces, lying in different universality classes.

In the presentation of scaling theory in Sect. 4.3 the approach used is that of Hankey and Stanley (1972). This method has the virtues of mathematical elegance and generality. The basic premises do, however, seem to be rather arbitrary without the provision of some intuitive insight. This we shall now try to provide, using the example of the ferromagnetic Ising model, following the ideas of Kadanoff (1966) (see also, Kadanoff et al. 1967).

4.2.2 Scaling for the Ising Model

Consider the spin- $\frac{1}{2}$ Ising model on a d -dimensional hypercubic lattice $\mathcal{N}$ of N sites and lattice spacing a with Hamiltonian given by (3.5.1). The density conjugate to L is the magnetization density $m := \mathcal{M}/N$, where $\mathcal{M} := \langle \hat{\mathcal{M}} \rangle$ and $\hat{\mathcal{M}}$ is given by (3.5.4). The ferromagnetic Ising model is believed to have, for all $d > 1$, a first-order phase transition along the zero-field axis ($L = 0$) in the interval $K_c < K \leq \infty$, where the critical point is $K = K_c$, $L = L_c = 0$.⁹ The order parameter for this transition region is m and the variance of $\hat{m}$ is given, from (2.5.16), by

$$\text{Var}[\hat{m}] = \frac{1}{N} \sum_{\{\mathbf{r}_1, \mathbf{r}_2\}} \Gamma_2(\sigma(\mathbf{r}_1), \sigma(\mathbf{r}_2)), \quad (4.2.2)$$

where

$$\Gamma_2(\sigma(\mathbf{r}_1), \sigma(\mathbf{r}_2)) = \langle \sigma(\mathbf{r}_1) \sigma(\mathbf{r}_2) \rangle - \langle \sigma(\mathbf{r}_1) \rangle \langle \sigma(\mathbf{r}_2) \rangle \quad (4.2.3)$$

is the *spin-spin (connected pair) correlation function*. When the system has translational and rotational symmetry this correlation function satisfies (2.5.25) (with $c_1 = c_2 = \sigma$) and is a measure of the amount of short-range order (spin alignment) between spins separated by a distance r . It is, therefore, proportional to the probability of a cluster of aligned spins forming with spatial extent r . The correlation length τ , given by

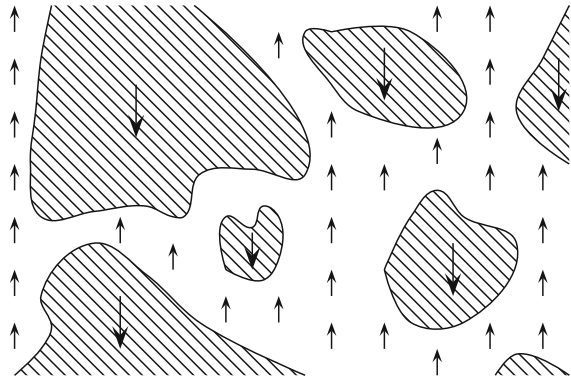
$$\tau^2 := c(\mathfrak{d}) \frac{\sum_{\{\mathbf{r}_1, \mathbf{r}_2\}} |\mathbf{r}_2 - \mathbf{r}_1|^2 \Gamma_2(\sigma(\mathbf{r}_1), \sigma(\mathbf{r}_2))}{\sum_{\{\mathbf{r}_1, \mathbf{r}_2\}} \Gamma_2(\sigma(\mathbf{r}_1), \sigma(\mathbf{r}_2))}, \quad (4.2.4)$$

is a measure of the largest clusters which will form. The prefactor $c(\mathfrak{d})$ depends on the number of dimensions $\mathfrak{d} \leq d$ in which the system is infinite (Sect. 4.10). Its exact form is unimportant for most purposes and it is sometimes omitted (Ziman 1979, p. 26). However, in Sect. 5.6.3, it is shown that, within the Gaussian approximation, the choice $c(\mathfrak{d}) = 1/(2\mathfrak{d})$ is necessary for consistency.

Consider now a path in the phase diagram with small positive field (L small and positive) beginning at a low temperature (K large). At this point thermal effects will be small and most spins will be aligned in the positive direction. As K is decreased thermal fluctuations will mean that random spin flips will show an increasing tendency to occur and the interaction energy between spins will encourage the formation of clusters based around these spins. As long as the correlation length is quite small these clusters will appear and disappear at random. However, as K continues to decrease towards the critical point, the spin-spin correlation function for

⁹ For a mean-field treatment which yields this property see Sect. 6.1.

Fig. 4.4 Clusters of down spins forming in a sea of up spins. Clusters form of all sizes up to lengths of the order of τ



a particular fixed r will increase. Correlations will grow stronger over larger distances and the correlation length and largest cluster size of down spins will become very large (Fig. 4.4), finally diverging at the critical point.

Near the critical point the correlation length is much greater than a , so it is possible to find a number λ , such that $\tau \gg \lambda a \gg a$. Ignoring boundaries or assuming periodic boundary conditions, let the lattice $\mathcal{N}$ be divided into cells, λa on each side and containing λ^d sites (Fig. 4.5), so that the set of cells can be represented by points $\tilde{\mathbf{r}} \in \tilde{\mathcal{N}}$, where $\tilde{\mathcal{N}}$ is of the same lattice structure as $\mathcal{N}$, but with lattice spacing $\tilde{a}$. This is a transformation from the lattice to the cell lattice given by

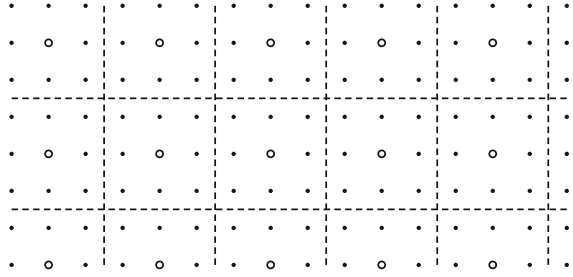
$$\begin{aligned} \mathbf{r} &\rightarrow \tilde{\mathbf{r}} := \mathbf{r}/\lambda, \\ N &\rightarrow \tilde{N} := N/\lambda^d, \\ a &\rightarrow \tilde{a} := a\lambda. \end{aligned} \tag{4.2.5}$$

Suppose that all interactions between different cells are turned off, so that each acts independently. Then there will be no correlations over distances larger than the cell size and no divergences in thermodynamic quantities. Correlations between cells are the necessary ingredient for thermodynamic behaviour near the critical point and this behaviour can be supposed to be independent of the details of the interactions within the cells. The interaction between the cell $\mathcal{B}(\tilde{\mathbf{r}})$ centred $\tilde{\mathbf{r}}$ and the other cells is modelled in terms of a single spin variable $\tilde{\sigma}(\tilde{\mathbf{r}}) = \pm 1$ related to the spins in the cell by

$$\sum_{\mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})} \sigma(\mathbf{r}) = w(\lambda) \tilde{\sigma}(\tilde{\mathbf{r}}), \tag{4.2.6}$$

for some $w(\lambda)$. The correlation length is the same for the spins $\tilde{\sigma}(\tilde{\mathbf{r}})$ on $\tilde{\mathcal{N}}$ as for $\sigma(\mathbf{r})$ on $\mathcal{N}$, if it is measured in units of a , but, if it is now measured in units of $\tilde{a}$, it has the value τ/λ . The thermodynamic state of the original model is given by values of

Fig. 4.5 Lattice of spacing a divided into cells, of linear dimension λa , containing λ^d spins



$\Delta K := K - K_c$ and L . The rescaled model, in which the correlation length is smaller by a factor λ^{-1} , will appear to be further from the critical point. It will, therefore, have effective values of ΔK and L larger than those in the original model. These rescaled values we denote by $\lambda^{y_K} \Delta K$ and $\lambda^{y_L} L$, for some unknown positive scaling exponents y_K and y_L . The important *assumption* made at this stage is that, for points sufficiently close to the critical point in phase space, y_K and y_L are constants. This means that τ acts as the single measure of the distance from the critical point and, under the sequence of transformations $\tau \rightarrow \tau/\lambda \rightarrow \tau/\lambda^2 \rightarrow \dots$, ΔK and L transform according to $\Delta K \rightarrow \lambda^{y_K} \Delta K \rightarrow \lambda^{2y_K} \Delta K \rightarrow \dots$ and $L \rightarrow \lambda^{y_L} L \rightarrow \lambda^{2y_L} L \rightarrow \dots$ respectively. Thus we have, in a neighbourhood of the critical point,

$$\tau(\lambda^{y_K} \Delta K, \lambda^{y_L} L) = \lambda^{-1} \tau(\Delta K, L). \quad (4.2.7)$$

As we saw in Sect. 4.1, the critical point will correspond to non-smooth behaviour of the free energy of the system. It will, however, itself be finite non-zero and contain a smooth component. The same will be true of ϕ_1 and we suppose that we can make the separation

$$\phi_1 = \phi_{\text{smth}} + \phi_{\text{sing}}, \quad (4.2.8)$$

where

$$\phi_{\text{sing}} = \phi_{\text{sing}}(\Delta K, L) \quad (4.2.9)$$

contains all the non-smooth behaviour of ϕ_1 and $\phi_{\text{sing}}(0, 0) = 0$.

That the free-energy density can be divided into smooth and singular parts is an *assumption* of scaling theory (Hankey and Stanley 1972, p. 3519), which is then justified in terms of the form of the Hamiltonian (Sect. 4.3.6). If we are concerned only with the critical behaviour of the system, we can, in most but not all cases (Sect. 4.3.4) ignore the smooth part of ϕ_1 .

Given that τ is the sole measure of the distance from the critical point, ϕ_{sing} can be given as a function of τ . In order for the critical properties of the system to remain unchanged under the scaling transformation the singular part of the free energy must be invariant and, since the number of lattice sites is reduced from N to N/λ^d ,

$$\phi_{\text{sing}}(\mathbf{r}) = \lambda^{-d} \phi_{\text{sing}}(\lambda^{-1} \mathbf{r}), \quad (4.2.10)$$

or, equivalently,

$$\phi_{\text{sing}}(\lambda^{y_K} \Delta K, \lambda^{y_L} L) = \lambda^d \phi_{\text{sing}}(\Delta K, L). \quad (4.2.11)$$

The corresponding expression for the magnetization density is obtained by differentiating (4.2.11) with respect to L and this can also be obtained by summing (4.2.6) over all $\mathcal{B}(\tilde{\mathbf{r}})$ and taking the expectation. Thus we conclude that $w(\lambda) = \lambda^{y_L}$.

4.3 General Scaling Formulation

Consider a system of the type described at the beginning of Sect. 4.1, with $n = n_f + n_e - 1$ independent couplings, which, for the sake of brevity, we denote by ζ_k , $k = 1, 2, \dots, n$, for which, where convenient, we also use $\boldsymbol{\zeta}$ to denote the vector of couplings.

Let there be a transition region $\mathcal{C}$ of dimension $n - s$ in the space of $\boldsymbol{\zeta}$ and suppose that, close to $\mathcal{C}$, there is a set of curvilinear coordinates

$$\theta_j = \theta_j(\Delta \boldsymbol{\zeta}), \quad j = 1, \dots, n, \quad \Delta \boldsymbol{\zeta} := \boldsymbol{\zeta} - \boldsymbol{\zeta}^*, \quad (4.3.1)$$

called *scaling fields*, which are smooth functions of the couplings oriented about an origin $\boldsymbol{\zeta}^* := \boldsymbol{\zeta}^* \in \mathcal{C}$, with $\theta_j^* := \theta_j(0) = 0$, and constructed in such a way that $\mathcal{C}$ is defined by

$$\theta_1 = \theta_2 = \dots = \theta_s = 0. \quad (4.3.2)$$

The variables $\theta_1, \dots, \theta_s$ and $\theta_{s+1}, \dots, \theta_n$ are called *relevant* and *irrelevant* scaling fields respectively. The latter give coordinates for points in $\mathcal{C}$.

4.3.1 The Kadanoff Scaling Hypothesis

In this section a version of the scaling hypothesis is introduced, which is essentially that due to Kadanoff (1966), with a treatment which follows the work of Hankey and Stanley (1972). This procedure is sufficient for an analysis of power-law singularities. However, certain systems, in particular the Ising model with $d = 2$ (Sect. 12.2.7.3), exhibit logarithmic singularities. It has been shown by Nightingale and 't Hooft (1974) that a slight generalization of the Kadanoff scaling hypothesis can be used to analyze these cases (Sect. 4.3.4).

As in the special case described in Sect. 4.2 it is assumed that ϕ can be divided into smooth and singular parts according to equation (4.2.8)¹⁰ with ϕ_{sing} a function of the scaling fields. For these it is supposed that, within a region around $\mathcal{C}$, for all real¹¹ $\lambda > 1$ and some set of *scaling exponents*¹² y_j , there is a transformation

$$\theta_j \rightarrow \tilde{\theta}_j := \lambda^{y_j} \theta_j, \quad (4.3.3)$$

for which

$$\phi_{\text{sing}} \rightarrow \tilde{\phi}_{\text{sing}} := \lambda^d \phi_{\text{sing}}. \quad (4.3.4)$$

The origin of the scaling fields is the *fixed point* of the scaling transformation (4.3.3) and (4.3.4) and by extension of $\mathcal{C}$, which is its *basin of attraction*. The scaling transformation is equivalent to the *Kadanoff scaling hypothesis*

$$\phi_{\text{sing}}(\lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) = \lambda^d \phi_{\text{sing}}(\theta_1, \dots, \theta_n), \quad (4.3.5)$$

which asserts that the singular part of ϕ is a *generalized homogeneous function* of the scaling fields. The scaling exponents fall into two subsets

$$\begin{aligned} y_j &> 0, & j &= 1, \dots, s, \\ y_j &< 0, & j &= s+1, \dots, n, \end{aligned} \quad (4.3.6)$$

called *relevant* and *irrelevant* respectively, like the corresponding scaling fields. A zero exponent is called *marginal*. We shall at this stage, for the sake of simplicity, assume that none of the exponents is marginal. The significance of marginal exponents in scaling theory is considered in Sect. 4.3.7. It will also be assumed that no exponent is complex. In practice, as will be seen in Sect. 15.5.2, this is not necessarily always the case, but situations arising from complex exponents are not difficult to interpret in particular examples.

Denoting by $\mathbf{k}$ a set of non-negative integers $(k_1, \dots, k_n)$, let $\psi(\mathbf{k})$ be the function obtained by differentiating ϕ k_j times with respect to θ_j , for $j = 1, \dots, n$. Then, from (4.3.5), assuming a division of $\psi(\mathbf{k})$ into smooth and singular parts as for ϕ ,

$$\psi_{\text{sing}}(\mathbf{k}; \lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) = \lambda^{d'(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n), \quad (4.3.7)$$

¹⁰ As we shall see later in this section, this analysis applies both to first-order transitions and to those for which some derivative of the free-energy density diverges in the transition region. The term ‘singular’ is taken, therefore, to include all types of non-smooth behaviour, including finite discontinuities in derivatives of ϕ .

¹¹ The parameter λ introduced in Sect. 4.2 is an integer greater than unity. If, however, (4.3.3) and (4.3.4) are defined for all real $\lambda > 1$ it is not difficult to see that (4.3.5) also holds for $1 \geq \lambda > 0$.

¹² The term ‘scaling exponent’ is used to make the distinction with the critical exponents defined in Sect. 4.4. When there is no danger of confusion they will be referred to simply as ‘exponents’.

where

$$d'(\mathbf{k}) := d - \sum_{j=1}^n k_j y_j. \quad (4.3.8)$$

Setting, for a particular relevant scaling field θ_m ,

$$\lambda = |\theta_m|^{-1/y_m} \quad (4.3.9)$$

in (4.3.5) and (4.3.7) gives

$$\phi_{\text{sing}}(\theta_1, \dots, \theta_n) = |\theta_m|^{\sigma_m} \phi_{\text{sing}}(g_1^{(m)}, \dots, g_n^{(m)}), \quad (4.3.10)$$

$$\psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n) = |\theta_m|^{\omega_m(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; g_1^{(m)}, \dots, g_n^{(m)}), \quad (4.3.11)$$

where

$$\omega_m(\mathbf{k}) := \frac{d'(\mathbf{k})}{y_m}, \quad \sigma_m := \omega_m(\mathbf{0}) = \frac{d}{y_m} \quad (4.3.12)$$

and

$$g_j^{(m)}(\theta_m, \theta_j) := \frac{\theta_j}{|\theta_m|^{\vartheta_j^{(m)}}}, \quad j = 1, \dots, n, \quad (4.3.13)$$

with

$$\vartheta_j^{(m)} := \frac{y_j}{y_m} = \frac{1}{\vartheta_m^{(j)}}, \quad j = 1, \dots, n. \quad (4.3.14)$$

From (4.3.8), (4.3.12) and (4.3.14)

$$\omega_m(\mathbf{k}) = \sigma_m - \sum_{j=1}^n k_j \vartheta_j^{(m)}. \quad (4.3.15)$$

The exponent $\vartheta_j^{(m)}$ gives the incremental change in the exponent $\omega_m(\mathbf{k})$ of the function $\psi_{\text{sing}}(\mathbf{k})$ on differentiation with respect to θ_j . These quantities are therefore referred to as *gap exponents*. By extension it is convenient to refer to the quantities $g_j^{(m)}$ as *gap fields*. They satisfy the conditions

$$|g_m^{(\ell)}|^{\vartheta_j^{(m)}} |g_j^{(m)}| = |g_j^{(\ell)}|, \quad j, \ell = 1, \dots, n, \quad (4.3.16)$$

with $g_m^{(m)} = \pm 1$, according as $\theta_m \gtrless 0$. From (4.3.10), (4.3.11) and (4.3.13),

$$\begin{aligned}\phi_{\text{sing}}(g_1^{(j)}, \dots, g_n^{(j)}) &= |g_m^{(j)}|^{\sigma_m} \phi_{\text{sing}}(g_1^{(m)}, \dots, g_n^{(m)}), \\ \psi_{\text{sing}}(\mathbf{k}; g_1^{(j)}, \dots, g_n^{(j)}) &= |g_m^{(j)}|^{\omega_m(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; g_1^{(m)}, \dots, g_n^{(m)}).\end{aligned}\quad (4.3.17)$$

Now consider the asymptotic forms derived from (4.3.10) and (4.3.11) as $\mathcal{C}$ is approached ($\theta_j \rightarrow 0$, $j = 1, \dots, s$) along some path $\mathcal{P}$. Suppose that

$$\lim_{\mathcal{P}} g_j^{(m)}(\theta_m, \theta_j) = \varpi_j^{(m)}. \quad (4.3.18)$$

It is not difficult to show that, for any path $\mathcal{P}$, it is always possible to find a relevant field θ_m such that $\varpi_j^{(m)}$ is finite for all $j = 1, \dots, s$.¹³ Conversely, as we see below, given a relevant field θ_m , there exists a part of the neighbourhood of $\mathcal{C}$ which contains paths for which $\varpi_j^{(m)}$ are finite for all $j = 1, \dots, n$. A more difficult problem is to be sure that the amplitudes

$$\begin{aligned}\lim_{\mathcal{P}} \phi_{\text{sing}}(g_1^{(m)}, \dots, g_n^{(m)}) &= \phi_{\text{sing}}(\varpi_1^{(m)}, \dots, \varpi_n^{(m)}), \\ \lim_{\mathcal{P}} \psi_{\text{sing}}(\mathbf{k}; g_1^{(m)}, \dots, g_n^{(m)}) &= \psi_{\text{sing}}(\mathbf{k}; \varpi_1^{(m)}, \dots, \varpi_n^{(m)}),\end{aligned}\quad (4.3.19)$$

are finite *and* non-zero. If this is indeed the case then it follows from (4.3.10) and (4.3.11) that

$$\begin{aligned}\phi_{\text{sing}}(\theta_1, \dots, \theta_n) &\simeq |\theta_m|^{\sigma_m} \phi_{\text{sing}}(\varpi_1^{(m)}, \dots, \varpi_n^{(m)}), \\ \psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n) &\simeq |\theta_m|^{\omega_m(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; \varpi_1^{(m)}, \dots, \varpi_n^{(m)}),\end{aligned}\quad (4.3.20)$$

as $\mathcal{C}$ is approached along $\mathcal{P}$. If $\varpi_j^{(m)} \neq 0$ for some j then, with the substitution $|\theta_m| = |\varpi_m^{(j)}| |\theta_j|^{\vartheta_m^{(j)}}$ in (4.3.20) and using (4.3.17), the formulae (4.3.20) are recovered with m and j interchanged. For the moment we shall assume that the amplitudes in (4.3.20), given by (4.3.19) are finite and non-zero. The corrections to scaling arising from a breakdown in this assumption are discussed in Sect. 4.8. Since ϕ_{sing} is continuous the amplitude for this asymptotic form is invariant under change of sign of θ_m . For ψ_{sing} this is not in general the case.

¹³ They are clearly zero for $j = s + 1, \dots, n$.

The geometry of the path $\mathcal{P}$ close to $\mathcal{C}$ can be understood more clearly if we consider the family of surfaces

$$|\theta_j| = \varpi |\theta_m|^{\vartheta_j^{(m)}}, \quad \forall \quad 0 < \varpi < \infty. \quad (4.3.21)$$

The members of this family are tangential to the surface $\theta_j = 0$ or $\theta_m = 0$ according as $y_j >, < y_m$. If $y_j = y_m$ the members of the family make angles greater than zero and less than $\pi/2$ with both surfaces. If $\varpi_j^{(m)} \neq 0$ the path $\mathcal{P}$ lies asymptotically on the member of the family with $\varpi = \varpi_j^{(m)}$ as $\mathcal{P}$ approaches $\mathcal{C}$. If $\varpi_j^{(m)} = 0$, for any ϖ and points sufficiently close to $\mathcal{C}$, the path $\mathcal{P}$ lies between that member of (4.3.21) and the surface $\theta_j = 0$. The family of surfaces (4.3.21) represent the border between one form of critical or coexistence behaviour and another, since they divide the phase space between paths on which the limiting value $\varpi_j^{(m)}$ of $g_j^{(m)}$ is finite or infinite. They are therefore referred to as *crossover surfaces* and in this context the gap exponent $\vartheta_j^{(m)}$ is called a *crossover exponent*. Looked at from a slightly different point of view the types of scaling behaviour can be classified by considering the different ranges of value of $\vartheta_j^{(m)}$ (Pfeuty and Toulouse 1977, Chap. 3).

4.3.2 First-Order Transitions

Densities, which are the first-order partial derivatives of ϕ with respect to the couplings, can be expressed in terms of its first-order partial derivatives with respect to the scaling fields. These are finite in the transition region and it therefore follows, from (4.3.8), (4.3.12) and (4.3.20), that

$$y_j \leq d, \quad j = 1, \dots, n. \quad (4.3.22)$$

In general we see from these same equations that, as $\mathcal{C}$ is approached along the path $\mathcal{P}$, ψ_{sing} is singular, or zero, according to whether

$$d < \sum_{j=1}^n k_j y_j, \quad (4.3.23)$$

$$d > \sum_{j=1}^n k_j y_j, \quad (4.3.24)$$

respectively. Consider now the possibility of there being a discontinuity in $\psi(\mathbf{k})$, as $\mathcal{C}$ is approached along $\mathcal{P}$. This must be in $\psi_{\text{sing}}(\mathbf{k}; \varpi_1^{(m)}, \dots, \varpi_n^{(m)})$, which must be finite and, on at least one side of the coexistence curve, non-zero. This in turn must

give a discontinuity in $\psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n)$. It follows from (4.3.20) that a necessary condition for this to be the case is that $\omega_m(\mathbf{k})$ and hence $d'(\mathbf{k})$ are zero; that is

$$d = \sum_{j=1}^n k_j y_j. \quad (4.3.25)$$

It must be emphasized that condition (4.3.25) is a necessary but not sufficient condition for the existence of a discontinuity in $\psi(\mathbf{k})$. Indeed it is also the case when the function exhibits a logarithmic singularity (Sect. 4.3.4). The most important case of (4.3.25) is $d = y_j$, for some j . It indicates the possibility of a first-order transition, that is a discontinuity in one or more¹⁴ of the densities (Nienhuis and Nauenberg 1975; Fisher and Berker 1982). In this case the ‘evidence’ for a discontinuity is rather stronger since the densities do not become infinite as $\mathcal{C}$ is approached, thereby excluding the possibility of a logarithmic singularity. The presence of a first-order transition does not impose any necessary asymptotic properties on the response functions. They may be finite or, as in the case of a *first-order critical point* (Sect. 4.5.6), have singularities.

4.3.3 Effective Exponents

In most cases transition regions do not exist in isolation in phase space. They are normally connected to other transition regions, as in the case of the critical point and coexistence curve discussed in Sects. 4.4 and 4.5. This means that in the neighbourhood of $\mathcal{C}$ governed by the scaling hypothesis there will be singular behaviour apart from that exhibited on $\mathcal{C}$ itself. For simplicity we consider the case where $\mathcal{C}$ (of dimension $n - s$) is connected to a transition region $\mathring{\mathcal{C}}$ of dimension $n - s + 1$ with the scaling fields θ_j for $\mathcal{C}$ and $\mathring{\theta}_j$ for $\mathring{\mathcal{C}}$ related by

$$\mathring{\theta}_j = \theta_j, \quad j \neq s, \quad \mathring{\theta}_s = 1 - \theta_s. \quad (4.3.26)$$

This means that θ_s is relevant for $\mathcal{C}$ and irrelevant for $\mathring{\mathcal{C}}$. Following Riedel and Wegner (1974) we choose a simple model, with parameter τ and *generalized scaling fields* Θ_j , satisfying

$$\begin{aligned} \frac{\partial \Theta_j(\tau)}{\partial \tau} &= \ln(\lambda)[y_j \Theta_j(\tau) + (\mathring{y}_j - y_j) \Theta_j(\tau) \Theta_s(\tau)], \quad j \neq s, \\ \frac{\partial \Theta_s(\tau)}{\partial \tau} &= \ln(\lambda) y_s \Theta_s(\tau) [1 - \Theta_s(\tau)]. \end{aligned} \quad (4.3.27)$$

¹⁴ A discontinuity in one first-order derivative with respect to a scaling field will, in general, lead to discontinuities in more than one ‘physical’ density (Eq. (4.3.119)).

Linearizing about $\Theta_j = 0$, for $j \neq s$ gives the scaling fields $\theta_j = \mathring{\theta}_j$, and about $\Theta_s = 0$ and $\Theta_s = 1$ the scaling fields θ_s and $\mathring{\theta}_s$, respectively, with $\mathring{y}_s = -y_s$ ¹⁵ satisfying

$$\frac{\partial \theta_j(\tau)}{\partial \tau} = y_j \ln(\lambda) \theta_j(\tau), \quad (4.3.28)$$

$$\frac{\partial \mathring{\theta}_j(\tau)}{\partial \tau} = \mathring{y}_j \ln(\lambda) \mathring{\theta}_j(\tau). \quad (4.3.29)$$

The solution to (4.3.27) is

$$\begin{aligned} \Theta_j(\tau) &= A_j \lambda^{y_j \tau} [1 + A_s \lambda^{y_s \tau}]^{x_j}, \quad j \neq s, \\ \Theta_s(\tau) &= A_s \lambda^{y_s \tau} [1 + A_s \lambda^{y_s \tau}]^{-1}, \end{aligned} \quad (4.3.30)$$

for some constants A_j , where, with the adaption $\vartheta_j := y_j/y_s$, $\mathring{\vartheta}_j := \mathring{y}_j/y_s$ of the notation of (4.3.14),

$$x_j := \frac{\mathring{y}_j - y_j}{y_s} = \mathring{\vartheta}_j - \vartheta_j. \quad (4.3.31)$$

Since $\theta_j(\tau) = A_j \lambda^{y_j \tau}$ is a solution to (4.3.28),

$$\begin{aligned} \Theta_j(\tau) &= \theta_j(\tau) [1 + \theta_s(\tau)]^{x_j}, \quad j \neq s, \\ \Theta_s(\tau) &= \theta_s(\tau) [1 + \theta_s(\tau)]^{-1}. \end{aligned} \quad (4.3.32)$$

As $\mathcal{C}$ and $\mathring{\mathcal{C}}$ are interchanged by the substitution $\Theta_s \rightarrow (1 - \Theta_s)$ in (4.3.27) the solution can also be written

$$\begin{aligned} \Theta_j(\tau) &= \mathring{\theta}_j(\tau) [1 + \mathring{\theta}_s(\tau)]^{x_j}, \quad j \neq s, \\ \Theta_s(\tau) &= [1 + \mathring{\theta}_s(\tau)]^{-1}. \end{aligned} \quad (4.3.33)$$

From (4.3.32) and (4.3.33),

¹⁵ Although this is not necessarily so, it is a good approximation in some cases. Thus, for example, as we see in Sect. 5.6.3, the Gaussian fixed point is unstable for $d < 4$. In the context of the ϵ -expansion version of renormalization group theory in $d = 4 - \epsilon$, this leads to the appearance of a neighbouring stable fixed point with the related exponents, for this and the Gaussian fixed point given by $\pm\epsilon + O(\epsilon^2)$ (Ma 1976a, Chap. 7).

$$\begin{aligned}\theta_j(\tau) &= \Theta_j(\tau)[1 - \Theta_s(\tau)]^{x_j}, & \dot{\theta}_j(\tau) &= \Theta_j(\tau)[\Theta_s(\tau)]^{x_j}, \\ \theta_s(\tau) &= [\dot{\theta}_s(\tau)]^{-1} = \Theta_s(\tau)[1 - \Theta_s(\tau)]^{-1}.\end{aligned}\tag{4.3.34}$$

The complete specification of all the scaling and generalized scaling fields is effected by setting values for the constants A_j . Since $\Theta_j(\tau) = 0$, for all j , are constants of the system (4.3.27), we impose the conditions that, for some $\tau = \tau_1$,

$$\Theta_1(\tau_1) = 1, \quad \Theta_s(\tau_1) = \alpha_1, \quad \Theta_j(\tau_1) = 0, \quad j \neq 1, s, \tag{4.3.35}$$

assuming that these values are outside the critical region. Then

$$\begin{aligned}\theta_s(\tau) &= \theta_s(\tau_1)\lambda^{y_s(\tau-\tau_1)}, & \theta_s(\tau_1) &= \alpha_1/(1 + \alpha_1), \\ \dot{\theta}_s(\tau) &= \dot{\theta}_s(\tau_1)\lambda^{y_s(\tau_1-\tau)}, & \dot{\theta}_s(\tau_1) &= (1 + \alpha_1)/\alpha_1, \\ \theta_1(\tau) &= \theta_1(\tau_1)\lambda^{y_1(\tau-\tau_1)}, & \theta_1(\tau_1) &= (1 - \alpha_1)^{x_1}, \\ \dot{\theta}_1(\tau) &= \dot{\theta}_1(\tau_1)\lambda^{y_1(\tau-\tau_1)}, & \dot{\theta}_1(\tau_1) &= \alpha_1^{x_1},\end{aligned}\tag{4.3.36}$$

with the corresponding formulae for the generalized scaling fields obtained by substituting in (4.3.32) or (4.3.33). Neglecting the zero fields, the formula

$$\psi_{\text{sing}}(\mathbf{k}; \Theta_1(\tau), \Theta_s(\tau)) = \lambda^{d'(\mathbf{k})(\tau-\tau_1)} \psi_{\text{sing}}(\mathbf{k}; \Theta_1(\tau_1), \Theta_s(\tau_1)) \tag{4.3.37}$$

extends the Kadanoff scaling hypothesis (4.3.7) for $\psi_{\text{sing}}(\mathbf{k})$. From (4.3.30),

$$\lambda^{\mathfrak{y}_1(\tau-\tau_1)} = \Theta_1(\tau)[\Theta_s(\tau)/\alpha_1]^{x_1} \tag{4.3.38}$$

and, substituting into (4.3.37), using the symmetry relationship between $\mathcal{C}$ and $\mathring{\mathcal{C}}$ and the notation of (4.3.12), gives

$$\psi_{\text{sing}}(\mathbf{k}; \Theta_1, \Theta_s) = \begin{cases} [\Theta_1]^{\mathring{\omega}_1(\mathbf{k})} [\Theta_s/\alpha_1]^{x_1 \mathring{\omega}_1(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; 1, \alpha_1), \\ [\Theta_1]^{\omega_1(\mathbf{k})} [(1 - \Theta_s)/\alpha_1]^{x_1 \omega_1(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; 1, \alpha_1). \end{cases} \tag{4.3.39}$$

We now consider the variation of this quantity along a line of constant Θ_2 . A constant of the motion of (4.3.32) is given by

$$(1 - \alpha_1)^{\mathfrak{y}_1} \alpha_1^{-\vartheta_1} = \Theta_1(1 - \Theta_s)^{\mathfrak{y}_1} \Theta_s^{-\vartheta_1}. \tag{4.3.40}$$

But, along the line of constant Θ_2 , α_1 must vary. To indicate this we replace α_1 with $\alpha(\Theta_1)$ giving

$$[1 - \alpha(\Theta_1)]^{\mathfrak{y}_1} \alpha(\Theta_1)^{-\vartheta_1} = \Theta_1[1 - \alpha(1)]^{\mathfrak{y}_1} \alpha(1)^{-\vartheta_1} \tag{4.3.41}$$

and

$$\frac{d \ln[\alpha(\Theta_1)]}{d \ln |\Theta_1|} = - \left\{ \vartheta_1 \left(\frac{\alpha(\Theta_1)}{1 - \alpha(\Theta_1)} \right) - \vartheta_1 \right\}^{-1}. \quad (4.3.42)$$

To measure the variation of $\psi_{\text{sing}}(\mathbf{k})$ we adopt the *effective exponent*

$$\Omega_1(\mathbf{k}; \Theta_1, \Theta_s) := \frac{\partial \ln \psi_{\text{sing}}(\mathbf{k})}{\partial \ln |\Theta_1|}. \quad (4.3.43)$$

introduced by Riedel and Wegner (1974). In the first line of (4.3.39), replacing α_1 with $\alpha(\Theta_1)$ and Θ_s with $\alpha(1)$ gives

$$\Omega_1(\mathbf{k}; \Theta_1, \alpha(1)) = \frac{\dot{\omega}_1(\mathbf{k}) \vartheta_1 [2\alpha(\Theta_1) - 1]}{\vartheta_1 \alpha(\Theta_1) - \vartheta_1 [1 - \alpha(\Theta_1)]} + Q(\mathbf{k}; \Theta_1), \quad (4.3.44)$$

where

$$Q(\mathbf{k}; \Theta_1) := \left\{ \frac{\vartheta_1}{1 - \alpha(\Theta_1)} - \frac{\vartheta_1}{\alpha(\Theta_1)} \right\}^{-1} \frac{d \ln[\psi_{\text{sing}}(\mathbf{k}; 1, \alpha(\Theta_1))]}{d \alpha(\Theta_1)}. \quad (4.3.45)$$

In the two limits $\alpha(\theta_1) \rightarrow 0$ and $\alpha(\theta_1) \rightarrow 1$ the limits of the first term on the right of (4.3.45) are $\omega_1(\mathbf{k})$ and $\dot{\omega}_1(\mathbf{k})$, respectively. In each case $Q(\Theta_1) \rightarrow 0$ if, as we assume, the second term on the right of (4.3.45) has a finite, or more weakly singular limit, than the first. From (4.3.41), $\Theta_1 \rightarrow 0$ gives $\alpha(\Theta_1) \rightarrow 1$, with $\Omega_1(\mathbf{k}; \Theta_1, \alpha(1)) \rightarrow \dot{\omega}_1(\mathbf{k})$, except when $\Theta_s = \alpha(\Theta_1) = \alpha_1 = 0$, an invariant line of (4.3.27), when $\Omega_1(\mathbf{k}; \Theta_1, \alpha(1)) \rightarrow \omega_1(\mathbf{k})$. That the whole of the interval $\Theta_s \in (0, 1]$ is attractive to $\Theta_s = 1$ is, of course, a consequence of the sign of y_s . However, it is of interest to investigate how close to $\Theta_1 = 0$ one needs to be before the attractive effects of $\Theta_s = 1$ become evident. An indicator of this is the extent of *crossover region*¹⁶ in the space $0 \leq \Theta_1 \leq 1$, $0 \leq \Theta_s \leq 1$, defined by Riedel and Wegner (1974) as the condition¹⁷

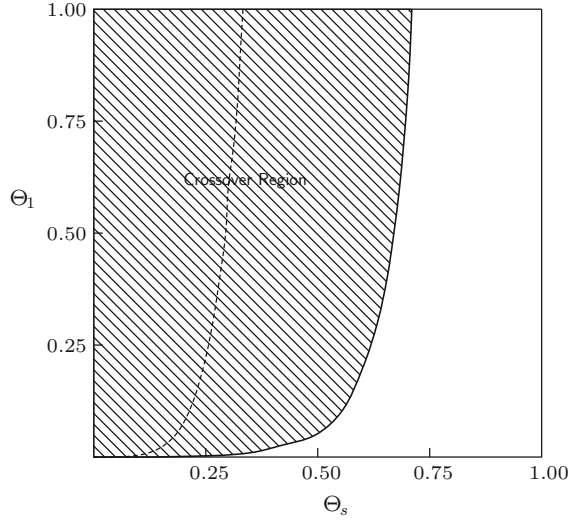
$$|\Omega_1(\mathbf{k}; \Theta_1, \alpha(1)) - \dot{\omega}_1(\mathbf{k})| = \left| \frac{\dot{\omega}(\mathbf{k})[1 - \alpha(\Theta_1)][\vartheta_1 - \vartheta_1]}{\vartheta_1 \alpha(\Theta_1) - \vartheta_1 [1 - \alpha(\Theta_1)]} \right| > \delta, \quad (4.3.46)$$

for some δ . An example of a crossover region is shown in Fig. 4.6.

¹⁶ This is different, but related, type of ‘crossover’ from that discussed in Sect. 4.3.1. There the change is between different types of asymptotic behaviour along different paths to a fixed point. Here it is a question of competition between two fixed points.

¹⁷ To compute this formula the second term in (4.3.44) has been neglected.

Fig. 4.6 The crossover region for $d = 2$, $k = 0$, $y_1 = 0.5$, $\hat{y}_1 = 1.0$, $y_s = 0.25$ and $\delta = 0.5$. The crossover curve with $y_1 = 0.9$ is shown as a broken line



4.3.4 The Nightingale–'T Hooft Scaling Hypothesis

Here we begin with the whole of the free-energy density ϕ and not just its singular part, replacing the Kadanoff scaling hypothesis (4.3.5) with the *Nightingale–'T Hooft scaling hypothesis* (Nightingale and 'T Hooft 1974) which supposes that, in a neighbourhood of the transition region $\mathcal{C}$

$$\phi(\lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n) = \lambda^d \phi(\theta_1, \dots, \theta_n) - \phi_\lambda(\theta_1, \dots, \theta_n), \quad (4.3.47)$$

for all real $\lambda > 1$. Iterating $\ell - 1$ times

$$\phi(\theta_1, \dots, \theta_n) = \lambda^{-\ell d} \phi(\lambda^{\ell y_1}\theta_1, \dots, \lambda^{\ell y_n}\theta_n) + \lambda^{-d} \phi_\lambda^{(\ell)}(\theta_1, \dots, \theta_n), \quad (4.3.48)$$

where

$$\phi_\lambda^{(\ell)}(\theta_1, \dots, \theta_n) := \sum_{r=0}^{\ell-1} \lambda^{-rd} \phi_\lambda(\lambda^{ry_1}\theta_1, \dots, \lambda^{ry_n}\theta_n) \quad (4.3.49)$$

and, of course, $\phi_\lambda^{(1)} = \phi_\lambda$.

We pause to note that (4.3.48) is the result of the repeated application of the transformation (4.3.3), resulting, for an initially chosen set of values, in a trajectory in the space of scaling fields. Of course, a similar iteration could be applied to the Kadanoff relationship (4.3.5). In the case of a relevant scaling field $\theta_j \lambda^{\ell y_j} \rightarrow \infty$ and for an irrelevant scaling field $\theta_j \lambda^{\ell y_j} \rightarrow 0$. The problem with the former is clear since the validity of the scaling hypotheses have been assumed only within some

neighbourhood of $\mathcal{C}$; this is discussed in Sect. 4.3.5. Problems in the latter category arise for so-called ‘*dangerous*’ *irrelevant exponents* which are discussed in Sect. 4.8.

Since $\phi_\lambda(\theta_1, \dots, \theta_n)$ is a smooth function of the scaling fields, it has a power series expansion of the form

$$\phi_\lambda(\theta_1, \dots, \theta_n) = \sum_{\{\mathbf{q}\}} B_\lambda(\mathbf{q}) \prod_{j=1}^n \frac{\theta_j^{q_j}}{q_j!}, \quad (4.3.50)$$

where the summation is over $q_j = 0, 1, 2, \dots$, for all $j = 1, \dots, n$, the set $(q_1, \dots, q_n)$ being represented by the vector $\mathbf{q}$. Substituting into (4.3.49) gives

$$\phi_\lambda^{(\ell)}(\theta_1, \dots, \theta_n) = \sum_{\{\mathbf{q}\}} B_\lambda(\mathbf{q}) \Lambda_\lambda^{(\ell)}(d'(\mathbf{q})) \prod_{j=1}^n \frac{\theta_j^{q_j}}{q_j!}, \quad (4.3.51)$$

where

$$\Lambda_\lambda^{(\ell)}(x) := \sum_{r=0}^{\ell-1} \lambda^{-rx} = \begin{cases} \ell, & x = 0, \\ \frac{1 - \lambda^{-x\ell}}{1 - \lambda^{-x}}, & x \neq 0, \end{cases} \quad (4.3.52)$$

and $d'(\mathbf{q})$ is as defined in (4.3.8).

For the function $\psi(\mathbf{k})$ obtained by differentiation ϕ , k_j times with respect to θ_j , for $j = 1, \dots, n$,

$$\begin{aligned} \psi(\mathbf{k}; \theta_1, \dots, \theta_n) &= \lambda^{-\ell d'(\mathbf{k})} \psi(\mathbf{k}; \lambda^{\ell y_1} \theta_1, \dots, \lambda^{\ell y_n} \theta_n) \\ &\quad + \lambda^{-d} \psi_\lambda^{(\ell)}(\mathbf{k}; \theta_1, \dots, \theta_n), \end{aligned} \quad (4.3.53)$$

where

$$\psi_\lambda^{(\ell)}(\mathbf{k}; \theta_1, \dots, \theta_n) = \sum'_{\{\mathbf{q}\}} B_\lambda(\mathbf{q}) \Lambda_\lambda^{(\ell)}(d'(\mathbf{q})) \prod_{j=1}^n \frac{\theta_j^{q_j - k_j}}{(q_j - k_j)!}, \quad (4.3.54)$$

with the summations in each case now restricted to $q_j \geq k_j$, for $j = 1, \dots, n$.

4.3.5 Constraints on Scaling

Having developed the formalism of scaling it is necessary to consider the limits of its applicability. As indicated above these arise when we consider the repeated application of the transformation (4.3.3). This will result in the arguments in the first

term on the right-hand side of (4.3.48) tending to zero and infinity for irrelevant and relevant scaling fields, respectively. The former will cause a problem if $\phi(\theta_1, \dots, \theta_n)$ is singular when an irrelevant scaling field is zero. For the present discussion we shall assume that this possibility, which is discussed in Sect. 4.8, does not arise and consider only the need to remain within the neighbourhood of validity of the scaling hypothesis. For any λ this imposes an upper bound on the number of iterations of (4.3.3). Suppose that the neighbourhood of validity of the scaling hypothesis (4.3.47) is given by

$$\mu(\theta_1, \dots, \theta_s) \leq R, \quad (4.3.55)$$

where

$$\mu(\theta_1, \dots, \theta_s) := [|\theta_1|^{2/y_1} + \dots + |\theta_s|^{2/y_s}]^{\frac{1}{2}} \quad (4.3.56)$$

is the metric which measures distances from $\mathcal{C}$ and which scales according to the formula

$$\mu(\lambda^{y_1}\theta_1, \dots, \lambda^{y_s}\theta_s) = \lambda\mu(\theta_1, \dots, \theta_s). \quad (4.3.57)$$

Then the iterative process remains in the scaling neighbourhood if

$$\lambda^\ell \leq \frac{R}{\mu(\theta_1, \dots, \theta_s)}. \quad (4.3.58)$$

Given a particular set of values for the relevant scaling fields, suppose that (4.3.57) is iterated until

$$\begin{aligned} \ell = L(\lambda; \theta_1, \theta_2, \dots, \theta_s) &:= \frac{1}{\ln(\lambda)} \ln \left[\frac{R}{\mu(\theta_1, \dots, \theta_s)} \right] \\ &= \frac{1}{\ln(\lambda)} \ln \left[\frac{\mu(\tilde{\theta}_1, \dots, \tilde{\theta}_s)}{\mu(\theta_1, \dots, \theta_s)} \right], \end{aligned} \quad (4.3.59)$$

when a point on the boundary is achieved with (4.3.58) an equality¹⁸ and the relevant scaling fields with values $\tilde{\theta}_j = \lambda^{Ly_j}\theta_j$, $j = 1, 2, \dots, s$.

In the interests of brevity it is convenient to remember that $\phi(\theta_1, \dots, \theta_n) = \psi(\mathbf{0}; \theta_1, \dots, \theta_n)$, so that the discussion can concentrate simply on the function $\psi(\mathbf{k}; \theta_1, \dots, \theta_n)$. Substituting into (4.3.53) for $\ell = L$ gives

¹⁸ The assumption here is that λ is such that equality in (4.3.58) can be achieved for an integer ℓ .

$$\begin{aligned} \psi(\mathbf{k}; \theta_1, \dots, \theta_n) &= \left[\frac{\mu(\theta_1, \dots, \theta_s)}{\mu(\tilde{\theta}_1, \dots, \tilde{\theta}_s)} \right]^{d'(\mathbf{k})} \psi(\mathbf{k}; \tilde{\theta}_1, \dots, \tilde{\theta}_n) \\ &\quad + \lambda^{-d} \psi_\lambda^{(L)}(\mathbf{k}; \theta_1, \dots, \theta_n). \end{aligned} \quad (4.3.60)$$

The first term on the right-hand side of this equation is a generalized homogeneous functions of the relevant scaling fields and $\psi(\mathbf{k}; \theta_1, \dots, \theta_n)$ must be independent of λ . To achieve this a specific form must be imposed on the coefficients $B_\lambda(\mathbf{q})$. To see what this must be we choose some relevant scaling field θ_m when, from (4.3.13), (4.3.56) and (4.3.59),

$$L(\lambda; \theta_1, \theta_2, \dots, \theta_s) = -\frac{\ln |\theta_m|}{y_m \ln(\lambda)} + \frac{1}{\ln(\lambda)} \ln \left[\frac{\mu(\tilde{\theta}_1, \dots, \tilde{\theta}_s)}{\mu(g_1^{(m)}, \dots, g_s^{(m)})} \right]. \quad (4.3.61)$$

Now consider the path $\mathcal{P}$ towards $\mathcal{C}$, described in Sect. 4.3.1, where the limiting values of the gap fields are all finite. From (4.3.52) and (4.3.61),

$$\Lambda_\lambda^{(L)}(x) = \begin{cases} -A_m^x |\theta_m|^{x/y_m} / (1 - \lambda^{-x}), & \text{if } x < 0, \\ -\ln |\theta_m| / [y_m \ln(\lambda)], & \text{if } x = 0, \\ 1/(1 - \lambda^{-x}), & \text{if } x > 0, \end{cases} \quad (4.3.62)$$

where

$$A_m := \frac{\mu(g_1^{(m)}, \dots, g_s^{(m)})}{\mu(\tilde{\theta}_1, \dots, \tilde{\theta}_s)}, \quad (4.3.63)$$

which is non-zero since $g_m^{(m)} = \pm 1$, according to the direction of approach to $\mathcal{C}$. Then substituting into (4.3.51) and using (4.3.62) it follows that $\lambda^{-d} \psi_\lambda^{(L)}(\mathbf{k}; \theta_1, \dots, \theta_n)$ is independent of λ if

$$B_\lambda(\mathbf{q}) := \bar{B}(\mathbf{q}) \times \begin{cases} \lambda^d (1 - \lambda^{-d'(\mathbf{q})}), & \text{if } d'(\mathbf{q}) \neq 0, \\ \lambda^d \ln(\lambda), & \text{if } d'(\mathbf{q}) = 0, \end{cases} \quad (4.3.64)$$

where the coefficients $\bar{B}(\mathbf{q})$ are independent of λ . Substituting into (4.3.4) for $\ell = 1$ gives

$$\begin{aligned} \lambda^{-d} \psi_\lambda(\mathbf{k}; \theta_1, \dots, \theta_n) &= \bar{\psi}(\mathbf{k}; \theta_1, \dots, \theta_n) - \lambda^{-d'(\mathbf{k})} \bar{\psi}(\mathbf{k}; \lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) \\ &\quad + \ln(\lambda) \xi(\mathbf{k}; \theta_1, \dots, \theta_n), \end{aligned} \quad (4.3.65)$$

where

$$\bar{\psi}(\mathbf{k}; \theta_1, \dots, \theta_n) := \sum'_{\{\mathbf{q}\}} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{\theta_j^{q_j - k_j}}{(q_j - k_j)!}, \quad (4.3.66)$$

$$\xi(\mathbf{k}; \theta_1, \dots, \theta_n) := \sum'_{\{\mathbf{q}\}} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{\theta_j^{q_j - k_j}}{(q_j - k_j)!}, \quad (4.3.67)$$

are smooth functions of the scaling fields.¹⁹ Substituting into (4.3.3), again with $\ell = 1$, and identifying

$$\psi_{\text{smth}}(\mathbf{k}; \theta_1, \dots, \theta_n) := \bar{\psi}(\mathbf{k}; \theta_1, \dots, \theta_n), \quad (4.3.68)$$

shows that

$$\psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n) := \psi(\mathbf{k}; \theta_1, \dots, \theta_n) - \bar{\psi}(\mathbf{k}; \theta_1, \dots, \theta_n), \quad (4.3.69)$$

for which

$$\begin{aligned} \psi_{\text{sing}}(\mathbf{k}; \lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) &= \lambda^{d'(\mathbf{k})} [\psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n) \\ &\quad - \ln(\lambda) \xi(\mathbf{k}; \theta_1, \dots, \theta_n)], \end{aligned} \quad (4.3.70)$$

will contain all the singular character of $\psi(\mathbf{k})$. The division of $\psi(\mathbf{k})$ into smooth and singular parts is represented by (4.3.68) and (4.3.69). The Kadanoff scaling (4.3.7) is replaced by (4.3.70) and is satisfied in its original form only when

$$\xi(\mathbf{k}; \theta_1, \dots, \theta_n) \equiv 0. \quad (4.3.71)$$

Setting λ according to (4.3.9), from (4.3.12)–(4.3.14), (4.3.69) and (4.3.70)

$$\begin{aligned} \psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n) &= |\theta_m|^{\omega_m(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; g_1^{(m)}, \dots, g_n^{(m)}) \\ &\quad - |\theta_m|^{\omega_m(\mathbf{k})} \ln |\theta_m| y_m^{-1} \sum'_{\{\mathbf{q}\}} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!}, \end{aligned} \quad (4.3.72)$$

replaces (4.3.11), with, from (4.3.66) and (4.3.69),

¹⁹ It should be noted that the subset of terms included in (4.3.67) can also be included in (4.3.66) since they cancel out in (4.3.65).

$$\begin{aligned}
\psi(\mathbf{k}; \theta_1, \dots, \theta_n) &= |\theta_m|^{\omega_m(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; g_1^{(m)}, \dots, g_n^{(m)}) \\
&+ |\theta_m|^{\omega_m(\mathbf{k})} \sum_{\{q\}}' \bar{B}(\mathbf{q}) |\theta_m|^{-\omega_m(\mathbf{q})} \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!} \\
&- |\theta_m|^{\omega_m(\mathbf{k})} \ln |\theta_m| y_m^{-1} \sum_{\{q\}}'^{[d'(\mathbf{q})=0]'} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!}.
\end{aligned} \tag{4.3.73}$$

An explicit asymptotic form for $\psi(\mathbf{k}; \theta_1, \dots, \theta_n)$ as $\mathcal{C}$ is approached along the path $\mathcal{P}$ can be derived from (4.3.60). Using (4.3.13)–(4.3.15) and (4.3.18),

$$\left[\frac{\mu(\theta_1, \dots, \theta_s)}{\mu(\tilde{\theta}_1, \dots, \tilde{\theta}_s)} \right]^{d'(\mathbf{k})} \psi(\mathbf{k}; \tilde{\theta}_1, \dots, \tilde{\theta}_n) \simeq |\theta_m|^{\omega_m(\mathbf{k})} A_m^{d'(\mathbf{k})} \psi(\mathbf{k}; \tilde{\theta}_1, \dots, \tilde{\theta}_n). \tag{4.3.74}$$

This term reproduces the power-law behaviour obtained in (4.3.20). The other contribution is, from (4.3.54) and (4.3.62)–(4.3.64),

$$\begin{aligned}
\lambda^{-d} \psi_L^{(\lambda)}(\mathbf{k}; \theta_1, \dots, \theta_n) &= -|\theta_m|^{\omega_m(\mathbf{k})} \sum_{\{q\}}'^{[d'(\mathbf{q})>0]'} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!} \\
&+ |\theta_m|^{\omega_m(\mathbf{k})} \ln |\theta_m| y_m^{-1} \sum_{\{q\}}'^{[d'(\mathbf{q})=0]'} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!} \\
&- |\theta_m|^{\omega_m(\mathbf{k})} \sum_{\{q\}}'^{[d'(\mathbf{q})<0]'} |\theta_m|^{-\omega_m(\mathbf{q})} \bar{B}(\mathbf{q}) \prod_{j=1}^n \frac{[g_j^{(m)}]^{q_j - k_j}}{(q_j - k_j)!}.
\end{aligned} \tag{4.3.75}$$

A term in these sums is zero if $q_j \neq k_j$, for any j for which $g_j^{(m)} = 0$. In particular to have a non-zero term q_j must be equal to k_j for all $j = s + 1, \dots, n$, since $g_j^{(m)} \rightarrow \varpi_j^{(m)} = 0$ for all irrelevant fields. The residual terms in the first sum give contributions of the same power-law type as (4.3.74) and can therefore, in general, be ignored, except in the case when the amplitudes in (4.3.74) are not non-zero and finite, when they may provide the only terms with this asymptotic behaviour. The exponent of $|\theta_m|$ for a term in the third sum is positive and less than $\omega_m(\mathbf{k})$, so this contribution will be negligible when $\psi(\mathbf{k})$ has a power-law singularity as $\mathcal{C}$ is approached along $\mathcal{P}$. The interesting and important terms are, as in (4.3.73), in the second sum. These will occur only when

$$d = \sum_{j=1}^n q_j y_j \tag{4.3.76}$$

has a solution, or solutions, for some set of non-negative integers $q_1, \dots, q_n$. If we choose $k_1, \dots, k_n$ to be such a solution, then $\psi(\mathbf{k})$ is such that (4.3.25) is satisfied and

$$\psi(\mathbf{k}; \theta_1, \dots, \theta_n) \simeq \bar{B}(\mathbf{k}) y_m^{-1} \ln |\theta_m|. \quad (4.3.77)$$

This particular derivative of ϕ exhibits a logarithmic singularity as $\mathcal{C}$ is approached along $\mathcal{P}$, as long as $\bar{B}(\mathbf{k}) \neq 0$.²⁰ The occurrence of logarithmic divergences in thermodynamic response functions are the first indication that the scaling ideas of Sect. 4.3 should be treated with some caution, and need modification in particular circumstances. Further modifications, arising from zero or singular amplitude factors, are considered in Sect. 4.8.

By construction

$$\mathcal{K}_0(\theta_1, \dots, \theta_n) := \lambda^{-d} \phi_L^{(\lambda)}(\theta_1, \dots, \theta_n) \quad (4.3.78)$$

is independent of λ except through the definition (4.3.59) of L . Substituting into (4.3.48)

$$\lambda^{-Ld} \phi(\lambda^{\ell_{y_1}} \theta_1, \dots, \lambda^{L_{y_n}} \theta_n) = \phi(\theta_1, \dots, \theta_n) - \mathcal{K}_0(\theta_1, \dots, \theta_n). \quad (4.3.79)$$

With

$$\lambda^L = \frac{R}{\mu(\theta_1, \dots, \theta_s)} > 1 \quad (4.3.80)$$

replaced by λ this is the linearized version of the scaling formula for the free-energy density derived from the renormalization group (Sect. 15.4 and Niemeijer and van Leeuwen 1976, p. 430).

4.3.6 Scaling Operators and Dimensions

For the sake of brevity we adopt the notation of Sect. 2.5, where ζ_j , $j = 1, 2, \dots, n$, represents the complete set of internal and external couplings. Suppose that the formulae (4.3.1) are invertible near $\mathcal{C}$ to give

$$\zeta_j = \zeta_j(\theta_1, \theta_2, \dots, \theta_n), \quad j = 1, 2, \dots, n, \quad (4.3.81)$$

²⁰ From a renormalization group point of view this divergence arises from the self-energy of the cells, whereas the more usual power-law singularities arise from the interaction between cells.

with

$$\zeta_{jc}(\theta_{s+1}, \dots, \theta_n) := \zeta_j(0, \dots, 0, \theta_{s+1}, \dots, \theta_n), \quad (4.3.82)$$

$$\zeta_j^* := \zeta_{jc}(0, \dots, 0). \quad (4.3.83)$$

Close to $\mathcal{C}$, to linear terms,

$$\zeta_j(\theta_1, \theta_2, \dots, \theta_n) = \zeta_{jc}(\theta_{s+1}, \dots, \theta_n) + \sum_{i=1}^s \theta_i \zeta_{jc}^{(i)}(\theta_{s+1}, \dots, \theta_n), \quad (4.3.84)$$

where

$$\zeta_{jc}^{(i)}(\theta_{s+1}, \dots, \theta_n) := \left(\frac{\partial \zeta_j}{\partial \theta_i} \right)_{\theta_1=\dots=\theta_s=0}. \quad (4.3.85)$$

Substituting from (4.3.84) into the Hamiltonian (2.5.2) gives, to linear terms in the relevant scaling fields,

$$\widehat{H}(N, \zeta_0, \theta_1, \dots, \theta_n; \boldsymbol{\sigma}) = \widehat{H}(N, \zeta_0, \zeta_{jc}; \boldsymbol{\sigma}) + \sum_{i=1}^s \Phi_i(\boldsymbol{\sigma}) \theta_i \quad (4.3.86)$$

where, from (2.5.1), the coefficients

$$\Phi_i(\boldsymbol{\sigma}) := \sum_{j=1}^n \widehat{C}_j(\boldsymbol{\sigma}) \zeta_{jc}^{(i)}(\theta_{s+1}, \dots, \theta_n), \quad (4.3.87)$$

are functions of the irrelevant scaling fields. Close to the fixed point these functions can be replaced by $\zeta_j^{(i)*} := \zeta_{jc}^{(i)}(0, \dots, 0)$ and the Hamiltonian in $\mathcal{C}$ expanded to give

$$\widehat{H}(N, \zeta_0, \zeta_{jc}; \boldsymbol{\sigma}) = \widehat{H}(N, \zeta_0, 0, \dots, 0, \theta_{s+1}, \dots, \theta_n; \boldsymbol{\sigma}) = N\zeta_0 + \sum_{i=s+1}^n \Phi_i(\boldsymbol{\sigma}) \theta_i. \quad (4.3.88)$$

Defining

$$\widehat{H}_{\text{sing}}(N, \theta_1, \dots, \theta_n; \boldsymbol{\sigma}) := \sum_{i=1}^n \Phi_i(\boldsymbol{\sigma}) \theta_i \quad (4.3.89)$$

yields a potential ϕ satisfying condition (2.5.3) with $\phi_{\text{smth}} = \zeta_0$. From (4.3.88) and (4.3.89),

$$\langle \hat{H} \rangle = N\zeta_0 + \sum_{i=1}^n \langle \Phi_i(\sigma) \rangle \theta_i = N\zeta_0 + \sum_{i=1}^n \theta_i \frac{\partial \phi_{\text{sing}}}{\partial \theta_i}. \quad (4.3.90)$$

Denoting by $\mathbf{i}_i$ the n -dimensional unit vector with one in the i th place, from (4.3.11)–(4.3.14),

$$\sum_{i=1}^n \theta_i \frac{\partial \phi_{\text{sing}}}{\partial \theta_i} = |\theta_m|^{d/y_m} \sum_{i=1}^n g_i^{(m)} \psi_{\text{sing}}(\mathbf{i}_i; g_1^{(m)}, \dots, g_n^{(m)}). \quad (4.3.91)$$

By approaching $\mathcal{C}$ along the path $\mathcal{P}$ where the limit (4.3.18) applies it can be seen that²¹

$$\phi_{\text{smth}}(\theta_{s+1}, \dots, \theta_n) := \zeta_0(\theta_{s+1}, \dots, \zeta_n) = \langle \hat{H} \rangle_c / N \quad (4.3.92)$$

with

$$\phi_{\text{sing}}(\theta_1, \dots, \theta_n) := \phi(0, \theta_1, \dots, \theta_n), \quad (4.3.93)$$

$$\phi(\zeta_0, \theta_1, \dots, \theta_n) = \phi_{\text{smth}}(\theta_1, \dots, \theta_n) + \phi_{\text{sing}}(\theta_1, \dots, \theta_n). \quad (4.3.94)$$

Substituting from (2.5.1) into (4.3.59) and then into (4.3.89) gives

$$\hat{H}_{\text{sing}}(N, \theta_1, \dots, \theta_n; \sigma, \{\mathbf{r}\}) := \sum_{i=1}^n \sum_{\{\mathbf{r}\}} \phi_i(\mathbf{r}; \sigma) \theta_i \quad (4.3.95)$$

where

$$\phi_i(\mathbf{r}; \sigma) := \sum_{j=1}^n \mathbf{c}_j(\mathbf{r}; \sigma) \zeta_j^{(i)\star} \quad (4.3.96)$$

are *scaling operators*. In (4.3.95) the explicit dependence of the Hamiltonian on the set of lattice sites has been exposed. The scaling of the lattice vector $\mathbf{r}$ is given by (4.2.5). Suppose that the scaling operators transform according to

$$\phi(\mathbf{r}; \sigma) \rightarrow \tilde{\phi}(\mathbf{r}; \sigma) := \phi(\tilde{\mathbf{r}}; \sigma) = \lambda^{x_i} \phi(\mathbf{r}; \sigma). \quad (4.3.97)$$

To determine the scaling behaviour of $\hat{H}_{\text{sing}}$, needed to give the scaling formula for ϕ_{sing} , it must be remembered that it is more directly related to $\Phi_{\text{sing}} = N\phi_{\text{sing}}$, and when the sum over lattice sites is transformed by the Jacobian λ^{-d} the Hamiltonian will be unchanged. Thus

²¹ It should be noted that, if (4.3.11) is replaced by (4.3.72) this result still applies since the scaling exponents satisfy (4.3.22).

$$\begin{aligned}
\widehat{H}_{\text{sing}}(N, \theta_1, \dots, \theta_n; \boldsymbol{\sigma}, \{\mathbf{r}\}) &\rightarrow \widehat{H}_{\text{sing}}(N, \lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n; \boldsymbol{\sigma}, \{\mathbf{r}/\lambda\}) \\
&= \sum_{i=1}^n \sum_{\{\mathbf{r}\}} \lambda^{x_i + y_i - d} \phi_i(\mathbf{r}; \boldsymbol{\sigma}) \theta_i = \widehat{H}_{\text{sing}}(N, \theta_1, \dots, \theta_n; \boldsymbol{\sigma}, \{\mathbf{r}\}), \quad (4.3.98)
\end{aligned}$$

giving the *scaling dimensions* of the scaling operators

$$x_j := d - y_j. \quad (4.3.99)$$

4.3.7 Correlation Functions

The development of Sect. 2.5 can be carried out for scaling fields and scaling operators just as well as for couplings and state operators. Thus for an arbitrary set of τ scaling operators²²

$$\langle \phi_1(\lambda^{-1} \mathbf{r}_1) \cdots \phi_\tau(\lambda^{-1} \mathbf{r}_\tau) \rangle = \lambda^z \langle \phi_1(\mathbf{r}_1) \cdots \phi_\tau(\mathbf{r}_\tau) \rangle, \quad (4.3.100)$$

$$\Gamma_\tau(\phi_1(\lambda^{-1} \mathbf{r}_1), \dots, \phi_\tau(\lambda^{-1} \mathbf{r}_\tau)) = \lambda^z \Gamma_\tau(\phi_1(\mathbf{r}_1), \dots, \phi_\tau(\mathbf{r}_\tau)), \quad (4.3.101)$$

$$\Gamma_\tau^*(\phi_1, \dots, \phi_\tau; \lambda \mathbf{k}_1, \dots, \lambda \mathbf{k}_\tau) = \lambda^{z - \tau d} \Gamma_\tau^*(\phi_1, \dots, \phi_\tau; \mathbf{k}_1, \dots, \mathbf{k}_\tau), \quad (4.3.102)$$

where

$$z = \sum_{j=1}^{\tau} x_j. \quad (4.3.103)$$

It is straightforward to check that these scaling relationships agree with the scaling for ϕ_{sing} through the forms

$$\sum_{\{\mathbf{r}_1, \dots, \mathbf{r}_\tau\}} \Gamma_\tau(\phi_1(\mathbf{r}_1), \dots, \phi_\tau(\mathbf{r}_\tau)) = -N \frac{\partial^\tau \phi_{\text{sing}}}{\partial \theta_1 \cdots \partial \theta_\tau}, \quad (4.3.104)$$

$$\Gamma_\tau^*(\phi_1, \dots, \phi_\tau; \mathbf{0}, \dots, \mathbf{0}) = -\frac{1}{N} \frac{\partial^\tau \phi_{\text{sing}}}{\partial \theta_1 \cdots \partial \theta_\tau} \quad (4.3.105)$$

of (2.5.14) and (2.5.22). When the system has translational symmetry the two-point and three-point correlation functions satisfy

²² For easy of presentation, we have omitted the dependence on the scaling fields.

$$\Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1/\lambda)) = \lambda^{x_1+x_2} \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1)), \quad (4.3.106)$$

$$\Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1/\lambda), \phi_3(\bar{\mathbf{r}}_2/\lambda)) = \lambda^{x_1+x_2+x_3} \Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1), \phi_3(\bar{\mathbf{r}}_2)), \quad (4.3.107)$$

in terms of the relative lattice vectors $\bar{\mathbf{r}}_1 := \mathbf{r}_2 - \mathbf{r}_1$ and $\bar{\mathbf{r}}_2 := \mathbf{r}_3 - \mathbf{r}_1$. Making the substitution $\lambda = r := |\bar{\mathbf{r}}_1|$ in (4.3.106) and (4.3.107) gives

$$\Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1)) = \frac{1}{r^{x_1+x_2}} \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1/r)), \quad (4.3.108)$$

$$\Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1), \phi_3(\bar{\mathbf{r}}_2)) = \frac{1}{r^{x_1+x_2+x_3}} \Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1/r), \phi_3(\bar{\mathbf{r}}_2/r)). \quad (4.3.109)$$

On the transition surface the asymptotic dependence of Γ_2 on r , in the limit of large r , is given by setting $\theta_1 = \dots = \theta_s = 0$. As long as the amplitude factor on the right-hand side of (4.3.108) is non-zero and finite we have

$$\Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1)) \simeq \frac{G_2(\phi_1, \phi_2)}{r^{x_1+x_2}}, \quad (4.3.110)$$

where $G_2(\phi_1, \phi_2)$ is a constant dependent only on the choice of scaling operators. A similar expression

$$\bar{\Gamma}_2^*(\phi_1, \phi_2; \bar{\mathbf{k}}_1) \simeq \frac{G_2^*(\phi_1, \phi_2)}{k^{y_1+y_2-d}} \quad (4.3.111)$$

can be obtained for the Fourier transform taken with respect to the relative lattice vectors and for the total pair correlation function

$$\langle \phi_1(\mathbf{0}) \phi_2(\bar{\mathbf{r}}_1) \rangle \simeq \frac{C_2(\phi_1, \phi_2)}{r^{x_1+x_2}}. \quad (4.3.112)$$

In this context it is a common terminology (Kadanoff 1969b) to say that ϕ_j *scales as* $1/r^{x_j}$. This simply means that, in the expectation value each scaling operator, ϕ_j contributes a scaling factor $1/r^{x_j}$.

4.3.8 Variable Scaling Exponents

Throughout the discussion in the previous sections we have assumed that the scaling exponents $y_1, \dots, y_n$ are constants. There are, however, a number of models, the

most well-known being the eight-vertex model, in which the critical exponents are functions of one of the couplings (Sects. 12.3.8 and 12.3.11). The way that this behaviour can be reconciled with the ideas of scaling has been explored by Kadanoff and Wegner (1971), using the *reduction hypothesis* of Kadanoff (1969a, b). We shall now present a simplified version of their argument.

If $|\bar{\mathbf{r}}_1| \sim |\bar{\mathbf{r}}_2|$, the length-scale dependence of Γ_3 on the right-hand side of (4.3.109) can be neglected, giving

$$\Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1), \phi_3(\bar{\mathbf{r}}_2)) \sim \frac{1}{r^{x_3}} \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1)). \quad (4.3.113)$$

From the formulae equivalent to (2.5.13) and (2.5.14) for scaling fields and operators

$$\frac{\partial \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1))}{\partial \theta_3} = \sum_{\{\bar{\mathbf{r}}_2\}} \Gamma_3(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1), \phi_3(\bar{\mathbf{r}}_2)), \quad (4.3.114)$$

and using (18.1.35) to replace the summation by an integral over $[0, r]$ we have

$$\frac{\partial \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1))}{\partial \theta_3} \sim \begin{cases} \frac{1}{r^{x_1+x_2-y_3}}, & \text{if } y_3 \neq 0, \\ \frac{\ln(r)}{r^{x_1+x_2}}, & \text{if } y_3 = 0. \end{cases} \quad (4.3.115)$$

The case for $y_3 \neq 0$ is easily understood on differentiating (4.3.108) with respect to θ_3 . On the right-hand side θ_3 appears in the combination $r^{y_3}\theta_3$ and the leading term in the derivative of $\Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1/r))$ contributes a factor r^{y_3} . When $y_3 = 0$, ϕ_3 scales as $1/r^d$. It would appear that in this case differentiation of (4.3.108) leaves the scaling dependence on r unchanged. Suppose however that y_1 and y_2 are functions of θ_3 . Then

$$\frac{\partial \Gamma_2(\phi_1(\mathbf{0}), \phi_2(\bar{\mathbf{r}}_1))}{\partial \theta_3} \sim \frac{\ln(r)}{r^{x_1+x_2}} \left(\frac{\partial y_1}{\partial \theta_3} + \frac{\partial y_2}{\partial \theta_3} \right) \quad (4.3.116)$$

and the required logarithmic dependence is obtained. A scaling field with zero exponent is marginal (Sect. 4.3) and in renormalization group analysis is often associated with a surface or line of fixed points. We have now shown that, if such a field is present, it is not inconsistent with scaling theory that the other scaling exponents should be functions of that field. A marginal scaling field or coupling represents an ‘underlying’ parameter of the system. Scaling theory can proceed without taking it into account. Once the phase diagram is understood the additional dimension represented by a marginal parameter simply increases the dimension of the phase space, without affecting the type of phenomena observed. The application of this analysis to the eight-vertex model is considered in Sect. 12.3.11.

4.3.9 Densities and Response Functions

Before considering the application of scaling to some particular cases it is useful to relate the general ideas of Sect. 4.3 to the scaling properties of more familiar thermodynamic functions. The differential form (4.1.3) contains two types of densities: ρ_i , $i = 1, \dots, n_f - 1$ are associated with external fields and are quantities like magnetization density or particle density; u_ℓ , $\ell = 1, \dots, n_e$ are the energy-densities associated with the types of inter-particle interaction. The linear sum in (2.3.13) gives the total dimensionless internal energy density v , and from (4.1.3)

$$u_\ell = -\frac{\partial \phi}{\partial K_\ell}, \quad \rho_i = -\frac{\partial \phi}{\partial L_i}. \quad (4.3.117)$$

Let

$$\rho_{ic} := -\frac{\partial \phi_{\text{smth}}}{\partial L_i}, \quad u_{\ell c} := -\frac{\partial \phi_{\text{smth}}}{\partial K_\ell}. \quad (4.3.118)$$

Then

$$\begin{aligned} \rho_i - \rho_{ic} &= -\frac{\partial \phi_{\text{sing}}}{\partial L_i} = -\sum_{j=1}^n \frac{\partial \phi_{\text{sing}}}{\partial \theta_j} \frac{\partial \theta_j}{\partial L_i}, \\ u_\ell - u_{\ell c} &= -\frac{\partial \phi_{\text{sing}}}{\partial K_\ell} = -\sum_{j=1}^n \frac{\partial \phi_{\text{sing}}}{\partial \theta_j} \frac{\partial \theta_j}{\partial K_\ell}. \end{aligned} \quad (4.3.119)$$

The scaling properties of the deviations of the densities from their transition values are determined by the scaling properties of the first-order partial derivatives of ϕ_{sing} with respect to the scaling fields θ_j . If $\mathcal{C}$ is approached along a path $\mathcal{P}$ such that, according to the second of equations (4.3.19),

$$\frac{\partial \phi_{\text{sing}}}{\partial \theta_j}(\varpi_1^{(m)}, \dots, \varpi_n^{(m)}) \quad \text{is finite and non-zero,} \quad (4.3.120)$$

it follows from (4.3.12), (4.3.15) and (4.3.20) that

$$\frac{\partial \phi_{\text{sing}}}{\partial \theta_j} \sim |\theta_m|^{(d-y_j)/y_m}. \quad (4.3.121)$$

The scaling behaviour in Eq. (4.3.119) is dominated by the term in the summation on the right-hand side which corresponds to the largest y_j for which the derivative of θ_j with respect to the appropriate coupling is non-zero on $\mathcal{C}$. We denote these exponents by y_i^f and y_ℓ^e respectively for the first and second of equations (4.3.119). Then, supposing that a path is chosen so that the finite non-zero amplitude condition (4.3.120) is satisfied,

$$\rho_i - \rho_{ic} \sim |\theta_m|^{(d-y_i^f)/y_m}, \quad u_\ell - u_{\ell c} \sim |\theta_m|^{(d-y_\ell^e)/y_m}. \quad (4.3.122)$$

The response functions, given by the second-order partial derivatives of ϕ with respect to the internal and external couplings, can be analyzed in the same way. If it is again assumed that the appropriate amplitudes are finite and non-zero, the scaling forms can be seen to be

$$\frac{\partial^2 \phi}{\partial L_i \partial L_j} \sim |\theta_m|^{(d-y_i^f - y_j^f)/y_m}, \quad (4.3.123)$$

$$\frac{\partial^2 \phi}{\partial K_\ell \partial K_j} \sim |\theta_m|^{(d-y_\ell^e - y_j^e)/y_m}, \quad (4.3.124)$$

$$\frac{\partial^2 \phi}{\partial L_i \partial K_\ell} \sim |\theta_m|^{(d-y_i^f - y_\ell^e)/y_m}. \quad (4.3.125)$$

4.4 Critical Point and Coexistence Curve

In Sect. 4.2 we discussed, in an intuitive way, the ideas of scaling as they relate to the critical point in the ferromagnetic Ising model. The case of a critical point, terminating a line of first-order transitions, is the situation which has received most attention both in discussions of scaling theory and more generally in the field of critical phenomena. The ferromagnetic Ising model has the virtue, from the point of view of simplicity, that the first-order transition and critical point lie on the zero-field axis, about which the phase diagram is symmetric. In this section we return to the case of a critical point, but assume no intrinsic symmetry. This means that the discussion is applicable to both fluid and magnetic models, these special cases being described in detail below.

The phase space is a plane ($n = 2$), the internal energy parameter is denoted by ε and the one independent external field is ξ , with couplings $K := \varepsilon/T$ and $L := \xi/T$. The free-energy density is $\phi_1 = \phi_1(K, L)$ with

$$d\phi_1 = u dK - \rho dL. \quad (4.4.1)$$

The first-order transition appears as a curve in the L - K plane terminating at the critical point (L_c, K_c) (Fig. 4.7). The phase transition makes its appearance in the ρ - K plane as a coexistence region bounded by two curves (Fig. 4.8)

$$\rho = \rho^{(\pm)}(K), \quad (4.4.2)$$

Fig. 4.7 A critical point (L_c, K_c) in the L - K plane. The first-order transition (coexistence curve) $L = L^*(K)$ is represented by a *broken line* and the critical isochore $L = L(K, \rho_c)$ by a *dotted line* (The fact that these two curves have a common tangent at the critical point is established by scaling theory)

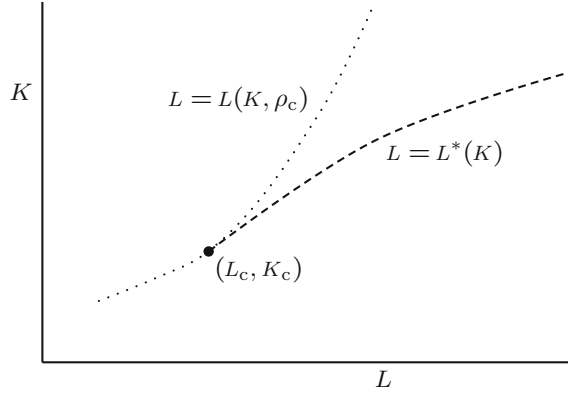
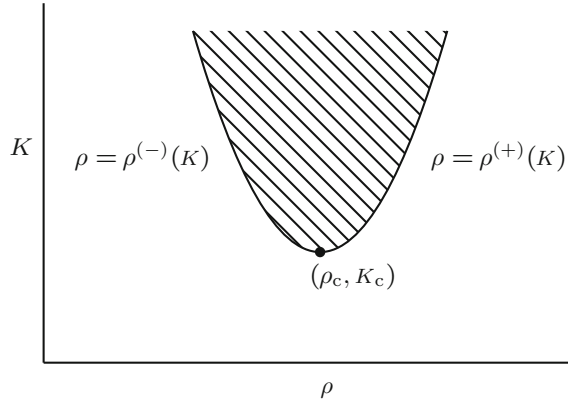


Fig. 4.8 The coexistence region (*shaded*) in the ρ - K plane



where

$$\rho^{(+)}(K) > \rho^{(-)}(K), \quad K > K_c, \quad (4.4.3)$$

$$\rho^{(+)}(K_c) = \rho^{(-)}(K_c) = \rho_c, \quad (4.4.4)$$

$$\left(\frac{d\rho^{(+)}}{dK} \right)_c = - \left(\frac{d\rho^{(-)}}{dK} \right)_c = \infty, \quad (4.4.5)$$

$$\left(\frac{d^2\rho^{(+)}}{dK^2} \right)_c = - \left(\frac{d^2\rho^{(-)}}{dK^2} \right)_c. \quad (4.4.6)$$

The Eqs. (4.4.3)–(4.4.5) are necessary consequences of the definition of the critical point. Equation (4.4.6) is not a necessary condition but is true in most cases of critical points. The equation of state of the system in terms of the variables K , L and ρ is given, from (4.4.1), by

$$\rho = -\frac{\partial \phi_1}{\partial L}, \quad (4.4.7)$$

yielding an equation of the form

$$L = L(K, \rho), \quad (4.4.8)$$

which is smooth outside the coexistence region. The projection of the coexistence curve into the L - K plane, which is the line of first-order transitions shown in Fig. 4.7, is given, for $K \geq K_c$, by

$$L = L^*(K) := L(K, \rho^{(\pm)}(K)). \quad (4.4.9)$$

From (4.4.9),

$$\frac{dL^*}{dK} = \left(\frac{\partial L}{\partial K} \right)_\rho + \left(\frac{\partial L}{\partial \rho} \right)_K \frac{d\rho^{(\pm)}}{dK}. \quad (4.4.10)$$

The critical isochore $\rho = \rho_c$ makes the curve

$$L = L(K, \rho_c) \quad (4.4.11)$$

in the L - K plane passing through the critical point. The gradients of the critical isochore and coexistence curve differ at the critical point by the critical point value of the second term on the right-hand side of (4.4.10). This term contains an infinite factor at the critical point (Eq. (4.4.5)). The fact that this term as a whole is zero will be established in Sect. 4.5.6. In terms of the scaling theory described in Sect. 4.3 the situation here is one in which we have two transition regions: the critical point of zero dimension, with $s = n = 2$ and two relevant scaling fields, and the coexistence curve, with $s = n - 1 = 1$ and one relevant and one irrelevant scaling field.

4.4.1 Critical Exponents

In the standardized notation of critical phenomena (Buckingham 1972) the critical exponents α , α' , β , δ , γ and γ' are defined by

$$c_\rho \sim \begin{cases} (T - T_c)^{-\alpha}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\alpha'}, & \text{along the coexistence curve, } T < T_c, \end{cases} \quad (4.4.12)$$

$$\rho - \rho_c \sim (T_c - T)^\beta, \text{ along the coexistence curve, } T < T_c, \quad (4.4.13)$$

$$\varphi_T \sim \begin{cases} (T - T_c)^{-\gamma}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\gamma'}, & \text{along the coexistence curve, } T < T_c, \end{cases} \quad (4.4.14)$$

$$\xi - \xi_c \sim (\rho - \rho_c)|\rho - \rho_c|^{\delta-1}, \text{ along the critical isotherm.} \quad (4.4.15)$$

It also convenient to define the exponents σ and σ' according to

$$c_\xi \sim \begin{cases} (T - T_c)^{-\sigma}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\sigma'}, & \text{along the coexistence curve, } T < T_c. \end{cases} \quad (4.4.16)$$

4.4.2 Exponent Inequalities

The heat capacities c_ρ and c_ξ in (1.4.8) are both positive. From this it follows that, if $\varphi_T > 0$, then c_ξ dominates both c_ρ and α_ξ^2/φ_T as $T \rightarrow T_c$. This means that

$$\sigma \geq \alpha, \quad \sigma' \geq \alpha', \quad (4.4.17)$$

$$\sigma' + 2\beta + \gamma' \geq 2. \quad (4.4.18)$$

The condition $\varphi_T > 0$ is true for a magnetic system and in this case the inequality (4.4.18) was first established by Rushbrooke (1963). The stronger condition

$$\alpha' + 2\beta + \gamma' \geq 2 \quad (4.4.19)$$

was obtained by Griffiths (1965) for both ferromagnetic and fluid systems using the convexity properties of the free energy. In fact, as we shall see in Sect. 4.5, for systems with the special symmetry of the ferromagnet $\alpha' = \sigma'$ and inequalities (4.4.18) and (4.4.19) become identical. Griffiths (1965) also derived a number of other inequalities. In particular

$$\gamma' \geq \beta(\delta - 1). \quad (4.4.20)$$

4.5 Scaling for a Critical Point

From the discussion in Sect. 4.3 it is clear that there is a degree of latitude in the definition of the singular part of ϕ . Given that it should contain all the non-smooth part of ϕ , it is convenient, for the critical point, to take

$$\phi_{\text{sing}}(\Delta K, \Delta L) := \phi_1(K, L) - \phi_1(K_c, L_c) - u_c \Delta K + \rho_c \Delta L, \quad (4.5.1)$$

where $\Delta K := K - K_c$ and $\Delta L := L - L_c$, so that

$$\phi_{\text{sing}}(0, 0) = 0, \quad (4.5.2)$$

$$\frac{\partial \phi_{\text{sing}}}{\partial \Delta K} = \Delta u, \quad (4.5.3)$$

$$\frac{\partial \phi_{\text{sing}}}{\partial \Delta L} = -\Delta \rho. \quad (4.5.4)$$

4.5.1 Scaling Fields for the Critical Point

In the light of Sect. 4.3, we see that (4.2.11) is the scaling hypothesis for a magnetic system using scaling fields $\Delta L = L$, ($L_c = 0$), and ΔK . The special feature of the magnetic system which is relevant here is that the first-order phase transition lies along the line $L = 0$. The scaling fields were chosen to pick out this special line. In the general case, before defining scaling fields, we must define new coordinates oriented to select the special direction represented by the tangent to the coexistence curve at the critical point. Assuming that the coexistence curve at the critical point is not parallel to the L -axis, $L^*(K)$ can be expressed in the form

$$L^*(K) = L_c + a\Delta K + \mathcal{L}(K), \quad (4.5.5)$$

where a is the (finite) slope of the coexistence curve at the critical point and

$$\mathcal{L}(K_c) = 0, \quad \mathcal{L}'(K_c) = 0, \quad \mathcal{L}(K) \simeq b|\Delta K|^\psi, \quad (4.5.6)$$

b being zero if the coexistence curve is linear, and $\psi (> 1)$ being called the *shift exponent*. We use the orthogonal coordinates

$$\Theta_1 := L - aK, \quad \Theta_2 := K + aL, \quad (4.5.7)$$

where $\Theta_1 = \Theta_{1c} = L_c - aK_c$ gives the tangent to the coexistence curve. Along the coexistence curve $\Delta K \geq 0$ and, assuming that $0 \leq a < \infty$, with $0 \leq \mathcal{L}(K) < \infty$ for $\Delta K \geq 0$, this curve will leave the critical point $(\Theta_{1c}, \Theta_{2c})$ and then satisfy the conditions $\Theta_1 \geq \Theta_{1c}$ and $\Theta_2 > \Theta_{2c}$. In terms of the orthogonal scaling fields

$$\theta_1 := \Theta_1 - \Theta_{1c} = \Delta L - a\Delta K, \quad \theta_2 := \Theta_2 - \Theta_{2c} = \Delta K + a\Delta L, \quad (4.5.8)$$

the scaling hypothesis takes the form

$$\phi_{\text{sing}}(\lambda^{y_1}\theta_1, \lambda^{y_2}\theta_2) = \lambda^d \phi_{\text{sing}}(\theta_1, \theta_2). \quad (4.5.9)$$

Following the procedure of Sect. 4.3, set $\lambda = |\theta_1|^{-1/y_1}$, giving

$$\phi_{\text{sing}}(\theta_1, \theta_2) = |\theta_1|^{d/y_1} \phi_{\text{sing}}(\pm 1, g_2), \quad (4.5.10)$$

and $\lambda = |\theta_2|^{-1/y_2}$, giving

$$\phi_{\text{sing}}(\theta_1, \theta_2) = |\theta_2|^{d/y_2} \phi_{\text{sing}}(g_1, \pm 1). \quad (4.5.11)$$

For the sake of brevity, the gap exponent is denoted as $\vartheta := y_1/y_2$ and the gap fields as

$$g_1(\theta_1, \theta_2) := \theta_1/|\theta_2|^\vartheta, \quad g_2(\theta_1, \theta_2) := \theta_2/|\theta_1|^{1/\vartheta}, \quad (4.5.12)$$

where $|g_1| = |g_2|^{-\vartheta}$. Both (4.5.10) and (4.5.11) are valid in a neighbourhood of the critical point and thus

$$\begin{aligned} \phi_{\text{sing}}(g_1, \pm 1) &= |g_1|^{d/y_1} \phi_{\text{sing}}(\pm 1, g_2), \\ \phi_{\text{sing}}(\pm 1, g_2) &= |g_2|^{d/y_2} \phi_{\text{sing}}(g_1, \pm 1), \end{aligned} \quad (4.5.13)$$

(cf. Eq. (4.3.17)). Along the coexistence curve $\phi_{\text{sing}}(\theta_1, \theta_2)$ exhibits non-smooth behaviour; specifically one, or both, of the first-order partial derivatives is discontinuous. Since $\theta_2 \neq 0$ along the coexistence curve, apart from at the critical point, it follows from (4.5.11) that the coexistence curve corresponds to singular behaviour in $\phi_{\text{sing}}(g_1, +1)$. Supposing that, in a neighbourhood of the critical point, the only singular behaviour corresponds to the critical point and coexistence curve, $\phi_{\text{sing}}(g_1, +1)$ will be singular at only one value

$$g_1(\theta_1, \theta_2) = \varpi_0, \quad (4.5.14)$$

where $0 \leq \varpi_0 < \infty$ and (4.5.14) is the equation of the coexistence curve in a neighbourhood of the critical point. The constant will be zero only if $\mathcal{L}(K) \equiv 0$, when the coexistence curve lies along the $\theta_1 = 0$ axis. Except in this special case the coexistence curve is tangential to $\theta_1 = 0$ and it follows that

$$\vartheta > 1, \quad y_1 > y_2. \quad (4.5.15)$$

The curves, defined by

$$|\theta_1| = \varpi |\theta_2|^\vartheta, \quad (4.5.16)$$

for fixed finite non-zero ϖ , are tangential to the $\theta_1 = 0$ axis at the critical point, picking it out as a special direction when (4.5.15) is satisfied. Experimental evidence indicates that, in all cases, there is one special direction at the critical point which is the direction tangential to the coexistence curve. This supports the conclusion that,

even when the coexistence curve lies along $\theta_1 = 0$, (4.5.15) must hold and that the curves (4.5.16), with non-zero, finite ϖ are tangential to $\theta_1 = 0$. The exponents y_1 and y_2 are referred to as the *strong* and *weak* exponents respectively.

Given that ϕ_{sing} is a continuous function we assume that the amplitudes

$$\lim_{g_2 \rightarrow 0} \phi_{\text{sing}}(\pm 1, g_2) := A_{\infty}^{(\pm)}, \quad \lim_{g_1 \rightarrow 0} \phi_{\text{sing}}(g_1, \pm 1) := A_0^{(\pm)} \quad (4.5.17)$$

are finite and non-zero.²³ Then, from (4.5.13),

$$\begin{aligned} \lim_{g_2 \rightarrow \infty} \phi_{\text{sing}}(\pm 1, g_2) &\simeq |g_2|^{d/y_2} A_0^{(\pm)}, \\ \lim_{g_1 \rightarrow \infty} \phi_{\text{sing}}(g_1, \pm 1) &\simeq |g_1|^{d/y_1} A_{\infty}^{(\pm)}. \end{aligned} \quad (4.5.18)$$

4.5.2 Approaches to the Critical Point

These can be divided into three classes:

- (i) Paths on which $|g_1| \rightarrow \infty$ ($|g_2| \rightarrow 0$), including the axis $\theta_2 = 0$, for which, from (4.5.10) and (4.5.17),

$$\phi_{\text{sing}}(\theta_1, \theta_2) \simeq |\theta_1|^{d/y_1} A_{\infty}^{(\pm)}. \quad (4.5.19)$$

- (ii) Paths on which $|g_1| \rightarrow 0$, including the axis $\theta_1 = 0$, for which, from (4.5.11) and (4.5.17),

$$\phi_{\text{sing}}(\theta_1, \theta_2) \simeq |\theta_2|^{d/y_2} A_0^{(\pm)}. \quad (4.5.20)$$

- (iii) Paths on which $|g_1| \rightarrow \pi_1$, where π_1 is finite and non-zero, when

$$\phi_{\text{sing}}(\theta_1, \theta_2) \simeq |\theta_2|^{d/y_2} A_{\pi_1}^{(\pm)}, \quad (4.5.21)$$

with

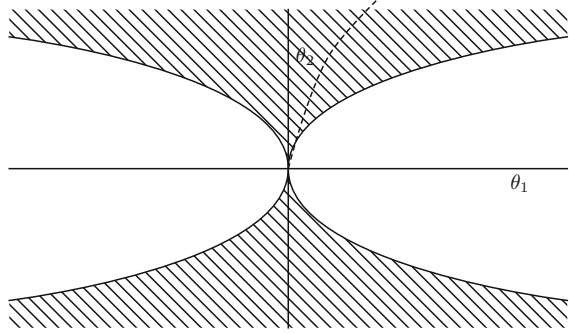
$$\lim_{g_1 \rightarrow \pi_1} \phi_{\text{sing}}(g_1, \pm 1) = A_{\pi_1}^{(\pm)} \quad (4.5.22)$$

assumed to be finite and non-zero.

It is, of course, the case that an alternative to (4.5.21) could be given in terms of θ_1 , when $\pi_1 \neq 0$. However, it is convenient to class cases (ii) and (iii) together with (4.5.21) and (4.5.22) applying for all finite values of π_1 . All these approaches will

²³ See Sect. 4.8 for a discussion of circumstances where this assumption breaks down.

Fig. 4.9 The θ_1 – θ_2 plane in a neighbourhood of the critical point. The line $\theta_1 = 0, \theta_2 \geq 0$ is tangential to the coexistence curve and the crossover curves $|\theta_1| = \varpi |\theta_2|^\vartheta$ separate weak (*shaded*) and strong (*unshaded*) scaling approaches



be called *weak* as they differ only in the value of the amplitude $A_{\pi_1}^{(\pm)}$. Approaches in class (i) are called *strong*. Equation (4.5.16), for some finite non-zero ϖ , gives two crossover curves in the θ_1 – θ_2 plane, which separate strong and weak scaling approaches (Fig. 4.9).²⁴

It is clear that an approach to the critical point along the coexistence curve itself is a weak approach with $\pi_1 = \varpi_0$. From (4.5.8),

$$\theta_1 \left(1 + a \frac{\theta_2}{\theta_1} \right) = (1 + a^2) \Delta L, \quad \theta_2 \left(1 - a \frac{\theta_1}{\theta_2} \right) = (1 + a^2) \Delta K \quad (4.5.23)$$

and since, by (4.5.16) with $\varpi = \varpi_0$, $\theta_1/\theta_2 \rightarrow 0$ as the critical point is approached along the coexistence curve

$$\theta_1 \simeq \varpi_0 (1 + a^2)^\vartheta |\Delta K|^\vartheta. \quad (4.5.24)$$

Then, from (4.5.5) and (4.5.8),

$$\mathcal{L}(K) \simeq \varpi_0 (1 + a^2)^\vartheta |\Delta K|^\vartheta, \quad (4.5.25)$$

showing that the gap exponent $\vartheta = \psi$, the shift exponent defined in (4.5.6).

4.5.3 Experimental Variables

To link this analysis with experiment it is necessary to translate the asymptotic formulae, given above, into formulae expressed in thermodynamic variables. Two approaches to the critical point are of particular interest:

²⁴ It will be seen from (4.5.16) that the actual value of ϖ , as long as it is finite and non-zero, is irrelevant to the asymptotic behaviour of ϕ_{sing} , representing only a multiplicative scaling of θ_1 .

- (a) The *critical isotherm* $T = T_c$, which in terms of the couplings, where $T = \varepsilon/K$ with ε constant, is the line $\Delta K = 0$. From (4.5.8) and (4.5.12) $|g_1| = 1/|a|^\vartheta |\Delta L|^{\vartheta-1}$. So, remembering that a is finite and possibly zero, it is clear that the critical isotherm gives a strong approach to the critical point.
- (b) The *critical isochamp*, that is the line $\xi = \xi_c$, where $\xi = TL$. It is straightforward to show that this is the line $K_c \Delta L = L_c \Delta K$ and this also corresponds to an approach to the critical point in a strong scaling direction unless

$$\Theta_{2c} = 0, \quad L_c = aK_c, \quad (4.5.26)$$

(including the case $a = 0, L_c = 0$).

For a weak scaling approach $\theta_1/\theta_2 \rightarrow 0$. So, from (4.5.20) and (4.5.23),

$$\phi_{\text{sing}}(\theta_1, \theta_2) \sim |\Theta_2 - \Theta_{2c}|^{d/y_2} \sim |\Delta K|^{d/y_2}. \quad (4.5.27)$$

Strong scaling approaches include all those where $\theta_2/\theta_1 \rightarrow 0$, for which (4.5.19) can be replaced by

$$\phi_{\text{sing}}(\theta_1, \theta_2) \sim |\Theta_1 - \Theta_{1c}|^{d/y_1} \sim |\Delta L|^{d/y_1}. \quad (4.5.28)$$

They also include approaches where θ_2/θ_1 remains finite. This affects the amplitude of the asymptotic form, but (4.5.28) still applies. We should, however, bear in mind the case where $\theta_1/\theta_2 \rightarrow 0$ with a curvature larger than that of the crossover curve. This is a somewhat pathological strong scaling approach. Equations (4.5.27) and (4.5.28) link the weak and strong exponents y_2 and y_1 with the couplings K and L and, as in Sect. 4.2, they could be denoted by y_K and y_L respectively. One could go further than this and link these exponents to thermodynamic variables. Since K is an internal coupling with ε fixed the only physically variable parameter is the temperature T . The exponent y_2 can, therefore, be referred to as the *thermal exponent* often denoted as y_T . In a similar way, L is an external coupling, which can be changed by altering the field ξ . So y_1 , can be referred to as the *field exponent* and denoted by y_ξ . We shall, however, in the remainder of this section continue to use the notation y_1 and y_2 for the exponents.

4.5.4 The Density and Response Functions

It is now necessary to express the density and the response functions in terms of the partial derivatives of ϕ_{sing} with respect to the scaling fields (Eqs. (4.3.119)). From (2.3.30), (2.3.31), (4.5.4), (4.5.7) and (4.5.8)

$$\rho - \rho_c = \Delta\rho = - \left[\frac{\partial \phi_{\text{sing}}}{\partial \theta_1} + a \frac{\partial \phi_{\text{sing}}}{\partial \theta_2} \right], \quad (4.5.29)$$

$$\varphi_T = - \left[\frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2} + 2a \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} + a^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} \right], \quad (4.5.30)$$

$$c_\xi = -k_B \left[\Theta_1^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2} + 2\Theta_1 \Theta_2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} + \Theta_2^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} \right]. \quad (4.5.31)$$

For c_ρ the expression (1.4.8) is used with

$$\alpha_\xi = \Theta_1 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2} + (a\Theta_1 + \Theta_2) \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} + a\Theta_2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2}. \quad (4.5.32)$$

Substituting from (4.5.30)–(4.5.32) into (1.4.8)

$$c_\rho = -\frac{k_B}{\varphi_T} (a\Theta_1 - \Theta_2)^2 \left[\left(\frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} \right)^2 - \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2} \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} \right]. \quad (4.5.33)$$

4.5.5 Asymptotic Forms

To obtain asymptotic forms it is necessary to determine the dominant contributions in (4.5.29)–(4.5.31) and (4.5.33), using the reasoning of Sect. 4.3.9. Since $y_1 > y_2$, these correspond to the terms with the largest number of differentiations with respect to θ_1 . On this basis we see that the terms in (4.5.29)–(4.5.31) are ordered in descending dominance and those in (4.5.33) have equal weight. Repeating the reasoning in (4.5.9)–(4.5.22) for the partial derivatives of ϕ_{sing} , replaces d by the corresponding $d - k_1 y_1 - k_2 y_2$. For the weak scaling direction the formulae corresponding to (4.5.22), when $\pi_1 = \varpi_0$, that is across the coexistence curve, do not give an unambiguous limit, yielding possibly different values as $g_1 \rightarrow \varpi_0$ from the two sides. This means that the amplitudes for approaches along the two branches of the coexistence curve can be different. Except in the special case where one is zero and corrections to scaling need to be considered (Sect. 4.8), this will not affect the exponent in the asymptotic form. Without further comment we shall now assume that all the appropriate amplitude factors given by the second of equations (4.3.19) are non-zero and, using the arguments of Sect. 4.3.9, it follows, from (4.5.29) and (4.5.30), that

$$\rho - \rho_c \sim \begin{cases} |\Delta L|^{(d-y_1)/y_1}, & \text{strong direction,} \\ |\Delta K|^{(d-y_1)/y_2}, & \text{weak direction,} \end{cases} \quad (4.5.34)$$

$$\varphi_T \sim \begin{cases} |\Delta L|^{(d-2y_1)/y_1}, & \text{strong direction,} \\ |\Delta K|^{(d-2y_1)/y_2}, & \text{weak direction.} \end{cases} \quad (4.5.35)$$

To consider the asymptotic properties of c_ξ it is necessary to take account of the possibility that $\Theta_1(L_c, K_c) = 0$, which is the condition given by (4.5.26). $\Theta_2(L_c, K_c) \neq 0$ if $\Theta_1(L_c, K_c) = 0$ and we have, from (4.5.31), two cases

$$c_\xi \sim \begin{cases} |\Delta L|^{(d-2y_1)/y_1}, & \text{strong direction,} \\ |\Delta K|^{(d-2y_1)/y_2}, & \text{weak direction,} \end{cases} \quad \text{if } L_c \neq aK_c \quad (4.5.36)$$

and

$$c_\xi \sim \begin{cases} |\Delta L|^{(d-2y_2)/y_1}, & \text{strong direction,} \\ |\Delta K|^{(d-2y_2)/y_2}, & \text{weak direction,} \end{cases} \quad \text{if } L_c = aK_c. \quad (4.5.37)$$

From (4.5.33) it can be seen that the asymptotic expressions for c_ρ will involve a combination of amplitude factors. Unless these cancel

$$c_\rho \sim \begin{cases} |\Delta L|^{(d-2y_2)/y_1}, & \text{strong direction,} \\ |\Delta K|^{(d-2y_2)/y_2}, & \text{weak direction.} \end{cases} \quad (4.5.38)$$

4.5.6 Critical Exponents and Scaling Laws

Given that an approach to the critical point along the coexistence curve is a weak approach and along the critical isotherm is a strong approach it follows from a comparison of (4.4.12)–(4.4.16) with (4.5.34)–(4.5.38) that

$$\alpha' = (2y_2 - d)/y_2, \quad (4.5.39)$$

$$\beta = (d - y_1)/y_2, \quad (4.5.40)$$

$$\gamma' = (2y_1 - d)/y_2, \quad (4.5.41)$$

$$\delta = y_1/(d - y_1), \quad (4.5.42)$$

$$\sigma' = \begin{cases} (2y_1 - d)/y_2 = \gamma', & \text{if } L_c \neq aK_c, \\ (2y_2 - d)/y_2 = \alpha', & \text{if } L_c = aK_c. \end{cases} \quad (4.5.43)$$

From (4.5.43) and (4.4.17) it follows that, when $L_c \neq aK_c$, $\gamma' \geq \alpha'$, a result which can be established without the aid of scaling theory (Buckingham 1972).

The magnetic system, described in Sect. 4.2 has the first-order transition line lying along the zero-field axis. So $L_c = 0$, $a = 0$ and $\sigma' = \alpha'$. It is for this reason that, when defining critical exponents for fluids and magnets, it is common practice (Stanley 1987; Yeomans 1992) to use α' for the heat capacity at constant density for the fluid and constant magnetic field for the magnet. We can now understand that the reason that the same exponent is appropriate to both cases is due to the special symmetry properties of the magnet, which also makes the analysis rather simpler.

Equations (4.5.39)–(4.5.42) give formulae for the exponents α' , β , γ' and δ in terms of y_1 and y_2 . They are, therefore, not independent and two relationships exist between them. These can be expressed in the form

$$\alpha' + 2\beta + \gamma' = 2, \quad (4.5.44)$$

called the *Essam–Fisher scaling law* (Essam and Fisher 1963), which is a strengthening of the inequality (4.4.19) and

$$\gamma' = \beta(\delta - 1), \quad (4.5.45)$$

called the *Widom scaling law* (Widom 1964) which is a strengthening of the inequality (4.4.20).

All the information necessary to consider the second term on the right-hand side of (4.4.10) is now available. Since this is the case of a weak direction approach to the critical point,

$$\frac{d\rho^{(\pm)}}{dK} \sim |\Delta K|^{-1+(d-y_1)/y_2}, \quad \left(\frac{\partial L}{\partial \rho}\right)_K \sim |\Delta K|^{-(d-2y_1)/y_2},$$

giving

$$\left(\frac{\partial L}{\partial \rho}\right)_K \frac{d\rho^{(\pm)}}{dK} \sim |\Delta K|^{\vartheta-1}. \quad (4.5.46)$$

As $\vartheta > 1$, this establishes the result that the critical isochore and coexistence curve share a common tangent (given by $\theta_1 = 0$) at the critical point. The critical isochore is thus a direction of weak approach to the critical point. This means that, in general,

$$\alpha = \alpha', \quad \gamma = \gamma', \quad \sigma = \sigma'. \quad (4.5.47)$$

However, of course, the amplitudes associated with the asymptotic forms for approaches to the critical point along the critical isochore and the coexistence curve are, in general different. If one of these were zero then this would lead to a correction to scaling as described in Sect. 4.8 with a corresponding breakdown in the equality given by (4.5.47).

The special case $y_1 = 0$, giving $\beta = 0$, $\delta = \infty$, allows for the possibility of discontinuity in ρ at the critical point. This is called a *first-order critical point* (Fisher and Berker 1982).

As indicated above the critical point and the coexistence curve can be regarded as two connected critical regions $\mathcal{C}$ and $\mathring{\mathcal{C}}$, of respective dimensions zero and one. The exponents y_1 and y_2 of $\mathcal{C}$ are both relevant and the exponent $\mathring{y}_1$ and $\mathring{y}_2$ of $\mathring{\mathcal{C}}$ are respectively relevant and irrelevant. This situation can be treated with the model of Sect. 4.3.3. There the assumption is made that $\mathring{y}_2 = -y_2$ and here the first-order transition across the coexistence curve implies the condition $\mathring{y}_1 = 1$. An effective exponent and the inequality (4.3.46) are used to defined the (shaded) crossover region shown in Fig. 4.6. This contracts to the axis $\Theta_2 = 0$ as $y_1 \rightarrow 1$, when the critical point becomes a first-order critical point.

4.5.7 Correlation Scaling at a Critical Point

In this section we consider scaling for the connected two-point correlation function $\Gamma_2(q(\mathbf{r}_1), q(\mathbf{r}_2))$ associated with the coupling L and defined by (2.5.9) with $c_1 = c_2 = q$. As before a d -dimensional hypercubic lattice is assumed with periodic boundary conditions and translational invariance so that the correlation function is dependent on $\mathbf{r}_1$ and $\mathbf{r}_2$ through their difference $\bar{\mathbf{r}} := \mathbf{r}_2 - \mathbf{r}_1$.

In Sect. 4.3.7 the scaling behaviour of correlation functions of scaling operators was developed. To translate these results into the scaling behaviour of state operators like q the reasoning of Sect. 4.3.9 is used to determine the leading scaling behaviour associated with the coupling L . With the definitions (4.5.8) for the scaling fields we showed in Sect. 4.5 that L scales like θ_1 . It follows that the two-point q correlation function scales like the two-point ϕ_1 - ϕ_1 correlation function. Thus from (4.3.99), (4.3.101) and (4.3.102)

$$\Gamma_2(\lambda^{y_1}\theta_1, \lambda^{y_2}\theta_2; q(\mathbf{0}), q(\lambda^{-1}\bar{\mathbf{r}})) = \lambda^{2x_1}\Gamma_2(\theta_1, \theta_2; q(\mathbf{0}), q(\bar{\mathbf{r}})) \quad (4.5.48)$$

and

$$\bar{\Gamma}_2^*(\lambda^{y_1}\theta_1, \lambda^{y_2}\theta_2; q; \lambda\bar{\mathbf{k}}) = \lambda^{d-2y_1}\bar{\Gamma}_2^*(\theta_1, \theta_2; q; \bar{\mathbf{k}}), \quad (4.5.49)$$

where $x_1 := d - y_1$ is the scaling dimension of q .

The correlation length is defined by (4.2.4) and it is easily seen that this can be expressed in the form

$$\tau^2(q) = -c(\partial) \frac{\nabla_{\vec{k}}^2 \bar{\Gamma}_2^*(q; \mathbf{0})}{\bar{\Gamma}_2^*(q; \mathbf{0})}. \quad (4.5.50)$$

Thus, from (4.5.50), the scaling form for the correlation length is

$$\tau(\lambda^{y_1} \theta_1, \lambda^{y_2} \theta_2; q) = \lambda^{-1} \tau(\theta_1, \theta_2; q). \quad (4.5.51)$$

According to the standard notation the exponents ν and ν' are defined by

$$\tau(q) \sim \begin{cases} (T - T_c)^{-\nu}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\nu'}, & \text{along the coexistence curve, } T < T_c, \end{cases} \quad (4.5.52)$$

and thus, taking into account that both the coexistence curve and the critical isochore are directions of weak approach,

$$\nu = \nu' = 1/y_2. \quad (4.5.53)$$

This result and (4.5.39) with (4.5.47) yields the *Josephson hyper-scaling law*

$$d\nu = 2 - \alpha, \quad (4.5.54)$$

which was proved by Josephson (1967) as an inequality with ‘ $\geq$ ’ replacing ‘ $=$ ’. At the critical point, for large $|\mathbf{r}_2 - \mathbf{r}_1|$, the standard asymptotic form for the connected two-point correlation function is given in terms of the exponent η by

$$\Gamma_2(q(\mathbf{r}_1), q(\mathbf{r}_2)) \sim \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|^{d-2+\eta}}. \quad (4.5.55)$$

From (4.3.110) this means that

$$\eta = 2 - 2y_1 + d = 2 + 2x_1 - d, \quad (4.5.56)$$

if it is assumed that the appropriate amplitude factor is non-zero and finite. It now follows from (4.5.49) that, at the critical point for small $|\vec{k}|$,

$$\bar{\Gamma}_2^*(q; \vec{k}) \sim |\vec{k}|^{\eta-2}. \quad (4.5.57)$$

The *Fisher scaling law*

$$(2 - \eta)\nu = \gamma, \quad (4.5.58)$$

follows from (4.5.41), (4.5.47), (4.5.53) and (4.5.56). This was proved by Fisher (1969) as an inequality with ‘ $\geq$ ’ replacing ‘ $=$ ’.

From the definition of the gap exponent ϑ , given below Eqs. (4.5.11), and (4.5.40)

$$\vartheta = \beta \delta . \quad (4.5.59)$$

From (4.5.16) and Fig. 4.9 it can be seen that the larger the value of ϑ the more of the θ_1 – θ_2 plane in a neighbourhood of the critical point corresponds to weak scaling approaches. We now see, from (4.5.59), that, for fixed δ , larger values of ϑ correspond to larger values of β , corresponding to a narrowing of the coexistence region in the ρ – K plane (Fig. 4.8) and a corresponding increase in the part of the plane corresponding to strong scaling approaches. This apparently paradoxical situation is understood if we project the crossover curves into the ρ – K plane where they share a common tangent with the coexistence region. So in the θ_1 – θ_2 (or equivalently L – K) plane the region of weak scaling approaches is wedge shaped at the critical point, whereas in the ρ – K plane it is the region of strong scaling approaches which is wedge shaped. This interesting point was discussed by Griffiths and Wheeler (1970), who observe that weak scaling approaches to the critical point will be more easily achieved by adjustments of ρ and K , than by adjustments of L and K .

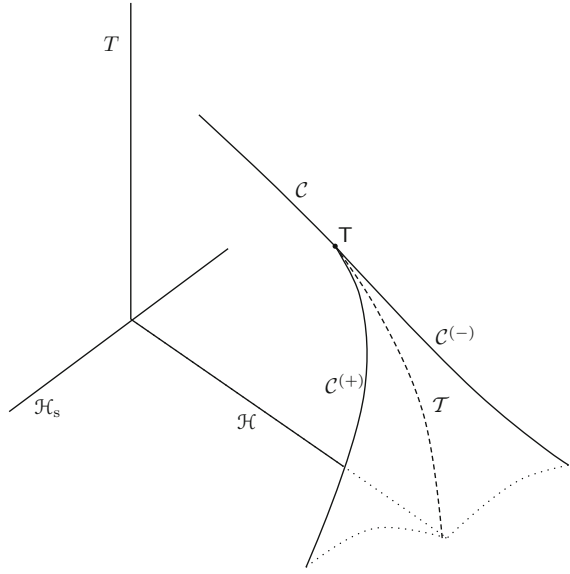
4.6 Tricritical Point

Section 4.1 contains a brief description of the kinds of critical phenomena which may be encountered in phase spaces of dimensions $n = 2$ and $n = 3$. We have described in Sects. 4.5 and 4.5.7 the application of scaling theory to the most well-known example of critical behaviour for $n = 2$, a critical point terminating a line of first-order transitions. We shall now consider the case of a tricritical point, an example of which occurs in the mean-field treatment of a metamagnet in Sect. 6.2.²⁵ For ease of discussion we shall in this section and Sect. 4.7 adopt a magnetic model with the curves of second-order transitions $\mathcal{C}$ and first-order transitions $\mathcal{T}$ lying in the T – $\mathcal{H}$ plane where they meet at the tricritical point T (Fig. 4.10). Of course, for this model there is no symmetry associated with $\mathcal{H} = 0$ axis, as there is with simple magnetic models. However, we shall assume the existence of a second *staggered field*²⁶ $\mathcal{H}_s$ with a phase diagram with symmetry with respect to the plane $\mathcal{H}_s = 0$. When the third dimension corresponding to the staggered field is included $\mathcal{C}$ becomes a critical line marking the edge of a surface of first-order transitions in the $\mathcal{H}_s = 0$ plane, and $\mathcal{T}$ is a triple line. Two further critical lines $\mathcal{C}^{(+)}$ and $\mathcal{C}^{(-)}$, bounding first-order surfaces, symmetrically placed about the plane $\mathcal{H}_s = 0$, also meet at the tricritical point T . For this system the free-energy density with no independent densities is $\phi_2(K, L, L_s)$ with

²⁵ In fact, as was indicated in Sect. 4.1, a tricritical point is a special case of a more general type of phase behaviour with a critical end-point and possibly a bicritical end-point (Fig. 4.3a). For the metamagnet an adjustable parameter is needed to achieve a tricritical point.

²⁶ The reason for this designation is made clear in Sect. 6.2.

Fig. 4.10 A tricritical point in a magnetic system



$$d\phi_2 = u dK - m dL - q dL_s, \quad (4.6.1)$$

where u , the (dimensionless) internal energy per lattice site, and m , the magnetization per lattice site, are the densities associated with the internal thermal coupling K and the external field coupling L . The magnetic density q measures ordering with respect to the staggered field coupling $L_s := \mathcal{H}_s/T$. The critical curve corresponds to the onset of this ordering. So $q = 0$ in the disordered large temperature or field region and also on the critical curve itself. The ordered ($q \neq 0$) region below the critical curve in the plane $\mathcal{H}_s = 0$ represents, as is the case on the zero-field axis below the Curie temperature for the Ising ferromagnet, coexistence of two phases with equal and opposite values of q .

The response function relations (2.3.29)–(2.3.32) need to be generalized to take into account the presence of two external couplings. The functions which measure the response of the magnetic densities to their conjugate fields are

$$\varphi_T := T \left(\frac{\partial m}{\partial \mathcal{H}} \right)_{T, \mathcal{H}_s} = \left(\frac{\partial m}{\partial L} \right)_{K, L_s} = - \frac{\partial^2 \phi_2}{\partial L^2}, \quad (4.6.2)$$

$$\psi_T := T \left(\frac{\partial q}{\partial \mathcal{H}_s} \right)_{T, \mathcal{H}} = \left(\frac{\partial q}{\partial L_s} \right)_{K, L} = - \frac{\partial^2 \phi_2}{\partial L_s^2}. \quad (4.6.3)$$

There are now four heat capacities corresponding to the choices of maintaining constant either the density or coupling in the pairs $(m, \mathcal{H})$ and $(q, \mathcal{H}_s)$. Two of these are the m and q constant heat capacity

$$c_{m,q} := k_B T \left(\frac{\partial s}{\partial T} \right)_{m,q} = -k_B K^2 \frac{\partial^2 \phi_0}{\partial K^2}, \quad (4.6.4)$$

where $\phi_0 = \phi_0(K, m, q)$ and the $\mathcal{H}$ and $\mathcal{H}_s$ constant heat capacity

$$\begin{aligned} c_{\mathcal{H}, \mathcal{H}_s} := k_B T \left(\frac{\partial s}{\partial T} \right)_{\mathcal{H}, \mathcal{H}_s} = & -k_B \left(K^2 \frac{\partial^2 \phi_2}{\partial K^2} + L^2 \frac{\partial^2 \phi_2}{\partial L^2} + L_s^2 \frac{\partial^2 \phi_2}{\partial L_s^2} \right. \\ & \left. + 2KL \frac{\partial^2 \phi_2}{\partial K \partial L} + 2KL_s \frac{\partial^2 \phi_2}{\partial K \partial L_s} + 2LL_s \frac{\partial^2 \phi_2}{\partial L \partial L_s} \right). \end{aligned} \quad (4.6.5)$$

There are in addition two heat capacities at one constant density and one constant field. Because of the symmetry of the system about the plane $\mathcal{H}_s = 0$ the only one of these which is of importance is

$$c_{m, \mathcal{H}_s} := k_B T \left(\frac{\partial s}{\partial T} \right)_{m, \mathcal{H}_s} = -k_B \left(K^2 \frac{\partial^2 \phi_1}{\partial K^2} + L_s^2 \frac{\partial^2 \phi_1}{\partial L_s^2} + 2KL_s \frac{\partial^2 \phi_1}{\partial K \partial L_s} \right), \quad (4.6.6)$$

where $\phi_1 = \phi_1(K, m, L_s)$. The generalizations of the standard formula (1.4.8) relating heat capacities now need

$$\vartheta_T := \left\{ \begin{aligned} T \left(\frac{\partial m}{\partial \mathcal{H}_s} \right)_{T, \mathcal{H}} &= \left(\frac{\partial m}{\partial L_s} \right)_{K, L} \\ T \left(\frac{\partial q}{\partial \mathcal{H}} \right)_{T, \mathcal{H}_s} &= \left(\frac{\partial q}{\partial L} \right)_{K, L_s} \end{aligned} \right\} = -\frac{\partial^2 \phi_2}{\partial L \partial L_s} \quad (4.6.7)$$

$$\begin{aligned} \alpha_{\mathcal{H}, \mathcal{H}_s} := T \left(\frac{\partial m}{\partial T} \right)_{\mathcal{H}, \mathcal{H}_s} &= -K \left(\frac{\partial m}{\partial K} \right)_{L, L_s} - L \left(\frac{\partial m}{\partial L} \right)_{K, L_s} - L_s \left(\frac{\partial m}{\partial L_s} \right)_{K, L} \\ &= K \frac{\partial^2 \phi_2}{\partial K \partial L} + L \frac{\partial^2 \phi_2}{\partial L^2} + L_s \frac{\partial^2 \phi_2}{\partial L \partial L_s}, \end{aligned} \quad (4.6.8)$$

$$\begin{aligned} \beta_{\mathcal{H}, \mathcal{H}_s} := T \left(\frac{\partial q}{\partial T} \right)_{\mathcal{H}, \mathcal{H}_s} &= -K \left(\frac{\partial q}{\partial K} \right)_{L, L_s} - L \left(\frac{\partial q}{\partial L} \right)_{K, L} - L_s \left(\frac{\partial q}{\partial L_s} \right)_{K, L_s} \\ &= K \frac{\partial^2 \phi_2}{\partial K \partial L_s} + L \frac{\partial^2 \phi_2}{\partial L_s \partial L} + L_s \frac{\partial^2 \phi_2}{\partial L_s^2}, \end{aligned} \quad (4.6.9)$$

and it can be shown that

$$c_{\mathcal{H}, \mathcal{H}_s} - c_{m, \mathcal{H}_s} = k_B \frac{\alpha_{\mathcal{H}, \mathcal{H}_s}^2}{\varphi_T}, \quad (4.6.10)$$

$$c_{\mathcal{H}, \mathcal{H}_s} - c_{m, q} = k_B \frac{\psi_T \alpha_{\mathcal{H}, \mathcal{H}_s}^2 + \varphi_T \beta_{\mathcal{H}, \mathcal{H}_s}^2 - 2\vartheta_T \alpha_{\mathcal{H}, \mathcal{H}_s} \beta_{\mathcal{H}, \mathcal{H}_s}}{\varphi_T \psi_T - \vartheta_T^2}. \quad (4.6.11)$$

We shall assume that the critical line $\mathcal{C}$ and the triple line $\mathcal{T}$ are asymptotically parallel at the tricritical point $\mathbf{T}$.²⁷ Let the critical curve $\mathcal{C}$ have the equation $\mathcal{H} = \mathcal{H}^*(T)$ with the tricritical point $\mathbf{T}$ given by the coordinates $T = T_t$, $\mathcal{H} = \mathcal{H}_t = \mathcal{H}^*(T_t)$ or alternatively, in terms couplings, $L = L^*(K)$ with the tricritical point $K = K_t$, $L = L_t = L^*(K_t)$.

Unlike the situation described in Sect. 4.5, the phase diagram here has reflectional symmetry; so the L_s axis is a special direction at the tricritical point. The only other special direction is the line of the tangent to the curves $\mathcal{C}$ and $\mathcal{T}$. Assuming that the curve $\mathcal{C}$ at the tricritical point is not parallel to the L -axis, $L^*(K)$ is expressed in the form

$$L^*(K) = L_t + a\Delta K + \mathcal{L}(K), \quad \Delta K = K - K_t, \quad (4.6.12)$$

where a is the (finite) slope of $\mathcal{C}$ at the tricritical point and

$$\mathcal{L}(K_t) = 0, \quad \mathcal{L}'(K_t) = 0, \quad \mathcal{L}(K) \simeq b|\Delta K|^{\psi_t}, \quad (4.6.13)$$

$\psi_t (>1)$ being the *tricritical shift exponent* (cf. Eqs. (4.5.5) and (4.5.6)). Then the orthogonal coordinates

$$\Theta_1 := L_s, \quad \Theta_2 := L - aK, \quad \Theta_3 := K + aL \quad (4.6.14)$$

are such that $\Theta_2 = \Theta_{2t} = L_t - aK_t$ is the equation of the tangent to $\mathcal{C}$ at $\mathbf{T}$ and $\Theta_3 = \Theta_{3t} = L_t + aK_t$ is the equation of an orthogonal line, with $\Theta_2 > \Theta_{2t} = L_t - aK_t$ in the disordered region and $\Theta_2 < \Theta_{2t} = L_t - aK_t$ in the ordered region. We now define critical exponents for the tricritical point. One exponent is defined for the direction orthogonal to the plane $\Theta_1 = L_s = 0$, namely

$$L_s \sim |q|^{|\delta_t|^{-1}}, \text{ along } \Theta_2 = \Theta_{2t}, \Theta_3 = \Theta_{3t} \quad (4.6.15)$$

and the rest are for directions in the plane $\Theta_1 = L_s = 0$ ²⁸:

²⁷ This is supported in Sect. 6.2.6 by mean-field calculations with the Hamiltonian (6.2.1).

²⁸ Lawrie and Sarbach (1984) define only two ‘ α -type’ tricritical exponents, adopting the commonly-used scaling ansatz for the free energy ((Fisher (1983) and Eq. (4.9.2)), whereby $2 - \alpha_t$ is the scaling exponent for the free energy in the direction of the temperature axis, with another exponent, which

$$c_{m,q} \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{-\alpha_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{-\alpha'_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{-\alpha_{tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.16)$$

$$c_{m,\mathcal{H}_s} \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{-\alpha_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{-\alpha'_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{-\alpha_{2tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.17)$$

$$c_{\mathcal{H},\mathcal{H}_s} \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{-\sigma_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{-\sigma'_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{-\sigma_{tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.18)$$

$$q \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{\beta_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{\beta_{tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.19)$$

$$(m - m_t) \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{\beta_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{\beta'_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{\beta_{2tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.20)$$

$$\psi_T \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{-\gamma_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{-\gamma'_t}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{-\gamma_{tc}}, & \text{along } \Theta_2 = \Theta_{2t}, \end{cases} \quad (4.6.21)$$

$$\varphi_T \sim \begin{cases} |\Theta_2 - \Theta_{2t}|^{-\gamma_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 > \Theta_{2t}, \\ |\Theta_2 - \Theta_{2t}|^{-\gamma'_{2t}}, & \text{along } \Theta_3 = \Theta_{3t}, \Theta_2 < \Theta_{2t}, \\ |\Theta_3 - \Theta_{3t}|^{-\gamma_{2tc}}, & \text{along } \Theta_2 = \Theta_{2t}. \end{cases} \quad (4.6.22)$$

(Footnote 28 continued)

they call simply α , giving the scaling exponent $2 - \alpha$ along the tangent to $\mathcal{C}$. Their calculations agree with (4.7.36) for α_t with $\alpha = \alpha_{2tc}$, given by (4.7.35).

4.7 Scaling for a Tricritical Point

Following a procedure similar to that of Sect. 4.5, the singular part of ϕ can be taken to be

$$\phi_{\text{sing}}(\Delta K, \Delta L, L_s) = \phi_2(K, L, L_s) - \phi_2(K_t, L_t, 0) - u_t \Delta K + m_t \Delta L, \quad (4.7.1)$$

where $\Delta K = K - K_t$, $\Delta L = L - L_t$, and u_t and m_t are the values of the internal energy and magnetization densities at the tricritical point.²⁹ Thus we have

$$\phi_{\text{sing}}(0, 0, 0) = 0, \quad (4.7.2)$$

$$\frac{\partial \phi_{\text{sing}}}{\partial \Delta K} = \Delta u, \quad (4.7.3)$$

$$\frac{\partial \phi_{\text{sing}}}{\partial \Delta L} = -\Delta m, \quad (4.7.4)$$

$$\frac{\partial \phi_{\text{sing}}}{\partial L_s} = -q. \quad (4.7.5)$$

4.7.1 Scaling Fields for the Tricritical Point

We use the orthogonal scaling fields

$$\theta_1 := \Theta_1, \quad \theta_2 := \Theta_2 - \Theta_{2t}, \quad \theta_3 := \Theta_3 - \Theta_{3t}, \quad (4.7.6)$$

with the triad of orthogonal unit vectors

$$\hat{\theta}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\theta}_2 := \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 0 \\ 1 \\ -a \end{pmatrix}, \quad \hat{\theta}_3 := \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} 0 \\ a \\ 1 \end{pmatrix}. \quad (4.7.7)$$

²⁹ At the tricritical point $L_s = 0$ and $q = 0$.

Setting $\theta_1 = 0$ gives the plane of reflectional symmetry, with $\hat{\theta}_3$ giving the direction of the tangent to $\mathcal{C}$ and $\mathcal{T}$ at the tricritical point. The scaling hypothesis takes the form

$$\phi_{\text{sing}}(\lambda^{y_1}\theta_1, \lambda^{y_2}\theta_2, \lambda^{y_3}\theta_3) = \lambda^d \phi_{\text{sing}}(\theta_1, \theta_2, \theta_3), \quad (4.7.8)$$

for $\lambda > 0$, (Riedel 1972; Hankey et al. 1972) and the general shapes of the critical curves in a neighbourhood of the tricritical point can be inferred from the relative magnitudes of y_1 , y_2 and y_3 . A consequence of (4.7.8) is that, if ϕ has non-smooth behaviour of a particular type at $(\theta_1, \theta_2, \theta_3) = (\kappa_1, \kappa_2, \kappa_3)$, then it will also have the same non-smooth behaviour at

$$(\theta_1, \theta_2, \theta_3) = (\kappa_1 \lambda^{y_1}, \kappa_2 \lambda^{y_2}, \kappa_3 \lambda^{y_3}). \quad (4.7.9)$$

It follows that the critical lines $\mathcal{C}$, $\mathcal{C}^{(+)}$ and $\mathcal{C}^{(-)}$ and the triple line $\mathcal{T}$ must all have parametric equations of the form (4.7.9), for some different sets of parameters $(\kappa_1, \kappa_2, \kappa_3)$ and $\lambda > 0$. We shall take the set of parameters for $\mathcal{C}$ to be $(\kappa_1, \kappa_2, \kappa_3)$, for $\mathcal{C}^{(\pm)}$ to be $(\kappa_1^{(\pm)}, \kappa_2^{(\pm)}, \kappa_3^{(\pm)})$ and for $\mathcal{T}$ to be (τ_1, τ_2, τ_3) . It follows from the symmetry of the model and the choice of scaling fields that $\kappa_1 = \tau_1 = 0$, $\kappa_1^{(+)} = -\kappa_1^{(-)} > 0$, $\kappa_2^{(+)} = \kappa_2^{(-)}$ and $\kappa_3^{(+)} = \kappa_3^{(-)}$. It will be assumed that in each of these triplets not more than one member is zero.³⁰ For $\mathcal{C}$ to approach the tricritical point in the direction of $\hat{\theta}_3$ it must be the case that the *tricritical gap exponent*

$$\vartheta_t := y_2/y_3 > 1. \quad (4.7.10)$$

It follows, by eliminating λ from (4.7.9), that the equation of $\mathcal{C}$ in a neighbourhood of the tricritical point is

$$\theta_2 = \varpi_0 |\theta_3|^{\vartheta_t}, \quad (4.7.11)$$

where $\varpi_0 = \kappa_2/|\kappa_3|^{\vartheta_t}$. From Eq. (4.7.6),

$$\theta_2 \left(1 + a \frac{\theta_3}{\theta_2}\right) = (1 + a^2) \Delta L, \quad \theta_3 \left(1 - a \frac{\theta_2}{\theta_3}\right) = (1 + a^2) \Delta K \quad (4.7.12)$$

and, by the same reasoning as that employed in Sect. 4.5, it follows, as in the case of a critical point, that the shift exponent $\psi_t = \vartheta_t$. The way in which the curves $\mathcal{C}^{(\pm)}$ approach the tricritical point is determined by the magnitude of y_1 relative to y_2 and y_3 and whether one pair of $\kappa_2^{(\pm)}$ or $\kappa_3^{(\pm)}$ is zero.

³⁰ Otherwise, with the definitions (4.7.6) for the scaling fields, the corresponding curve would be linear and in the direction of one of the unit vectors (4.7.7).

4.7.2 Tricritical Exponents and Scaling Laws

We now consider the asymptotic forms for the densities m and q and the response functions φ_T , ψ_T , $c_{m,q}$, $c_{m,\mathcal{H}_s}$ and $c_{\mathcal{H},\mathcal{H}_s}$ at the tricritical point. Since we are, for all except the exponent δ_t defined by (4.6.15), considering asymptotic forms in the plane $\Theta_1 = 0$ we shall, for simplicity, in the following formulae apply this condition.³¹ From (4.6.2), (4.6.3), (4.6.5) and (4.7.4)–(4.7.6),

$$m - m_t = - \left[\frac{\partial \phi_{\text{sing}}}{\partial \theta_2} + a \frac{\partial \phi_{\text{sing}}}{\partial \theta_3} \right], \quad (4.7.13)$$

$$q = - \frac{\partial \phi_{\text{sing}}}{\partial \theta_1}, \quad (4.7.14)$$

$$\varphi_T = - \left[\frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} + 2a \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2 \partial \theta_3} + a^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_3^2} \right], \quad (4.7.15)$$

$$\psi_T = - \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2}, \quad (4.7.16)$$

$$c_{\mathcal{H},\mathcal{H}_s} = -k_B \left[\Theta_2^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} + 2\Theta_2 \Theta_3 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2 \partial \theta_3} + \Theta_3^2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_3^2} \right]. \quad (4.7.17)$$

For $c_{m,\mathcal{H}_s}$ and $c_{m,q}$ we use (4.6.10) and (4.6.11). Thus we need the expressions

$$\vartheta_T = a \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} - \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_3}, \quad (4.7.18)$$

$$\alpha_{\mathcal{H},\mathcal{H}_s} = \Theta_2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} + (a\Theta_2 + \Theta_3) \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2 \partial \theta_3} + a\Theta_3 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_3^2}, \quad (4.7.19)$$

$$\beta_{\mathcal{H},\mathcal{H}_s} = \Theta_2 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_2} + \Theta_3 \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1 \partial \theta_3} \quad (4.7.20)$$

giving

$$c_{m,\mathcal{H}_s} = - \frac{k_B}{\varphi_T} (a\Theta_2 - \Theta_3)^2 \left[\left(\frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2 \partial \theta_3} \right)^2 - \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_2^2} \frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_3^2} \right]. \quad (4.7.21)$$

The exponents associated with m , q and the response functions for different paths of approach are given by Eqs. (4.6.15)–(4.6.22). Following the procedure used for the

³¹ The formula (4.7.14), for q is unaffected by this assumption.

general scaling formulation in Sect. 4.3 and for the critical point in Sect. 4.5 we take $\lambda = |\theta_1|^{-1/y_1}$, $\lambda = |\theta_2|^{-1/y_2}$ and $\lambda = |\theta_3|^{-1/y_3}$ in Eq. (4.7.8) to give respectively

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) = |\theta_1|^{d/y_1} \phi_{\text{sing}}(\pm 1, g_2^{(1)}, g_3^{(1)}), \quad (4.7.22)$$

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) = |\theta_2|^{d/y_2} \phi_{\text{sing}}(g_1^{(2)}, \pm 1, g_3^{(2)}), \quad (4.7.23)$$

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) = |\theta_3|^{d/y_3} \phi_{\text{sing}}(g_1^{(3)}, g_2^{(3)}, \pm 1), \quad (4.7.24)$$

where the gap fields $g_j^{(m)}(\theta_m, \theta_j)$ and exponents $\vartheta_j^{(m)}$, $m, j = 1, 2, 3$, are given by (4.3.13) and (4.3.14). It follows from (4.7.11) that, in a neighbourhood of the tricritical point, the equation of $\mathcal{C}$ in the plane $\theta_1 = 0$ is $g_2^{(3)}(\theta_2, \theta_3) = \varpi_0$. From (4.7.22), with $\theta_2 = \theta_3 = 0$, using similar reasoning to that used in Sect. 4.5, it follows that the exponent δ_t , defined in (4.6.15), is given by

$$\delta_t = y_1/(d - y_1). \quad (4.7.25)$$

A classification of the paths of approach to the tricritical point in the plane $L_s = \Theta_1 = 0$ can now be made in a similar way to that for the critical point in Sect. 4.5. The two types of approach are:

- (i) Paths on which $|g_3^{(2)}| \rightarrow 0$, including the direction of the vector $\hat{\theta}_2$, for which, from (4.7.23),

$$\phi_{\text{sing}}(0, \theta_2, \theta_3) \simeq |\theta_2|^{d/y_2} A_{\infty}^{(\pm)}, \quad (4.7.26)$$

where

$$\lim_{g_3^{(2)} \rightarrow 0} \phi_{\text{sing}}(0, \pm 1, g_3^{(2)}) = A_{\infty}^{(\pm)} \quad (4.7.27)$$

is assumed to be finite and non-zero.

- (ii) Paths on which $|g_2^{(3)}| \rightarrow \pi_2$, where π_2 is finite (possibly zero), when

$$\phi_{\text{sing}}(0, \theta_2, \theta_3) \simeq |\theta_3|^{d/y_3} A_{\pi_2}^{(\pm)}, \quad (4.7.28)$$

where

$$\lim_{g_2^{(3)} \rightarrow \pi_2} \phi_{\text{sing}}(0, g_2^{(3)}, \pm 1) = A_{\pi_2}^{(\pm)} \quad (4.7.29)$$

is assumed to be finite and non-zero.

Since $\vartheta_t > 1$, all curves in class (ii), including $\mathcal{C}$, are tangential to the direction of the unit vector $\hat{\theta}_3$. Some curves in class (i) are also tangential to $\hat{\theta}_3$, but all curves not tangential to this vector are certainly in class (i). The asymptotic forms in equations

(4.6.16)–(4.6.21) with $\Theta_3 = \Theta_{3t}$ fall into class (i) and those with $\Theta_2 = \Theta_{2t}$ into class (ii). Thus from (4.6.19), (4.6.21), (4.7.14) and (4.7.16)

$$\beta_t = (d - y_1)/y_2, \quad \beta_{tc} = \vartheta_t \beta_t = (d - y_1)/y_3, \quad (4.7.30)$$

$$\gamma'_t = (2y_1 - d)/y_2, \quad \gamma_{tc} = \vartheta_t \gamma_t = (2y_1 - d)/y_3. \quad (4.7.31)$$

Since the density m and response functions φ_T and $c_{\mathcal{H}, \mathcal{H}_s}$ arise as derivatives of ϕ with respect to couplings K and L , which appear as linear combinations in the scaling field θ_2 and θ_3 , we must consider, as we did in Sect. 4.3.9, the relative importance of different terms. From (4.6.18), (4.6.20), (4.6.22), (4.7.13), (4.7.15), (4.7.17) and (4.7.21), since $y_2 > y_3$,

$$\beta'_{2t} = (d - y_2)/y_2, \quad \beta_{2tc} = \vartheta_t \beta_{2t} = (d - y_2)/y_3, \quad (4.7.32)$$

$$\gamma'_{2t} = (2y_2 - d)/y_2, \quad \gamma_{2tc} = \vartheta_t \gamma_{2t} = (2y_2 - d)/y_3, \quad (4.7.33)$$

$$\sigma'_{2t} = (2y_2 - d)/y_2, \quad \sigma_{tc} = \vartheta_t \sigma_t = (2y_2 - d)/y_3, \quad (4.7.34)$$

$$\alpha'_{2t} = (2y_3 - d)/y_2, \quad \alpha_{2tc} = \vartheta_t \alpha_{2t} = (2y_3 - d)/y_3. \quad (4.7.35)$$

That $\sigma'_t = \gamma'_{2t}$ and $\sigma_{tc} = \gamma_{2tc}$ could be anticipated from our calculations for the critical point, in Sect. 4.5.6, since the system has no symmetry with respect to the conjugate variables m and $\mathcal{H}$. The formula we used for deriving (4.7.21) and thus the exponents (4.7.35) was (4.6.10). The difference between the exponent for $c_{\mathcal{H}, \mathcal{H}_s}$ and $c_{m, \mathcal{H}_s}$ is a consequence of the cancellation of the most singular term in $c_{\mathcal{H}, \mathcal{H}_s}$ with that contained in the right-hand side of (4.6.10). We apply a similar procedure, using (4.6.11), to obtain the exponents for $c_{m, q}$. In this case, some calculation shows that no cancellation occurs. The exponents are the same as those for $c_{\mathcal{H}, \mathcal{H}_s}$, namely,

$$\alpha'_t = (2y_2 - d)/y_2, \quad \alpha_{tc} = \vartheta_t \alpha'_t = (2y_2 - d)/y_3. \quad (4.7.36)$$

From these formulae relating critical exponents to scaling exponents it follows that

$$2 = \alpha'_t + 2 \beta_t + \gamma'_t, \quad 2 = \alpha_{2tc} + 2 \beta_{2tc} + \gamma_{2tc}, \quad (4.7.37)$$

which are versions of the Essam–Fisher scaling law (4.5.44), and

$$\gamma'_t = \beta_t(\delta_t - 1), \quad \gamma_{tc} = \beta_{tc}(\delta_t - 1), \quad (4.7.38)$$

which are versions of the Widom scaling law (4.5.45). We also have

$$\gamma'_{2t} = \beta_{2t}(\delta_{2t} - 1), \quad \gamma_{2tc} = \beta_{2tc}(\delta_{2t} - 1), \quad (4.7.39)$$

where $\delta_{2t} = 1/\beta_{2t}$ is the exponent which would appear in a formula like (4.6.15) with $L - L_t$ replacing L_s and $m - m_t$ replacing q .

Since, in the plane $\Theta_3 = \Theta_{3t}$, both the critical isochore $q = 0$ and the coexistence curve lie along the line $\Theta_1 = 0$, scaling predicts the equality of the exponents in the ordered and disordered regions. That is³²

$$\gamma_t = \gamma'_t, \quad \beta_{2t} = \beta'_{2t}, \quad \gamma_{2t} = \gamma'_{2t}, \quad \alpha_{2t} = \alpha'_{2t}. \quad (4.7.40)$$

In terms of the scaling formulation of Sect. 4.3, the tricritical point is a transition region connected to seven other transition regions; the critical curves $\mathcal{C}$, $\mathcal{C}^{(+)}$ and $\mathcal{C}^{(-)}$, the triple curve $\mathcal{T}$ and the three first-order surfaces. An analysis of scaling along these curves can be made using effective exponents as described in Sect. 4.3.3.

4.8 Corrections to Scaling

In Sect. 4.3.4 we considered one case where the simple power-law asymptotic behaviour of equations (4.3.20) broke down. This arose from the exponent ω_m being zero, and led to a logarithmic divergence in the corresponding derivative of ϕ . In this section modifications to scaling which arise from the amplitude factors are considered. For simplicity we concentrate on the case $n = 3$ with a transition region $\mathcal{C}$ with $s = 2$. That is to say, we have a system with two relevant fields θ_1 and θ_2 and one irrelevant field θ_3 . From (4.3.10)–(4.3.13),

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) = |\theta_1|^{\sigma_1} \phi_{\text{sing}}\left(\pm 1, \frac{\theta_2}{|\theta_1|^{\vartheta_2^{(1)}}}, \frac{\theta_3}{|\theta_1|^{\vartheta_3^{(1)}}}\right), \quad (4.8.1)$$

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) = |\theta_2|^{\sigma_2} \phi_{\text{sing}}\left(\frac{\theta_1}{|\theta_2|^{\vartheta_1^{(2)}}}, \pm 1, \frac{\theta_3}{|\theta_2|^{\vartheta_3^{(2)}}}\right). \quad (4.8.2)$$

In Sect. 4.3 we considered the asymptotic forms which can be derived from (4.8.1) and (4.8.2). In each case it was necessary to choose a path of approach to the transition region so that the arguments of the functions on the right-hand sides tend to finite values. Even then, in order to obtain the simple form of scaling described in Sect. 4.3 and utilized for a critical point and tricritical point in Sects. 4.5 and 4.7 respectively, it is necessary that the amplitude factors are finite and non-zero. We begin by considering the functions

$$f^{(\pm)}(u, v) := \phi_{\text{sing}}(\pm 1, u, v), \quad (4.8.3)$$

³² However, of course, the same warning applies as that given for the critical point in Sect. 4.5.6. A zero amplitude in an asymptotic form can cause the equality of the exponents to break down. It is also the case that if the model were generalized to eliminate the symmetry about $\Theta_1 = 0$ the equality of the exponents would need to be established analytically as it was for the critical point.

which appear on the right-hand side of (4.8.1). The limit $u \rightarrow 0$ with v small and non-zero can be achieved by taking $\theta_2 \rightarrow 0$ with θ_1 and θ_3 non-zero. Excluding the possibility that $\theta_2 = 0$ is some transition region connected to $\mathcal{C}$, $f^{(\pm)}(u, v)$ have power-series expansions of the form

$$f^{(\pm)}(u, v) = w_0^{(\pm)}(v) + uw_1^{(\pm)}(v) + \dots \quad (4.8.4)$$

If $w_0^{(\pm)}(0)$ is finite and non-zero then the power-law behaviour is that predicted by simple scaling theory, although the magnitude of $w_1^{(\pm)}(0)$ will affect the range over which the leading scaling behaviour can be observed. If $w_0^{(\pm)}(0) = 0$ there is a breakdown of simple scaling and the power-law predictions rest on the first non-zero term in the expansion. If this is given by $w_1^{(\pm)}(0)$ being finite and non-zero, then the singular behaviour is not only weaker but will depend on the next layer of detail of the path of approach $\mathcal{P}$. A possible $\mathcal{P}$ consistent with the condition $u \rightarrow 0$ is $\theta_2 \sim |\theta_1|^\tau$, where $\tau > \vartheta_2^{(1)}$. Then from (4.8.1)

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) \sim |\theta_1|^{\sigma_1 + \tau - \vartheta_2^{(1)}}, \quad (4.8.5)$$

which is a weaker power-law behaviour. The same analysis can be applied to (4.8.2).

Consider now the possibility of $f^{(\pm)}(0, 0)$ being infinite. It is clear that, for an isolated transition region, $f^{(\pm)}(0, v)$ cannot be infinite and thus, any singular behaviour must be arise from $w_0^{(\pm)}(v)$, therefore involving the irrelevant variable θ_3 . That this is a possibility was pointed out by Fisher (1983), who classified irrelevant variables as *dangerous* or *harmless* according as they do or do not yield such divergent behaviour. Suppose that

$$w_0^{(\pm)}(v) \sim |v|^{-\tau}, \quad (4.8.6)$$

where $\tau > 0$. Then, from (4.8.1)–(4.8.4),

$$\phi_{\text{sing}}(\theta_1, \theta_2, \theta_3) \sim |\theta_1|^{\bar{\sigma}_1}, \quad (4.8.7)$$

where

$$\bar{\sigma}_1 := \frac{\bar{d}}{y_1}, \quad (4.8.8)$$

and

$$\bar{d} := d - \tau|y_3|. \quad (4.8.9)$$

The effect of the dangerous irrelevant variable θ_3 can be viewed as an effective reduction in the dimension of the system from d to $\bar{d}$, provided that all paths to the

transition region produce the same change. This means that we must be satisfied that a parallel analysis of (4.8.2) yields the same formula (4.8.9). In Sect. 5.6.3 the significance of the presence of a dangerous irrelevant variable for the interpretation of Gaussian critical exponents in relation to scaling theory is discussed.

4.9 Scaling and Universality

The universality hypothesis, which is described in Sect. 4.2.1, and the scaling hypotheses of Sects. 4.3.1 and 4.3.4 both apply to a particular transition region rather than to a physical system or theoretical model as a whole. Different transition regions of the same system may be in different universality classes and may have different scaling functions and exponents. Discussion of the relationship between scaling and universality is made easier if we take a particular type of transition region. The simplest example to choose is a critical point with power-law scaling. In the weak scaling direction, from (4.5.16), $\theta_1/\theta_2 \rightarrow 0$ as the critical point is approached and thus, from (4.5.8) and (4.5.23),

$$\theta_1 = \Delta L - a\Delta K, \quad \theta_2 \simeq (1 + a^2)\Delta K. \quad (4.9.1)$$

From Eqs. (4.5.11), (4.5.12), (4.5.40), (4.5.39) and (4.5.59), the forms for the singular part of ϕ , the density difference from its critical value and the response function φ_T can be expressed as

$$\phi_{\text{sing}} = A_0 |\Delta K|^{2-\alpha} W_0^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}} \right), \quad (4.9.2)$$

$$\Delta\rho = -\frac{\partial\phi_{\text{sing}}}{\partial\Delta L} = A_1^{(\pm)} |\Delta K|^{\beta} W_1^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}} \right), \quad (4.9.3)$$

$$\varphi_T = -\frac{\partial^2\phi_{\text{sing}}}{\partial\Delta L^2} = A_2^{(\pm)} |\Delta K|^{-\gamma} W_2^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}} \right), \quad (4.9.4)$$

where C , A_0 , $A_1^{(\pm)}$ and $A_2^{(\pm)}$ are constants and

$$A_s^{(\pm)} W_s^{(\pm)}(x) = -C^s A_0 \frac{d^s W_0^{(\pm)}(x)}{dx^s}, \quad s = 1, 2. \quad (4.9.5)$$

By imposing the conditions

$$W_0^{(\pm)}(0) = W_1^{(\pm)}(0) = W_2^{(\pm)}(0) = 1, \quad (4.9.6)$$

we ensure that A_0 , $A_1^{(\pm)}$ and $A_2^{(\pm)}$ are the critical amplitudes with

$$A_s^{(\pm)} = C^s A_0 Q_s^{(\pm)},$$

$$Q_s^{(\pm)} = - \left(\frac{d^s W_0^{(\pm)}(x)}{dx^s} \right)_{x=0} \quad s = 1, 2. \quad (4.9.7)$$

Each of the Eqs.(4.9.2)–(4.9.4) represents a family of functions in two variables, $(\phi_{\text{sing}}, \Delta L)$, $(\Delta \rho, \Delta L)$, and $(\varphi_T, \Delta L)$ respectively, parameterized by ΔK . The scaling-hypothesis implication of the form of these equations is that each family of isotherms (ΔK constant) ‘collapses’ onto one curve when plotted in the space of $(\Delta L - a\Delta K)/|\Delta K|^{\beta\delta}$ against $\phi_{\text{sing}}/|\Delta K|^{2-\alpha}$, $\Delta \rho/|\Delta K|^\beta$ and $\varphi_T/|\Delta K|^\gamma$ respectively. Equation (4.9.3) is the equation of state. For a ferromagnetic critical point $a = 0$, $\Delta L := \mathcal{H}/T \simeq \mathcal{H}/T_c$, $\Delta K := K - K_c \simeq J(T_c - T)/T_c^2$ and $\Delta \rho = m$, the magnetization per lattice site. The validity of an equation of the form (4.9.3) has been investigated for nickel by Arrott and Noakes (1967), Kouvel and Rodbell (1967) and Kouvel and Comly (1968), for CrO_2 by Kouvel and Rodbell (1967) and for CrBr_3 by Ho and Litster (1969). Similar tests for a number of fluid systems have been made by Green et al. (1967). In all cases the experimental data gives strong support for the scaling hypothesis.

Although the critical exponent equalities (4.5.47) and the Essam–Fisher and Widom scaling laws (4.5.44) and (4.5.45) are analytic consequences of (4.9.2)–(4.9.4), they can be used as a direct test of scaling for theoretical models and experimental data. The well-known classical values for the purely thermodynamic exponents given in (5.3.30) satisfy these scaling laws. They arise for systems of any dimension when the free energy is assumed to be smooth in a neighbourhood of the critical point. In particular they follow from Landau theory, as we shall see in Sect. 5.3. Exact calculations for the two-dimensional Ising model give $\alpha = \alpha' = 0$, $\beta = \frac{1}{8}$ and $\gamma = \gamma' = \frac{7}{4}$, where the value of α corresponds to a logarithmic singularity (Sects. 12.2.7.3 and 12.3.9). These values satisfy the Essam–Fisher scaling law (4.5.44). If the Widom scaling law (4.5.45) is assumed then $\delta = 15$, which is the generally accepted value (Baxter 1982b). Ho and Litster (1969) obtained the values $\gamma = \gamma' = 1.215 \pm 0.015$, $\beta = 0.368 \pm 0.005$ and $\delta = 4.28 \pm 0.1$ for CrBr_3 , which yields $\gamma' - \beta(\delta - 1) = 0.00746 \pm 0.0682$.

The universality hypothesis was presented in Sect. 4.2.1 as a list of criteria for dividing critical problems into universality classes. This hypothesis can be tested when it is clear what is understood by the critical properties of two problems being ‘the same’. This is usually taken to mean that the problems have the same values for critical exponents and the same scaling functions to within metric factors (Fisher 1983). For Eq. (4.9.2)–(4.9.4), this means that the functions $W_s^{(\pm)}(x)$, $s = 0, 1, 2$ and the exponents α , β , γ and δ are the same for all problems in the same universality class but the constants C , $A_s^{(\pm)}$, $s = 0, 1, 2$ may be different. As indicated above, the work of Green et al. (1967) represents a test of scaling. It can also be taken as a test of universality, which is clearly supported by their results which show that the branches of the coexistence curves of a number of fluids lie on a single universal curve.

So far in this section we have concentrated the discussion on thermodynamic functions and variables. Now we consider the correlation function and correlation length for the statistical mechanical variable $q(\mathbf{r})$ associated with the coupling L . Let $\Delta q(\mathbf{r}) := q(\mathbf{r}) - \rho_c$. Then

$$\langle \Delta q(\mathbf{r}) \rangle = \langle q(\mathbf{r}) \rangle - \rho_c = \rho - \rho_c = \Delta \rho, \quad (4.9.8)$$

$$\langle \Delta q(\mathbf{r}_1) \Delta q(\mathbf{r}_2) \rangle = \langle q(\mathbf{r}_1) q(\mathbf{r}_2) \rangle - 2\rho\rho_c + \rho_c^2 \quad (4.9.9)$$

and, from (2.5.9) with $c_1 = c_2 = q$,

$$\Gamma_2(q(\mathbf{r}_1), q(\mathbf{r}_2)) = \langle \Delta q(\mathbf{r}_1) \Delta q(\mathbf{r}_2) \rangle - \langle \Delta q(\mathbf{r}_1) \rangle \langle \Delta q(\mathbf{r}_2) \rangle. \quad (4.9.10)$$

As in Sect. 4.5.7 we assume that the system is translational invariant so that the connected two-point correlation function is dependent on the single spatial variable $\bar{\mathbf{r}} = \mathbf{r}_2 - \mathbf{r}_1$. Then, from (4.5.48), (4.5.49) and (4.9.1),

$$\Gamma_2(q(\mathbf{0}), q(\bar{\mathbf{r}})) = B^{(\pm)} |\Delta K|^{d\nu-\gamma} X^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}}, \frac{\bar{\mathbf{r}} |\Delta K|^\nu}{R_0} \right), \quad (4.9.11)$$

$$\bar{\Gamma}_2^*(q; \bar{\mathbf{k}}) = B^{(\pm)} (R_0^d/v_0) |\Delta K|^{-\gamma} X^{*(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}}, \frac{R_0 \bar{\mathbf{k}}}{|\Delta K|^\nu} \right), \quad (4.9.12)$$

where v_0 is the volume per lattice site and

$$X^{*(\pm)}(x; \bar{\mathbf{k}}) = \sum_{\{\bar{\mathbf{r}}\}} X^{(\pm)}(x; \bar{\mathbf{r}}) \exp(-i\bar{\mathbf{k}} \cdot \bar{\mathbf{r}}). \quad (4.9.13)$$

That the metric factor C in (4.9.11) and (4.9.12) is the same as that in (4.9.2)–(4.9.4) follows from (2.5.16) and (2.5.24), from which it can also be seen that

$$B^{(\pm)} = A_2^{(\pm)} v_0 / R_0^d, \quad (4.9.14)$$

$$X^{*(\pm)}(x; \mathbf{0}) = W_2^{(\pm)}(x). \quad (4.9.15)$$

From (4.5.50),

$$\tau(q) = R_0 Q_3^{(\pm)} |\Delta K|^{-\gamma} Y^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}} \right), \quad (4.9.16)$$

where

$$Q_3^{(\pm)} Y^{(\pm)}(x) = \sqrt{-c(d) \frac{\nabla_{\bar{\mathbf{k}}}^2 X^{*(\pm)}(x, \bar{\mathbf{0}})}{W_2^{(\pm)}(x)}}. \quad (4.9.17)$$

By imposing the condition

$$Y^{(\pm)}(0) = 1, \quad (4.9.18)$$

we have

$$Q_3^{(\pm)} = \sqrt{-c(d)\nabla_{\vec{k}}^2 X^{*(\pm)}(0, \vec{0})}. \quad (4.9.19)$$

The functions $X^{(\pm)}(x; \vec{r})$ are supposed to be universal making $X^{*(\pm)}(x; \vec{k})$ and $Y^{(\pm)}(x)$ also universal. It thus follows, from (4.9.19), that $Q_3^{(\pm)}$, like $Q_1^{(\pm)}$ and $Q_2^{(\pm)}$ defined in (4.9.7), are universal constants.

It would seem that in each of the regions $\Delta K >, < 0$ there are three independent non-universal metric parameters C , A_0 and R_0 . It has, however, been argued by Stauffer et al. (1972) that, at the critical point, the singular part of the dimensionless free energy, in a region of the dimension of the correlation length, is a universal constant. In our notation this quantity is the critical point value of $\tau^d(q)\phi_{\text{sing}}/v_0$. From (4.9.2), (4.9.6), (4.9.16), (4.9.18) and (4.9.19)

$$\frac{\tau^d(q)\phi_{\text{sing}}}{v_0} \sim A_0 R_0^d v_0^{-1} |\Delta K|^{2-\alpha-d\nu} [Q_3^{(\pm)}]^d \quad (4.9.20)$$

at the critical point. When the Josephson hyper-scaling law (4.5.54) is satisfied the right-hand side of (4.9.20) is a constant and the proposition, referred to as the *two-scale factor* or *hyper-universality hypothesis*, is that the quantity $A_0 R_0^d / v_0$ is universal. In Sect. 4.8 we discussed the idea of dangerous irrelevant variables. In Sect. 5.6.3 these ideas will be amplified in relation to the Gaussian critical exponents to show that the hyper-scaling law is satisfied with the ‘real’ physical dimension d when $d \leq d_u$, the upper borderline dimension, which is four for a critical point. For $d > d_u$ the hyper-scaling law is satisfied (at least for the Gaussian exponents) when d is replaced by $\bar{d}$, the dimension reduced by the presence of a dangerous irrelevant variable (Eq. (4.8.9)). The hyper-universality hypothesis is thus expected to be true for $d \leq 4$. This is supported by the results for theoretical models and experimental systems given by Stauffer et al. (1972). Later work (Ferer et al. 1973) shows good agreement for systems with different lattice structures, which on other grounds are taken to be in the same universality class, and Aharony (1974) used a renormalization group ϵ -expansion to establish its validity to order ϵ^2 .

The argument of Privman and Fisher (1984) for the truth of the hyper-universality hypothesis is also based on renormalization group ideas. They conclude that $\langle \Delta q(\mathbf{r}_1) \Delta q(\mathbf{r}_2) \rangle$ and $\Gamma_2(q; \vec{r})$ have scaling forms which may differ in their universal scaling functions, but will have the same non-universal metric factors. Thus

$$\langle \Delta q(\mathbf{r}_1) \Delta q(\mathbf{r}_2) \rangle = B^{(\pm)} |\Delta K|^{d\nu-\gamma} Z^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}}, \frac{\vec{r}|\Delta K|^\nu}{R_0} \right), \quad (4.9.21)$$

and, taking the limit $|\bar{r}| \rightarrow \infty$, it follows that

$$(\Delta\rho)^2 = B^{(\pm)} |\Delta K|^{d\nu-\gamma} Z^{(\pm)} \left(C \frac{[\Delta L - a\Delta K]}{|\Delta K|^{\beta\delta}}, \infty \right). \quad (4.9.22)$$

Comparing Eqs. (4.9.3) and (4.9.22) gives

$$B^{(\pm)} = [A_1^{(\pm)}]^2 \quad (4.9.23)$$

and, from (4.9.7) and (4.9.14),

$$A_0 R_0^d / v_0 = Q_2^{(\pm)} / [Q_1^{(\pm)}]^2. \quad (4.9.24)$$

In Sect. 5.6.3 we compute $A_0 R_0^d / v_0$ in the Gaussian approximation.

The universality hypothesis for the equation of state (4.9.3) is rather similar to the law of corresponding states (Sect. 4.2). One difference is that now the universal scaling function is not analytic in the thermodynamic variables but in non-analytic combinations of them. It also contains non-universal metric parameters which are not simply the critical point parameters. Another important difference is that (4.9.3) applies to the locality of the critical point rather than to the system as a whole. In line with the quote from Levelt Sengers (1974) on p. 94, one may nevertheless regard the universality hypothesis as a result of the same ‘unifying’ impulse which led to the earlier formulation of the law of corresponding states.

4.10 Finite-Size Scaling

All the discussion in this chapter until now has assumed that the system of interest is one for which the thermodynamic limit has been applied, as described in Sect. 1.3 and defined with greater mathematical precision for a one-component lattice fluid in Sect. 8.1. Here we consider a system on the finite d -dimensional hypercubic lattice $\mathcal{N}$ defined in Sect. 18.1.1. The thermodynamic limit when $\mathcal{N} \rightarrow \mathcal{L}$, the corresponding d -dimensional infinite system, is achieved by taking $N_\ell \rightarrow \infty$ simultaneously for every $\ell = 1, 2, \dots, d$, in such a way that the van Hove criterion (8.1.11) is satisfied. Consider now the case where the thermodynamic limit is taken only for the variables N_ℓ , $\ell \leq \mathfrak{d} < d$. This is a system which is infinite in $\mathfrak{d}$ dimensions and finite in $d - \mathfrak{d}$ dimensions. The parameter

$$\aleph := \left[\prod_{\ell=\mathfrak{d}+1}^d N_\ell \right]^{1/(d-\mathfrak{d})} \quad (4.10.1)$$

can be understood as the *thickness* of the system. We denote this system by $\mathcal{L}_{\mathfrak{d}}(\aleph)$. The thermodynamic limit $\mathcal{L}_{\mathfrak{d}}(\aleph) \rightarrow \mathcal{L}$ now means that $\aleph \rightarrow \infty$ with the *shape ratios* $N_\ell / \aleph$, $\ell = \mathfrak{d} + 1, \dots, d$ kept constant.

Two distinguishable phenomena are of interest for $\mathcal{L}_\mathfrak{d}(\aleph)$, surface effects and finite-size effects. The former, which are present only in the absence of periodic boundary conditions, concern the effect on thermodynamic and statistical mechanical properties of the presence of a surface, including the occurrence of surface phase transitions.³³ This is not, however, the subject of this section which will be solely concerned with the changes to bulk scaling which arise from the finite value of $\aleph$. This subject was initiated by Fisher (1971) and Fisher and Barber (1972).³⁴ We have seen in Sect. 4.5.7 that bulk scaling ideas can, with suitable modification, be applied to first-order transitions. Similar developments can also be made to finite-size scaling methods (Fisher and Berker 1982; Privman and Fisher 1983; Binder and Landau 1984; Fisher and Privman 1985). We shall, however, in this section restrict our attention to continuous transitions.

If $\mathfrak{d} = 0$ the system $\mathcal{L}_0(\aleph) = \mathcal{N}$ is completely finite. In terms of some suitable variable like the fugacity the partition function will be a polynomial with positive coefficients and no phase transitions can occur (see p. 253). In spite of this such systems play an important role as approximations to infinite systems in computer simulation methods and in real-space renormalization group finite-lattice methods (Sect. 15.8). When $\mathfrak{d} > 0$, $\mathcal{L}_\mathfrak{d}(\aleph)$ can be regarded as a perhaps somewhat complicated infinite lattice system of $\mathfrak{d}$ dimensions and the type of phase transition behaviour will be characterized by $\mathfrak{d}$ rather than d . For $\mathfrak{d} = 1$ and short-range interactions transfer matrix methods of the type described in Chap. 11 can be applied yielding no phase transitions at non-zero temperatures. The correlation length may, however, (for reasonably larger values of $\aleph$) attain a steep maximum at some values of the couplings. These are termed *pseudocritical values*, (Fisher 1971; Fisher and Ferdinand 1967) associated with an *incipient phase transition* (Sect. 11.2). Many authors (Ree and Chesnut 1966; Runnels and Combs 1971; Lavis 1976) have used such maxima and similar maxima in response functions as indicating the approximate location of the phase transition in the corresponding infinite system. The pseudocritical values of the couplings will be functions of the thickness $\aleph$ of the system. When $\mathfrak{d} > 1$ the maxima associated with pseudocritical coupling values will be replaced by genuine critical values. These will be dependent on both $\aleph$ and $\mathfrak{d}$, although, according to the discussion of universality in Sects. 4.2 and 4.9, the corresponding exponents will be dependent solely on $\mathfrak{d}$.

4.10.1 The Finite-Size Scaling Field

Consider the situation described in Sect. 4.3 of a system with n couplings and a transition region $\mathcal{C}$ of dimension $n - s$, defined in terms of the scaling fields $\theta_1, \theta_2, \dots, \theta_n$ by (4.3.2). The basic assumption of finite-size scaling theory is that, for any specified boundary conditions, the variable $\theta_{n+1} = 1/\aleph$ can be treated as another scaling field.

³³ For a review, see Binder (1983).

³⁴ For a review see Barber (1983) and, for a collection of papers on finite-size scaling, Cardy (1988).

The defining condition (4.3.2) for $\mathcal{C}$ must then be augmented by

$$\theta_{n+1} = 0. \quad (4.10.2)$$

It follows from the block-spin idea of Kadanoff, described in Sect. 4.2, that the scaling exponent y_{n+1} associated with the finite-size scaling field θ_{n+1} is unity. It is therefore a relevant scaling field. In the case $\mathfrak{d} > 1$ there is a transition region $\mathcal{C}_{\mathfrak{d}}(\aleph)$ for $\mathcal{L}_{\mathfrak{d}}(\aleph)$, which can be regarded as the projection of $\mathcal{C}$ into the extended phase space. Since $\mathfrak{d} \neq d$, the critical exponents for this projection will have different critical exponents from those of $\mathcal{C}$, which is consistent with θ_{n+1} being a relevant scaling field.

For a d -dimensionally infinite system, with $d > 1$, the singular part of ϕ can be identified unambiguously. For a d -dimensional system with $\mathfrak{d} < 2$ this is not the case. To cover both situations Privman and Fisher (1984) proposed that $\phi_{\text{sing}}(K_1, \dots, L_1, \dots, \aleph)$, the singular part of the free-energy density of a system of thickness $\aleph$, is *defined* by

$$\phi_{\text{sing}}(K_1, \dots, L_1, \dots, \aleph) = \phi(K_1, \dots, L_1, \dots, \aleph) - \phi_{\text{smth}}(K_1, \dots, L_1, \dots), \quad (4.10.3)$$

where the second term on the right is the smooth component of ϕ for the $\aleph$ -infinite system. For this singular part of ϕ , and $\psi(\mathbf{k})$ obtained by differentiating it k_j times with respect to θ_j , for $j = 1, 2, \dots, n$, (4.3.5) and (4.3.7) must be modified to give

$$\phi_{\text{sing}}(\lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n, \lambda\theta_{n+1}) = \lambda^d \phi_{\text{sing}}(\theta_1, \dots, \theta_n, \theta_{n+1}), \quad (4.10.4)$$

$$\psi_{\text{sing}}(\mathbf{k}; \lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n, \lambda\theta_{n+1}) = \lambda^{d'(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_n, \theta_{n+1}), \quad (4.10.5)$$

where d' is given by (4.3.8). In a similar way the formula (4.5.51) for the correlation length generalizes to

$$\tau(\mathbf{q}; \lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n, \lambda\theta_{n+1}) = \lambda^{-1} \tau(\mathbf{q}; \theta_1, \dots, \theta_n, \theta_{n+1}). \quad (4.10.6)$$

Substituting $\lambda = 1/\theta_{n+1} = \aleph$ in Eqs. (4.10.4)–(4.10.6) gives

$$\phi_{\text{sing}}(\theta_1, \dots, \theta_{n+1}) = \aleph^{-d} \phi_{\text{sing}}(\theta_1 \aleph^{y_1}, \dots, \theta_n \aleph^{y_n}, 1), \quad (4.10.7)$$

$$\psi_{\text{sing}}(\mathbf{k}; \theta_1, \dots, \theta_{n+1}) = \aleph^{-d'(\mathbf{k})} \psi_{\text{sing}}(\mathbf{k}; \theta_1 \aleph^{y_1}, \dots, \theta_n \aleph^{y_n}, 1), \quad (4.10.8)$$

$$\tau(\mathbf{q}; \theta_1, \dots, \theta_{n+1}) = \aleph \tau(\mathbf{q}; \theta_1 \aleph^{y_1}, \dots, \theta_n \aleph^{y_n}, 1) \quad (4.10.9)$$

and it is also useful to note that

$$\frac{\partial \tau}{\partial \theta_m}(\mathbf{q}; \theta_1, \dots, \theta_{n+1}) = \aleph^{1-y_m} \frac{\partial \tau}{\partial \theta_m}(\mathbf{q}; \theta_1 \aleph^{y_1}, \dots, \theta_n \aleph^{y_n}, 1),$$

$$m = 1, 2, \dots, n. \quad (4.10.10)$$

Equations (4.10.8) and (4.10.10) encapsulate the important practical value of finite-size scaling. If the transition region $\mathcal{C}$ is approached along any path for which the function on the right-hand side has a finite non-zero limit then, from (4.3.8),

$$\lim_{\aleph \rightarrow \infty} \frac{\ln[\psi_{\text{sing}}(\mathbf{k})]}{\ln(\aleph)} = -d'(\mathbf{k}) = \sum_{j=1}^n k_j y_j - d, \quad (4.10.11)$$

$$\lim_{\aleph \rightarrow \infty} \frac{\ln\left(\frac{\partial \tau}{\partial \theta_m}\right)}{\ln(\aleph)} = 1 - y_m, \quad m = 1, 2, \dots, n. \quad (4.10.12)$$

From the second of equations (4.3.12) it can be seen that the variable $d'(\mathbf{k})$ is the numerator of the critical exponent $\omega_m(\mathbf{k})$ associated with the response function (or density) $\psi(\mathbf{k})$ as $\mathcal{C}$ is approached along a path $\mathcal{P}$ which satisfies (4.3.18) with finite values of $\varpi_j^{(m)}$, $j = 1, 2, \dots, s$. The corresponding denominator y_m appears in formula (4.10.12). This means that the critical exponents for the infinite system $\mathcal{L}$ can be obtained approximately from investigating systems of finite thickness. Just as in the case of scaling for infinite systems, modifications to this analysis are needed for the case of logarithmic singularities (Sect. 4.3.4) and when zero amplitude factors imply the need for considering corrections to scaling (Sect. 4.8).

4.10.2 The Shift and Rounding Exponents

We now investigate in more detail the relationship between the transition region $\mathcal{C}$ for the d -dimensionally infinite system and the transition region $\mathcal{C}_0(\aleph)$ for the system of thickness $\aleph$. For simplicity we confine our attention to the case of a critical point, with two couplings L and K , where the approach to the critical point is in the weak scaling direction. For the infinite system $\mathcal{L}$ the scaling fields θ_1 and θ_2 are given by Eqs. (4.9.1) and from (4.9.4) the response function φ_T is given by

$$\varphi_T = -\frac{\partial^2 \phi_{\text{sing}}}{\partial \theta_1^2}. \quad (4.10.13)$$

Thus, from (4.3.8) and (4.10.8),

$$\varphi_T(\theta_1, \theta_2, \theta_3) = \aleph^{2y_1-d} \varphi_T(\theta_1 \aleph^{y_1}, \theta_2 \aleph^{y_2}, 1). \quad (4.10.14)$$

When $\mathfrak{d} > 1$, φ_T will exhibit some kind of singular behaviour in the transition region $\mathcal{C}_{\mathfrak{d}}(\aleph)$. Since $\aleph$ is finite and non-zero on $\mathcal{C}_{\mathfrak{d}}(\aleph)$ it follows that this transition region for systems of finite width must be given by some finite values ϖ_j for the arguments $\theta_j \aleph^{y_j}$, $j = 1, 2$ of the function on the right-hand side of (4.10.14). The transition region is, therefore, given by

$$[L_c(\mathfrak{d}; \aleph) - L_c] - a[K_c(\mathfrak{d}; \aleph) - K_c] \simeq \varpi_1 \aleph^{-y_1}, \quad (4.10.15)$$

$$K_c(\mathfrak{d}; \aleph) - K_c \simeq (1 + a^2)^{-1} \varpi_2 \aleph^{-y_2}, \quad (4.10.16)$$

where $L_c(\mathfrak{d}; \aleph)$ and $K_c(\mathfrak{d}; \aleph)$ are the critical values of L and K for the system of dimension $\mathfrak{d}$ and thickness $\aleph$. In this context the exponent y_2 which appears on the right-hand side of (4.10.16) is called the *shift exponent* (Fisher and Ferdinand 1967; Fisher 1971) and is usually denoted by λ . According to (4.5.53)

$$\lambda = y_2 = 1/\nu. \quad (4.10.17)$$

In the case $\mathfrak{d} = 1$, $L_c(\mathfrak{d}; \aleph)$ and $K_c(\mathfrak{d}; \aleph)$ are replaced by $L_{\max}(\mathfrak{d}; \aleph)$ and $K_{\max}(\mathfrak{d}; \aleph)$, the values where φ_T attains a maximum. A second value of K which is of interest (Fisher and Ferdinand 1967) is the *rounding value* $K^*(\mathfrak{d}; \aleph)$ at which φ_T for $\mathcal{L}_{\mathfrak{d}}(\aleph)$ differs significantly from its value for $\mathcal{L}$. A *rounding exponent* w can then be defined by³⁵

$$|K^*(\mathfrak{d}; \aleph) - K_c| \sim \aleph^{-w}. \quad (4.10.18)$$

If, following Fisher and Ferdinand (1967), we suppose that this will occur when $\tau(q) \sim \aleph$ then, from (4.5.52) and (4.10.18),

$$w = 1/\nu. \quad (4.10.19)$$

The relationship (4.10.17) between the shift exponent and ν is a consequence of the finite-size scaling assumption (4.10.4) for the singular part of the free-energy density as long as the coefficient ϖ_2 in (4.10.16) is non-zero. If we include the possibility that $\varpi_2 = 0$, then corrections to scaling lead to

$$\lambda \geq 1/\nu. \quad (4.10.20)$$

For the two-dimensional Ising model (Ferdinand and Fisher 1969; Au-Yang and Fisher 1975) $\lambda = w = 1/\nu = 1$. Fisher and Ferdinand (1967) have, however, argued for the possibility that $\lambda > 1/\nu$ in three dimensions and the results of Barber and Fisher (1973a, b) for the spherical model and ideal Bose gas respectively in d dimensions with layer geometries and open boundary conditions gave $\lambda = 1 > 1/\nu$.

³⁵ The rounding exponent is usually denoted by θ . We have used w to avoid confusion with the use of θ for scaling fields.

This was later understood to be a pathological result arising from constraints which give an effective long-range interaction (Brézin 1983) and (4.10.17) is now thought to be generally true (Barber 1983).

As indicated above, when $\mathfrak{d} > 1$, the transition regions $\mathcal{C}$ and $\mathcal{C}_0(\mathfrak{N})$ are distinct but connected in phase space. In this sense the situation is somewhat similar to the relationship between the first-order line and the critical point in the ferromagnetic Ising model. A situation of this kind is analyzed using effective exponents in Sect. 4.3.3.

4.10.3 Universality and Finite-Size Scaling

As in Sect. 4.9 we concentrate on the case of a critical point. From (4.10.4) and (4.10.6), the scaling forms (4.9.2) and (4.9.16) are replaced by

$$\phi_{\text{sing}} = A_0 |\Delta K|^{2-\alpha} \mathring{W}_0^{(\pm)} \left(C \frac{[\Delta L - a \Delta K]}{|\Delta K|^{\beta \delta}}, \frac{D}{|\Delta K|^{\nu \mathfrak{N}}} \right), \quad (4.10.21)$$

$$\mathfrak{r}(\mathfrak{q}) = R_0 Q_3^{(\pm)} |\Delta K|^{-\nu} \mathring{Y}^{(\pm)} \left(C \frac{[\Delta L - a \Delta K]}{|\Delta K|^{\beta \delta}}, \frac{D}{|\Delta K|^{\nu \mathfrak{N}}} \right), \quad (4.10.22)$$

with the bulk forms being recovered, in the limit $\mathfrak{N} \rightarrow \infty$, with

$$\mathring{W}_0^{(\pm)}(x, 0) := W_0^{(\pm)}(x), \quad \mathring{Y}^{(\pm)}(x, 0) := Y^{(\pm)}(x). \quad (4.10.23)$$

When the Josephson hyper-scaling law (4.5.54) is valid, (4.10.21) can be rescaled.

$$\begin{aligned} \phi_{\text{sing}} &= \frac{A_0 D^d}{\mathfrak{N}^d} \left[\frac{|\Delta K| \mathfrak{N}^{\frac{1}{\nu}}}{D^{\frac{1}{\nu}}} \right]^{2-\alpha} \\ &\times \mathring{W}_0^{(\pm)} \left(\left[\frac{C}{D^{\frac{\beta \delta}{\nu}}} \right] [(\Delta L - a \Delta K) \mathfrak{N}^{\frac{\beta \delta}{\nu}}] \left[\frac{|\Delta K| \mathfrak{N}^{\frac{1}{\nu}}}{D^{\frac{1}{\nu}}} \right]^{-\beta \delta}, \left[\frac{|\Delta K| \mathfrak{N}^{\frac{1}{\nu}}}{D^{\frac{1}{\nu}}} \right]^{-\nu} \right), \end{aligned} \quad (4.10.24)$$

gives

$$\phi_{\text{sing}} = A_0 D^d \mathfrak{N}^{-d} U(C' [\Delta L - a \Delta K] \mathfrak{N}^{\frac{\beta \delta}{\nu}}, D' \Delta K \mathfrak{N}^{\frac{1}{\nu}}), \quad (4.10.25)$$

where

$$U(x, \pm y) := y^{2-\alpha} \mathring{W}_0^{(\pm)}(xy^{-\beta \delta}, y^{-\nu}), \quad (4.10.26)$$

$$C' := C/D^{\frac{\beta\delta}{\nu}}, \quad D' := 1/D^{\frac{1}{\nu}}. \quad (4.10.27)$$

In a similar way (4.10.22) can be re-expressed in the form

$$\tau(q) = \frac{R_0 Q_3^{(\pm)}}{D} \aleph V(C'[\Delta L - a\Delta K] \aleph^{\frac{\beta\delta}{\nu}}, D' \Delta K \aleph^{\frac{1}{\nu}}), \quad (4.10.28)$$

where

$$V(x, \pm y) := y^{-\nu} \mathring{Y}^{(\pm)}(xy^{-\beta\delta}, y^{-\nu}). \quad (4.10.29)$$

Equations (4.10.25) and (4.10.28) could be derived directly from (4.10.7) and (4.10.9) except that then no relationship would be established with the metric factors in (4.10.21) and (4.10.22).

As in the case of the d -dimensionally infinite system discussed in Sect. 4.9 the scaling functions $\mathring{W}_0^{(\pm)}(x, y)$, $\mathring{Y}^{(\pm)}(x, y)$, $U(x, y)$ and $V(x, y)$ are supposed to be universal with non-universal metric factors A_0 , C , D and R_0 . However, we can in this case go further if we adopt a standard assumption of finite-size scaling theory (Barber 1983) that all lengths which diverge at bulk criticality should scale like the bulk correlation length. This means that the second argument in (4.10.21) and (4.10.22) $D/|\Delta K|^{\nu} \aleph \simeq \tau(q)/v_0^{\frac{1}{d}} \aleph$. From (4.9.16) $D \simeq R_0 Q_3^{(\pm)}/v_0^{\frac{1}{d}}$. Substituting into (4.10.25) and (4.10.28) and using the hyper-universality formula (4.9.24) gives

$$\phi_{\text{sing}} = E_0 \aleph^{-d} U(C'[\Delta L - a\Delta K] \aleph^{\frac{\beta\delta}{\nu}}, D' \Delta K \aleph^{\frac{1}{\nu}}), \quad (4.10.30)$$

$$\tau(q) = v_0^{\frac{1}{d}} \aleph V(C'[\Delta L - a\Delta K] \aleph^{\frac{\beta\delta}{\nu}}, D' \Delta K \aleph^{\frac{1}{\nu}}), \quad (4.10.31)$$

where

$$E_0 := A_0 D^d = Q_2^{(\pm)} [Q_3^{(\pm)}]^d / [Q_1^{(\pm)}]^2. \quad (4.10.32)$$

That the scaling forms (4.10.30) and (4.10.31) have leading universal amplitude factors is a result derived by Privman and Fisher (1984). At the bulk critical point $\Delta L = \Delta K = 0$ (4.10.30) and (4.10.31) becomes

$$\phi_{\text{sing}} = \aleph^{-d} E_0 U(0, 0), \quad (4.10.33)$$

$$\tau(q)/v_0^{\frac{1}{d}} = \aleph V(0, 0). \quad (4.10.34)$$

As we shall see in Sect. 4.11 the value of the universal constants $V(0, 0)$ and $E_0 U(0, 0)$ can in the case $d = 2$, $\nu = 1$ be derived from conformal invariance.

4.11 Conformal Invariance

As indicated in the introduction this part of the book the application of conformal invariance to critical phenomena began with the work of Polyakov (1970). Since then it has developed into a powerful tool which provides results beyond those given by scaling theory. In this section we indicate briefly some of the main features of the theory and prove a famous result due to Cardy (1984), which establishes the value of the universal constant $V(0, 0)$ of Eq. (4.10.34) in two dimensions.³⁶

4.11.1 From Scaling to the Conformal Group

As we have seen in Sects. 4.5 and 4.7, scaling theory yields scaling laws for critical and tricritical points. Stronger assumptions are needed to give the possibility of extracting further information. Suppose that the elements of the scaling transformation given by (4.2.5), (4.3.3) and (4.3.97) are ‘localized’ to

$$\begin{aligned} \mathbf{r} &\rightarrow \tilde{\mathbf{r}} &= [\lambda(\mathbf{r})]^{-1} \mathbf{r}, \\ \theta_j &\rightarrow \tilde{\theta}_j &= [\lambda(\mathbf{r})]^{y_j} \theta_j, \\ \phi_j(\mathbf{r}) &\rightarrow \phi_j(\tilde{\mathbf{r}}) &= [\lambda(\mathbf{r})]^{x_j} \phi_j(\mathbf{r}). \end{aligned} \quad (4.11.1)$$

This modification to scaling, which can be justified with the same kind of argument that was used in Sect. 4.2, will leave (4.3.98) intact with $\lambda(\mathbf{r})$ replacing λ as long as the spatial transformation has the Jacobian $[\lambda(\mathbf{r})]^{-d}$. It is not difficult to confirm that this will be the case if either $\lambda(\mathbf{r})$ is a constant, which is the usual dilatation transformation of scaling, or $\lambda(\mathbf{r}) = 1/|\mathbf{r}|^2$, which is the inversion transformation. As we describe in Sect. 17.4 the group generated by translations, rotations, dilatations and inversions is the *conformal group*.

4.11.2 Correlation Functions for $d \geq 2$

Given that the transformation (4.11.1) belongs to the conformal group, the analysis of Sect. 4.3.7 leading to the scaling form (4.3.100) for the total correlation function is not affected by the Jacobian having a spatial dependence and thus we now have the generalization

$$\langle \phi_1(\tilde{\mathbf{r}}_1) \cdots \phi_\tau(\tilde{\mathbf{r}}_\tau) \rangle = \left\{ \prod_{j=1}^{\tau} [\lambda(\mathbf{r}_j)]^{x_j} \right\} \langle \phi_1(\mathbf{r}_1) \cdots \phi_\tau(\mathbf{r}_\tau) \rangle \quad (4.11.2)$$

³⁶ For more detailed discussions of conformal invariance the reader is referred to the review of Cardy (1987) and the books of Itzykson (1988), Christe and Henkel (1993) and Henkel (1999).

of (4.3.100). In the case of the two-point correlation function, rotation and translation can be used to interchange $\mathbf{r}_1$ and $\mathbf{r}_2$, giving

$$\begin{aligned}\langle \phi_1(\tilde{\mathbf{r}}_1) \phi_2(\tilde{\mathbf{r}}_2) \rangle &= [\lambda(\mathbf{r}_1)]^{x_1} [\lambda(\mathbf{r}_2)]^{x_2} \langle \phi_1(\mathbf{r}_1) \phi_2(\mathbf{r}_2) \rangle, \\ \langle \phi_1(\tilde{\mathbf{r}}_2) \phi_2(\tilde{\mathbf{r}}_1) \rangle &= [\lambda(\mathbf{r}_2)]^{x_1} [\lambda(\mathbf{r}_1)]^{x_2} \langle \phi_1(\mathbf{r}_2) \phi_2(\mathbf{r}_1) \rangle,\end{aligned}\tag{4.11.3}$$

with

$$\begin{aligned}\langle \phi_1(\tilde{\mathbf{r}}_1) \phi_2(\tilde{\mathbf{r}}_2) \rangle &= \langle \phi_1(\tilde{\mathbf{r}}_2) \phi_2(\tilde{\mathbf{r}}_1) \rangle, \\ \langle \phi_1(\mathbf{r}_1) \phi_2(\mathbf{r}_2) \rangle &= \langle \phi_1(\mathbf{r}_2) \phi_2(\mathbf{r}_1) \rangle.\end{aligned}\tag{4.11.4}$$

For a non-zero two-point correlation function it would appear that these results must be inconsistent unless $x_1 = x_2$, leading to the conclusion that the two-point total correlation function of two operators with different scaling dimensions is zero. In fact it is clear that this conclusion must be modified to take into account two-point correlation functions like

$$\langle \phi(\mathbf{r}_1) \nabla_{\mathbf{r}_2} \phi(\mathbf{r}_2) \rangle = \nabla_{\mathbf{r}_2} \langle \phi(\mathbf{r}_1) \phi(\mathbf{r}_2) \rangle,\tag{4.11.5}$$

which will not, in general, be zero in spite of the fact that the scaling dimensions differ by two. This argument has been analyzed by Schäfer (1976) (see also Cardy 1987) who concluded that

$$\langle \phi_1(\mathbf{r}_1) \phi_2(\mathbf{r}_2) \rangle = 0, \quad \text{if } x_1 - x_2 \neq \text{integer}.\tag{4.11.6}$$

When x_1 and x_2 differ by an integer, the two-point correlation function is sometimes but not always non-zero. Analysing the case $x_1 = x_2 = x$, Schäfer (1976) derived conditions for this to be the case. Then, within the context of conformal invariance, (4.3.112) in the form

$$\langle \phi(\mathbf{r}_1) \phi(\mathbf{r}_2) \rangle = \frac{C_2(\phi, \phi)}{|\mathbf{r}_2 - \mathbf{r}_1|^{2x}}\tag{4.11.7}$$

is exact. In the case of the three-point correlation function, the operations of the conformal group can be used to map three arbitrary points $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ into any three chosen points α , β , γ . It has been shown by Polyakov (1970) that

$$\begin{aligned}\langle \phi_1(\mathbf{r}_1) \phi_2(\mathbf{r}_2) \phi_3(\mathbf{r}_3) \rangle \\ = \frac{C_3(\phi_1, \phi_2, \phi_3)}{|\mathbf{r}_1 - \mathbf{r}_2|^{x_1+x_2-x_3} |\mathbf{r}_2 - \mathbf{r}_3|^{x_2+x_3-x_1} |\mathbf{r}_3 - \mathbf{r}_1|^{x_3+x_1-x_2}},\end{aligned}\tag{4.11.8}$$

where $C_3(\phi_1, \phi_2, \phi_3)$ is a constant dependent only on the choice of scaling operators. An expression can also be obtained for the four-point correlation function, although

since it is not possible to map four arbitrary points into four prescribed points with the operations of the conformal group an arbitrary function is involved (Polyakov 1970).

4.11.3 Universal Amplitudes for $d = 2$

We now consider the square lattice with lattice spacing a . In two dimensions the conformal group is infinite dimensional and consists, for $\mathbf{r} = a(x, y)$, of transformations given by analytic functions³⁷ of the complex variable $z = x + iy$ (Sect. 17.4). For the transformation $z \rightarrow Z(z)$ the Jacobian is $|Z'(z)|^2$. So, for the two-point correlation function of the scaling operator ϕ_1 , (4.11.2) gives

$$\langle \phi(z_1) \phi(z_2) \rangle = |Z'(z_1)|^x |Z'(z_2)|^x \langle \phi(Z(z_1)) \phi(Z(z_2)) \rangle. \quad (4.11.9)$$

The transformation given by

$$Z(z) = \frac{\aleph}{2\pi} \ln(z), \quad (4.11.10)$$

is analytic for all $z \neq 0$ and maps the z -plane onto the strip $0 \leq \tilde{y} < \aleph$. Thus, from (4.11.7), (4.11.9) and (4.11.10), the two-point correlation function on the strip of width $\aleph$ lattice sites with periodic boundary conditions is

$$\langle \phi(\tilde{\mathbf{r}}) \phi(\tilde{\mathbf{r}}_2) \rangle = \frac{C_2(\phi, \phi) (2\pi/\aleph a)^{2x}}{\left\{ 2 \cosh \left[\frac{2\pi(\tilde{x}_2 - \tilde{x}_1)}{\aleph} \right] - 2 \cos \left[\frac{2\pi(\tilde{y}_2 - \tilde{y}_1)}{\aleph} \right] \right\}^x}, \quad (4.11.11)$$

where $\tilde{\mathbf{r}} = a(\tilde{x}, \tilde{y})$. When $|\tilde{z}_2 - \tilde{z}_1| \ll \aleph$,

$$\langle \phi(\tilde{\mathbf{r}}) \phi(\tilde{\mathbf{r}}_2) \rangle \simeq \frac{C_2(\phi, \phi)}{|\tilde{\mathbf{r}}_1 - \tilde{\mathbf{r}}_2|^{2x}}, \quad (4.11.12)$$

but, when $|\tilde{x}_2 - \tilde{x}_1| \gg \aleph$,

$$\langle \phi(\tilde{\mathbf{r}}) \phi(\tilde{\mathbf{r}}_2) \rangle \simeq C_2(\phi, \phi) \left(\frac{2\pi}{\aleph a} \right)^{2x} \exp \left[-\frac{2\pi x(\tilde{x}_2 - \tilde{x}_1)}{\aleph} \right]. \quad (4.11.13)$$

³⁷ In complex variable theory a *holomorphic function* is one which is complex differentiable in a neighbourhood of every point in its domain and this should more strictly be our usage here. The term ‘analytic’ which has already been used in the context of real-valued functions denotes a function (real or complex) which is equal to its Taylor series in a neighbourhood of each point in its domain. The fact that the class of complex analytic functions coincides with the class of holomorphic functions is a theorem of complex analysis. A function which is holomorphic/analytic throughout the complex plane is an *entire function*.

We now consider the case of finite-size scaling at a critical point discussed in Sect. 4.10. If the leading behaviour of the state operator q is given by the scaling operator ϕ with $\langle \phi(r) \rangle = 0$ at the critical point, (4.11.13) can be substituted into (4.2.4) (with σ replaced by q) to give the correlation length $\tau(q)$. In the limit of small lattice spacing the sums become integrals and

$$\tau^2(q) = \frac{1}{2} a^2 \frac{\int_0^\infty r^2 \exp\left(-\frac{2\pi x r}{\mathfrak{s}}\right) dr}{\int_0^\infty \exp\left(-\frac{2\pi x r}{\mathfrak{s}}\right) dr}, \quad (4.11.14)$$

giving

$$\tau(q) = \frac{\mathfrak{s}a}{2\pi x}. \quad (4.11.15)$$

This result was first obtained by Cardy (1984). Substituting into (4.10.34) with $v_0^{\frac{1}{d}} = a$ gives

$$V(0, 0) = \frac{1}{2\pi x}. \quad (4.11.16)$$

The detailed analysis which leads to the determination of the critical amplitude $E_0 U(0, 0)$ in (4.10.33), for $d = 2$, is beyond the scope of this book. We simply state the result, obtained by Blöte et al. (1986) and Affleck (1986), that

$$E_0 U(0, 0) = -\frac{1}{6} \pi c, \quad (4.11.17)$$

where c is the *conformal anomaly number* defined by $\frac{1}{2}c$ being the coefficient of the leading term of the operator product expansion of the stress tensor with itself.³⁸ From this it can be shown that c is the *central charge in the Virasoro algebra* of the generators of the conformal transformation. For models where the transfer matrix is symmetric (or can be made symmetric at criticality) the Virasoro algebra is unitary and, for $c < 1$, the only allowed values are given by

$$c(m) = 1 - \frac{6}{m(m+1)}, \quad m = 3, 4, 5, \dots \quad (4.11.18)$$

If a particular model can be identified with a value of m and c the scaling dimensions of its primary scalar operators, indexed by i and j with $1 \leq j \leq i \leq m-1$, are given by the *Kac formula*³⁹

³⁸ For a full explanation of this and all the following discussion see Cardy (1987).

³⁹ These formulae are usually stated (Henkel 1999, p. 93) in terms of $\Delta_{ij} := \frac{1}{2}x_{ij}$.

$$x_{ij}(m) = \frac{[i(m+1) - jm]^2 - 1}{2m(m+1)}. \quad (4.11.19)$$

The (critical) Ising model has been identified with the case $m = 3$, $c = \frac{1}{2}$, the tricritical Ising model with $m = 4$, $c = \frac{7}{10}$, and the 3-state Potts model with $m = 5$, $c = \frac{4}{5}$. The known models with continuously varying exponents correspond to the limiting case $m \rightarrow \infty$, $c = 1$. These include the two-dimensional XY model (Sect. 3.6.1.1) and the Ashkin–Teller Model (Sect. 3.6.2), which is related to the eight-vertex model by a duality transformation.

For the Ising model the three allowed pairs of values for the parameters are:

- $i = j = 1$ giving $x_{11} = 0$.
This corresponds to an operator present in all models. It is similar to the trivial coupling in renormalization group theory.
- $i = 2, j = 1$ giving $x_{21} = 1$.
This gives the scaling exponent $y_{21} = d - x_{21} = 1$. By identifying y_{21} with the scaling exponent y_2 of Sect. 4.5 and using (4.5.39) we obtain the value $\alpha = 0$, in agreement with the exact result (12.3.280). Since by the fluctuation-response function relation (2.5.15) the heat capacity is associated with the energy-energy correlation function and we are able to identify the operator as the energy density.
- $i = j = 2$ giving $x_{22} = \frac{1}{8}$.
This gives the scaling exponent $y_{22} = d - x_{22} = \frac{15}{8}$. By identify y_{22} with the exponent y_1 of Sect. 4.5 and using (4.5.40) we obtain the value $\gamma = \frac{7}{4}$ in agreement with the exact result (12.3.280). In this case the operator is the magnetization density.

In the cases of the tricritical Ising model and 3-state Potts model there are a larger number of allowed parameter values. It is only with the use of additional information that it becomes possible to identify the physical nature of the operators. Thus, for example, for the 3-state Potts model the known values of α and γ given in (12.3.322) allow the identification of the cases $i = 2, j = 1$ $x_{21} = \frac{4}{5}$ and $i = j = 3$, $x_{33} = \frac{2}{15}$ with the energy density and magnetization density respectively.

4.11.4 Schramm–Loewner Evolution

Schramm–Loewner evolution (SLE) originated with the application by Schramm (2000) of *Loewner’s equation* (Löwner 1923) to the study of the continuum limit of a loop-erased random walk and developed into a general method for obtaining rigorous information about two-dimensional critical behaviour. Here we give a brief outline of the ideas involved. For more detailed accounts the reader is referred to Cardy (2005) or Lawler (2011) and the bibliographies included therein.

The starting point for SLE for a particular model is the identification of a type of non-crossing path passing from one point on the boundary of the lattice to another.

For the Ising model on the honeycomb lattice such a path is a domain boundary between regions of aligned spins on the dual triangular lattice. An example for the Potts model on a square lattice is given by Fig. 3.8, where the paths are drawn on the medial lattice as the boundaries of the shaded regions containing connected clusters of spins. Of course, the method of constructing these curves, after the spins have been fixed, will not just produce curves passing between points on the boundary. It will, as can be seen in Fig. 3.8, also produce closed loops. Part of the insight of SLE is that the universal properties of the system can be extracted by concentrating solely on the open paths, which for each of the examples given must terminate on the boundary. The other initiating aspect of the method is to regard the curves as being constructed step-by-step using a random process. Consider for simplicity the finite honeycomb lattice $\mathcal{N}_{\text{HC}}$ and suppose that a path γ is constructed beginning at a boundary site $\mathbf{r}_s$ and ending at a boundary site $\mathbf{r}_t$. As each site is reached there is a choice between two for the next step. With an initial degeneracy of two, and as the path is generated, spin states are assigned to sites of the dual triangular lattice $\mathcal{N}_{\text{T}}$ making γ a domain boundary, and the probability of turning left or right at a lattice site is calculated from the expected values for the spin on the site of $\mathcal{N}_{\text{T}}$ immediately ahead, based on the states already assigned to sites of that lattice. Thus the relative probabilities of turning left or right at step n are determined by the lattice and γ_{n-1} , the path up to that point. Let $p(\gamma; \mathbf{r}_s, \mathbf{r}_t, \mathcal{N}_{\text{HC}})$ be the probability of the path γ being constructed from $\mathbf{r}_s$ to $\mathbf{r}_t$ by this procedure and suppose that γ_n , the path in place after n steps, has reached the point $\mathbf{r}_n$. $\mathcal{N}_{\text{HC}} \setminus \gamma_n$ is the part of $\mathcal{N}_{\text{HC}}$ from which the edges of γ_n (and all dangling edges and exposed spins of $\mathcal{N}_{\text{T}}$) have been removed. Then the method of path-generation implies that

$$p(\gamma - \gamma_n; \mathbf{r}_n, \mathbf{r}_t, \mathcal{N}_{\text{HC}} \setminus \gamma_n) = p(\gamma; \mathbf{r}_n, \mathbf{r}_t, \mathcal{N}_{\text{HC}} \setminus \gamma_n). \quad (4.11.20)$$

SLE is concerned with the continuum limit of this situation when the lattice is replaced by a domain $\mathcal{D} \subset \mathbb{C}$ with $\mathbf{r}_s$ and $\mathbf{r}_t$ points on the boundary⁴⁰ and $\gamma_n \in \mathcal{D}$ a non-crossing curve from $\mathbf{r}_s$ to $\mathbf{r}_n$. The domain boundary γ on the lattice neither crosses *nor touches* itself. In the continuum case the latter situation can occur. Then the curve creates a *hull* $\tilde{\gamma}$, which is γ together with any regions of $\mathcal{D}$ which are surrounded by the curve by virtue of it touching itself. That any such curve γ can be described by a process of continuous growth (*called Loewner evolution*) was proved by Löwner (1923)

The relationship (4.11.20), with the probability $p(\cdot)$ replaced by a measure $\mu(\cdot)$ and $\mathcal{N}_{\text{HC}}$ replaced by $\mathcal{D}$, is one of the defining properties of an SLE. The other concerns a conformal invariance property which must hold for the system. Let $\mathfrak{M}$ be a conformal mapping from the interior of $\mathcal{D}$ to the interior of another domain $\mathcal{D}'$ and suppose that the pair $[\mathbf{r}_s, \mathbf{r}_t]$ maps to $[\mathbf{r}'_s, \mathbf{r}'_t]$ on the boundary of $\mathcal{D}'$. The measure μ

⁴⁰ This is *chordal SLE*. There is another version, called *radial SLE*, which will not be discussed, where $\mathbf{r}_t$ is an interior point of $\mathcal{D}$.

induces a measure $\mathfrak{M} \star \mu$ on the image curves $\mathfrak{M}(\gamma)$ in $\mathcal{D}'$. This must satisfy

$$(\mathfrak{M} \star \mu)(\gamma; \mathbf{r}_s, \mathbf{r}_t, \mathcal{D}) = \mu(\mathfrak{M}(\gamma); \mathbf{r}'_s, \mathbf{r}'_t, \mathcal{D}'). \quad (4.11.21)$$

The importance of this conformal property is that $\mathcal{D}$ can be chosen to be $\mathbb{C}^{(+)}$, the upper half of $\mathbb{C}$, with $\mathbf{r}_s = 0$ and $\mathbf{r}_t = \infty$, when the distribution of paths has the reflectional symmetry $x \rightarrow -x$. The curve $\gamma(t)$ is now a function of a continuous parameter t with final point $z(t)$. Since $\mathbb{C}^{(+)} \setminus \bar{\gamma}(t)$ is simply connected it can, by the Riemann mapping theorem, be mapped by a real analytic function $g(t; z)$ into $\mathbb{C}^{(+)}$, while preserving the real axis outside $\gamma(t)$. When $\gamma(t)$ is a non-touching curve ($\bar{\gamma}(t) = \gamma(t)$) it is a cut in $\mathbb{C}^{(+)}$ which is ‘opened up’ into a section of the real axis bounded at its ends by the two images of the origin and containing a point $a(t)$ which is the image of $z(t)$. The development of $\gamma(t)$ is determined by $a(t)$, which is itself a random function.

Subject to the condition that $g(t; z) \sim z + O(1/z)$ as $z \rightarrow \infty$, the function $g(t; z)$ which describes this situation satisfies *Loewner’s equation*

$$\frac{dg(t; z)}{dt} = \frac{2}{g(t; z) - a(t)}, \quad (4.11.22)$$

(Löwner 1923). Given an evolving path, $\bar{\gamma}(t)$ can be determined and, in principle, $g(t; z)$ and $a(t)$ be obtained. Conversely, the equation can be integrated, with more or less difficulty (Kager et al. 2004) for given $a(t)$. The point given by $a(t)$ moves randomly on the real axis. A crucial ingredient of SLE is the result due to Schramm (2000) that, if (4.11.20) in its continuum version, (4.11.21) and reflectional symmetry hold then the movement of $a(t)$ is proportional to a *standard Brownian motion* with diffusion constant κ . The nature of $\gamma(t)$ varies with κ , which is now the only parameter of the system. For $\kappa \leq 4$ it is a simple curve, for $4 < \kappa < 8$ it intersects itself so that every point is contained in a loop and for $\kappa \geq 8$ it is space filling. These results can be illuminated by a formula, valid for $\kappa \leq 8$ that γ has the fractal dimension $d_f := 1 + \frac{1}{2}\kappa$, which increases with κ reaching the (space-filling) value two when $\kappa = 8$. The central charge in the Virasoro algebra of the generators of the conformal transformation can be obtain in terms of κ as

$$c(\kappa) = \frac{(3\kappa - 8)(6 - \kappa)}{2\kappa}. \quad (4.11.23)$$

Solving this formula with (4.11.18) gives

$$\kappa = \begin{cases} \frac{4m}{m+1}, & \kappa \leq 4, \\ \frac{4(m+1)}{m}, & \kappa \geq 4 \end{cases} \quad (4.11.24)$$

and, from (4.11.19),

$$x(\kappa) := \frac{6 - \kappa}{\kappa} = \begin{cases} x_{21}(\kappa), & \kappa \leq 4, \\ x_{12}(\kappa), & \kappa \geq 4. \end{cases} \quad (4.11.25)$$

Although there are few direct results, these formulae allow identification of values of κ with the models with specified values of c on p. 162. Given the nature of the method it is not surprising that the proven and strongly suggested cases of connections to particular models are those related to random walks and percolation. In particular rigorous results have been obtained for the continuum limit of critical site percolation on a triangular lattice, which is identified with $\kappa = 6$ (Smirnov 2001), and for loop-erased random walks and the boundaries of uniform spanning trees, identified respectively with $\kappa = 2$ and $\kappa = 8$ (Lawler et al. 2004).⁴¹

⁴¹ For $\kappa = 2$ the SLE is radial rather than chordal.

Part II

Classical Approximation Methods

Introduction

As is clear from the discussion in Sect. 2.1 the ultimate aim, given any statistical mechanical model, is to obtain an exact expression for the partition function. In the absence of this, the oldest tool of statistical mechanics is the use of classical approximations.¹ As we see in Chap. 5 the first step in such methods is the replacement of the Hamiltonian with an approximation which leads to a form for the free-energy density ϕ_0 which is a smooth function of the density or densities. This results in the elimination of all reference to spatially fluctuating variables.² The equation of state, like that of van der Waals (Sect. 5.2), becomes a smooth function of the thermodynamic variables. Landau theory (Sects. 5.3–5.5), where the thermodynamic potential has a power-series expansion in some order parameter is a natural extension of this. The absence of spatially fluctuating variables prevents an investigation of the critical properties of any quantities, like the correlation length and pair correlation function, which depend on such fluctuations. In Sect. 5.6 a generalization of Landau theory, due to Ginzburg and Landau (1950), is introduced for a critical point in a magnetic system. In this model spatial fluctuations are incorporated and the partition function is expressed as a functional integral. At the level of the *Gaussian approximation*, considered here, only harmonic fluctuations about Landau values are retained. This is the simplest way to include some fluctuations and, in the context of quantum field theory, it is equivalent to the ‘tree approximation’ (Amit 1978).³ In the mean-field approximations of Chap. 6 the summation in the partition function is replaced by a single term with the microstates replaced by their mean values. The further summation then needed to obtain the

¹ The methods described in this part are often generically referred to as ‘mean-field methods’. For an historic perspective on their place in the theory of phase transitions see Kadanoff (2009).

² Although, of course, overall density fluctuations are still present (Sect. 6.1.1).

³ For a more detailed treatment of generalizations to Landau theory the reader is referred to Luban (1976).

free-energy density per lattice site with no independent densities is replaced by selecting the largest term in the summation. Chapter 6 illustrates this approximation procedure with a number of examples. In these, phases are characterized by densities, associated in the case of lattice systems with the proportion of the lattice sites having a microsystem in a particular microstate. The higher order cluster approximations of Chap. 7 refine this programme by distinguishing the different microstates of small clusters of sites. Increasing the size of the clusters yields a hierarchy of approximation methods which, with increasing difficulty and the need for ingenuity, yields increasingly accurate predictions for some characteristics of the system like critical curves. However, these remain classical approximations, which continue to yield classical (in general, incorrect) values for critical exponents and in some cases give incorrect predictions for the orders of the phase transitions.

Chapter 5

Phenomenological Theory and Landau Expansions

5.1 Classical Methods

In Sect. 4.1, as a preliminary to the discussion of scaling theory, we gave a general heuristic treatment of the geometry of phase transitions. This was predicated on the assumption that, for a system with independent internal couplings $K_1, \dots, K_{n_e}$ and external couplings $L_1, \dots, L_{n_f-1}$, the free-energy density $\phi(K_1, \dots, K_{n_e}, L_1, \dots, L_{n_f-1})$ was known and thus could be represented as a continuous surface in the $(n_e + n_f)$ -dimensional space of the variables $\phi, K_1, \dots, K_{n_e}, L_1, \dots, L_{n_f-1}$. As we have indicated above this ideal situation rarely occurs and some approximate form for ϕ must be derived. Beginning with the work of Bragg and Williams (1934), for the order-disorder transition in alloys, a class of approximation methods was developed, which has been very successful in making qualitative predictions for phase transitions. These *classical methods* include not only mean-field approximations, of the type employed by Bragg and Williams, which are discussed in Chap. 6, but also the more sophisticated cluster-variation approximations, described in Chap. 7, of which the improved approximation used by Bethe (1931, 1935) was the first instance.

For simplicity we here concentrate on the case of a system with $n_f = 2$ and fields $\xi = \xi_1$ and ξ_2 with corresponding extensive variables Q_1 and Q_2 . Given that there is a set of internal energy parameters ε_ℓ , $\ell = 1, 2, \dots, n_e$ we shall adopt the device, described in Sect. 2.3.1, of identifying $\varepsilon := \varepsilon_1$ and setting $\kappa_\ell := \varepsilon_\ell/\varepsilon_1$. Then the system has effectively two couplings $K := \varepsilon/T$ and $L := \xi/T$. At this stage it is convenient to frame the discussion in the coupling-extensive variable representation, with dimensionless free energies $\Phi_0(Q_1, Q_2, K)$ and $\Phi_1(Q_2, K, L)$, where, from (2.3.5) and (2.3.6),

$$\Phi_0(Q_1, Q_2, K) = -\ln Z_0(Q_1, Q_2, K), \quad (5.1.1)$$

and

$$Z_0(Q_1, Q_2, K) = \sum_{\{\sigma\}} \exp[-\hat{H}_0(Q_1, Q_2, K; \sigma)], \quad (5.1.2)$$

the summation being over all microstates compatible with a particular value for Q_1 . From (2.1.12),

$$\Phi_1(Q_2, K, L) = -\ln \left\{ \sum_{\{\hat{Q}_1\}} \exp[L \hat{Q}_1 - \Phi_0(\hat{Q}_1, Q_2, K)] \right\}. \quad (5.1.3)$$

The common feature of all classical methods is that some approximation is made to the Hamiltonian in (5.1.2) leading to an approximation for $\Phi_0(Q_1, Q_2, K)$, denoted by $\Phi_0^{(\text{cl})}(Q_1, Q_2, K)$, which is, for all K , a smooth function of Q_1 .¹ Given that the thermodynamic limit can be applied to this approximation with size parameter $Q := Q_2$

$$\Phi_0^{(\text{cl})}(Q_1, Q, K) = Q\phi_0^{(\text{cl})}(\rho, K) + O(Q^0), \quad (5.1.4)$$

where $\rho := Q_1/Q$. Then, from (5.1.3),

$$\phi_1(K, L) = -\frac{1}{Q} \ln \left\{ \sum_{\{\hat{\rho}\}} \exp[Q\{L \hat{\rho} - \phi_0^{(\text{cl})}(\hat{\rho}, K)\} + O(Q^0)] \right\}. \quad (5.1.5)$$

For large Q the summation is dominated by its maximum term and thus, in the thermodynamic limit, $Q \rightarrow \infty$,

$$\phi_1^{(\text{cl})}(K, L; \rho) = \phi_0^{(\text{cl})}(\rho, K) - \rho L, \quad (5.1.6)$$

where ρ is given by minimizing $\phi_1^{(\text{cl})}(K, L; \rho)$.² Thus we have the conditions

$$\frac{\partial \phi_1^{(\text{cl})}}{\partial \rho} = \frac{\partial \phi_0^{(\text{cl})}}{\partial \rho} - L = 0, \quad (5.1.7)$$

$$\frac{\partial^2 \phi_1^{(\text{cl})}}{\partial \rho^2} = \frac{\partial^2 \phi_0^{(\text{cl})}}{\partial \rho^2} = \frac{\partial L}{\partial \rho} > 0. \quad (5.1.8)$$

The last condition in (5.1.8) is understood in the sense that L is given as a function of ρ from (5.1.7). This condition, expressed in the form $L = L(\rho, K)$, is the *equation of state* of the system and, from (5.1.7),

$$\phi_0^{(\text{cl})}(\rho, K) = C(K) + \int L(\rho, K) d\rho, \quad (5.1.9)$$

where $C(K)$ is some undetermined function of K .

¹ See Definition 17.1.1.

² Cf. (1.1.13), (1.1.14) and the comment following the latter.

5.1.1 A First-Order Transition

As we have indicated in Sect. 4.1.1, the curve derived from the equation of state $L = L(\rho, K)$, for a system undergoing a first-order phase transition, will have a horizontal section connecting points representing phases in coexistence.³ However, of course, the classical equation of state derived from (5.1.7) has been assumed to give L as a smooth function of ρ for all K , excluding the possibility of a horizontal section (apart from in the form of a single point).

To understand the resolution of this problem it is pertinent to outline briefly the historic development of the theory of phase transitions. Early work on the subject was mainly concerned with the liquid-vapour transition in fluids and Andrews (1869) was the first to measure pressure and temperature along isotherms, both above and below the critical temperature.⁴ He concluded that “the gaseous and liquid states are only distant stages of the same condition of matter, and are capable of passing into one another by a process of continuous change” (Andrews 1869, p. 589). Commenting on this work Thomson (1871, p. 2) suggested “that, although there be a practical breach of continuity in crossing the line of boiling-points from liquid to gas or from gas to liquid, there may exist, in the nature of things, a theoretical continuity across the breach having some real and true significance”. He supported this view by observing “that the ordinary liquid state does not necessarily cease abruptly at the line of boiling-points, . . . liquids may, with due precautions, be heated considerably beyond the boiling temperature” (op. cit., p. 3.). A smooth continuation between the branches below A and above B in Fig. 4.2c, resulting from a classical approximation of the equation of state would, of course, be dependent on the nature of the approximation. The simplest would be the case where the isotherm crossed the line AB at one intermediate point. Such is the case for van der Waals’ equation, which is explored in some detail in Sect. 5.2.⁵ There the isotherm has a single loop with a range of pressures for which lines of constant pressure cut the isotherm at three points. The corresponding situation for a classical isotherm in the coupling–density plane ρ – L is shown in Fig. 5.1a.⁶ However, now the question arises: if we begin, as van der Waals did, with an equation of state (rather than the free energy), at what pressure (or value of L) does the transition between phases occur? For the van der Waals equation this question was answered by Maxwell (1875), who derived, in terms of the transformation of heat into work, the condition now known as *Maxwell’s equal area rule*. In terms of Fig. 5.1a this simply means that the points A and B, between

³ See Fig. 4.2c, where the points representing coexistent phases are denoted by A and B.

⁴ For an account of this work see Rowlinson (1969).

⁵ The reader is recommended to refer to this example as an illustration of the development of Landau theory with one order parameter.

⁶ Ignore for the moment the shaded regions. Remember also that isotherms for the van der Waals equation will be mirror images about a vertical line of those given in Fig. 5.1a, because increase in pressure yields a decrease in v : see the comment in the footnote on p. 6.

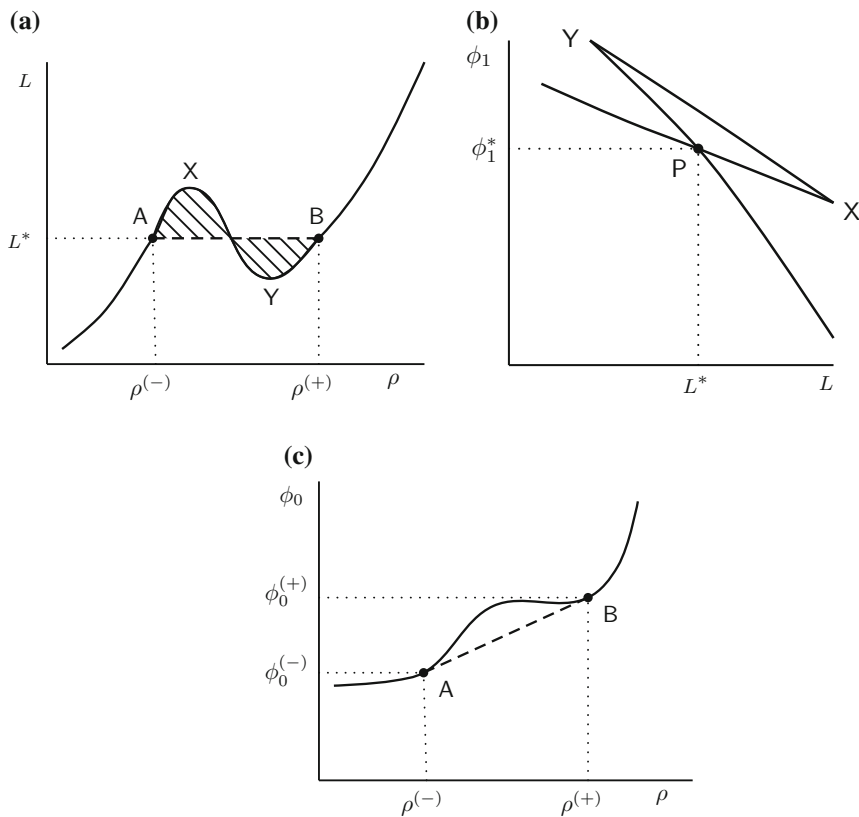


Fig. 5.1 **a** A first-order transition showing as a *horizontal* part of an isotherm of L plotted against ρ . **b** A first-order transition showing as a discontinuity of slope in an isothermal section of the ϕ_1 surface. **c** A first-order transition showing as the *double tangent* AB in an isothermal section of the ϕ_0 surface

which a transition occurs are determined by the equality of the shaded areas. That is

$$\int_{\rho^{(-)}}^{\rho^{(+)}} L(\rho, K) d\rho = L^*[\rho^{(+)} - \rho^{(-)}]. \quad (5.1.10)$$

From (5.1.9),

$$\int_{\rho^{(-)}}^{\rho^{(+)}} L(\rho, K) d\rho = \phi_0^{(\text{cl})}(\rho^{(+)}, K) - \phi_0^{(\text{cl})}(\rho^{(-)}, K), \quad (5.1.11)$$

which then gives, from (5.1.6) and (5.1.10),

$$\phi_1^{(\text{cl})}(K, L^*; \rho^{(+)}) = \phi_1^{(\text{cl})}(K, L^*; \rho^{(-)}), \quad (5.1.12)$$

where now the densities are not independent variables but are given by the equation of state (5.1.7). Equality of ϕ_1 , which from (2.3.15) is equivalent to equality of the second external coupling L_2 , appears in the discussion of the appearance of a first-order transition in Sect. 4.1.1 in a slightly different but equivalent way. There ϕ_1 was assumed to be a continuous function of K and L with a first-order transition signalled by a discontinuity of slope (density). Now the discontinuity of slope in Fig. 4.2a is replaced by a loop in Fig. 5.1b.

A locally stable state, according to (5.1.7)–(5.1.8), corresponds to a minimum of $\phi_1^{(\text{cl})}$ with respect to ρ and this applies to the isotherm in Fig. 5.1a to the left of X and to the right of Y. It can then be seen from Fig. 5.1b that, in the range of L where there are two minima, the lower of the two is given by the isotherm to the left of A and the right of B. A single-phase system in a locally stable but globally unstable state represented by a point on AX or YB is said to be *metastable*. The segment XY consists of *intrinsically unstable* states. The double tangent in Fig. 4.2b appears in Fig. 5.1c as the completion of the convex envelope of $\phi_0^{(\text{cl})}(K, \rho)$; the part of the curve not belonging to the envelope comprises both metastable and intrinsically unstable states. The demarcation between them is given, from (5.1.8), by points where $(\partial^2 \phi_0^{(\text{cl})} / \partial \rho^2)_K = 0$.

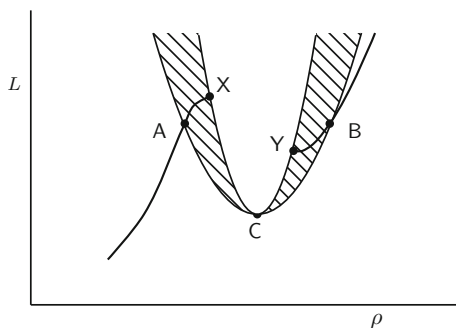
If we assume that the equation of state (5.1.7) has the same general properties as van der Waals' equation then an isotherm of the form shown in Fig. 5.1a will occur for K greater than that a critical K_c ($T < T_c$) and on the critical isotherm $K = K_c$ there will be a critical point (ρ_c, L_c) where

$$\left(\frac{\partial L}{\partial \rho} \right)_K = 0, \quad \left(\frac{\partial^2 L}{\partial \rho^2} \right)_K = 0. \quad (5.1.13)$$

Plots of the points A and B as K approaches the critical value from above in the $\rho - L$ plane meet at a critical point C and define the boundaries of the coexistence region. In principle we can obtain (at least numerically, if not analytically) a plot for the first-order transition curve $L = L^*(K)$ in the $L - K$ plane similar to that shown in Fig. 4.7. The two branches of the coexistence curve can then be plotted either against K as shown in Fig. 4.8 or against L . We choose the latter, so that we can also include part of the isotherm shown in Fig. 5.1a. The coexistence curve will nevertheless still have the general appearance of that shown in Fig. 4.8.⁷ But now we have a feature which did not arise in our analysis in Sect. 4.4 of the exact free-energy density, namely the possibility of a one-phase system being metastable. The region in the $\rho - L$ plane in which this occurs (shaded in Fig. 5.2) lies between the coexistence curve and the *spinodal* curve, which is the line traced out by X and Y as K is varied.

⁷ Although K increases along the first-order transition curve it is not necessarily the case that L does. However, this will be assumed in line with Fig. 4.7.

Fig. 5.2 Coexistence curve and spinodal lines derived from Fig. 5.1a



5.1.2 Metastability

As indicated in the quote from Thomson (1871) given above, metastable states are not simply artifacts of classical theory, but are frequently encountered experimentally. Let us consider, for example, the case of a fluid system where, in Fig. 5.2, the low-density states to the left of X and the high-density states to the right of Y correspond respectively to a vapour and a liquid. The density of the vapour is increased (by increasing L) along the isotherm from the left towards A. If the system remains in thermodynamic equilibrium, any further increase in L will cause a transition to the liquid state. However, when Fig. 5.2 represents the behaviour of a real fluid, a dust-free vapour may progress along the continuous segment AX into the metastable (shaded) region. It is then termed *supersaturated*. Metastable states can also be reached by cooling a liquid below the equilibrium temperature of transition to the crystalline form. Such liquids are called *supercooled*.

We now discuss briefly the physical reasons for the existence of metastable states in real substances, using the supersaturated vapour as an example. For the latter to go over to the liquid state, localized (low-wavelength) density fluctuations of large amplitude must occur, resulting in the formation of nuclei, each containing a large number of molecules. Such a nucleus has a lower free energy than an equal mass of vapour and will grow by accretion. However, a nucleus created by a spontaneous local density fluctuation is likely to be small and because of its relatively large surface area will have a higher free energy than an equal mass of vapour. This is another aspect of the local stability of metastable states. Nevertheless thermodynamically metastable states are different in character from static mechanical metastable states, like that of a ball on a surface at the lowest point of a depression separated by a ridge from a deeper depression.

The degree of permanence of a thermodynamically metastable state depends on the *relaxation time*, the mean waiting time for a sufficiently large fluctuation to initiate the transformation to the stable state. For the mechanical analogy to be appropriate, in fact, the ball must be regarded as subject to impulses of random strength at random intervals. However, relaxation times can be very large. For instance, glasses are supercooled liquids with higher free energies than the crystalline forms whereas

diamond is a metastable crystalline form of carbon.⁸ In a supersaturated vapour dust particles drastically reduce the relaxation time by facilitating the formulation of large nuclei. Charged particles (ions) have the same effect and this is used in the cloud chamber apparatus. Here the tracks of charged particles through supersaturated water vapour become visible because of the formation of liquid water droplets. Work on first-order transition mechanisms is reviewed by Gunton et al. (1983) and Binder (1987).

5.2 The Van der Waals Equation

In real fluids there are strong repulsive forces between molecules when they are very close together and attractive forces at larger distances. Van der Waals in his thesis of 1873⁹ proposed a modification of the perfect gas law $PV = MT$ to take these effects into account. He included a term proportional to¹⁰ M , giving an effective reduction of the volume, and a term proportional to $(M/V)^2$ to increase the pressure, yielding

$$\left(P + \frac{\varepsilon M^2}{V^2}\right)(V - M) = TM. \quad (5.2.1)$$

For $T < T_c := \frac{8}{27}\varepsilon$ the isotherm in the P - V plane has a single loop with a range of pressures for which lines of constant pressure intersect the isotherm at three points and the critical isotherm with $T = T_c$ has a point of inflexion at $P = P_c := \frac{1}{27}\varepsilon$, $V = V_c := 3M$. We note, in reference to the law of corresponding states, discussed in Sect. 4.2, that (5.2.1) can be expressed in the form

$$\left(\frac{P}{P_c} + \frac{3V_c^2}{V^2}\right)\left(\frac{V}{V_c} - \frac{1}{3}\right) = \frac{8T}{3T_c}. \quad (5.2.2)$$

Using $K := \varepsilon/T$, $L' := -P/T$ and $v := V/M$, for which $K_c = \frac{27}{8}$, $L'_c = -\frac{1}{8}$ and $v_c = 3$, (5.2.1) can also be expressed in the coupling-density form

$$L'(K; v) = \frac{K}{v^2} - \frac{1}{v-1}. \quad (5.2.3)$$

⁸ Thus, contrary to the common belief, diamonds are not 'for ever', just for a very long time.

⁹ For a translation see Rowlinson (1988) and for an assessment of the legacy of van der Waals see Rowlinson (1973).

¹⁰ In fact, in terms of the variables introduced in Sect. 1.1, exactly equal to M .

From (5.1.9)¹¹

$$\phi_0^{(\text{cl})'}(v, K) = C(K) - \ln(v - 1) - \frac{K}{v}. \quad (5.2.4)$$

The use of primes in (5.2.3) and (5.2.4) follows the usage in (2.3.26) and is motivated by the intention of using $L := \mu/T$ as independent coupling with density $\rho := 1/v$. From (2.3.22) and (2.3.26),

$$L(K; \rho) = C(K) - 2K\rho - \ln\left(\frac{1-\rho}{\rho}\right) + \frac{1}{1-\rho}, \quad (5.2.5)$$

$$\phi_1^{(\text{cl})}(K, L; \rho) = \rho C(K) - K\rho^2 - \rho \ln\left(\frac{1-\rho}{\rho}\right) - L\rho. \quad (5.2.6)$$

Expanding in powers of $\Delta\rho := \rho - \rho_c$ gives

$$\phi_1^{(\text{cl})}(K, L; \rho) = B_0(K, L) + B_1(K, L)\Delta\rho + \sum_{i=2}^{\infty} B_i(K)(\Delta\rho)^i, \quad (5.2.7)$$

where the first few coefficients in the expansion are

$$\begin{aligned} B_0(K, L) &:= \frac{1}{3}C(K) - \frac{1}{3}\ln(2) - \frac{1}{9}K - \frac{1}{3}L, \\ B_1(K, L) &:= C(K) + \frac{3}{2} - \ln(2) - \frac{2}{3}K - L = L(K; \rho_c) - L, \\ B_2(K) &:= \frac{27}{8} - K = K_c - K, \quad B_3(K) := 0, \quad B_4(K) := \frac{243}{64}. \end{aligned} \quad (5.2.8)$$

5.3 Landau Expansions with One Order Parameter

In Chap. 4 we considered the form of ϕ in a neighbourhood of a critical region. The idea underlying the Landau (1937) expansion method for a classical approximation is very similar. Because ϕ is now a smooth function of the density (or densities) it can be expanded as a Taylor series in the differences of the densities from their critical values. These expansion variables are the *order parameters* of the system and the nature of the series (the presence or absence of particular powers in the expansion)

¹¹ By taking the perfect gas $v \rightarrow \infty$ limit, when the entropy is given by the Sackur-Tetrode equation, it can be shown (Lavis and Bell 1977) that

$$C(K) = -1 - \ln\left[v_0 \left(\frac{2m \varepsilon \pi}{Kh^2}\right)^{\frac{3}{2}}\right],$$

where m is the mass of a particle and h is Planck's constant.

can be deduced from the symmetry group of the order parameters. In this section we continue the case, discussed in the previous section, of a system with a critical point and a single density, and thus a single order parameter. In Sect. 5.4.2 we consider a case with two order parameters.¹²

The Taylor expansion of $\phi_1^{(\text{cl})}(K, L; \rho)$ about some $\rho = \rho_0$ takes the form¹³

$$\begin{aligned} \phi_1^{(\text{cl})}(K, L; \rho) := & \phi_1^{(\text{cl})}(K, L; \rho_0) + b_1(K, L; \rho_0)(\rho - \rho_0) \\ & + \sum_{i=2}^{\infty} b_i(K; \rho_0)(\rho - \rho_0)^i, \end{aligned} \quad (5.3.1)$$

where $b_1(K, L; \rho_0)$ is linear in L . The condition $b_1(K, L; \rho) = 0$ is equivalent to (5.1.7), yielding an equilibrium solution $\rho = \rho_e(K, L)$. If $b_2(K; \rho_e) > 0$ then $\rho = \rho_e$ is a minimum (stable equilibrium point) and if $b_2(K; \rho_e) < 0$ then it is a maximum (unstable equilibrium point). If $b_2(K; \rho_e) = 0$ and $b_3(K; \rho_e) \neq 0$, $\rho = \rho_e$ is a point of inflexion. If $b_2(K; \rho_e) = b_3(K; \rho_e) = 0$ the criteria are repeated with the index increased by two. However the values of K and L must, from Eqs. (5.1.6)–(5.1.8) and (5.1.13), now be the critical values. Thus the *critical parameters* K_c, L_c and ρ_c for a critical point terminating a line of first-order transitions in the classical approximation described in Sect. 5.1.1 satisfy the conditions

$$b_1(K, L; \rho) = 0, \quad (5.3.2)$$

$$b_2(K; \rho) = 0, \quad (5.3.3)$$

$$b_3(K; \rho) = 0. \quad (5.3.4)$$

Whether a known ϕ like (5.2.6) is expanded as a Taylor series, or the expansion is of a general form, one important question is immediately evident. Given that we truncate after a certain number of terms, how many are needed to reproduce the local properties of the full free-energy density near to the critical point?¹⁴

This is a problem which can be discussed most fruitfully in the context of *catastrophe theory* (Sect. 17.1). Let $\rho_0 := \rho_c$ with $\Delta\rho := \rho - \rho_c$ in (5.3.1). Then

$$\begin{aligned} \phi_1^{(\text{cl})}(K, L; \rho) = & B_0(K, L) + B_1(K, L)\Delta\rho + B_2(K)(\Delta\rho)^2 \\ & + B_3(K)(\Delta\rho)^3 + \sum_{i=4}^{\infty} B_i(K)(\Delta\rho)^i, \end{aligned} \quad (5.3.5)$$

¹² Order parameters are often given a vector form, the case of two order parameters being referred to as that of a two-dimensional order parameter.

¹³ That the coefficients $b_i(K; \rho_0)$ are not functions of L for $i > 1$ follows from the fact that L appears in $\phi_1^{(\text{cl})}(K, L; \rho)$ only in a term linear in ρ and L ; see (5.1.6) and, for the van der Waals equation, (5.2.6).

¹⁴ And, of course, the same question can be posed in relation the neighbourhood of other critical regions.

where, for brevity, we have adopted the notation

$$\begin{aligned} B_0(K, L) &:= \phi_1^{(\text{cl})}(K, L; \rho_c), & B_1(K, L) &:= b_1(K, L; \rho_c), \\ B_i(K) &:= b_i(K; \rho_c), & i &= 2, 3, \dots \end{aligned} \quad (5.3.6)$$

With ρ fixed at the critical value ρ_c we have the two equations (5.3.3) and (5.3.4) which must agree on the critical value $K = K_c$. In some examples like the van der Waals equation $B_3(K)$ is identically zero, but this is not necessarily the case and, in order to formulate the problem in the context of catastrophe theory, the order parameter in (5.3.5) is changed to

$$x := \Delta\rho - \delta, \quad (5.3.7)$$

where δ is chosen so that the cubic term is zero and

$$\phi_1^{(\text{cl})}(K, L; x) = C_0(K, L) + C_1(K, L)x + C_2(K)x^2 + G(K; x), \quad (5.3.8)$$

with

$$G(K; x) := \sum_{i=4}^{\infty} C_i(K)x^i. \quad (5.3.9)$$

Both δ and the coefficients $C_i(K)$ can be derived as series expansions in the parameter

$$y := \frac{B_3(K)}{4B_4(K)} \quad (5.3.10)$$

and it is not difficult to show that the required expression for δ gives

$$x = \Delta\rho + y + \frac{5B_5(K)}{2B_4(K)}y^2 + O(y^3), \quad (5.3.11)$$

with

$$\begin{aligned} C_0(K, L) &= B_0(K, L) - B_1(K, L)y \\ &\quad - \frac{5B_1(K, L)B_5(K) - 2B_2(K)B_4(K)}{2B_4(K)}y^2 + O(y^3), \end{aligned} \quad (5.3.12)$$

$$C_1(K, L) = B_1(K, L) - 2B_2(K)y - \frac{5B_2(K)B_5(K)}{B_4(K)}y^2 + O(y^3), \quad (5.3.13)$$

$$C_2(K) = B_2(K) - 6B_4(K)y^2 + O(y^3), \quad (5.3.14)$$

$$C_4(K) = B_4(K) - 5B_5(K)y + \frac{5(6B_4(K)B_6(K) - 5[B_5(K)]^2)}{2B_4(K)}y^2 + O(y^3), \quad (5.3.15)$$

with the remaining coefficients of the form $C_i(K) = B_i(K) + O(y)$. Suppose that $C_4(K) > 0$ in a neighbourhood of $K = K_c$. Then $G(K; x)$ has a degenerate minimum and a critical point¹⁵ at $x = 0$ for all K in a neighbourhood of K_c . It was shown in Example 17.1.3 that a function of this form has strong determinacy of 4 and codimension 2. $G(K; x)$ is strongly equivalent to $C_4(K)x^4$ and the universal unfolding (5.3.8) of $G(K; x)$ is strongly equivalent to

$$\tilde{\phi}_1^{(cl)}(K, L; x) := C_0(K, L) + C_1(K, L)x + C_2(K)x^2 + C_4(K)x^4. \quad (5.3.16)$$

All the local properties of $\phi_1^{(cl)}$ are present in $\tilde{\phi}_1^{(cl)}$. In particular, from Theorem 17.1.3, $\tilde{\phi}_1^{(cl)}$ has at most three equilibrium points near $x = 0$ for small $\Delta K := K - K_c$ and $\Delta L := L - L_c$. These are given by

$$C_1(K, L) = -2C_2(K)x - 4C_4(K)x^3, \quad (5.3.17)$$

with a solution being stable or unstable according as

$$2C_2(K) + 12C_4(K)x^2 >, < 0. \quad (5.3.18)$$

The equal-areas rule is satisfied for the cubic (5.3.17) by the coexistent non-zero solutions on the line

$$C_1(K, L) = 0, \quad (5.3.19)$$

which gives the coexistence curve in the K - L plane. From (5.3.11), (5.3.14) and (5.3.15) the solutions $\rho^{(0)}$ and $\rho^{(\pm)}$ to (5.3.17) and (5.3.19) are given by

$$\rho^{(0)} = \rho_c - y - \frac{5B_5(K)}{2B_4(K)}y^2 + O(y^3), \quad (5.3.20)$$

$$\rho^{(\pm)} = \rho^{(0)} \pm \sqrt{-\frac{C_2(K)}{2C_4(K)}}. \quad (5.3.21)$$

From (5.3.18), the solution $\rho^{(0)}$ is stable and the solutions $\rho^{(\pm)}$ unstable if $C_2(K) > 0$ with the converse being the case if $C_2(K) < 0$. We also see that the states $\rho^{(\pm)}$ can exist only if $C_2(K)$ and $C_4(K)$ have opposite signs. Thus, for a line of first-order

¹⁵ In both the statistical mechanical and catastrophe sense (Definition 17.1.11 together with the footnote).

transitions to develop from the critical point with a phase coexistence diagram of the form of Fig. 4.8, we must have $C_2(K) <, > 0$ for $K >, < K_c$ and $C_4(K) > 0$ for $K > K_c$ and $< K_c$. The simplest, although not unique way, to achieve this is to suppose that

$$\begin{aligned} B_0(K, L) &\simeq -L\rho_c + B_{00} + B_{01}\Delta K + B_{02}(\Delta K)^2, \\ B_1(K, L) &\simeq -\Delta L + B_{11}\Delta K + B_{12}(\Delta K)^2, \\ B_2(K) &\simeq -B_{21}\Delta K + B_{22}(\Delta K)^2, \\ B_3(K) &\simeq B_{31}\Delta K, \quad B_4(K) \simeq B_{40} + B_{41}\Delta K, \end{aligned} \quad (5.3.22)$$

where B_{21} and B_{40} (but not necessarily the other constants) are positive.¹⁶ Then, from (5.3.10), (5.3.14), (5.3.15) and (5.3.20)–(5.3.22),

$$\rho^{(0)} \simeq \rho_c - \frac{B_{31}}{B_{40}}\Delta K, \quad (5.3.23)$$

$$\rho^{(\pm)} \simeq \rho^{(0)} \pm \sqrt{\frac{B_{21}}{2B_{40}}}(\Delta K)^{\frac{1}{2}}, \quad (5.3.24)$$

along the coexistence curve ($K > K_c$) and, from (5.3.11), (5.3.13)–(5.3.15), (5.3.17) and (5.3.22),

$$\Delta L \simeq 4B_{40}(\Delta\rho)^3, \quad (5.3.25)$$

along the critical isotherm ($K = K_c$). From (2.3.29)–(2.3.32), (5.3.11), (5.3.13), (5.3.17) and (5.3.22)–(5.3.24)

$$\varphi_T \simeq \begin{cases} -\frac{1}{2B_{21}}\Delta K, & \text{along the critical isochore, } K < K_c, \\ \frac{1}{4B_{21}}\Delta K, & \text{along the coexistence curve, } K > K_c, \end{cases} \quad (5.3.26)$$

$$c_\rho \simeq \begin{cases} -k_B K_c^2 [2B_{02} + O(\Delta K)], & \text{along the critical isochore, } K < K_c, \\ -k_B K_c^2 \left[2B_{02} + O([\Delta K]^{\frac{1}{2}}) \right], & \text{along the coexistence curve, } K > K_c, \end{cases} \quad (5.3.27)$$

¹⁶ We have explicitly displayed only those coefficients needed in the calculations of the asymptotic forms for the response functions.

$$c_\xi \simeq \begin{cases} -k_B K_c^2 \left[\frac{Q^2}{2B_{21}} (\Delta K)^{-1} + 2B_{02} + \frac{f_1(Q)}{2B_{21}^2} + O(\Delta K) \right], \\ \text{along the critical isochore, } K < K_c, \\ -k_B K_c^2 \left[-\frac{Q^2}{4B_{21}} (\Delta K)^{-1} + \frac{f_2(Q)}{16B_{21}^2} (\Delta K)^{-\frac{1}{2}} \right. \\ \left. + 2B_{02} + \frac{f_3(Q)}{32B_{40}B_{21}^2} + O([\Delta K]^{\frac{1}{2}}) \right], \\ \text{along the coexistence curve, } K > K_c, \end{cases} \quad (5.3.28)$$

where

$$\begin{aligned} Q &:= \frac{B_{11}K_c - L_c}{K_c}, & f_1(Q) &:= Q(4B_{12}B_{21} + QB_{22}), \\ f_2(Q) &:= Q\sqrt{\frac{2B_{21}}{B_{40}}}(8B_{21}^2 + 3QB_{31}), \\ f_3(Q) &:= -16B_{21}^4 - 16Q(3B_{21}^2B_{31} + 2B_{21}B_{12}B_{40}) \\ &\quad + Q^2(4B_{22}B_{40} + 12B_{21}B_{41} - 9B_{31}^2). \end{aligned} \quad (5.3.29)$$

Comparing (5.3.24)–(5.3.27) with (4.4.12)–(4.4.15) we obtain the *classical critical exponent values*¹⁷

$$\alpha = \alpha' = 0, \quad \beta = \frac{1}{2}, \quad \gamma = \gamma' = 1, \quad \delta = 3, \quad (5.3.30)$$

with c_ρ finite and continuous with a discontinuous gradient at the critical point. Since $f_1(0) = f_2(0) = 0$, $f_3(0) = -16B_{21}^4$, it follows from (4.4.16) and (5.3.28) that

$$\sigma = \sigma' = \begin{cases} \gamma, & \text{if } Q \neq 0, \\ \alpha, & \text{if } Q = 0. \end{cases} \quad (5.3.31)$$

From (5.3.19) and (5.3.22), a which appears in (4.5.5) is equal to B_{11} and (5.3.31) is the result given by (4.5.43).

¹⁷ Sometimes referred to as the mean-field or Landau values.

5.3.1 The Spin- $\frac{1}{2}$ Ising Model

The spin- $\frac{1}{2}$ Ising model, with Hamiltonian (3.5.1) has magnetization density m conjugate to $L := \mathcal{H}/T$. As has been discussed in Sect. 3.5.1 the symmetry group of the Hamiltonian is $\mathcal{S}_2^{(mL)} := \{I, Q^{(m)}Q^{(L)}\}$, where $Q^{(m)}$ and $Q^{(L)}$ denote spin and field inversion respectively. The free-energy density $\phi_1^{(\text{cl})}(K, L; m)$ inherits the symmetry properties of the Hamiltonian, with $Q^{(m)} : m \rightarrow -m$. Its most general form would therefore be a linear combination of the invariants $m^i L^j$ of $\mathcal{S}_2^{(mL)}$, for all $i + j$ even. However, from (3.5.1), L can appear only linearly. So apart from a term containing mL , $\phi_1^{(\text{cl})}(K, L; m)$ is an even function of m (see Eqs. (6.1.9) and (6.1.10)). As we showed in Example 17.1.5 the catastrophe theory analysis remains valid, but now $B_{2i+1}(K) = 0$ for $i = 1, 2, \dots$ with $B_1(K) = -L$. The calculation to remove the cubic term in the expansion is not required, $C_i(K) = B_i(K)$, $i = 0, 1, 2, \dots$ and, from (5.3.16),

$$\phi_1^{(\text{cl})}(K, L; m) := B_0(K, L) - Lm + B_2(K)m^2 + B_4(K)m^4. \quad (5.3.32)$$

The solutions (5.3.20) and (5.3.21) reduce to $m = 0$ and

$$m^2 = -\frac{B_2(K)}{2B_4(K)} \simeq \frac{B_{21}\Delta K}{B_{40}}. \quad (5.3.33)$$

From (5.3.19), the critical curve in the K - L plane lies along the line $L = 0$.

Landau theory and scaling theory can both be seen as methods which extract information about a whole class of physical systems, based on a minimum of assumptions about their physical structure. The difference between the two is, of course, that, while scaling theory is capable of determining only scaling law relationships between critical exponents, Landau theory allows actual values for the exponents to be obtained. It is an illuminating exercise to apply the procedures of scaling theory to the free-energy density (5.3.32), in order to obtain the critical exponents. The scaling fields in this case are, as for the example of Kadanoff (1966) described in Sect. 4.2 and for the mean-field calculation of Sect. 6.1, $\theta_1 = L$ and $\theta_2 = \Delta K$. Let the smooth part of $\phi_1^{(\text{cl})}(K, L; m)$ be identified as $B_0(K)$. Then

$$\phi_{\text{sing}}(\theta_1, \theta_2; m) = -\theta_1 m - B_{21}\theta_2 m^2 + B_{40}m^4. \quad (5.3.34)$$

Suppose the homogeneity condition, with exponents associated with θ_1 and θ_2 but also with m . Then

$$\phi_{\text{sing}}(\lambda^{y_1}\theta_1, \lambda^{y_2}\theta_2; \lambda^z m) = \lambda^d \phi_{\text{sing}}(\theta_1, \theta_2; m). \quad (5.3.35)$$

It follows from (5.3.34) and (5.3.35) that

$$y_1 = \frac{3}{4}d, \quad y_2 = \frac{1}{2}d, \quad z = \frac{1}{4}d. \quad (5.3.36)$$

The exponent z for the magnetization satisfies, as it must (Eq. (4.3.8)), the condition $z = d - y_1$. Substituting into (4.5.40) and (4.5.39) the classical critical exponent values given in (5.3.30) are recovered. The exponent γ applies to the divergence of the susceptibility as the critical point is approached from above and below along the zero-field axis. However, it can be easily seen that these directions yield different amplitudes. Since $\alpha = 0$ there is no divergence, at the critical point, in the heat capacity $c_{\mathcal{H}}$ at constant field along the zero-field axis. However, there is a discontinuity, given, from (5.3.28) (with $B_{11} = 0$, $L_c = 0$, giving $Q = 0$) by

$$\Delta c_{\mathcal{H}} = \frac{k_B K_c^2 B_{21}^2}{2B_{40}}. \quad (5.3.37)$$

5.4 Landau Expansions with Two Order Parameters

5.4.1 The Spin-1 Ising Model

The Hamiltonian of the spin-1 Ising model is given by (3.5.29) and there are now two densities $m := \langle \sigma(\mathbf{r}) \rangle$, conjugate to $\mathcal{H}$ and $\rho := \langle \sigma^2(\mathbf{r}) \rangle$ conjugate to μ . The free-energy density is a function of these densities and four couplings, $K_1 := J/T$, $K_2 := \varepsilon/T$, $L_1 := \mathcal{H}/T$ and $L_2 := \mu/T$. The symmetry group is in general the same as the spin- $\frac{1}{2}$ case. The parameter ρ , is unaffected by spin inversion and is thus related to the one-dimensional symmetric representation of S_2 (Sect. 17.3.4) giving

$$\begin{aligned} \phi_1^{(\text{cl})}(K_1, K_2, L_1, L_2; m, \rho) = & -L_1 m - L_2 \rho + B_0(K_1, K_2; \rho) \\ & + \sum_{i=1}^{\infty} B_{2i}(K_1, K_2; \rho) m^{2i}. \end{aligned} \quad (5.4.1)$$

Since, in general, no new symmetry is associated with the parameter ρ there is no way, within the framework of Landau theory, of obtaining information about a possible transition with ρ as order parameter. There are two exceptions to this. One is the case of the 3-state Potts model, discussed in Sect. 5.4.2, and the other is when $J = \varepsilon' = \mathcal{H} = 0$. Then, as we have indicated in Sect. 3.5.2, the Hamiltonian is equivalent to the lattice gas Hamiltonian (3.5.20). The order parameter $n := \langle \sigma^2(\mathbf{r}) \rangle - \frac{1}{2}$ is invariant under the operator $Q^{(n)}$ which changes all site states $|\sigma| = 1$ to zero and sites states zero to $+1$ or -1 with equal probability. The first-order transition line and critical point lie, from (3.5.22), on $\mu = \mu_c = -\frac{1}{2}z\varepsilon$ and the Landau expansion has the same form as that for the spin- $\frac{1}{2}$ model with n replacing m , $(\mu - \mu_c)/T$ replacing L and ε/T replacing K .

5.4.2 The 3-State Potts Model

The ν -state Potts model is described in Sect. 3.6.4. Here we use Landau theory to consider the case $\nu = 3$ and $\mu = 0$, which was shown (Sect. 3.6.4.2) to be a special case of the spin-1 Ising model, with Hamiltonian (3.5.29), in zero field. This means that, if the 3-state Potts model has a transition, this will occur in a subspace of the phase space of the spin-1 Ising model.

We first need to reformulate the problem in terms of two order parameters which are zero at high temperatures. This is true for $m := \langle \sigma(\mathbf{r}) \rangle$, since the states $+1$ and -1 will be equally probable. But now we have three equally probable states with $\langle \sigma^2(\mathbf{r}) \rangle = \frac{2}{3}$ at high temperatures. So we choose the two order parameters¹⁸

$$\begin{aligned} p &:= \frac{2}{\sqrt{3}} \left(1 - \frac{3}{2} \langle \sigma^2(\mathbf{r}) \rangle \right), \\ m &:= \langle \sigma(\mathbf{r}) \rangle. \end{aligned} \quad (5.4.2)$$

The group $\mathcal{S}_3$ has six elements and three generators I , $Q^{(m)}$ and

$$Q^{(*)} : -1 \rightarrow 0 \rightarrow +1 \rightarrow -1. \quad (5.4.3)$$

The group $\mathcal{S}_2^{(m)}$, which is a subgroup of $\mathcal{S}_3$, has the two-dimensional representation

$$R(I) = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R(Q^{(m)}) = Q^{(m)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (5.4.4)$$

for operations on the vector

$$\mathbf{x} := \begin{pmatrix} p \\ m \end{pmatrix}. \quad (5.4.5)$$

This representation is, of course, reducible, corresponding to the symmetric and antisymmetric representations A and B with operations on the order parameters p and m respectively. From (5.4.2) and (5.4.3),

$$\begin{aligned} Q^{(*)} : p &\rightarrow -\frac{1}{2}p + \frac{\sqrt{3}}{2}m \rightarrow -\frac{1}{2}p - \frac{\sqrt{3}}{2}m \rightarrow p, \\ Q^{(*)} : m &\rightarrow -\frac{\sqrt{3}}{2}p - \frac{1}{2}m \rightarrow \frac{\sqrt{3}}{2}p - \frac{1}{2}m \rightarrow m, \end{aligned} \quad (5.4.6)$$

¹⁸ The leading factor of $2/\sqrt{3}$ in p is included in order to make the matrix representations of $Q^{(m)}$ and $Q^{(*)}$ unitary. It is omitted in series calculations (see Enting and Wu 1982, and Sect. 14.2.1).

Table 5.1 The character table for $\mathcal{S}_3$

$\mathcal{S}_3$	$\mathcal{Q}_1$	$\mathcal{Q}_2$	$\mathcal{Q}_3$
A	1	1	1
B	1	1	-1
E	2	-1	0

giving the unitary matrix

$$R(Q^{(*)}) = Q^{(*)} = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \quad (5.4.7)$$

$\mathcal{S}_3$ is generated by $Q^{(m)}$ and $Q^{(*)}$ and since $Q^{(m)}Q^{(*)} = Q^{(*)2}Q^{(m)}$ and $Q^{(m)}Q^{(*)2} = Q^{(*)}Q^{(m)}$ the conjugacy classes of $\mathcal{S}_3$ are $\mathcal{Q}_1 := I$, $\mathcal{Q}_2 := \{Q^{(*)}, Q^{(*)2}\}$ and $\mathcal{Q}_3 := \{Q^{(m)}, Q^{(m)}Q^{(*)}, Q^{(m)}Q^{(*)2}\}$ with character Table 5.1 (see e.g. Leech and Newman 1969, p. 109).¹⁹ It follows from (17.3.2) that the matrices (5.4.4) and (5.4.7) are generators of the two-dimensional irreducible representation E.

To obtain the Landau expansion for the 3-state Potts model we need to determine the functions of p and m which are invariant under the operations of $\mathcal{S}_3$. Since

$$\mathbf{x}^T Q^{(*)T} Q^{(*)} \mathbf{x} = \mathbf{x}^T Q^{(m)T} Q^{(m)} \mathbf{x} = \mathbf{x}^T \mathbf{x} = p^2 + m^2, \quad (5.4.8)$$

a set of even-order invariants are given by powers of

$$v_E(p, m) := p^2 + m^2. \quad (5.4.9)$$

Odd-order invariants under the operation $Q^{(*)}$ are given, from (5.4.6), by observing that the product of the elements of each of these three-cycles will give a third-order invariant under $Q^{(*)}$. The second of these is odd with respect to m and thus not an invariant under $Q^{(m)}$. However, the first gives the function

$$v_O(p, m) := p(p^2 - 3m^2), \quad (5.4.10)$$

which is an odd-order invariant. All other invariants are products of powers of $v_E(p, m)$ and $v_O(p, m)$ (Gaite 1992).

¹⁹ $\mathcal{S}_3$ is isomorphic to the dihedral group $\mathcal{D}_3$ and to $\mathcal{C}_{3v}$ (see the character table in Example 17.3.2).

Applying the analysis of catastrophe theory (Example 17.1.6), the Landau expansion for the 3-state Potts model is equivalent to

$$\begin{aligned} \phi_1^{(\text{cl})}(K; p, m) := & B_0(K) + B_2(K)(p^2 + m^2) + B_3(K)p(p^2 - 3m^2) \\ & + B_4(K)(p^2 + m^2)^2, \end{aligned} \quad (5.4.11)$$

as long as $B_4(K) \neq 0$. Stable equilibrium states are given by minimizing this free-energy density with respect to p and m . The equilibrium conditions are

$$\begin{aligned} 0 &= 2B_2(K)p + 3B_3(K)(p^2 - m^2) + 4B_4(K)p(p^2 + m^2), \\ 0 &= 2B_2(K)m - 6B_3(K)p m + 4B_4(K)m(p^2 + m^2), \end{aligned} \quad (5.4.12)$$

and the stability conditions are obtained by considering the determinant

$$\Phi(K; p, m) := \begin{vmatrix} \phi_{pp}(K; p, m) & \phi_{pm}(K; p, m) \\ \phi_{mp}(K; p, m) & \phi_{mm}(K; p, m) \end{vmatrix}, \quad (5.4.13)$$

where

$$\begin{aligned} \phi_{pp}(K; p, m) &:= 2B_2(K) + 6B_3(K)p + 4B_4(K)(3p^2 + m^2), \\ \phi_{pm}(K; p, m) &:= \phi_{mp}(K; p, m) = -6B_3(K)m + 8B_4(K)p m, \\ \phi_{mm}(K; p, m) &:= 2B_2(K) - 6B_3(K)p + 4B_4(K)(p^2 + 3m^2). \end{aligned} \quad (5.4.14)$$

A solution to (5.4.12) is a maximum or a minimum of $\tilde{\phi}^{(\text{cl})}(K; p, m)$ if $\Phi(K; p, m) > 0$, otherwise it is a saddle point. In the former case the solution is a maximum or minimum according as $\phi_{pp}(K; p, m)$ is negative or positive. The high-temperature solution $p = m = 0$ is a maximum if $B_2(K) < 0$ and a minimum if $B_2(K) > 0$. Suppose that, to leading order in ΔK ,

$$B_2(K) \simeq -B_{21}\Delta K, \quad B_3(K) \simeq B_{30}, \quad B_4(K) \simeq B_{40}, \quad (5.4.15)$$

where B_{21} and B_{40} (but not necessarily B_{30}) are positive. Then the high-temperature solution is stable if $K < K_c$ ($T > T_c$). In general, the derivation of non-zero solutions to (5.4.12) is quite complicated. We shall, however, be concerned only with solutions at the critical temperature, which are not in a neighbourhood of the high-temperature solution. In this case it is not difficult to show that, when $K = K_c$ ($B_2(K_c) = 0$), there are three such solutions

$$\mathbf{x}_1^{(\pm)} := \frac{3B_{30}}{4B_{40}} \begin{pmatrix} \frac{1}{2} \\ \pm \frac{\sqrt{3}}{2} \end{pmatrix} \quad \mathbf{x}_2 := -\frac{3B_{30}}{4B_{40}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (5.4.16)$$

Given any solution to (5.4.12), other solutions can be obtained by the operations of the group $\mathcal{S}_3$. In general solutions will occur in sets of six, corresponding to the six elements of $\mathcal{S}_3$. However, in this case

$$\begin{aligned} Q^{(m)}x_1^{(+)} &= x_1^{(-)}, & Q^{(*)}x_1^{(-)} &= x_1^{(+)}, \\ Q^{(m)}x_2 &= x_2, & Q^{(*)}x_1^{(+)} &= x_2, \\ Q^{(m)}x_1^{(-)} &= x_1^{(+)}, & Q^{(*)}x_2 &= x_1^{(-)}, \end{aligned} \quad (5.4.17)$$

and, since the group is generated by $Q^{(m)}$ and $Q^{(*)}$, the solutions (5.4.16) form a closed set under the operations of the group. Since

$$\begin{aligned} \Phi(K_c; p_1^{(\pm)}, m_1^{(\pm)}) &= \frac{243B_{30}^4}{16B_{40}^2}, & \phi_{pp}(K_c; p_1^{(\pm)}, m_1^{(\pm)}) &= \frac{45B_{30}^2}{8B_{40}}, \\ \tilde{\phi}_1^{(\text{cl})}(K_c; p_1^{(\pm)}, m_1^{(\pm)}) &= B_0(K_c) - \frac{27B_{30}^4}{256B_{40}^3}, \end{aligned} \quad (5.4.18)$$

$$\begin{aligned} \Phi(K_c; p_2, m_2) &= \frac{243B_{30}^4}{16B_{40}^2}, & \phi_{pp}(K_c; p_2, m_2) &= \frac{9B_{30}^2}{4B_{40}}, \\ \phi_1^{(\text{cl})}(K_c; p_2, m_2) &= B_0(K_c) - \frac{27B_{30}^4}{256B_{40}^3}, \end{aligned} \quad (5.4.19)$$

it follows that, when $B_{40} > 0$, the solutions (5.4.16) are lower minima of $\phi_1^{(\text{cl})}(K; p, m)$ than $p = m = 0$ at, and in a neighbourhood of, the critical temperature. This means that the system will have already passed, with equal probability, to one of these states at some temperature above the critical value. The phase transition will be first-order since it involves a discontinuous change in p and m .

The prediction by Landau theory of a first-order phase transition in the 3-state Potts model is an interesting example of the limitations of this method. The results of Landau theory for the order of phase transitions are the same as those obtained by any mean-field calculation and, for the general ν -state Potts model, these have been shown (Mittag and Stephen 1974) to give a first-order transition for $\nu > 2$ in any dimension. In contradiction to this it has been shown rigorously by Baxter (1973d), for the square lattice and by Baxter et al. (1978), for the triangular lattice, that the ferromagnetic ν -state Potts model has a first-order transition for $\nu > 4$, but a higher-order transition for $\nu \leq 4$. For the antiferromagnetic model the situation is in general more complicated (Wang et al. 1989) because of the presence of infinite ground-state degeneracy. There are two exceptions to this. On the square lattice the antiferromagnetic 3-state Potts model is, like the ferromagnetic case, equivalent to a staggered six-vertex model (Sect. 3.8.2.1) with a higher-order transition although in this case it occurs at zero temperature. On the triangular lattice the ground state is six-fold degenerate (Schick and Griffiths 1977) like the triangular spin- $\frac{1}{2}$ Ising

antiferromagnet in a field. Monte Carlo simulations (Grest 1981; Saito 1981; Adler et al. 1995), series methods (Enting and Wu 1982) and analytical investigations (Park 1994) indicate that the transition is first-order in agreement with the prediction of Landau theory.

5.5 Landau Theory for a Tricritical Point

In Sect. 4.6 we described a magnetic model which exhibits a critical line meeting a triple line at a tricritical point in the symmetry plane with the staggered field $\mathcal{H}_s = 0$ (Fig. 4.10). In Sect. 6.2 we show, using mean-field theory, that this model can be realized by a metamagnetic generalization of the spin- $\frac{1}{2}$ Ising model for which we have discussed a Landau treatment in Sect. 5.3.1. Points in the symmetry plane $L_s := \mathcal{H}_s/T = 0$ are given in terms of the coupling $K := J/T$ and $L := \mathcal{H}/T$ and ϕ_2 is an even function of q , the order parameter associated with the staggered field. Thus we have

$$\begin{aligned} \phi_2^{(\text{cl})}(K, L, L_s; q) := & -L_s q + B_0(K, L) + B_2(K, L)q^2 \\ & + B_4(K, L)q^4 + G(L, L_s; q), \end{aligned} \quad (5.5.1)$$

where

$$G(L, L_s; q) := \sum_{i=3}^{\infty} B_{2i}(K, L)q^{2i}. \quad (5.5.2)$$

We consider only the case $B_6(K, L) > 0$, for which, as it was shown in Example 17.1.5, (5.5.1) is strongly equivalent to

$$\begin{aligned} \tilde{\phi}_2^{(\text{cl})}(K, L, L_s; q) := & -L_s q + B_0(K, L) + B_2(K, L)q^2 \\ & + B_4(K, L)q^4 + B_6(K, L)q^6. \end{aligned} \quad (5.5.3)$$

The conditions for minimizing $\phi_2^{(\text{cl})}$ as a function of q are

$$2B_2(K, L)q + 4B_4(K, L)q^3 + 6B_6(K, L)q^5 - L_s = 0, \quad (5.5.4)$$

$$2B_2(K, L) + 12B_4(K, L)q^2 + 30B_6(K, L)q^4 > 0. \quad (5.5.5)$$

In the symmetry plane when $L_s = 0$ the solutions of (5.5.4) are $q = 0$, which is stable when $B_2(K, L) > 0$, and

$$q^{(\pm)2} = -\frac{B_4(K, L)}{3B_6(K, L)} \left[1 \pm \sqrt{1 - w(K, L)} \right], \quad (5.5.6)$$

where

$$w(K, L) := \frac{3B_2(K, L)B_6(K, L)}{B_4^2(K, L)}. \quad (5.5.7)$$

From (5.5.4) and (5.5.5) the stability condition for the solutions (5.5.6) becomes

$$\mp B_4(K, L)\sqrt{1 - w(K, L)} > 0. \quad (5.5.8)$$

It follows that the possible stable solutions from (5.5.6) are $\pm q^{(-)}$ or $\pm q^{(+)}$ according as $B_4(K, L) >, < 0$ respectively.

- In the case $B_4(K, L) > 0$, $q^{(-)}$ is real only when $w(K, L) < 0$. The curve

$$L = L^*(K), \quad (5.5.9)$$

given as a solution of

$$B_2(K, L) = 0, \quad (5.5.10)$$

is, as in Sect. 4.7.1, the critical line $\mathcal{C}$ of continuous transitions in the plane $L_s = 0$, with the ordered and disordered phases stable in the regions $B_2(K, L) < 0$ and $B_2(K, L) > 0$ respectively. On physical grounds we expect the ordered phase to occur on the side of $\mathcal{C}$ for which $L < L^*(K)$.

- In the case $B_4(K, L) < 0$, (5.5.6) gives a real value of $q^{(+)}$ as long as $w(K, L) < 1$ and, from (5.5.3),

$$\begin{aligned} \tilde{\phi}_2^{(\text{cl})}(K, L, 0; q^{(+)}) &= B_0(K, L) \\ &\quad + \frac{2}{3} B_4(K, L) q^{(+)^4} \left[\sqrt{1 - w(K, L)} - \frac{1}{2} \right]. \end{aligned} \quad (5.5.11)$$

The two minima $\pm q^{(+)}$ are therefore metastable, with the stable solution given by $q = 0$, when $w(K, L) < \frac{3}{4}$. A first-order phase transition occurs at $w(K, L) = \frac{3}{4}$, or

$$B_2(K, L) = \frac{B_4^2(K, L)}{4B_6(K, L)}, \quad (5.5.12)$$

from the disordered state to the ordered state, with a discontinuous change of the order parameter to

$$q^{(+)} = \pm \sqrt{\frac{-B_4(K, L)}{2B_6(K, L)}}, \quad (5.5.13)$$

the two values occurring with equal probability. The solution curve $\mathcal{T}$ to the equation (5.5.12) is a line of first-order transitions, which becomes a triple line in the full phase diagram in which the staggered coupling L_s is included (Fig. 4.10). Suppose that $\mathcal{T}$ takes the form

$$L = L^\dagger(K). \quad (5.5.14)$$

Then the critical curve $\mathcal{C}$ given by (5.5.9) and the first-order curve $\mathcal{T}$ given by (5.5.14) meet at a tricritical point $\mathbf{T}$ with coordinates $(K_t, L_t, L_s = 0)$ given as the solution of²⁰

$$B_2(K, L) = 0, \quad B_4(K, L) = 0, \quad (5.5.15)$$

where

$$L_t = L^*(K_t) = L^\dagger(K_t). \quad (5.5.16)$$

It is straightforward to show that $\mathcal{C}$ and $\mathcal{T}$ together form a curve in the K – L plane with a continuous gradient at $\mathbf{T}$.

5.5.1 Tricritical Exponents

In Sect. 4.6 exponents for the tricritical point $\mathbf{T}$ were introduced and orthogonal coordinates defined to represent the directions along and perpendicular to the tangent to $\mathcal{C}$ at $\mathbf{T}$. Here we repeat this procedure but alter the point of interest to an arbitrarily chosen point $\mathbf{P}$ on $\mathcal{C}$, whilst also considering the limit as $\mathbf{P} \rightarrow \mathbf{T}$. The set of exponents at $\mathbf{P}$ are defined as in (4.6.15)–(4.6.22) with the ‘t’ omitted. Let

$$a_p := \left(\frac{dL^*}{dK} \right)_p, \quad (5.5.17)$$

with orthogonal coordinates

$$\Theta_2^{(p)} := L - a_p K, \quad \Theta_3^{(p)} := K + a_p L. \quad (5.5.18)$$

At the tricritical point the label ‘p’ is replaced by ‘t’.²¹ Now, following (4.6.12) and (4.6.13), let

$$L^*(K) := L_p + a_p \Delta_p K + \mathcal{L}_p(K), \quad \Delta_p K := K - K_p, \quad (5.5.19)$$

where

$$\mathcal{L}_p(K_p) = 0, \quad \mathcal{L}'_p(K_p) = 0. \quad (5.5.20)$$

²⁰ The free-energy density (5.5.3) is, therefore, the universal unfolding of $B_6(KL)q^6$ at the tricritical point.

²¹ It is clear that $(\Theta_2^{(t)} \Theta_3^{(t)})$ are the coordinates (Θ_2, Θ_3) defined in (4.6.14) with $a = a_t$. It is for this reason that we have retained the labels 2,3 rather than the more obvious 1,2.

In order to satisfy the critical curve condition (5.5.10), the tricritical condition (5.5.15), and with $B_4(K, L)$ and $B_6(K, L)$ having the appropriate signs, we suppose that, to leading order,

$$\begin{aligned} B_2(K, L) &\simeq B_{21}[L - L^*(K)] \simeq B_{21}[\Theta_2^{(p)} - \Theta_{2p}^{(p)}], \\ B_4(K, L) &\simeq -B_{41}\Delta_t K, \\ B_6(K, L) &\simeq B_{60}, \end{aligned} \quad (5.5.21)$$

where B_{21} , B_{41} and B_{60} are positive constants.²² From (5.5.7),

$$w(K, L) \simeq \frac{3B_{21}[\Theta_2^{(p)} - \Theta_{2p}^{(p)}]}{B_{41}^2(\Delta_t K)^2}. \quad (5.5.22)$$

At the point $\mathbf{P}$ consider an approach to $\mathcal{C}$ in the orthogonal direction $\Theta_3^{(p)} = \Theta_{3p}^{(p)}$ within the ordered region ($\Theta_2^{(p)} < \Theta_{2p}^{(p)}$). From (5.5.6), (5.5.7) and (5.5.21),

$$q^2 = q^{(-)2} \simeq -\frac{B_{21}[\Theta_2^{(p)} - \Theta_{2p}^{(p)}]}{2B_{41}\Delta_t K} \quad (5.5.23)$$

and it follows from (5.5.23) that $\beta = \frac{1}{2}$. At the tricritical point $\Theta_i^{(p)} = \Theta_i^{(t)}$, $i = 2, 3$ and with $\Theta_3^{(t)} = \Theta_{3t}^{(t)}$

$$\Delta_t K = -\frac{a_t[\Theta_2^{(t)} - \Theta_{2t}^{(t)}]}{1 + a_t^2}, \quad (5.5.24)$$

$w(K, L)$ diverges at the tricritical point and

$$q^2 = q^{(-)2} \simeq \sqrt{\frac{B_{21}[\Theta_2^{(t)} - \Theta_{2t}^{(t)}]}{3B_{60}}}. \quad (5.5.25)$$

Thus, from (4.6.19), $\beta_t = \frac{1}{4}$. Consider now the asymptotic form for q^2 as the tricritical point is approached along the critical curve. From (5.5.5), (5.5.6) and (5.5.10), $w(K, L)$ is identically zero and, from (5.5.18),

$$\Delta_t K = \frac{[\Theta_3^{(t)} - \Theta_{3t}^{(t)}]}{1 + a_t^2}, \quad (5.5.26)$$

²² On $\mathcal{C}$, $K < K_t$, and, on $\mathcal{T}$, $K > K_t$ (Fig. 4.10).

giving

$$q^2 = q^{(-)2} \simeq -\frac{B_{41}[\Theta_3^{(t)} - \Theta_{3t}^{(t)}]}{3(1 + a_t^2)B_{60}}. \quad (5.5.27)$$

It follows from (5.5.26) that $\Theta_3^{(t)} < \Theta_{3t}^{(t)}$ as the tricritical point is approached along the critical curve and, from (4.6.19), $\beta_{tc} = \frac{1}{2}$. From (4.6.3) and (5.5.4)

$$\psi_T := \left(\frac{\partial q}{\partial L_s} \right)_{K,L} \simeq \frac{1}{2B_2(K, L) + B_4(K, L)q^2 + 30B_6q^4}. \quad (5.5.28)$$

In the disordered region where $q = 0$, from (4.6.21) and (5.5.21), $\gamma = \gamma_t = 1$ and in the ordered region, using asymptotic forms for q^2 given by (5.5.23), (5.5.25) and (5.5.27) $\gamma' = \gamma'_t = 1$, $\gamma_{tc} = 2$. From (5.5.4) and (5.5.21), when $K = K_p$

$$L_s \simeq 4B_{41}(K_t - K_p)q^3 + 6B_{60}q^5, \quad (5.5.29)$$

giving, from (4.6.15) $\delta = 3$, $\delta_t = 5$. The exponents β , γ and δ on $\mathcal{C}$ away from the tricritical point are the same as those given by (5.3.30). This means that, if the coincidence of values (but not, of course, the values themselves) applies to the exact and not just classical calculations, the second-order transition on the critical curve is in the universality class of a ordinary critical point with $\alpha' = 0$ (Sect. 4.2.1).

The classical values of the remaining tricritical exponents, defined in (4.6.16)–(4.6.18), and their equivalents on the critical curve could be derived in a similar way.²³ However, if scaling is assumed, it follows from the discussion in Sect. 4.7.2 that, once we have expressions for y_1 , y_2 and y_3 , all the remaining exponents along $\mathcal{C}$ and in the ordered region can be derived. Using the values we have just derived for δ_t , β_t and β_{tc} and (4.7.25) and (4.7.30)

$$y_1 = \frac{d\delta_t}{1 + \delta_t} = \frac{5}{6}d, \quad y_2 = \frac{d - y_1}{\beta_t} = \frac{2}{3}d, \quad y_3 = \frac{d - y_1}{\beta_{tc}} = \frac{1}{3}d. \quad (5.5.30)$$

These expressions then, from (4.7.31), return our values for γ'_t and γ_{tc} and, using (4.7.32)–(4.7.36), we can derive the set of classical values for the tricritical exponents defined in (4.6.16)–(4.6.18).

²³ For the heat capacity exponents we should need also to assume an asymptotic form for $B_0(K, L)$ as we did for the critical point in (5.3.22).

$$\begin{aligned}
\alpha'_t &= \frac{1}{2}, & \alpha_{tc} &= -1, \\
\beta_t &= \frac{1}{4}, & \beta_{tc} &= \frac{1}{2}, \\
\beta'_{2t} &= \frac{1}{2}, & \beta_{2tc} &= 1, \\
\gamma'_t &= 1, & \gamma_{tc} &= 2, \\
\gamma'_{2t} &= \frac{1}{2}, & \gamma_{2tc} &= 1, \\
\delta_t &= 5.
\end{aligned} \tag{5.5.31}$$

5.6 Ginzburg–Landau Theory

5.6.1 A Critical Point

Let $\mathcal{V}$ be a d -dimensional hypercube of volume R^d . In terms of some reference volume v_0 , $V = R^d/v_0$ is the dimensionless volume of the system. At every point $\mathbf{r}$ of $\mathcal{V}$ an n -dimensional vector field $\boldsymbol{\sigma}(\mathbf{r})$ is defined, in terms of which the statistical mechanical magnetization density is

$$\hat{\mathbf{m}} := \frac{1}{R^d} \int_{\mathcal{V}} \boldsymbol{\sigma}(\mathbf{r}) d\mathcal{V}, \tag{5.6.1}$$

(cf. (3.5.4), (3.5.5)). The vector field $\boldsymbol{\sigma}(\mathbf{r})$ is the *local* magnetization density with conjugate vector coupling $\mathbf{L} := \mathcal{H}/T$, where $\mathcal{H}$ is an n -dimensional magnetic field. The local free-energy density is given by generalizing (5.3.32) to²⁴

$$\begin{aligned}
\hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \boldsymbol{\sigma}(\mathbf{r})) &:= -\mathbf{L} \cdot \boldsymbol{\sigma}(\mathbf{r}) + B_0(K) + B_2(K) |\boldsymbol{\sigma}(\mathbf{r})|^2 \\
&\quad + B_4(K) |\boldsymbol{\sigma}(\mathbf{r})|^4 + C(K) \sum_{j=1}^n |\nabla \sigma_j(\mathbf{r})|^2.
\end{aligned} \tag{5.6.2}$$

Then

$$\phi_n(K, \mathbf{L}) := -\frac{1}{V} \ln Z_n(K, \mathbf{L}), \tag{5.6.3}$$

²⁴ Suppose, in the case $n = 1$, we select the form of σ which minimizes $\hat{\phi}_1^{(\text{cl})}$. For $C(K) > 0$ this will occur when σ is uniform and equal to m over $\mathcal{V}$ and, assuming that this distribution dominates the functional integral in (5.6.4) we recover (5.3.32).

where the partition function is given by the functional integral

$$Z_n(K, \mathbf{L}) := \int_{\Sigma} \exp \left[-\frac{1}{v_0} \int_{\mathcal{V}} \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \boldsymbol{\sigma}(\mathbf{r})) d\mathcal{V} \right] d\Sigma \quad (5.6.4)$$

and Σ is the space of values of $\boldsymbol{\sigma}$. Suppose that the magnetic field $\mathcal{H}$ is in the direction of the first component σ_1 of $\boldsymbol{\sigma}$. The first term in (5.6.2) is now replaced by $-L\sigma_1$. The first component of $\boldsymbol{\sigma}$ is referred to as *longitudinal*, the other components as *transverse*. Let $\hat{\sigma}$ be defined according to the high-temperature minimization conditions for $\hat{\phi}_n^{(\text{cl})}$, when the distribution is uniform. Thus

$$\begin{aligned} 0 &= 2B_2(K)\hat{\sigma}_1 + 4B_4(K)\hat{\sigma}_1^3 - L, \\ 0 &= \hat{\sigma}_j, \quad j \neq 1. \end{aligned} \quad (5.6.5)$$

5.6.2 The Gaussian Approximation

This involves retaining only harmonic fluctuations around the uniform values given by (5.6.5). With

$$\sigma'_j(\mathbf{r}) := \sigma_j(\mathbf{r}) - \hat{\sigma}_j, \quad j = 1, \dots, n, \quad (5.6.6)$$

we have

$$\begin{aligned} \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \boldsymbol{\sigma}(\mathbf{r})) &= \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \hat{\boldsymbol{\sigma}}) + 4B_4(K)\hat{\sigma}_1^2\sigma_1'^2(\mathbf{r}) + D(K)|\boldsymbol{\sigma}'(\mathbf{r})|^2 \\ &\quad + C(K) \sum_{j=1}^n |\nabla \sigma'_j(\mathbf{r})|^2, \end{aligned} \quad (5.6.7)$$

where

$$D(K) := B_2(K) + 2B_4(K)\hat{\sigma}_1^2. \quad (5.6.8)$$

With periodic boundary conditions on the hypercube $\mathcal{V}$, Fourier transforms are defined (Eqs. (18.1.14) and (18.1.15)) by²⁵

$$\psi_j(\mathbf{k}) := \frac{1}{R^d} \int_{\mathcal{V}} \sigma'_j(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathcal{V}, \quad (5.6.9)$$

²⁵ Not to be confused with the derivatives of the free-energy density appearing in other chapters.

with

$$\sigma'_j(\mathbf{r}) = \sum_{\{\mathbf{k}\}} \psi_j(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (5.6.10)$$

where the wave vectors

$$\mathbf{k} := \frac{2\pi}{R} \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix} \quad m_i = 0, \pm 1, \pm 2, \dots \quad (5.6.11)$$

It can be seen from (5.6.10) that $\psi_j(\mathbf{k})$ corresponds to the amplitude fluctuation of $\sigma_j(\mathbf{r})$ with wave vector $\mathbf{k}$. The larger the magnitude of $\mathbf{k}$ the shorter the wave length of the fluctuation. The coefficients $\psi_1(\mathbf{k})$ represent fluctuations in the direction of the external field, called *longitudinal fluctuations*; $\psi_j(\mathbf{k})$, for $j \neq 1$, correspond to *transverse fluctuations*. Since all the variables $\sigma'_j(\mathbf{k})$ are real, $\psi_j(-\mathbf{k})$ is the conjugate complex of $\psi_j(\mathbf{k})$, with $\psi_j(\mathbf{0})$ real. From (5.6.1), (5.6.6) and (5.6.9)

$$\hat{\mathbf{m}} = \hat{\sigma} + \boldsymbol{\psi}(\mathbf{0}), \quad (5.6.12)$$

where $\boldsymbol{\psi}(\mathbf{0})$ is the n -dimensional vector with components $\psi_j(\mathbf{0})$. This term in (5.6.12) represents the fluctuations in magnetization per unit volume. For $\mathbf{k} \neq \mathbf{0}$ let

$$\psi_j(\pm\mathbf{k}) := \psi_j^{(\Re)}(\mathbf{k}) \pm i\psi_j^{(\Im)}(\mathbf{k}). \quad (5.6.13)$$

From (5.6.7) and (5.6.10),

$$\begin{aligned} \frac{1}{R^d} \int_{\mathcal{V}} \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \boldsymbol{\sigma}(\mathbf{r})) d\mathcal{V} &= \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \hat{\sigma}) + 4B_4(K) \hat{\sigma}_1^2 \psi_1^2(\mathbf{0}) + D(K) |\boldsymbol{\psi}(\mathbf{0})|^2 \\ &+ 2 \sum_{\{\mathbf{k} > \mathbf{0}\}} [4B_4(K) \hat{\sigma}_1^2 + D(K) + C(K)k^2] [\psi_1^{(\Re)2}(\mathbf{k}) + \psi_1^{(\Im)2}(\mathbf{k})] \\ &+ 2 \sum_{j=2}^n \sum_{\{\mathbf{k} > \mathbf{0}\}} [D(K) + C(K)k^2] [\psi_j^{(\Re)2}(\mathbf{k}) + \psi_j^{(\Im)2}(\mathbf{k})], \end{aligned} \quad (5.6.14)$$

where ' $\mathbf{k} > \mathbf{0}$ ' means all the wave vectors in half the $\mathbf{k}$ -vector space excluding $\mathbf{k} = \mathbf{0}$. The partition function is now given by substituting from (5.6.14) into (5.6.4). The functional integral over the space Σ is replaced by integrals over the variables $\psi_j(\mathbf{0})$, $\psi_j^{(\Re)}(\mathbf{k})$ and $\psi_j^{(\Im)}(\mathbf{k})$, for $j = 1, \dots, n$, which we take to extend over the range $(-\infty, \infty)$. This gives

$$Z_n(K, \mathbf{L}) = \exp[-V \hat{\phi}_n^{(\text{cl})}(K, \mathbf{L}; \hat{\sigma})] \prod_{j=1}^n \left[\Psi_j(\mathbf{0}) \prod_{\{\mathbf{k} > \mathbf{0}\}} \Psi_j(\mathbf{k}) \right], \quad (5.6.15)$$

where

$$\Psi_j(\mathbf{0}) := \begin{cases} \int_{-\infty}^{\infty} d\psi_1(\mathbf{0}) \exp\{-V[4B_4(K)\hat{\sigma}_1^2 + D(K)]\psi_1^2(\mathbf{0})\}, & j = 1, \\ \int_{-\infty}^{\infty} d\psi_j(\mathbf{0}) \exp\{-V D(K)\psi_j^2(\mathbf{0})\}, & j \neq 1, \end{cases} \quad (5.6.16)$$

and, for $\mathbf{k} \neq \mathbf{0}$,

$$\Psi_j(\mathbf{k}) := \begin{cases} \int_{-\infty}^{\infty} d\psi_1^{(\Re)}(\mathbf{k}) \int_{-\infty}^{\infty} d\psi_1^{(\Im)}(\mathbf{k}) \exp\{-V[4B_4(K)\sigma_1^{*2} + D(K) \\ \quad + C(K)k^2][\psi_1^{(\Re)2}(\mathbf{k}) + \psi_1^{(\Im)2}(\mathbf{k})]\}, & j = 1, \\ \int_{-\infty}^{\infty} d\psi_j^{(\Re)}(\mathbf{k}) \int_{-\infty}^{\infty} d\psi_j^{(\Im)}(\mathbf{k}) \exp\{-V[D(K) \\ \quad + C(K)k^2][\psi_j^{(\Re)2}(\mathbf{k}) + \psi_j^{(\Im)2}(\mathbf{k})]\}, & j \neq 1. \end{cases} \quad (5.6.17)$$

The variables $\psi_j(\mathbf{0})$, $\psi_j^{(\Re)}(\mathbf{k})$ and $\psi_j^{(\Im)}(\mathbf{k})$, for all j and $\mathbf{k}$ are normally distributed and uncorrelated with

$$\langle \psi_j(\mathbf{k}) \rangle = 0, \quad (5.6.18)$$

$$\langle |\psi_j(\mathbf{k})|^2 \rangle = \begin{cases} g_{\parallel}(K, \mathbf{L}; k), & j = 1, \\ g_{\perp}(K, \mathbf{L}; k), & j \neq 1, \end{cases} \quad (5.6.19)$$

where

$$\begin{aligned} g_{\parallel}(K, \mathbf{L}; k) &:= \{2V[4B_4(K)\hat{\sigma}_1^2 + D(K) + C(K)k^2]\}^{-1}, \\ g_{\perp}(K, \mathbf{L}; k) &:= \{2V[D(K) + C(K)k^2]\}^{-1}. \end{aligned} \quad (5.6.20)$$

From Eqs. (5.6.5), (5.6.10), (5.6.12) and (5.6.18),

$$\langle \boldsymbol{\sigma}'(\mathbf{r}) \rangle = 0, \quad (5.6.21)$$

$$\langle \boldsymbol{\sigma}(\mathbf{r}) \rangle = \langle \hat{\boldsymbol{\sigma}} \rangle = \hat{\boldsymbol{\sigma}},$$

$$\mathbf{m} = \langle \hat{\mathbf{m}} \rangle = \hat{\boldsymbol{\sigma}} = \hat{\sigma}_1 \hat{\mathbf{h}}, \quad (5.6.22)$$

where $\hat{\mathbf{h}}$ is the unit vector in the direction of the external magnetic field. It follows that the thermodynamic magnetization per unit volume is the same as that per lattice site in Landau theory.²⁶

Formula (4.2.3) for the pair correlation function applies to the continuum, as well as to a lattice system, and from (5.6.10) using the notation introduced in (2.5.25) for a system with translational and rotational invariance,

$$\begin{aligned}\Gamma_2(\sigma; |\mathbf{r} - \mathbf{r}'|) &:= \langle \sigma'(\mathbf{r}) \cdot \sigma'(\mathbf{r}') \rangle \\ &= \sum_{j=1}^n \sum_{\{\mathbf{k}\}} \langle |\psi_j(\mathbf{k})|^2 \rangle \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')].\end{aligned}\quad (5.6.23)$$

With $r := |\mathbf{r} - \mathbf{r}'|$, the Fourier transform $\Gamma_2^*(\sigma; k)$ of $\Gamma_2(\sigma; r)$ is defined by the continuum form of (2.5.17) (Eq. (5.6.9)) and thus, from (5.6.20) and (5.6.23),

$$\Gamma_2^*(\sigma; k) = R^d \sum_{j=1}^n \langle |\psi_j(\mathbf{k})|^2 \rangle = R^d [g_{\parallel}(K, \mathbf{L}; k) + (n-1)g_{\perp}(K, \mathbf{L}; k)]. \quad (5.6.24)$$

From (4.5.50),²⁷ (5.6.20) and (5.6.24) the correlation length τ is given by

$$\tau^2 = 4c(d)dVC(K) \frac{g_{\parallel}^2(K, \mathbf{L}; 0) + (n-1)g_{\perp}^2(K, \mathbf{L}; 0)}{g_{\parallel}(K, \mathbf{L}; 0) + (n-1)g_{\perp}(K, \mathbf{L}; 0)}. \quad (5.6.25)$$

To evaluate ϕ , given by (5.6.3) and (5.6.15), it is necessary to insert explicit expressions for the Gaussian integrals (5.6.16) and (5.6.17). From (5.6.19) and (5.6.20) this gives

$$\begin{aligned}\phi(K, \mathbf{L}) &= \hat{\phi}_L(K, \mathbf{L}; \hat{\sigma}) \\ &= -\frac{1}{2V} \sum_{\{\mathbf{k}\}} \{ \ln[2\pi g_{\parallel}(K, \mathbf{L}; k)] + (n-1) \ln[2\pi g_{\perp}(K, \mathbf{L}; k)] \}.\end{aligned}\quad (5.6.26)$$

5.6.3 Gaussian Critical Exponents

Adopting the forms for $B_2(K)$ and $B_4(K)$ given by the leading terms in (5.3.22), it can be shown from (5.6.5) and (5.6.22) that the exponents β , γ and δ take their classical values (5.3.30). We now consider the situation along the zero-field axis, $\mathbf{L} = \mathbf{0}$, and suppose that, to leading order in $\Delta K := K - K_c$,

$$C(K) \simeq C_0 > 0. \quad (5.6.27)$$

²⁶ Cf. (5.6.5) and (5.3.17) with $\Delta L = L$, $\Delta \rho = m$, $B_1(K) = B_3(K) = 0$.

²⁷ In this case $\mathfrak{d} = d$, the thermodynamic limit is taken in each of the d dimensions.

From (5.3.22) and (5.6.20),

$$g_{\parallel}(K, \mathbf{0}; k) = g_{\perp}(K, \mathbf{0}; k) = [2V(-B_{21}\Delta K + C_0k^2)]^{-1}, \quad (5.6.28)$$

for $K < K_c$, when $\sigma_1 = 0$, and

$$\begin{aligned} g_{\parallel}(K, \mathbf{0}; k) &= [2V(2B_{21}\Delta K + C_0k^2)]^{-1}, \\ g_{\perp}(K, \mathbf{0}; k) &= (2VC_0k^2)^{-1}, \end{aligned} \quad (5.6.29)$$

for $K > K_c$, when, from (5.3.33) and (5.6.22), $\sigma_1^2 = B_{21}\Delta K/(2B_{40})$. Thus, from (5.6.25), it can be seen that throughout the magnetized, low-temperature, phase the correlation length is infinite when $n \neq 1$. This behaviour arises from the transverse singular amplitude of the long-wavelength contribution to the pair correlation function. The magnetized phase is not well-defined in the Gaussian approximation. We shall, for the rest of this section, concentrate on the high-temperature phase. From (5.6.25) and (5.6.28)

$$\tau = \left[\frac{2c(d)dC_0}{B_{21}(K_c - K)} \right]^{\frac{1}{2}} = \left[\frac{C_0}{B_{21}(K_c - K)} \right]^{\frac{1}{2}}, \quad (5.6.30)$$

with $c(d) = 1/(2d)$, (a choice which is justified below). From (4.5.52) this gives the value for the exponent $\nu = \frac{1}{2}$. From (5.6.24), (5.6.28) and (5.6.30),

$$\bar{\Gamma}_2^*(\sigma; k) = \frac{nv_0}{2C_0(k^2 + \tau^{-2})} \quad (5.6.31)$$

and thus

$$\bar{\Gamma}_2^*(\sigma; k) = \frac{nv_0}{2C_0k^2} \quad (5.6.32)$$

at the critical point. Comparing this formula with the asymptotic form (4.5.57) yields the result $\eta = 0$. As in Sect. 3.6.1 we need to introduce the high-frequency wavelength cut-off k_0 given in terms of the minimum resolution length a_0 by (3.6.16). Then, from (18.1.38),

$$\begin{aligned} \Gamma_2(\sigma; r) &= \lim_{k_0 \rightarrow \infty} r \int_0^{k_0} \frac{nv_0}{2C_0(k^2 + \tau^{-2})} \left(\frac{k}{2\pi r} \right)^{\frac{d}{2}} J_{\frac{d}{2}-1}(kr) dk \\ &= \frac{nv_0 r^{2-d}}{2C_0(2\pi)^{\frac{d}{2}}} \lim_{k_0 \rightarrow \infty} \mathfrak{G}\left(k_0; \frac{d}{2}, \frac{d}{2} - 1; r/\tau\right), \end{aligned} \quad (5.6.33)$$

where

$$\mathfrak{G}(k_0; x, y, z) := \int_0^{k_0} \frac{k^x J_y(k)}{k^2 + z^2} dk. \quad (5.6.34)$$

From (18.1.41),

$$\Gamma_2(\sigma; r) = \frac{nv_0}{2C_0(2\pi)^{\frac{d}{2}}}(r\tau)^{1-\frac{d}{2}}K_{\frac{d}{2}-1}(r/\tau), \quad \text{for } d < 5. \quad (5.6.35)$$

Substituting into the continuum form of (4.2.4) and using (18.1.35) for the surface area of a d -dimensional hypersphere, gives

$$\tau^2 = c(d) \frac{\int_0^\infty r^{\frac{d}{2}+2} K_{\frac{d}{2}-1}(r/\tau) dr}{\int_0^\infty r^{\frac{d}{2}} K_{\frac{d}{2}-1}(r/\tau) dr}. \quad (5.6.36)$$

The integrals in (5.6.36) are of a standard form, yielding the result $c(d) = 1/(2d)$. In the limit $r \rightarrow \infty$, from (18.1.42),

$$\Gamma_2(\sigma; r) \simeq \frac{nv_0}{2^{\frac{d}{2}+\frac{3}{2}}C_0(\pi r\tau)^{\frac{d}{2}-\frac{3}{2}}} \frac{\exp(-r/\tau)}{r}. \quad (5.6.37)$$

Since (18.1.42) is exact for $x = \pm \frac{1}{2}$, (5.6.37) is exact for $d = 1$ and $d = 3$, when it includes the limiting behaviour as $\tau \rightarrow \infty$. From (5.6.35) and (18.1.43),

$$\begin{aligned} \Gamma_2(\sigma; r) &\simeq -\frac{nv_0}{2\pi C_0} \ln\left(\frac{r}{2\tau}\right), & \text{if } d = 2, \\ \Gamma_2(\sigma; r) &\simeq \frac{nv_0}{8\pi^2 C_0} \frac{1}{r^2}, & \text{if } d = 4, \end{aligned} \quad (5.6.38)$$

in the limit $\tau \rightarrow \infty$. From (5.6.33) and (18.1.48)

$$\begin{aligned} \Gamma_2(\sigma; r) &\simeq \frac{nv_0}{2C_0 r^{d-2} (2\pi)^{\frac{d}{2}}} \left\{ \Omega\left(\frac{d}{2}, \frac{d}{2} - 1\right) \right. \\ &\quad \left. + \sqrt{\frac{2}{\pi}} k_0^{\frac{d}{2}-\frac{5}{2}} \cos\left(k_0 - \frac{\pi}{4}[d+1]\right) \right\}, \quad \text{for } d > 4. \end{aligned} \quad (5.6.39)$$

So for $d \geq 5$, $\Gamma_2(\sigma; r)$ diverges in the $k_0 \rightarrow \infty$, high-frequency limit. From the point of view of critical phenomena it is the long-wavelength, low-frequency, fluctuations which are important. This short-wavelength, high-frequency, divergence is dealt with by supposing that the system has a minimum resolution length a_0 , so that fluctuations of frequency greater than $k_0 := 2\pi/a_0$ are dampened by the system. From (5.6.37)–(5.6.39),

$$\Gamma_2(\sigma; r) \sim \frac{1}{r^{d-2}}, \quad (5.6.40)$$

for $d \geq 3$. Comparison with (4.5.55) again shows that $\eta = 0$, but now we have additional information about the cases $d = 1$ and $d = 2$. In the limit $\tau \rightarrow \infty$, from (5.6.37) and (5.6.38) respectively,

$$\Gamma_2(\sigma; r) \sim \begin{cases} \tau \exp(-r/\tau), & d = 1, \\ \ln(r/\tau), & d = 2. \end{cases} \quad (5.6.41)$$

The classical values for the exponents β , γ and δ satisfy the Widom scaling law (4.5.45) and these together with the values, derived above for η and ν satisfy the Fisher scaling law (4.5.58).

It is interesting to see from (5.6.35) that, according to this analysis for $d < 5$, the correlation function can be expressed in the form

$$\Gamma_2(\sigma; r) = \frac{f(\tau/r)}{r^{d-2-\eta}}, \quad (5.6.42)$$

given by (4.5.55), where in this case $\eta = 0$ and

$$f(x) := \frac{nv_0 x^{\frac{d}{2}-1}}{2C_0(2\pi)^{\frac{d}{2}}} K_{\frac{d}{2}-1}(x). \quad (5.6.43)$$

It is now necessary to determine the exponent associated with $c_{\mathcal{H}}$, which because of the magnetic symmetry is α . Again we consider the thermodynamic limit. The sum in (5.6.26) becomes an integral with the wave vectors forming an effective continuum with $\mathcal{S}_d(k)V(2\pi)^{-d}dk$ states in the hypershell $k < |\mathbf{k}| < k + dk$, where $\mathcal{S}_d(k)$ is the surface area of a d -dimensional hypersphere of radius k , given by (18.1.35). Introducing the high-frequency cut off k_0 , from (5.6.26) and (5.6.28),

$$\begin{aligned} \phi_n(K, \mathbf{0}) &= \hat{\phi}_n^{(\text{cl})}(K, \mathbf{0}; \mathbf{0}) \\ &+ \frac{nv_0}{\Gamma(d/2)2^d\pi^{d/2}} \int_0^{k_0} k^{d-1} \ln \left\{ \frac{V[A_2(K_c - K) + C_0k^2]}{\pi} \right\} dk, \end{aligned} \quad (5.6.44)$$

for $K < K_c$. The heat capacity $c_{\mathcal{H}}$ at constant field along the zero-field axis is given, from (2.3.30), by

$$c_{\mathcal{H}} = -k_B K^2 \frac{\partial^2 \phi_n}{\partial K^2} \quad (5.6.45)$$

The first term on the right-hand side of (5.6.44) is the contribution arising from Landau theory and gives the discontinuity $\Delta c_{\mathcal{H}}$ of (5.3.37). In differentiating the integral in (5.6.44) only the most potentially singular contribution is retained to give

$$c_{\mathcal{H}} \simeq \frac{k_B n K_c^2 v_0 B_{21}^2}{\Gamma(d/2)2^d\pi^{d/2}} \int_0^{k_0} \frac{k^{d-1} dk}{[B_{21}(K_c - K) + C_0k^2]^2}. \quad (5.6.46)$$

Transforming to the variable

$$\kappa(k) := k\tau = k \left[\frac{C_0}{B_{21}(K_c - K)} \right]^{\frac{1}{2}} \quad (5.6.47)$$

with $\kappa_0 = \kappa(k_0)$ gives

$$c_{\mathcal{H}} \simeq \frac{k_B n K_c^2 v_0}{\Gamma(d/2) 2^d \pi^{d/2}} \left(\frac{B_{21}}{C_0} \right)^2 \left[\frac{C_0}{B_{21}(K_c - K)} \right]^{\frac{4-d}{2}} \int_0^{\kappa_0} \frac{\kappa^{d-1} d\kappa}{(1 + \kappa^2)^2}. \quad (5.6.48)$$

A singularity in $c_{\mathcal{H}}$, at the critical point can arise, either from the integral, or, when $d < 4$, from the prefactor. The integral is bounded above by its limiting value, given by taking $\kappa_0 \rightarrow \infty$. This is

$$\int_0^\infty \frac{\kappa^{d-1} d\kappa}{(1 + \kappa^2)^2} = \Gamma\left(\frac{1}{2}d\right) \Gamma\left(2 - \frac{1}{2}d\right). \quad (5.6.49)$$

The gamma function is singular at 0, -1 , -2 , $\dots$ and the integral works as a finite upper bound only when $d < 4$. In this case the prefactor diverges giving $\alpha = (4 - d)/2$. When $d = 4$ the prefactor is absent and any singular behaviour must come from the integral. Since the integrand is zero at the lower limit, for K sufficiently close to K_c ,

$$\int_0^{\kappa_0} \frac{\kappa^3 d\kappa}{(1 + \kappa^2)^2} \sim \int_\epsilon^{\kappa_0} \frac{d\kappa}{\kappa} \sim -\frac{1}{2} \ln(K_c - K). \quad (5.6.50)$$

The heat capacity has a logarithmic divergence at the critical point. For $d > 4$ the prefactor does not diverge and, as in the case of $d = 4$, any singularity must arise from the integral in a neighbourhood of its upper limit. But

$$\int_0^{\kappa_0} \frac{\kappa^{d-1} d\kappa}{(1 + \kappa^2)^2} \sim \int_\epsilon^{\kappa_0} \kappa^{d-5} d\kappa = \frac{\kappa_0^{d-4} - \epsilon^{d-4}}{d-4}. \quad (5.6.51)$$

The heat capacity does not diverge, but there is a discontinuity from the Landau term. The exponent $\alpha = 0$.

The results of the combination of Landau theory and the Gaussian approximation applied to Ginzburg–Landau theory can now be summarized as follows:

- When $d \geq 4$, the values for the critical exponents are

$$\begin{aligned} \alpha &= 0, & \beta &= \frac{1}{2}, & \delta &= 3, \\ \gamma &= 1, & \eta &= 0, & \nu &= \frac{1}{2}. \end{aligned} \quad (5.6.52)$$

These values satisfy the three scaling laws (4.5.45) (Widom), (4.5.44) (Essam–Fisher) and (4.5.58) (Fisher), but not the Josephson hyper-scaling law (4.5.54), except at $d = 4$.

- When $d < 4$, η and ν are still given by (5.6.52) and to satisfy the Fisher scaling law (4.5.58) we need also to retain the value of γ given in (5.6.52). Since we have shown that now

$$\alpha = \frac{4 - d}{2}, \quad (5.6.53)$$

the Josephson hyper-scaling law (4.5.54) is satisfied. However, the Widom and Essam–Fisher scaling laws are not satisfied by α from (5.6.53) and β and δ from (5.6.52). To satisfy these two laws we need the new expressions

$$\beta = \frac{d - 2}{4}, \quad \delta = \frac{d + 2}{d - 2}. \quad (5.6.54)$$

The classical values for β and δ derived from Landau theory in Sect. 5.3 are intended to be true for all values of d . The analysis, carries over to the Ginzburg–Landau theory and this seems, therefore, to contradict the validity of the values (5.6.54) (or alternatively of scaling theory) unless a reconciliation can be achieved. The key to this can be found in the treatment of Ginzburg–Landau using the renormalization group (see, for example Ma 1976a). In this the parameters B_{21} and B_{40} of Eq. (5.3.34) are also treated as renormalizable (rescalable) parameters. If we modify the scaling approach use in (5.3.34) to include exponents u_2 and u_4 for B_{21} and B_{40} respectively then it follows that

$$z = d - y_1, \quad u_2 = 2y_1 - y_2 - d, \quad u_4 = 4y_1 - 3d. \quad (5.6.55)$$

It is not possible to obtain unique exponent values from these expressions. They are, however, satisfied by

$$y_2 = 2, \quad y_1 = \frac{1}{2}d + 1, \quad u_2 = 0, \quad u_4 = 4 - d \quad (5.6.56)$$

and these are the values associated with the *Gaussian fixed point* of renormalization theory (Ma 1976a). From (4.5.40) and (4.5.39) the values of $\gamma = 1$ and α , β and δ given by (5.6.53) and (5.6.54) are then obtained. These values are called the *Gaussian critical exponents*. In this new context we can view the derivation of the exponent values (5.3.36) as based on the tacit assumption that both B_{21} and B_{40} are marginal parameters ($u_2 = u_4 = 0$). The modified derivation leading to (5.6.56) still leaves B_{21} marginal. Indeed, if this were not the case, the derivation of $\nu = \frac{1}{2}$ from (5.6.30) would not be valid. However, for $d < 4$, B_{40} is relevant, for $d = 4$ it is marginal and for $d > 4$ it is irrelevant. A discussion of the case $d < 4$, for which the Gaussian fixed point is unstable requires the full renormalization group analysis. The case of

interest here is $d > 4$. Taking the singular part of the free-energy density (5.3.34) on the zero-field axis ($\theta_1 = 0$) and substituting from (5.3.33) and (5.3.22) we have

$$\phi_{\text{sing}}(0, \theta_2, B_{40}) = \frac{3B_{21}^2 \theta_2^2}{4B_{40}}. \quad (5.6.57)$$

Using the homogeneity condition

$$\phi_{\text{sing}}(0, \lambda^{y_2} \theta_2, \lambda^{u_4} B_{40}) = \lambda^d \phi_{\text{sing}}(0, \theta_2, B_{40}) \quad (5.6.58)$$

and substituting $\lambda = |\theta_2|^{-1/y_2}$ we have

$$\phi_{\text{sing}}(0, \theta_2, B_{40}) = |\theta_2|^{d/y_2} w_0^{(\pm)}(v), \quad (5.6.59)$$

where $v := B_{40}/|\theta_2|^{u_4/y_2}$ and

$$w_0^{(\pm)}(v) := \frac{3B_{21}^2}{4v} \quad (5.6.60)$$

is a particular realization of the function appearing in Eq. (4.8.4). It follows that, in terms of the discussion in Sect. 4.8, B_{40} is a dangerous irrelevant variable, with $\tau = 1$ and, from (4.8.9), an effective change of dimension in the system to

$$\bar{d} := d - |d - 4| = 4. \quad (5.6.61)$$

Using $\bar{d}$ in place of d in Eqs. (5.6.53) and (5.6.54) gives the mean-field exponents (5.6.52). We have, therefore, obtained a consistent scaling analysis using the Gaussian exponents, with the proviso that, for $d > 4$, account is taken of the dangerous irrelevant variable B_{40} , which leads to the effective dimension of the system being fixed at the value four.²⁸ Thus, for a simple critical point, the upper critical dimension (Sect. 3.1) $d_{\text{uc}} = 4$.

Although the existence of a singularity in the heat capacity at the critical point, superimposed upon a discontinuity, is dependent only on the dimension d , the ability to detect the singularity experimentally will depend on the range of temperature over which the singularity dominates the discontinuity. This in turn will depend on the relative size of the parameters B_{21} , B_{40} and C_0 and on the dimension n of the order parameter. The *width of the critical region* $\mathcal{W}_c$ can be defined by the condition that

$$\text{msp}\{c_{\mathcal{H}}\} \geq \Delta c_{\mathcal{H}} \quad \text{if and only if} \quad \left| 1 - \frac{T}{T_c} \right| \leq \mathcal{W}_c, \quad (5.6.62)$$

²⁸ The discussion presented here is not quite complete. As pointed out in Sect. 4.8 it is necessary to consider other approaches to the critical point to determine whether the change of effective dimension can be consistently applied. This involves consideration of the case $\theta_1 \neq 0$, when the magnetization is given, as in Sect. 6.1, by a root of a cubic equation.

where ‘msp’ stands for ‘most singular part’. For $T > T_c$ the most singular part of $c_{\mathcal{H}}$ is given by substituting from (5.6.49) into (5.6.48) and $\Delta c_{\mathcal{H}}$ is given by (5.3.37). This yields the result

$$\mathcal{W}_c = \frac{1}{K_c B_{21}} \left[\frac{nv_0 \Gamma(2 - \frac{1}{2}d) B_{40}}{2^{d-1} (C_0 \pi)^{\frac{d}{2}}} \right]^{2/(4-d)}. \quad (5.6.63)$$

Equation (5.6.63) is a form of the *Ginzburg criterion*, (Ginzburg 1960). The singularity in $c_{\mathcal{H}}$ is caused by fluctuations and the critical region is the range of temperature around the critical point within which fluctuations are important. The motivation for Ginzburg’s work was to understand why a finite discontinuity was seen at some critical points, like those in superconductors, whereas in others, like the λ -point in liquid helium, a singularity was observed in addition to the discontinuity. On the basis of (5.6.63) it can now be seen that, when $\mathcal{W}_c$ is too small, the critical region cannot be resolved experimentally and the singularity will not be seen. Estimates of $\mathcal{W}_c$ for various systems are given by Ginzburg (1960). He concludes that, for superconductors $\mathcal{W}_c \sim 3 \times 10^{-16}$, whereas in helium $\mathcal{W}_c \sim 0.3$.

An analysis of tricritical behaviour using the Gaussian approximation to Ginzburg–Landau theory can be made on the lines of Sects. 5.6.1 and 5.6.3. This yields $\nu = \frac{1}{2}$ and $\eta = 0$, the same values as for a critical point. The Fisher scaling law (4.5.58) is therefore satisfied. The Josephson hyper-scaling law (4.5.54) is satisfied only when $d = d_{uc} = 3$, which is the upper critical dimension for a tricritical point. It may, however, be shown (Sarbach and Fisher 1978) that the Gaussian critical exponents (5.6.53) and (5.6.54) apply to the tricritical point with the crossover from the classical values governed by a parameter with exponent $3 - d$. This parameter is thus relevant for $d < 3$, leading to the crossover, and marginal for $d = 3$. For $d > 3$ it is dangerously irrelevant, leading to an effective change of dimension in the system to $\bar{d} = 3$.

Chapter 6

Mean-Field Theory

6.1 The Ising Ferromagnet

The zero-field Ising Hamiltonian $\hat{H}_0(N, K, \mathcal{M}; \sigma)$ is just $\hat{H}(K, 0; \sigma)$, given by (3.5.1), with the sum of the microstates constrained by the condition

$$\mathcal{M} = \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}). \quad (6.1.1)$$

This can be re-expressed in the form

$$\hat{H}_0(N, K, \mathcal{M}; \sigma) = -\frac{1}{2}K \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}) \sum_{\{\mathbf{r}'\}}^{(\mathbf{r})} \sigma(\mathbf{r}'), \quad (6.1.2)$$

where the inner summation over $\mathbf{r}'$ is restricted to the z first-neighbours of $\mathbf{r}$. In the canonical distribution ($\eta = 0$) considered here the magnetization density $m := \mathcal{M}/N$ is an independent variable with $\langle \sigma(\mathbf{r}) \rangle = m$ for every lattice site. If in (6.1.2) $\sigma(\mathbf{r}')$ is replaced by its expectation value m it follows from (6.1.1) that (6.1.2) becomes

$$\hat{H}_0(N, K, \mathcal{M}) = -\frac{1}{2N} z K \mathcal{M}^2. \quad (6.1.3)$$

Equivalently this result can be achieved in two steps. First, the first-neighbour interaction terms $\sigma(\mathbf{r})\sigma(\mathbf{r}')$ in (6.1.2) are replaced by weighted expectation values $\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle$ over all pairs of sites. Except at the critical point, spin states on widely separated sites are more-or-less independent. So taking

$$\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle = \langle \sigma(\mathbf{r}) \rangle \langle \sigma(\mathbf{r}') \rangle = m^2 \quad (6.1.4)$$

is a reasonable approximation. Using this assumption we obtain (6.1.3). However, with $\mathbf{r}$ and $\mathbf{r}'$ close together (in this case first-neighbour sites), the situation is different. For the Ising ferromagnet the spins tend to align, with

$$\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle > m^2. \quad (6.1.5)$$

This represents *short-range order*. As we saw in Sect. 2.5.2,

$$\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle - m^2 = \langle [\langle \sigma(\mathbf{r}) \rangle - \langle \sigma(\mathbf{r}') \rangle][\langle \sigma(\mathbf{r}') \rangle - \langle \sigma(\mathbf{r}') \rangle] \rangle \quad (6.1.6)$$

is the (fluctuation) pair correlation function. So the assumption (6.1.4), which is applied to achieve the mean-field approximation, is equivalent to assuming that *all* spins are uncorrelated; not just those which are widely separated.

The number of terms in the the partition function $Z_0(N, K, \mathcal{M})$ is

$$W(N, \mathcal{M}) := \frac{N!}{\left\{\frac{1}{2}(N + \mathcal{M})\right\}! \left\{\frac{1}{2}(N - \mathcal{M})\right\}!}, \quad (6.1.7)$$

the number of spin arrangement consistent with (6.1.1). From (5.1.1)–(5.1.6), by applying Stirling's formula¹ to $W(N, \mathcal{M})$,

$$\begin{aligned} \Phi_1^{(\text{cl})}(N, K, L; Nm) &= \frac{1}{4}[2N(1 + m) + 1] \ln(1 + m) \\ &\quad + \frac{1}{4}[2N(1 - m) + 1] \ln(1 - m) \\ &\quad - (N + 1) \ln(2) - \frac{1}{2} \ln(2\pi N) \\ &\quad - \frac{1}{2} N z K m^2 - N L m + O(1/N) \end{aligned} \quad (6.1.8)$$

and taking the thermodynamic limit, $N \rightarrow \infty$, gives

$$\phi_1^{(\text{cl})}(K, L; m) = \mathcal{G}(zK; m) - \ln(2) - Lm, \quad (6.1.9)$$

where

$$\mathcal{G}(X; m) := \frac{1}{2}[(1 + m) \ln(1 + m) + (1 - m) \ln(1 - m)] - \frac{1}{2} X m^2. \quad (6.1.10)$$

This function and its derivatives, for which we introduce the notation

$$\mathcal{G}^{(i)}(X; m) := \frac{\partial^i \mathcal{G}}{\partial m^i}, \quad (6.1.11)$$

¹ For large n

$$\ln(n!) = \left(n + \frac{1}{2}\right) \ln(n) - n + \frac{1}{2} \ln(2\pi) + O(1/n).$$

giving

$$\mathcal{G}^{(1)}(X; m) = \frac{1}{2} \ln \left[\frac{1+m}{1-m} \right] - Xm, \quad (6.1.12)$$

$$\mathcal{G}^{(2)}(X; m) = \frac{1}{1-m^2} - X, \quad (6.1.13)$$

appears in most mean-field calculations. From (5.1.7) and (5.1.8) the equilibrium and stability conditions are

$$\frac{\partial \phi_1^{(\text{cl})}}{\partial m} = \mathcal{G}^{(1)}(zK; m) - L = 0, \quad (6.1.14)$$

$$\frac{\partial^2 \phi_1^{(\text{cl})}}{\partial m^2} = \mathcal{G}^{(2)}(zK; m) > 0. \quad (6.1.15)$$

Substituting from (6.1.14) back into (6.1.9) shows that, at equilibrium,

$$\phi_1^{(\text{cl})}(K, L; m) = \frac{1}{2} \ln(1-m^2) + \frac{1}{2} Kzm^2 - \ln(2). \quad (6.1.16)$$

Curves derived from (6.1.14) are shown in Fig. 6.1. For $K < z^{-1}$ Eq. (6.1.14) has a single real root $m = m(K, L)$ (the broken line) where, $m(K, 0) = 0$. The coupling $K_c := z^{-1}$ is equivalent to $T_c := Jz$, which is the mean-field expression for the critical or *Curie temperature* in the ferromagnetic Ising model. For $K > K_c$ the equation exhibits the characteristic looping effect found in all classical approximations (Fig. 5.1a). Because of the magnetic symmetry it follows from Maxwell's equal areas rule, described in Sect. 5.1.1, that the first-order transition is along the line $L = 0$, where the minima in $\phi_1^{(\text{cl})}(K, 0; m) = \phi_0^{(\text{cl})}(K, m)$ have equal depth

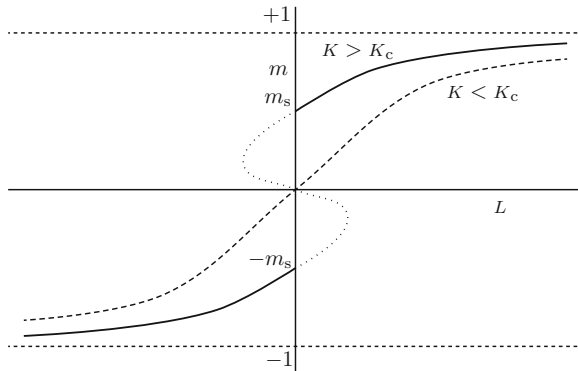


Fig. 6.1 Isotherms in the L - m plane derived from (6.1.14)

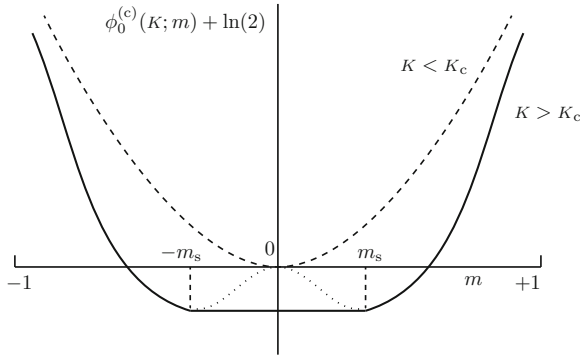


Fig. 6.2 Free energy–magnetization isotherms at zero field

(Fig. 6.2). The unbroken part of the isotherm $K > K_c$ in Fig. 6.1 corresponds to the solution to (6.1.14) which is the lower minimum of the free energy. It follows from (6.1.14) and (6.1.15) that the portion of the broken part of the curve $K > K_c$ with $\partial L / \partial m < 0$ represents unstable solutions and that with $\partial L / \partial m > 0$ metastable solutions (Sect. 5.1.2).

Along the critical isotherm $K = K_c$, from (6.1.14),

$$L \simeq \frac{2}{3}m^3 + O(m^5). \quad (6.1.17)$$

The implicit form for the magnetization, derived from (6.1.14), with

$$\lim_{L \rightarrow \pm 0} m(K, L) := \pm m_s(K), \quad (6.1.18)$$

has a discontinuity of $2m_s$ along the zero-field axis for $K > K_c$ with

$$m_s^2(K) = 3 \left(\frac{K - K_c}{K_c} \right) - \frac{3}{5} \left(\frac{K - K_c}{K_c} \right)^2 + O \left[\left(\frac{K - K_c}{K_c} \right)^3 \right], \quad \text{as } K \rightarrow K_c + \quad (6.1.19)$$

and

$$m_s(K) = 1 - 2 \exp \left(-\frac{2K}{K_c} \right) + O \left[\frac{2K}{K_c} \exp \left(-\frac{4K}{K_c} \right) \right], \quad \text{as } K \rightarrow \infty. \quad (6.1.20)$$

The curve of m_s as a function of K is shown in Fig. 3.1, where it is compared with the corresponding results derived from higher-order cluster expansions. From (2.3.31) and (6.1.14),

$$\varphi_\tau = \frac{\partial m}{\partial L} = \frac{1}{\mathcal{G}^{(2)}(zK; m)} = \frac{K_c}{K_c(1 - m^2)^{-1} - K}, \quad (6.1.21)$$

giving, from (6.1.19),

$$\varphi_T = \begin{cases} \frac{1}{\mathcal{G}^{(2)}(zK; 0)} = \frac{K_c}{K_c - K}, & \text{on the zero-field axis, for } K < K_c, \\ \frac{1}{\mathcal{G}^{(2)}(zK; m_s)} = \frac{K_c}{2(K - K_c)} \left\{ 1 - \frac{21}{5} \left(\frac{K - K_c}{K_c} \right) + O \left[\left(\frac{K - K_c}{K_c} \right)^2 \right] \right\}, & \text{on the coexistence curve, for } K > K_c. \end{cases} \quad (6.1.22)$$

From (2.3.30), (6.1.9), (6.1.14) and (6.1.21)

$$c_{\mathcal{H}} = \frac{k_B(Km + LK_c)^2}{K_c \mathcal{G}^{(2)}(zK; m)}, \quad (6.1.23)$$

giving, from (6.1.19) and (6.1.22),

$$c_{\mathcal{H}} = \begin{cases} 0, & \text{on the zero-field axis, for } K < K_c, \\ \frac{3}{2} k_B K_c \left\{ 1 - \frac{21}{5} \left(\frac{K - K_c}{K_c} \right) + O \left[\left(\frac{K - K_c}{K_c} \right)^2 \right] \right\}, & \text{on the coexistence curve, for } K > K_c. \end{cases} \quad (6.1.24)$$

Comparing (6.1.17), (6.1.19), (6.1.22) and (6.1.24) with (4.4.12)–(4.4.16) (and remembering that in this case because of the magnetic symmetry $\sigma = \alpha$) we recover, as we might expect, the classical values of the critical exponents given by (5.3.30).

We also note that (6.1.14) can be re-expressed in the form

$$m = \tanh(L + Kzm). \quad (6.1.25)$$

With $J = 0$ (giving $K = 0$) this formula gives the magnetization of a paramagnet in a field $\mathcal{H} = LT$. It can be obtained exactly by summing the partition function with the Hamiltonian (3.5.1) with $J = 0$. For the ferromagnet considered here (6.1.25) is the *Weiss field equation*. It can be understood as the consistency equation resulting from an approximation in which the interaction between a spin and its neighbours is replaced by a molecular field Jzm .

6.1.1 Mean-Field Fluctuations

As was indicated in the discussion in Sect. 4.2.2, the onset of a phase transition is related to the growth in the size of regions of local order. These long-range correlations lead to divergences in some or all of the response functions. In the

case of a ferromagnet the susceptibility divergence is thus related to the divergence in the size of the regions of parallel spins. Although the overall numbers of up and down spins remain equal there are large regions of local order or ‘patches’ of parallel spins. As $K \rightarrow K_c$ the patches become infinitely large which means that the correlation length, defined in Sect. 4.2.2, is approaching infinity. In mean-field theory, only overall fluctuations, which change the magnetization density m from its equilibrium value, are recognized, although (6.1.22) shows that these are still sufficient to produce a susceptibility divergence as $K \rightarrow K_c \pm 0$.²

From (5.1.5) the mean-field probability distribution for $\hat{m}$, along the zero-field axis, is given by

$$p(K; \hat{m}) = \frac{\exp[-N\phi_0^{(\text{cl})}(K, \hat{m})]}{\sum_{\{\hat{m}\}} \exp[-N\phi_0^{(\text{cl})}(K, \hat{m})]}. \quad (6.1.26)$$

Expanding (6.1.9) about $\hat{m} = 0$ gives

$$\begin{aligned} \phi_0^{(\text{cl})}(K, \hat{m}) &= \phi_1^{(\text{cl})}(K, 0; \hat{m}) \\ &= -\ln(2) + \left(\frac{K_c - K}{2K_c}\right) \hat{m}^2 + \frac{1}{12} \hat{m}^4 + O(\hat{m}^6). \end{aligned} \quad (6.1.27)$$

If terms of degree greater than four are neglected, the turning point at $\hat{m} = 0$ and the leading term of (6.1.19) are still correctly reproduced by (6.1.27) when $(K - K_c)$ is small and positive. Using (6.1.26) and (6.1.27) the mean-field probability distribution is shown in Fig. 6.3. For $K < K_c$ the distribution shows one peak at $\hat{m} = 0$ and, for $K > K_c$, two peaks at

$$\hat{m} = m^{(\pm)} = \pm \sqrt{\frac{3(K - K_c)}{K_c}}, \quad (6.1.28)$$

which corresponds to the leading terms in $\pm m_s(K)$.

For $K < K_c$ the distribution is concentrated around $\hat{m} = 0$ and the quartic term in (6.1.27) can be neglected. Since the distribution is symmetric $\langle \hat{m} \rangle = 0$. To determine the variance we make the substitution

$$x = \hat{m} \sqrt{\frac{N(K_c - K)}{2K_c}} \quad (6.1.29)$$

and, for large N , the sum in the denominator of (6.1.26) can be replaced by an integral over the range $(-\infty, \infty)$. Thus, from (6.1.22),

² In an accurate treatment, which must include the effect of local fluctuations, the susceptibility divergence is sharper.

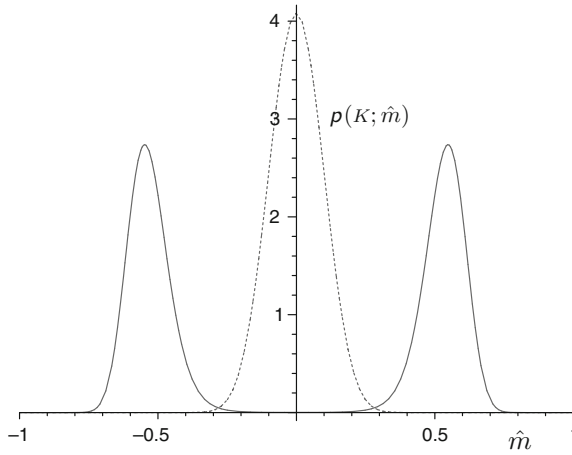


Fig. 6.3 Plots of $p(K; \hat{m})$ with $N = 1000$, $K = 1.1K_c$ (continuous curve) and $K = 0.9K_c$ (broken curve)

$$\text{Var}[\hat{m}] = \frac{2K_c}{N(K_c - K)} \frac{\int_{-\infty}^{\infty} x^2 \exp(-x^2) dx}{\int_{-\infty}^{\infty} \exp(-x^2) dx} = \frac{K_c}{N(K_c - K)} = \frac{1}{N} \varphi_T. \quad (6.1.30)$$

For $K > K_c$ the quartic term must be retained and, from (6.1.27) and (6.1.28),

$$\phi_0^{(\text{cl})}(K, \hat{m}) \simeq -\ln(2) + \frac{(K - K_c)}{K_c} [\hat{m} - m^{(\pm)}]^2, \quad \hat{m} \gtrless 0. \quad (6.1.31)$$

The method of evaluation of the mean and variance is now similar to that for $K < K_c$ except that the range of the integral is divided into two parts

$$\begin{aligned} (-\infty, 0) \quad & \text{for which the variable is } x_1 = [\hat{m} - m^{(-)}] \sqrt{\frac{N(K - K_c)}{K_c}}, \\ (0, \infty) \quad & \text{for which the variable is } x_2 = [\hat{m} - m^{(+)}] \sqrt{\frac{N(K - K_c)}{K_c}}. \end{aligned} \quad (6.1.32)$$

Each integral is then extended over the whole range $(-\infty, \infty)$ and it is straightforward to show that

$$\langle \hat{m} \rangle = m^{(+)} + m^{(-)} = 0. \quad (6.1.33)$$

Just as in the case $K < K_c$ the mean magnetization is zero, but here the reason is because the two peaks of the probability distribution make equal and opposite contributions. As we have seen, the zero-field line represents a coexistence region between phases with equal and opposite magnetizations. In a similar way

$$\begin{aligned}
\text{Var}[\hat{m}] &\simeq \frac{K_c}{N(K - K_c)} \frac{\int_{-\infty}^{\infty} y_1^2 \exp(-x_1^2) dx_1 + \int_{-\infty}^{\infty} x_2^2 \exp(-x_2^2) dx_2}{\int_{-\infty}^{\infty} \exp(-x_1^2) dx_1 + \int_{-\infty}^{\infty} \exp(-x_2^2) dx_2} \\
&\simeq \frac{K_c}{2N(K - K_c)} \simeq \frac{1}{N} \varphi_T.
\end{aligned} \tag{6.1.34}$$

Equations (6.1.30) and (6.1.34) are examples of fluctuation-response function relations (Sect. 2.5.2).

6.2 A Model for Metamagnetism

Metamagnetism arises in situations where a lattice of magnetic dipoles is divided into a layered sublattice structure with ferromagnetic interactions within layers and antiferromagnetic interactions between layers (Martin 1967, p. 58). Such a model can be developed as a generalization of the Ising ferromagnet on a d -dimensional hypercubic lattice. Suppose that one particular axis direction is selected and the $(d - 1)$ -dimensional hyperplanes orthogonal to this axis are labelled in sequence $\mathfrak{a}, \mathfrak{b}, \mathfrak{a}, \mathfrak{b}, \dots$ to form a layered sublattice structure with two equivalent sublattices $\mathfrak{a}$ and $\mathfrak{b}$. The Hamiltonian for such a model is given by

$$\begin{aligned}
\hat{H}(N, J_1, J_2, \mathcal{H}, \mathcal{H}_s; \sigma) &= J_1 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(\mathfrak{a}, \mathfrak{b})} \sigma(\mathbf{r}) \sigma(\mathbf{r}') - J_2 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(\mathfrak{a}, \mathfrak{a})} \sigma(\mathbf{r}) \sigma(\mathbf{r}') \\
&\quad - J_2 \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(\mathfrak{b}, \mathfrak{b})} \sigma(\mathbf{r}) \sigma(\mathbf{r}') - (\mathcal{H} + \mathcal{H}_s) \sum_{\{\mathbf{r}\}}^{(\mathfrak{a})} \sigma(\mathbf{r}) - (\mathcal{H} - \mathcal{H}_s) \sum_{\{\mathbf{r}\}}^{(\mathfrak{b})} \sigma(\mathbf{r}), \tag{6.2.1}
\end{aligned}$$

where the first-neighbour pairs and site summations are restricted as indicated by the superscripts, with $J_1, J_2 > 0$. The field $\mathcal{H}$ is uniform over the lattice, but $\mathcal{H}_s$ is a staggered field with a different sign on the two sublattices. The system can be thought of as having four independent fields $T, \mathcal{H}, \mathcal{H}_s$ and one of J_1 or J_2 . Of these only T and $\mathcal{H}$ are the theoretical counterparts of experimentally controllable quantities. Here we concentrate mainly on the case where the staggered field is zero,³ although it will be included in the initial stages of the discussion. The number of first-neighbours to a site which are situated in the other sublattice are denoted by z_1 and in the same sublattice by z_2 .⁴

The mean-field free-energy density is now derived using a generalization of the procedure of Sect. 6.1. The Hamiltonian $\hat{H}_0(N, K_1, K_2, \mathcal{M}_a, \mathcal{M}_b; \sigma)$ for the

³ We are interested in deriving the part of the phase diagram in the plane $\mathcal{H}_s = 0$ shown in Fig. 4.3a, b.

⁴ Of course, with this layered model on a d -dimensional hypercubic lattice $z_1 = 2, z_2 = 2(d - 1)$, but z_1 and z_2 are used in the interests of generality.

metamagnet is just $\widehat{H}(N, K_1, K_2, 0, 0; \sigma)$, given by (6.2.1), with the sum of the microstates on the two sublattices constrained by the conditions

$$\mathcal{M}_a = \sum_{\{r\}}^{(a)} \sigma(r), \quad \mathcal{M}_b = \sum_{\{r\}}^{(b)} \sigma(r), \quad (6.2.2)$$

and $K_1 := J_1/T$, $K_2 := J_2/T$. Approximating the interactions in the Hamiltonian in the manner exemplified by (6.1.2) and (6.1.3) gives

$$\widehat{H}_0(N, K_1, K_2, \mathcal{M}_a, \mathcal{M}_b) = \frac{1}{N} \{2K_1 z_1 \mathcal{M}_a \mathcal{M}_b - K_2 z_2 (\mathcal{M}_a^2 + \mathcal{M}_b^2)\}. \quad (6.2.3)$$

The number of terms in the partition function $Z_0(N, K_1, K_2, \mathcal{M}_a, \mathcal{M}_b)$ is the number

$$\begin{aligned} W(N, \mathcal{M}_a, \mathcal{M}_b) &:= \frac{(\frac{1}{2}N)!}{\{\frac{1}{2}(\frac{1}{2}N + \mathcal{M}_a)\}! \{\frac{1}{2}(\frac{1}{2}N - \mathcal{M}_a)\}!} \\ &\quad \times \frac{(\frac{1}{2}N)!}{\{\frac{1}{2}(\frac{1}{2}N + \mathcal{M}_b)\}! \{\frac{1}{2}(\frac{1}{2}N - \mathcal{M}_b)\}!} \end{aligned} \quad (6.2.4)$$

of spin arrangements consistent with (6.2.2). Applying Stirling's formula and taking the thermodynamic limit, as we did in Sect. 6.1, gives

$$\begin{aligned} \phi_2^{(\text{cl})}(K, L, L_s; m_a, m_b) &= \frac{1}{2} [\mathcal{G}(K(1 + \lambda); m_a) + \mathcal{G}(K(1 + \lambda); m_b)] \\ &\quad + \frac{1}{4} K(m_a + m_b)^2 - \frac{1}{2} L(m_a + m_b) - \frac{1}{2} L_s(m_a - m_b), \end{aligned} \quad (6.2.5)$$

where $m_a := 2\mathcal{M}_a/N$, $m_b := 2\mathcal{M}_b/N$. The function $\mathcal{G}$ and its first two derivatives are given by (6.1.10)–(6.1.13) and, for convenience, we now take $K := z_1 K_1 = z_1 J_1/T$ and $\lambda := z_2 K_2/z_1 K_1 = z_2 J_2/z_1 J_1 > 0$, which measures the relative strengths of the ferromagnetic and antiferromagnetic interaction energies.

The Hamiltonian (3.5.17), for the Ising model with first-neighbour antiferromagnetic interactions and second-neighbour ferromagnetic interactions, and the Hamiltonian (6.2.1), for our present metamagnetic model, will be seen to be of identical form when $\mathcal{H}_s = 0$. It follows that the mean-field free-energy densities for these two models will be identical with, in the first case, the parameters J_1 , J_2 and the corresponding z_1 and z_2 being reinterpreted as applying to first-neighbour and second-neighbour pairs.⁵ As indicated above, for the metamagnetic model λ is positive. However, as we shall see below, the analysis is applicable to the case $\lambda = 0$, which is the first-neighbour antiferromagnetic model. Thus we shall henceforth assume that $\lambda \geq 0$.

It is convenient to reformulate the free-energy density in terms of the magnetization m and sublattice order parameter q , given respectively by

⁵ For the simple cubic lattice $z_1 = 6$, $z_2 = 12$.

$$m := \frac{1}{2}(m_a + m_b), \quad (6.2.6)$$

$$q := \frac{1}{2}(m_a - m_b), \quad (6.2.7)$$

when

$$\begin{aligned} \phi_2^{(\text{cl})}(K, L, L_s; m, q) &= \frac{1}{2} [\mathcal{G}(K(1 + \lambda); m + q) + \mathcal{G}(K(1 + \lambda); m - q)] \\ &\quad + Km^2 - Lm - L_s q. \end{aligned} \quad (6.2.8)$$

The generalizations of the equilibrium and stability conditions (5.1.7) and (5.1.8) are now

$$\begin{aligned} 2 \frac{\partial \phi_2^{(\text{cl})}}{\partial m} &= \mathcal{G}^{(1)}(K(1 + \lambda); m + q) + \mathcal{G}^{(1)}(K(1 + \lambda); m - q) \\ &\quad + 4Km - 2L = 0, \end{aligned} \quad (6.2.9)$$

$$2 \frac{\partial \phi_2^{(\text{cl})}}{\partial q} = \mathcal{G}^{(1)}(K(1 + \lambda); m + q) - \mathcal{G}^{(1)}(K(1 + \lambda); m - q) - 2L_s = 0, \quad (6.2.10)$$

and

$$2 \frac{\partial^2 \phi_2^{(\text{cl})}}{\partial m^2} = \mathcal{G}^{(2)}(K(1 + \lambda); m + q) + \mathcal{G}^{(2)}(K(1 + \lambda); m - q) + 4K > 0, \quad (6.2.11)$$

$$2 \frac{\partial^2 \phi_2^{(\text{cl})}}{\partial q^2} = \mathcal{G}^{(2)}(K(1 + \lambda); m + q) + \mathcal{G}^{(2)}(K(1 + \lambda); m - q) > 0, \quad (6.2.12)$$

$$\begin{aligned} \Delta &:= \frac{\partial^2 \phi_2^{(\text{cl})}}{\partial m^2} \frac{\partial^2 \phi_2^{(\text{cl})}}{\partial q^2} - \left(\frac{\partial^2 \phi_2^{(\text{cl})}}{\partial m \partial q} \right)^2 \\ &= \mathcal{G}^{(2)}(K(1 + \lambda); m + q) \mathcal{G}^{(2)}(K(1 + \lambda); m - q) \\ &\quad + K[\mathcal{G}^{(2)}(K(1 + \lambda); m + q) + \mathcal{G}^{(2)}(K(1 + \lambda); m - q)] > 0, \end{aligned} \quad (6.2.13)$$

where the first and second derivatives of the function $\mathcal{G}$ are given by (6.1.12) and (6.1.13). Substituting conditions (6.2.9) into (6.2.8), it follows that, at equilibrium,

$$\begin{aligned} \phi_2^{(\text{cl})}(K, L, L_s; m, q) &= \frac{1}{4} \ln[(1 - m^2 - q^2)^2 - 4m^2 q^2] - \ln(2) \\ &\quad - \frac{1}{4} K[(1 - \lambda)m^2 - (1 + \lambda)q^2], \end{aligned} \quad (6.2.14)$$

and from (4.6.2), (4.6.3), (6.2.9) and (6.2.13),

$$\varphi_T = \left(\frac{\partial m}{\partial L} \right)_{K, L_s} = \frac{\mathcal{G}^{(2)}(K(1+\lambda); m+q) + \mathcal{G}^{(2)}(K(1+\lambda); m-q)}{2\Delta}, \quad (6.2.15)$$

$$\psi_T = \left(\frac{\partial q}{\partial L_s} \right)_{K, L} = \frac{\mathcal{G}^{(2)}(K(1+\lambda); m+q) + \mathcal{G}^{(2)}(K(1+\lambda); m-q) + 4K}{2\Delta}. \quad (6.2.16)$$

We shall now investigate the phase diagram in the plane $\mathcal{H}_s = L_s = 0$.

6.2.1 The Paramagnetic State

The paramagnetic state in the plane $L_s = 0$, where $q = 0$, $m_a = m_b = m$, is stable for large temperatures (K small) or fields (L large). Condition (6.2.10) is identically satisfied and, from (6.2.9), the magnetization, as a function of K and L , is given by

$$\mathcal{G}^{(1)}(K(1+\lambda); m) + 2Km = L. \quad (6.2.17)$$

From (6.2.11)–(6.2.13), this state is locally stable when

$$\mathcal{G}^{(2)}(K(1+\lambda); m) > 0. \quad (6.2.18)$$

As K increases (T decreases), with $m(K)$ given by (6.2.17), $\mathcal{G}^{(2)}(K(1+\lambda); m)$ decreases monotonically (Fig. 6.4) and the paramagnetic state will become unstable with respect to sublattice ordering when

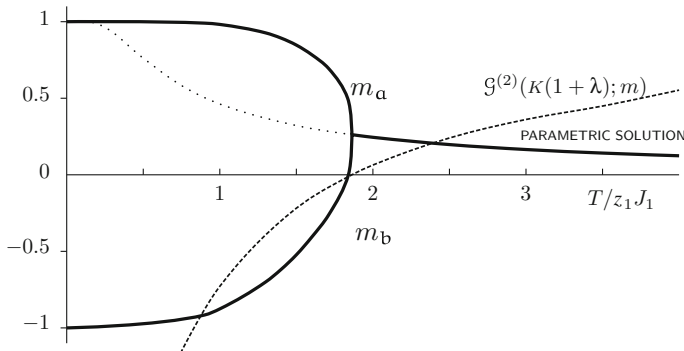


Fig. 6.4 Magnetization plotted against temperature for the metamagnetic for $\lambda = 1$, $\mathcal{H} = 0.5z_1 J_1$. The point on the critical curve separating the paramagnetic and antiferromagnetic solutions is at $T = 1.8625z_1 J_1$ where $m = 0.2622$. Below the critical curve the unstable paramagnetic solution is denoted by a dotted line

$$\mathcal{G}^{(2)}(K(1+\lambda); m) = 0. \quad (6.2.19)$$

The paramagnetic solution as a function of temperature along an isochamp (line of constant field) is shown in Fig. 6.4⁶ with the unstable part of the solution shown by a dotted line. By eliminating m between (6.2.19) and (6.2.17) we can obtain the equation $\mathcal{H} = \mathcal{H}^*(T)$ of the *critical curve* $\mathcal{C}$ of points, where the paramagnetic state becomes unstable, in the $\mathcal{H}$ – T plane or equivalently its form $L = L^*(K)$ in the L – K plane. It follows from (6.1.13) and (6.2.19) that the magnetizations on the two branches of the critical curve are given by

$$m = \pm m^*(\lambda, K), \quad m^*(\lambda, K) := \left(\frac{K(1+\lambda) - 1}{K(1+\lambda)} \right)^{\frac{1}{2}} \quad (6.2.20)$$

and, from (6.2.17),

$$\mathcal{G}^{(1)}(K(1+\lambda); m^*(\lambda, K)) + 2Km^*(\lambda, K) = L^*(K) \quad (6.2.21)$$

gives the critical curve. From (6.2.17), $m = 0$ on the zero-field axis in the paramagnetic state and thus, from (6.2.19)–(6.2.21), the critical curve cuts the zero-field axis orthogonally at the critical (Néel) temperature given by

$$K_c = \frac{1}{1+\lambda}, \quad T_c = z_1 J_1 (1+\lambda) = z_1 J_1 + z_2 J_2. \quad (6.2.22)$$

This is the point where the paramagnetic state becomes unstable when temperature is reduced at zero field.

From (6.1.12), (6.2.20) and (6.2.21) it can be shown that on the critical curve, as $T \rightarrow 0$, $\mathcal{H}^*(T) \rightarrow \pm z_1 J_1 (1-\lambda)$, which are not equal to the critical field values $\mathcal{H}_c = \pm z_1 J_1$, given by (3.5.19), unless $\lambda = 0$. The implications of this are discussed below.

6.2.2 The Antiferromagnetic State

Consider now the antiferromagnetic state in the $L_s = 0$ plane. Here there is sublattice order, $q \neq 0$ and $m_a \neq m_b$.

On the zero-field axis ($L = 0$) the equilibrium equations (6.2.9) and (6.2.10) are satisfied by $m = 0$, $q = \pm m_s$, (Fig. 6.7a) with

⁶ Here we encounter the usual problem in statistical mechanics that, whereas the equations take a simpler form in terms of couplings K and L , physical insight is helped by presenting figures in terms of the fields T and $\mathcal{H}$. We shall, therefore, in the discussion make free use of both sets of variables.

$$\mathcal{G}^{(1)}(K(1+\lambda); m_s) = \frac{1}{2} \ln \left(\frac{1+m_s}{1-m_s} \right) - K(1+\lambda)m_s = 0. \quad (6.2.23)$$

With the replacement of Jz by $z_1 J_1 + z_2 J_2$, this equation is identical to (6.1.14)⁷ for the spontaneous magnetization in zero field of the simple ferromagnetic Ising model. It is not difficult to show that this replacement is exactly that which occurs when second-neighbour interactions are added to the ferromagnetic model of Sect. 6.1. This together with the correspondence, discussed in Sect. 3.5.1.1, between that model and the one with antiferromagnetic first-neighbour interactions and ferromagnetic second-neighbour interactions and the further correspondence, noted above, between the latter and the metamagnetic model, allows us to understand why the Curie temperature for the ferromagnet is equal to the Néel temperature given by (6.2.22).

To investigate the stability of the antiferromagnetic state it is convenient to use the response function φ_T given by (6.2.15). Since the numerator of φ_T is positive it can be seen, by comparing (6.2.15) with (6.2.11)–(6.2.13), that the homogeneous antiferromagnetic phase is stable if and only if

$$\varphi_T = \left(\frac{\partial m}{\partial L} \right)_K > 0. \quad (6.2.24)$$

6.2.3 A Neighbourhood of the Critical Curve

Here we calculate some but, for the sake of brevity, not all of the critical exponents at a point $\mathbf{P}$ with coordinates (K_p, L_p) and $m_p = m^*(\lambda, K_p)$ on the critical curve $\mathcal{C}$, defined by (6.2.20) and (6.2.21), and at the tricritical point $\mathbf{T}$. Let

$$\Delta_p K := K - K_p, \quad \Delta_p L := L - L_p, \quad \Delta_p m := m - m_p. \quad (6.2.25)$$

Any second-order transitions, that is ones for which m and q are continuous, from the paramagnetic to the antiferromagnetic state, occur on this curve as it is approached along any line in the plane $L_s = 0$ which is not tangential to the curve. In terms of the orthogonal coordinates defined by (5.5.18)

$$a_p = \left(\frac{dL^*}{dK} \right)_p = (1-\lambda)m^*(\lambda, K_p). \quad (6.2.26)$$

So, in the asymptotic forms (4.6.16)–(4.6.22), with ‘t’ replaced by ‘p’,⁸ approaches along $\Theta_3 = \Theta_{3t}$ can be replaced by either an approach along the isotherm $K = K_p$ with $\Theta_2 - \Theta_{2p}$ replaced by $\Delta_p L$ or along the line $L = L_p$ with $\Theta_2 - \Theta_{2p}$ replaced by $\Delta_p K$.

⁷ With $\mathcal{H} = 0$, and remembering that in Sect. 6.1 $K := J/T$ not zJ/T .

⁸ As indicated in Sect. 5.5, for the exponents we simply omit the ‘p’.

Taking the plane $K = K_p$ and expanding (6.2.9) and (6.2.10) about $\mathbf{P}$ gives

$$\begin{aligned} \Delta_p L = & W_1(K_p)(\Delta_p m) + W_2(\lambda, K_p)[(\Delta_p m)^2 + q^2] + W_3(\lambda, K_p)(\Delta_p m) q^2 \\ & + W_4(\lambda, K_p) q^4 + O[(\Delta_p m)^3, q^6], \end{aligned} \quad (6.2.27)$$

$$\begin{aligned} L_s = & 2W_2(\lambda, K_p)(\Delta_p m) q + \frac{1}{3} W_3(\lambda, K_p)[3(\Delta_p m)^2 q + q^3] \\ & + 4W_4(\lambda, K_p)(\Delta_p m) q^3 + W_5(\lambda, K_p) q^5 + O[(\Delta_p m)^3, q^6], \end{aligned} \quad (6.2.28)$$

where

$$\begin{aligned} W_1(K) &:= 2K, \\ W_2(\lambda, K) &:= \frac{1}{2} \mathcal{G}^{(3)}(K(1+\lambda); m^*(\lambda, K)) = m^*(\lambda, K) K^2 (1+\lambda)^2, \\ W_3(\lambda, K) &:= \frac{1}{2} \mathcal{G}^{(4)}(K(1+\lambda); m^*(\lambda, K)) = K^2 (1+\lambda)^2 [4K(1+\lambda) - 3], \\ W_4(\lambda, K) &:= \frac{1}{24} \mathcal{G}^{(5)}(K(1+\lambda); m^*(\lambda, K)) \\ &= m^*(\lambda, K) K^3 (1+\lambda)^3 [2K(1+\lambda) - 1], \\ W_5(\lambda, K) &:= \frac{1}{120} \mathcal{G}^{(6)}(K(1+\lambda); m^*(\lambda, K)) \\ &= \frac{1}{5} K^3 (1+\lambda)^3 [16K^2 (1+\lambda)^2 - 20K(1+\lambda) + 5]. \end{aligned} \quad (6.2.29)$$

In the paramagnetic phase ($q = 0$) along the line $L_s = 0$, (6.2.10) and (6.2.28) are identically satisfied and, from (6.2.27) with $\Delta_p m < 0$,

$$\Delta_p L = W_1(K_p) \Delta_p m + W_2(\lambda, K_p) (\Delta_p m)^2 + O[(\Delta_p m)^3]. \quad (6.2.30)$$

In the antiferromagnetic phase ($q \neq 0$) along the line $L_s = 0$, (6.2.28) gives

$$q^2 = X_1(\lambda, K_p) \Delta_p m + X_2(\lambda, K_p) (\Delta_p m)^2 + O[(\Delta_p m)^3], \quad (6.2.31)$$

where

$$\begin{aligned} X_1(\lambda, K) &:= -6 \frac{W_2(\lambda, K)}{W_3(\lambda, K)}, \\ X_2(\lambda, K) &:= -3 + 72 \frac{W_2(\lambda, K) W_4(K)}{[W_3(\lambda, K)]^2} - 108 \frac{[W_2(\lambda, K)]^2 W_5(\lambda, K)}{[W_3(\lambda, K)]^3}. \end{aligned} \quad (6.2.32)$$

Then, from (6.2.27) with $\Delta_p m > 0$,

$$\Delta_p L = W'_1(\lambda, K_p) \Delta_p m + W'_2(\lambda, K_p) (\Delta_p m)^2 + O[(\Delta_p m)^3], \quad (6.2.33)$$

where

$$\begin{aligned} W'_1(\lambda, K) &:= W_1(K) + X_1(\lambda, K)W_2(\lambda, K) \\ &= \frac{6\lambda K K_c(K_t - K)}{K_t(4K - 3K_c)}, \end{aligned} \quad (6.2.34)$$

$$\begin{aligned} W'_2(\lambda, K) &:= [1 + X_2(\lambda, K)]W_2(\lambda, K) \\ &\quad + X_1(\lambda, K)W_3(\lambda, K) + [X_1(\lambda, K)]^2W_4(\lambda, K), \end{aligned}$$

K_c is given by (6.2.22),

$$K_t = \frac{3\lambda}{(1 + \lambda)(3\lambda - 1)}, \quad T_t = \frac{(3\lambda - 1)}{3\lambda}T_c, \quad (6.2.35)$$

and

$$W'_2(\lambda, K_t) = \frac{24\lambda^3(3\lambda)^{\frac{1}{2}}(3 - 5\lambda)}{5(1 + \lambda)^3(3\lambda - 1)^2}. \quad (6.2.36)$$

Comparing (4.6.20) with (6.2.30) and (6.2.33) gives

$$\beta_2 = 1, \quad \beta'_2 = 1, \quad \beta_{2t} = 1, \quad \beta'_{2t} = \frac{1}{2}. \quad (6.2.37)$$

Along the line $L = L_p$, (6.2.27) gives

$$\Delta_p m = Y_1(\lambda, K_p)q^2 + Y_2(\lambda, K_p)q^4 + O[q^6], \quad (6.2.38)$$

where

$$\begin{aligned} Y_1(\lambda, K) &:= -\frac{W_2(\lambda, K)}{W_1(K)}, \\ Y_2(\lambda, K) &:= -\frac{W_2(\lambda, K)}{[W_1(K)]^3} \{ [W_2(\lambda, K)]^2 - W_1(K)W_3(\lambda, K) \\ &\quad + W_2(\lambda, K)W_4(\lambda, K) \}. \end{aligned} \quad (6.2.39)$$

Then, from (6.2.28),

$$L_s = W''_1(\lambda, K_p)q^3 + W''_2(\lambda, K_p)q^5 + O[q^7], \quad (6.2.40)$$

where

$$\begin{aligned} W_1''(\lambda, K) &:= 2Y_1(\lambda, K)W_2(\lambda, K) + \frac{1}{3}W_3(\lambda, K) \\ &= \frac{\lambda K^2(K_t - K)}{K_t K_c^2}, \end{aligned} \quad (6.2.41)$$

$$\begin{aligned} W_2''(\lambda, K) &:= 2Y_2(\lambda, K)W_2(\lambda, K) + [Y_1(\lambda, K)]^2 W_3(\lambda, K) \\ &\quad + 4Y_1(\lambda, K)W_4(\lambda, K) + W_5(\lambda, K), \end{aligned}$$

and

$$W_2''(\lambda, K_t) = \frac{54\lambda^5(5\lambda - 3)}{5(3\lambda - 1)^5}. \quad (6.2.42)$$

It follows from (4.6.15) and (6.2.40)–(6.2.42) that

$$\delta = 3, \quad \delta_t = 5. \quad (6.2.43)$$

Now consider the response function φ_T close to the critical curve. From (6.2.30) and (6.2.33),

$$\varphi_T \simeq \begin{cases} [W_1(K_p) + 2W_2(\lambda, K_p)\Delta_p m]^{-1}, \\ \quad \text{in the paramagnetic phase,} \\ [W_1'(\lambda, K_p) + 2W_2'(\lambda, K_p)\Delta_p m]^{-1}, \\ \quad \text{in the antiferromagnetic phase.} \end{cases} \quad (6.2.44)$$

When $\frac{3}{4}K_c < K < K_t$, ($T_t < T < \frac{4}{3}T_c$), from (6.2.30) and (6.2.34), $W_1'(\lambda, K) > 0$ and

$$\varphi_T \simeq \begin{cases} \frac{1}{2K_p}, & \text{in the paramagnetic region,} \\ \frac{K_t(4K - 3K_c)}{6\lambda K K_c(K_t - K)}, & \text{in the antiferromagnetic region.} \end{cases} \quad (6.2.45)$$

The interesting case, from the point of view of the metamagnet, is when $\lambda > \frac{1}{3}$ and $K_t > 0$. Then, as K is increased through K_t , $W_1'(\lambda, K)$ changes sign and, from (6.2.24) and (6.2.44), the antiferromagnetic state becomes unstable near the critical curve. The situation for $K > K_t$ will be explored in Sect. 6.2.5 where we show that a first-order transition occurs between the antiferromagnetic and paramagnetic states meaning that $K = K_t$, $L = L^*(K_t)$ is a tricritical point. At this point, from (6.2.36), $W_2'(\lambda, K_t) > 0$, if $\lambda > \frac{3}{5}$. The significance of this value of λ will be discussed in

Sect. 6.2.5. For the moment we confine attention to $\frac{3}{5} < \lambda < 1$ where, from (6.2.44) and (6.2.33), in the antiferromagnetic region

$$\varphi_T \simeq \frac{1}{2|W'_2(\lambda, K_t)||\Delta_t m|} \simeq \frac{1}{2|W'_2(\lambda, K_t)|^{\frac{1}{2}}|\Delta_t L|^{\frac{1}{2}}}. \quad (6.2.46)$$

Comparing (4.6.22) with (6.2.45) and (6.2.46) gives

$$\gamma_2 = 0, \quad \gamma'_2 = 0, \quad \gamma_{2t} = 0, \quad \gamma'_{2t} = \frac{1}{2}. \quad (6.2.47)$$

In the antiferromagnetic phase, from (6.2.31) and (6.2.33),

$$\Delta_p L = Y'_1(\lambda, K_p)q^2 + Y'_2(\lambda, K_p)q^4 + O(q^6), \quad (6.2.48)$$

where

$$\begin{aligned} Y'_1(\lambda, K) &:= \frac{W'_1(\lambda, K)}{X_1(\lambda, K)} = -\frac{m^*(\lambda, K)\lambda K^2(K_t - K)}{K_t(K - K_c)} \\ Y'_2(\lambda, K_t) &:= \frac{W'_2(\lambda, K_t)X_1(\lambda, K_t) - W'_1(\lambda, K_t)X_2(\lambda, K_t)}{[X_1(\lambda, K_t)]^3} \\ &= -\frac{9\lambda^2(3\lambda)^{\frac{1}{2}}(2\lambda^3 + 15\lambda^2 + 54\lambda + 9)}{(3\lambda - 1)^4(1 + \lambda)} \end{aligned} \quad (6.2.49)$$

and comparing (6.2.48) with (4.6.19) gives

$$\beta = \frac{1}{2}, \quad \beta_t = \frac{1}{4}. \quad (6.2.50)$$

The values of the exponents β'_{2t} , γ'_{2t} , δ , δ_t , β and β_t obtained here agree with those derived by a mixture of Landau theory and scaling in Sect. 5.5. However, it is interesting to observe that this mean-field calculation gives $\beta_{2t} \neq \beta'_{2t}$ and $\gamma_{2t} \neq \gamma'_{2t}$. In each case this is a consequence of a vanishing amplitude factor ($W_1(K_t) \neq 0$ but $W'_1(\lambda, K_t) = 0$).

6.2.4 The First-Neighbour Antiferromagnet: $\lambda = 0$

Here we consider briefly the case $\lambda = 0$, which is a first-neighbour antiferromagnetic system. As noted in Sect. 6.2.1, this case is the only one in which the critical curve meets the zero-temperature axis at the critical field. Since the tricritical temperature is now negative the antiferromagnetic state is stable close to the critical temperature for the entire range $0 < T < T_c$ and the transition line is identical to the critical curve which is the $\lambda = 0$ line in Fig. 6.5. Although the finite discontinuity in φ_T at the transition may be an artifact of mean-field theory, the general picture of the phase

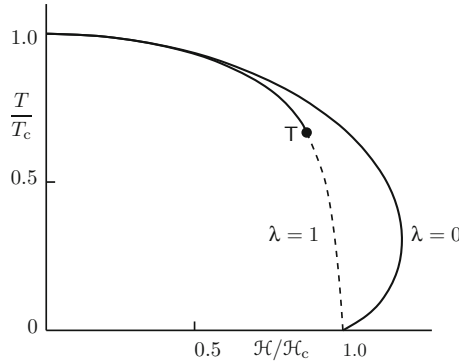


Fig. 6.5 Second-order (*full lines*) and first-order (*broken line*) transition curves in field–temperature plane for the antiferromagnet ($\lambda = 0$) and metamagnet ($\lambda = 1$), where the first and second-order transitions meet at a tricritical point T

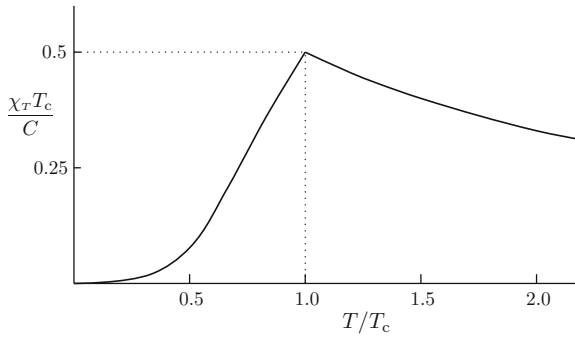


Fig. 6.6 The zero-field susceptibility for the first-neighbour antiferromagnet, ($\lambda = 0$)

diagram for the simple antiferromagnet, with the antiferromagnetic and paramagnetic regions separated by a higher-order transition line, is confirmed by series analysis (Domb 1974c).

Consider now the situation as K (or T) is varied along $L = 0$, where (6.2.9) is satisfied by $m = 0$ for any value of q . From (1.4.9), (6.1.13), (6.2.13) and (6.2.15) the zero-field susceptibility is

$$\chi_T := \frac{C\psi_T}{T} = \frac{C}{T(1 - q^2)^{-1} + T_c}. \quad (6.2.51)$$

In the paramagnetic region ($T > T_c$), $q = 0$ and in the antiferromagnetic region ($T < T_c$) q is given by (6.2.10). It follows that χ_T and $d\chi_T/dT \rightarrow 0$, as $T \rightarrow 0$. Whereas $d\chi_T/dT > 0$, for $T \leq T_c$, from (6.2.51), $d\chi_T/dT < 0$, for $T > T_c$. While χ_T is continuous at $T = T_c$, $d\chi_T/dT$ has a finite discontinuity, giving a cusp in the χ_T – T curve (Fig. 6.6). This contrasts with series results (Fisher and Sykes 1962;

Domb 1974c) which indicate that $d\chi_T/dT$ is continuous at all T and that χ_T has a maximum at a temperature which is about $1.5 T_c$ for $d = 2$ and $1.08 T_c$ for $d = 3$.

When an experimental plot of χ_T^{-1} against T is asymptotic at high temperatures to a line $A(T - T_{pc})$, for some constant A , a *Curie–Weiss law* is said to be obeyed, and T_{pc} is called a *paramagnetic Curie temperature*. From (6.1.22), $T_{pc} = T_c$ in the mean-field approximation for the Ising model ferromagnet. In a real ferromagnet, the paramagnetic Curie temperature is 10 or 20 K above the ferromagnetic Curie temperature (Martin 1967). From (6.2.51), $T_{pc} = -T_c$ for the mean-field antiferromagnet. A negative paramagnetic Curie temperature is a well-known characteristic of real antiferromagnetic substances for which the susceptibility below the Néel point depends on the direction in which the external magnetic field is applied. The curve shown in Fig. 6.6 corresponds qualitatively to the response to a field along the zero-field magnetization axis of the sublattices. For responses to fields perpendicular to the axis of zero-field magnetization see Fisher (1963), Martin (1967), Mattis (1985).

6.2.5 The First-Order Transition

As indicated above, a positive λ is essential for the metamagnetic model and in that case, on the zero-temperature axis, the end points $\mathcal{H}^*(0) = \pm z_1 J_1(1 - \lambda)$ of the critical curves are not coincident with the critical field $\mathcal{H}_c = \pm z_1 J_1$; at low temperatures there cannot be a second-order transition from the antiferromagnetic state to the paramagnetic state. It has already been shown in Sect. 6.2.3 that, when $\lambda > \frac{1}{3}$, the antiferromagnetic state is unstable close to the critical curve when $T < T_t$. However, there remains the range of λ values $(0, \frac{1}{3})$ when $T_t < 0$. Now on the zero-field axis $m_a = -m_b = q \geq 0$, $m = 0$, and on the critical curve $m_a = m_b = m > 0$, $q = 0$. Since m_b is a monotonically increasing function of L , or $\mathcal{H}$, as the field is increased from zero to its critical value, at constant temperature, there is an intermediate field value where $m_b = 0$ ($m = q$), and at that value it can be shown that, for sufficiently small temperature, $\varphi_T < 0$ for all $0 < \lambda \leq 1$. There is thus an instability loop in the plot of L (or $\mathcal{H}$) against m at constant K similar to that shown in Fig. 5.1a and resulting in phase separation, with the point where $m_b = 0$ lying in the region of intrinsic instability (Fig. 6.7a).⁹ Hence a line of first-order transitions exists in the low-temperature part of the $\mathcal{H}$ – T plane, terminating at $T = 0$, where we would expect it to meet the $\mathcal{H}$ -axis at $\mathcal{H} = \mathcal{H}_c$. This is confirmed by the fact that, if the values of $\mathcal{H}$ and T where $m_b = 0$ are plotted in the $\mathcal{H}$ – T plane, this curve meets the $\mathcal{H}$ -axis at $\mathcal{H} = \mathcal{H}_c$. The remainder of this section will be concerned with how this first-order transition line terminates at its upper end.¹⁰

⁹ The case shown here is for $\lambda = 1$, $T_t > 0$; the differences that occur when $\lambda < \frac{1}{3}$, $T_t < 0$ are discussed below.

¹⁰ There is also a mirror-image transition line in the $\mathcal{H} < 0$ half-plane, but we need consider only the $\mathcal{H} > 0$ line explicitly.

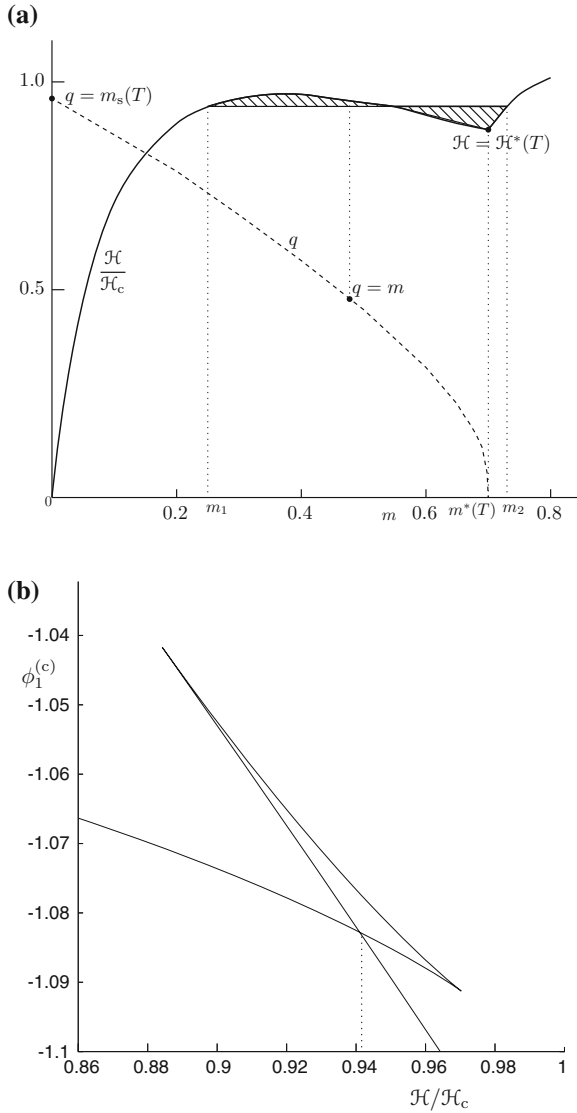


Fig. 6.7 **a** Field-magnetization and q -magnetization plots for $\lambda = 1$, $T = \frac{1}{2}T_c$. The field-magnetization plot shows an instability loop and the equal-areas tie line. **b** A plot, for $\lambda = 1$, $T = \frac{1}{2}T_c$ ($K = 1$) of the equilibrium free-energy density given by (6.2.14) against $\mathcal{H}/\mathcal{H}_c = L$. The crossing point of the curve is used to locate the equal-areas tie line in Fig. 6.7a

Since, from (6.2.44), $(\partial L/\partial m)_K = 1/\varphi_T = W'_1(\lambda, K)$ at $m = m^* - 0$, the stability condition (6.2.24) is not satisfied on the antiferromagnetic side of the critical curve if $W'_1(\lambda, K) < 0$. From (6.2.34) and (6.2.35), $W'_1(\lambda, K) = 2K_c$ at $T = T_c$ and decreases steadily with T , passing through the value zero at $T = T_t$. For $T_c > T > T_t$

there is a second-order transition on the critical curve, as discussed at the end of Sect. 6.2.2. For $T_t > T > 0$ and λ sufficiently large (in fact, $\lambda > \frac{3}{5}$, as shown below), the plot of $\mathcal{H}$ against m is similar to that in Fig. 6.7a where the loop implies a first-order transition.

The magnetizations m_1 and m_2 in the conjugate phases, correspond to the ends of the equal-areas tie line shown. The value of $\mathcal{H}$ for the tie line can be most easily calculated by using (6.2.14), with m_a and m_b given by (6.2.9), to plot $\phi_1^{(cl)}$ against $\mathcal{H}/\mathcal{H}_c$ (or L) as shown in Fig. 6.7b. The crossing point of the curve locates the value of $\mathcal{H}$ for the tie line (see Fig. 5.1a, b). The slope discontinuity at $m = m^*$ falls in the instability interval (m_1, m_2) and, for the range $T_t > T > 0$, there is a line of first-order transitions, as shown for the $\lambda = 1$ case in Fig. 6.5. This line ends at the *tricritical point* $T = T_t$, $\mathcal{H} = \mathcal{H}^*(T_t)$, where it is continued by a second-order transition line (see Sects. 4.1.1, 4.6 and 5.5). The full phase diagram for this case (including the staggered field) has been obtained by Kincaid and Cohen (1975) and is shown in Fig. 4.3b, where it can be seen that the line of first-order phase transitions can now be understood as a line of coexistence of three phases, two antiferromagnetic and one paramagnetic.

For small values of λ the first-order transition line starting at $\mathcal{H} = \mathcal{H}_c$ on the zero-temperature axis cannot end at a tricritical point. In fact from (6.2.35), $T_t < 0$ for $\lambda < 1/3$ and there is an critical point given by

$$\left(\frac{\partial L}{\partial m}\right)_K = 0, \quad \left(\frac{\partial^2 L}{\partial m^2}\right)_K = 0. \quad (6.2.52)$$

Since the condition (6.2.24) is satisfied everywhere in the paramagnetic region this critical point must lie in the antiferromagnetic region, that is inside the critical curve in the $\mathcal{H}$ - T plane and it is referred to as a *bicritical end-point*. Figure 6.8 shows part of the $\mathcal{H}$ - T plane for $\lambda = 0.3$. For any T between the values corresponding to points B and E, the plot of $\mathcal{H}$ against m will show a loop, but the point $m = m^*$, where a slope discontinuity occurs, lies outside the instability interval associated with the loop. Hence, when $\mathcal{H}$ is increased at constant temperature, there is a first-order transition on the segment BE, and then an antiferromagnetic-paramagnetic second-order transition at $\mathcal{H} = \mathcal{H}^*$ on the critical line. At the temperature corresponding to E the point $m = m^*$ enters the instability interval of the first-order transition, and the line of second-order transitions on the critical curve terminates at the *critical end-point* E (see Sect. 4.1.1). The full phase diagram for this case is shown in Fig. 4.3a, where it can be seen that the line of first-order phase transitions BE is a line of coexistence of four antiferromagnetic phases.

It is now easy to find the value $\lambda = \lambda_1$ where the low- λ critical end-point phase pattern goes over to the high- λ tricritical point phase pattern. Consider Fig. 6.8 and suppose that λ increases steadily. The segment BE will shrink until eventually the critical end-point E coalesces with the bicritical end-point B at $\lambda = \lambda_1$. When coalescence occurs, the conditions (6.2.52) defining E are satisfied on the antiferromagnetic

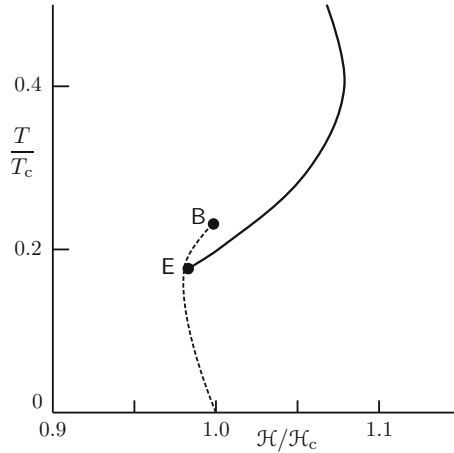


Fig. 6.8 As in Fig. 6.5 for $\lambda = 0.3$, with critical end-point E and bicritical end-point B

side of the critical curve so that $W'_2(\lambda, K_t) = 0$, giving $\lambda_1 = \frac{3}{5}$. (This result was derived by Bidaux et al. (1967), who gave a comprehensive set of diagrams.)

6.2.6 A Neighbourhood of the Tricritical Point

We now consider the tricritical regime again and in particular the situation just below the tricritical temperature where

$$T = T_t - \theta z_1 J_1, \quad \text{or equivalently,} \quad K = \frac{K_t}{1 - K_t \theta}. \quad (6.2.53)$$

From (6.2.35) the tricritical coupling $K_t = 3\lambda/(3\lambda - 1)(1 + \lambda)$ and $\theta \ll 1$ and the conjugate antiferromagnetic and paramagnetic phases existing for $T < T_t$ are labelled by the indices 1 and 2. From (5.1.6), (5.1.7) and (5.1.12) the phase equilibrium conditions become

$$L(K, m_1) = L(K, m_2), \quad (6.2.54)$$

$$L(K, m_1)m_1 - L(K, m_2)m_2 = - \int_{m_1}^{m_2} L(K, m) dm.$$

Using the variables

$$\delta := m - m^*(\lambda, K), \quad \varepsilon(K, \delta) := L(K, m) - L(K, m^*(\lambda, K)), \quad (6.2.55)$$

these conditions give

$$\varepsilon(K, \delta_1) = \varepsilon(K, \delta_2),$$

$$\varepsilon(K, \delta_1) \delta_1 - \int_0^{\delta_1} \varepsilon(K, \delta) d\delta = \varepsilon(K, \delta_2) \delta_2 - \int_0^{\delta_2} \varepsilon(K, \delta) d\delta, \quad (6.2.56)$$

with $\delta_1 < 0$ and $\delta_2 > 0$. On the antiferromagnetic side of the critical curve where $\delta < 0$, transcription of (6.2.33) gives

$$\varepsilon(K, \delta) = W'_1(\lambda, K)\delta + W'_2(\lambda, K)\delta^2 + O(\delta^3). \quad (6.2.57)$$

Since $\theta \ll 1$ we can, from (6.2.34) and (6.2.36) use the leading-term approximations

$$W'_1(\lambda, K) \simeq -K_t W_{11} \theta \quad W'_2(\lambda, K) \simeq K_t W_{20}, \quad (6.2.58)$$

where

$$W_{11} := \frac{6\lambda^2}{(1+\lambda)^2}, \quad W_{20} := \frac{8\lambda^2(3\lambda)^{\frac{1}{2}}(3-5\lambda)}{5(1+\lambda)^2(3\lambda-1)}. \quad (6.2.59)$$

For $\delta > 0$ and close to the critical curve, from (6.2.45),

$$\varepsilon(K, \delta) \simeq \frac{\delta}{\varphi_T} = 2K_t \delta \quad (6.2.60)$$

and since, for the tricritical range $\lambda > \frac{3}{5}$, $W_{20} < 0$, (6.2.57) yields a plot of the same general form as Fig. 6.7a. Substitution in the two relations of (6.2.56) yields

$$-W_{11}\theta\delta_1 + W_{20}\delta_1^2 = 2\delta_2, \quad (6.2.61)$$

$$-\frac{1}{2}W_{11}\theta\delta_1^2 + \frac{2}{3}W_{20}\delta_1^3 = \delta_2^2.$$

The conjugate magnetizations and coupling

$$m_1(\lambda, K) = m^*(\lambda, K) + \delta_1(\theta) = m^*(\lambda, K) - \frac{3W_{11}}{4|W_{20}|}\theta + O(\theta^2), \quad (6.2.62)$$

$$m_2(\lambda, K) = m^*(\lambda, K) + \delta_2(\theta) = m^*(\lambda, K) - \frac{3W_{11}^2}{32|W_{20}|}\theta^2 + O(\theta^3),$$

$$L(K) = L^*(K) + \frac{3K_t W_{11}^2}{16|W_{20}|}\theta^2 + O(\theta^3), \quad (6.2.63)$$

can then be derived. From (6.2.63) the first-order transition line and the critical curve have a common tangent at the tricritical point ($K = K_t, \theta = 0$), as shown for $\lambda = 1$ in Fig. 6.5, and from (6.2.62) the paramagnetic branch $m = m_2(K)$ of the coexistence curve in the m - K plane has a common tangent with the critical magnetization curve

$m = m^*(\lambda, K)$ at $K = K_t$, but the antiferromagnetic branch $m = m_1(K)$ meets it at a non-zero angle. Hence the coexistence curve has a cusp at the tricritical point, in contrast to the continuous tangent at a critical point. From (6.2.57) and (6.2.58), at the tricritical point

$$L - L^* \simeq K_t W_{20} (m^* - m)^2, \quad (6.2.64)$$

for $m < m^*$ (that is, on the antiferromagnetic side of the critical curve). Now, from (6.2.59), $W_{20} < 0$ for $\lambda < \frac{3}{5}$, giving $L > L^*$. This confirms that, for $\lambda < \frac{3}{5}$, the tricritical point falls in the instability range of the first-order transition. As we have seen, T_t becomes negative for $\lambda < \frac{1}{3}$.

Chapter 7

Cluster-Variation Methods

7.1 Improving Mean-Field Theory

As we saw in the application of mean-field theory to the Ising model in Sect. 6.1, the assumption that the spins on the lattice sites are uncorrelated is equivalent to replacing the interactions on a single spin with a *mean or molecular field* proportional to the magnetization. Minimising the free energy within the approximation then gives Eq. (6.1.25). Of course, if the partition function with an independent magnetic field were derived exactly, the magnetization would arise as the first derivative of the free energy with respect to the field. However, using the mean-field approximation, the free energy *is a function of the magnetization*. So (6.1.25) can also be regarded as a consistency condition.

It is not difficult to see how to improve on this method. Instead of breaking the correlations between single sites we could divide the lattice into larger uncorrelated clusters of sites. Methods based on a pair of sites (*first-order methods*) were invented independently by various authors, notably Guggenheim (1935), (the *quasi-chemical method*), Bethe (1931, 1935), (the *Bethe-pair method*, see Sect. 7.3)¹ and Kasteleyn and van Kranendonk (1956a, b, c), (the *constant-coupling method*) and many schemes for proceeding to larger groups have been proposed. The generalization of the quasi-chemical method, due to Guggenheim and McGlashan (1951), which allows any suitable group of sites to be used, enables quite complicated forms of ordering and intermolecular attraction to be incorporated in a natural way. However it remains a first-order method in the sense that it does not deal in a systematic way with all the sub-clusters within the chosen group of sites. Kikuchi (1951) developed a method for carrying the process of approximation to a higher level by minimizing the free energy with respect to the occupation probabilities of subgroups of sites, including first-neighbour pairs, as well as those of the main group and this procedure was placed

¹ The ‘original’ Bethe method was based on a cluster consisting of a site and all its z first neighbours. However, it can be shown that this yields no greater accuracy than a cluster consisting of a pair of first-neighbour sites.

on a systematic basis by Hijmans and de Boer (1955a, b, c). We shall, therefore, refer to these methods as the Kikuchi-Hijmans-De Boer (KHDeB) hierarchy of approximations. The cluster counting procedures used by Hijmans and de Boer have now been used in the finite-lattice series expansion method of de Neef and Enting (1977) described in Sect. 14.5.

Mean-field approximations, based on clusters consisting of a single site, take their place in the KHDeB hierarchy as *zeroth-order methods*. However, the way that contributions to the Hamiltonian (or enthalpy) are incorporated, both at the mean-field and higher levels of approximation is worthy of some comment. In Sect. 2.3 we derived the coupling-density formulation of thermodynamics, which involved the introduction of a set $L_i := \xi_i/T$, $i = 1, \dots, n_f$ of external couplings and $K_\ell := \varepsilon_\ell/T$, $\ell = 1, \dots, n_e$, of internal couplings. The fields ξ_i give contributions to the Hamiltonian (and enthalpy) from single microsystems, either because ξ_i is a single-site energy like the chemical potential, or because ξ_i is a magnetic or electric field representing an interaction with the environment. The internal energies ε_ℓ represent interactions between clusters of sites and one of the principles guiding the use of an appropriate approximation for a lattice model is to try to ensure that each group of sites having an irreducible interaction energy ε_ℓ is small enough to be a subset of a cluster within the chosen cluster approximation. This criterion is satisfied for the simple Ising model by the Bethe-pair approximation (Sect. 7.3). However, it is not satisfied by mean-field methods for any multi-site interactions. In these terms the only ‘satisfactory’ mean-field results are for models like the paramagnet, when it is exact² and, as we shall see in Sect. 7.4, the ‘reduction’ of a cluster-variation approximation to the corresponding mean-field analysis for a system with more than single-site energies needs an additional assumption. Of course applying the full KHDeB method to a model with irreducible interactions in large groups of sites will be very complicated and one simplification, which is still an advance on the zeroth-order method, is the Guggenheim and McGlashan approximation (see Sects. 7.6 and 16.4.8).

Because they incorporate short-range ordering, cluster-variation methods improve on the mean-field methods as regards the accuracy of phase diagrams and critical parameters. Sometimes they can treat effects which the mean-field method cannot and which are not yet accessible to exact analysis. However, the behaviour they predict near critical points is still ‘classical’, giving the same critical exponents as in mean-field theory. In fact, this is bound to be true for any method in which the densities depend analytically on thermodynamic variables and order parameters.

² An interesting twist to the story is that, though first-order theory is approximate for $d = 2$ or $d = 3$ regular lattices, like those shown in Sect. 16.2.1, it is exact for the interior sites of more loosely connected arrays, interacting only through first-neighbour pair interactions, (Baxter 1982b, Chap. 4).

7.2 The KHDeB Hierarchy of Approximations

Consider a lattice $\mathcal{N}$ of N sites and assume it to be so large that boundary effects can be neglected. The KHDeB cluster method begins by taking a particular *connected section graph*³ $\mathbf{g}$ on $\mathcal{N}$ and generating the set of section graphs obtained as distinct overlapping subgraphs of two different copies of $\mathbf{g}$. If any one of these subgraphs is not connected then all of its components are included in the sequence; making each member of the sequence connected. This is an example of the overlap procedure described in Sect. 16.4.6. Suppose that the number of *non-isomorphic* (distinct) graphs generated in this way (including $\mathbf{g}$) is n and label them as a partially-ordered sequence $\mathcal{S}$ so that if $\mathbf{g}_i \subset \mathbf{g}_j$ then $i < j$, with $\mathbf{g}_n$ being the generating graph $\mathbf{g}$ and $\mathbf{g}_1$ being a single lattice point. As shown in Theorem 16.4.4, $\mathcal{S}$ is *complete* (Definition 16.4.1).

7.2.1 Distribution Numbers

A *configuration* σ on $\mathcal{N}$ is a particular assignment of microsystem states to the sites of $\mathcal{N}$ and a configuration on $\mathbf{g}_j$ is defined in the same way. The configurations on two copies of $\mathbf{g}_j$ are said to be *isomorphic* if, when they are brought into coincidence by some combination of translations, rotations and reflections, the state on each coinciding site is the same. The non-isomorphic configurations on $\mathbf{g}_j$ are labelled with the index⁴ ℓ and we denote the graph $\mathbf{g}_j$ in configuration ℓ by $\mathbf{g}_j(\ell)$ with $\lambda_j(\ell)$ denoting the number of isomorphic configurations on $\mathbf{g}_j$ with the same ℓ . Let

$$\pi_j(\ell | \sigma) := \text{Prob}[\text{configuration of } \mathbf{g}_j \text{ is } \ell \mid \text{configuration of } \mathcal{N} \text{ is } \sigma]. \quad (7.2.1)$$

Then, for large N , the total number of $\mathbf{g}_j(\ell)$ in each of the $\lambda_j(\ell)$ isomorphic configurations is $N \omega_j \pi_j(\ell | \sigma)$. The set of probabilities $\{\pi_j(\ell | \sigma)\}$, also known as *distribution numbers*, must satisfy, for all σ (as will be implied in all relevant places in the following equations), the normalization conditions

$$1 = \sum_{\{\ell\}} \lambda_j(\ell) \pi_j(\ell | \sigma), \quad j = 1, 2, \dots, n. \quad (7.2.2)$$

As defined in Sect. 16.4.5, the lattice constant Ω_{ij} is the number of ways of arranging $\mathbf{g}_i$ on $\mathbf{g}_j$ and the lattice constant per site ω_i is such that that $N\omega_i$ is the number of

³ For definitions of emphasized lattice graph theoretic terms see Sect. 16.4.

⁴ We use a vector variable to label the configurations since they may, in general, contain more than one designating integer. There is further discussion of the meaning of isomorphic, when applied to configurations on graphs in Sect. 7.2.6.

ways of arranging $\mathbf{g}_i$ on $\mathcal{N}$.⁵ We define Λ_{ij} to be the number of $\mathbf{g}_j$ of which a particular $\mathbf{g}_i$ is a subgraph. By counting the number of $\mathbf{g}_i$ on the lattice it follows that

$$\omega_i \Lambda_{ij} = \Omega_{ij} \omega_j. \quad (7.2.3)$$

Consistency relationships between the distribution numbers can now be obtained by a counting procedure similar to that used to derive (7.2.3). Let $\vartheta_{ij}(\ell' | \ell)$ be the number of $\mathbf{g}_i(\ell')$ in $\mathbf{g}_j(\ell)$. Then

$$N \omega_i \lambda_i(\ell') \pi_i(\ell' | \sigma) \Lambda_{ij} = N \omega_j \sum_{\{\ell\}} \vartheta_{ij}(\ell' | \ell) \lambda_j(\ell) \pi_j(\ell | \sigma) \quad (7.2.4)$$

is obtained by counting the number of $\mathbf{g}_i(\ell')$ on the lattice directly (on the right) and as a subgraph of each of the $\mathbf{g}_j(\ell)$ (on the left), when each $\mathbf{g}_i(\ell')$ is counted Λ_{ij} times. Thus, from (7.2.3), we have the $i - j$ consistency conditions

$$\lambda_i(\ell') \pi_i(\ell' | \sigma) \Omega_{ij} = \sum_{\{\ell\}} \vartheta_{ij}(\ell' | \ell) \lambda_j(\ell) \pi_j(\ell | \sigma), \quad \forall \ell'. \quad (7.2.5)$$

Summing (7.2.5) over all ℓ' and using (7.2.2) gives

$$\Omega_{ij} = \sum_{\{\ell'\}} \vartheta_{ij}(\ell' | \ell), \quad \forall \ell. \quad (7.2.6)$$

For later use note that the important equations among those given here are the normalization condition (7.2.2), with $j = n$,

$$1 = \sum_{\{\ell\}} \lambda_n(\ell) \pi_n(\ell | \sigma) \quad (7.2.7)$$

and, from (7.2.5), the $j - n$ consistency conditions

$$\lambda_j(\ell') \pi_j(\ell' | \sigma) \Omega_{jn} = \sum_{\{\ell\}} \vartheta_{jn}(\ell' | \ell) \lambda_n(\ell) \pi_n(\ell | \sigma), \quad \forall \ell'. \quad (7.2.8)$$

Using these equations all the distribution numbers can be expressed in terms of the set $\{\pi_n(\ell | \sigma)\}$.

⁵ As $\mathbf{g}_i$ is connected ω_i is independent of N .

7.2.2 Extensive Quantities

Let $\varepsilon(\mathbf{g}_j(\ell))$ be a property (like, for example, the energy) of $\mathbf{g}_j(\ell)$ which is additive on non-overlapping copies of $\mathbf{g}_j$ and which includes all contributions from proper subgraphs of $\mathbf{g}_j$ in configurations compatible with ℓ . We now wish to calculate $\mathcal{E}(\mathcal{N}(\sigma))$, which is the sum of ε -contributions from all the graphs of $\mathcal{N}$ in the configurations implied by the configuration σ of $\mathcal{N}$. The KHDeB method provides a hierarchy of methods for achieving this to a corresponding level of approximation. In Sect. 7.2.1 we introduced the distribution numbers $\{\pi_i(\ell | \sigma)\}$, $i = 1, 2, \dots, n$ and showed that it is necessary to specify only the set $\{\pi_n(\ell | \sigma)\}$, since the distribution numbers for graphs further down the sequence are given by the consistency conditions (7.2.8). Suppose that n is increased with the sequence $\mathcal{S}$ including larger and larger graphs. Ultimately it will embrace the whole lattice; then $\ell \equiv \sigma$ and $\pi_n(\ell | \sigma) = 1$, and in a trivial sense for this limiting sequence the configurations $\{\sigma\}$ can be replaced by the distribution numbers $\{\pi_n(\sigma)\}$.

The steps of the KHDeB method, for some more restricted sequence $\mathcal{S}$ and some (usually quite small) n , are:

- (1) To suppose that the configuration σ can be approximately represented by the distribution numbers $\{\pi_i(\ell)\}$, $i = 1, 2, \dots, n$, from which dependence on σ is now removed so that $\mathcal{E}(\mathcal{N}(\sigma))$ becomes a function of the distribution numbers and is written $\mathcal{E}(\{\pi_n(\ell)\})$.
- (2) To suppose that n is sufficiently large so that every graph $\mathbf{g}_j$, which makes a contribution $\varepsilon(\mathbf{g}_j(\ell))$ when its configuration is ℓ and which is not already contributed by a proper subgraph of $\mathbf{g}_j$, is included in $\mathcal{S}$.
- (3) The total number of $\mathbf{g}_j(\ell)$ on $\mathcal{N}$ is $N \omega_j \lambda_j(\ell) \pi_j(\ell)$. So if we were simply to count contributions only from members of $\mathcal{S}$ in their various configurations and ignore any overlaps we would have

$$\mathcal{E}(\{\pi_n(\ell)\}) = N \sum_{j=1}^n \omega_j \sum_{\{\ell\}} \varepsilon(\mathbf{g}_j(\ell)) \lambda_j(\ell) \pi_j(\ell). \quad (7.2.9)$$

Formula (7.2.9) is clearly incorrect, since some contributions from $\mathbf{g}_j(\ell)$ will have already been counted in a contribution from $\mathbf{g}_i(\ell)$, where $i > j$. The correct expression must be of the form

$$\mathcal{E}(\{\pi_n(\ell)\}) = N \sum_{j=1}^n \xi_j \sum_{\{\ell\}} \varepsilon(\mathbf{g}_j(\ell)) \lambda_j(\ell) \pi_j(\ell), \quad (7.2.10)$$

for some set of numbers ξ_j , $j = 1, 2, \dots, n$, with $\xi_j \leq \omega_j$, $j < n$. These numbers now need to be determined. A reasonable assumption is that they are independent of the property ε . In particular we can choose, for some fixed $\mathbf{g}_i$,

$\varepsilon(\mathfrak{g}_j(\ell))$ to be equal, for all ℓ , to the number Ω_{ij} of arrangements of $\mathfrak{g}_i$ on $\mathfrak{g}_j$. Then $\mathcal{E}(\{\pi_n(\ell)\}) = N\omega_i$ and (7.2.10) gives

$$\omega_i = \sum_{j=1}^n \Omega_{ij} \xi_j, \quad i = 1, 2, \dots, n. \quad (7.2.11)$$

Thus, from (16.4.26), $\xi_j = \chi_j^{(n)}$, $j = 1, 2, \dots, n$, and

$$\mathcal{E}(\{\pi_n(\ell)\}) = N \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \varepsilon(\mathfrak{g}_j(\ell)) \lambda_j(\ell) \pi_j(\ell). \quad (7.2.12)$$

The numbers $\chi_j^{(n)}$, $j = 1, 2, \dots, n$, are the weights for the sequence $\mathcal{S}$ (Sect. 16.4.5).

This method of deriving (7.2.12) is not that given by Hijmans and de Boer (1955a) and it is instructive to give a brief account of their method, which consists of disconnecting the lattice into its separate graphs $\mathfrak{g}_j$ at each of the levels $j = 1, 2, \dots, n$. The set of separate graphs $\mathfrak{g}_j$ is called a j -graph assembly. If disconnection occurs at level j of the lattice $\mathcal{N}$ then there will be $N\omega_j$ members of the j -graph assembly. If disconnection at level j occurs for a lattice which has already been disconnected at a higher level then the number must be made smaller to take into account the graphs already counted at a higher level. We denote by $\mathcal{N}_j(\xi_j)$ a j -graph assembly of $N\xi_j$ members. The approximation of Hijmans and de Boer, as given in (1) above, is to take the largest member of $\mathcal{S}$ as the highest level of disconnection. Thus we begin with an n -assembly $\mathcal{N}_n(\xi_n)$ with, from (16.4.26), $\xi_n = \omega_n = \chi_n^{(n)}$. Now consider $j = n-1$. The number of members $N\xi_{n-1}$ in the $(n-1)$ -graph assembly must take account of the fact that $N\Omega_{n-1n}\chi_n^{(n)}$ have already been counted in the n -graph assembly. Thus, again using (16.4.26), $\xi_{n-1} = \omega_{n-1} - \Omega_{n-1n}\chi_n^{(n)} = \chi_{n-1}^{(n)}$. Repeating this process gives

$$\xi_j = \omega_j - \sum_{i=j+1}^n \Omega_{ji} \chi_i^{(n)} = \chi_j^{(n)}, \quad j = 1, 2, \dots, n. \quad (7.2.13)$$

As indicated above, the quantity $\varepsilon_j(\ell)$ contains, in general, not only a contribution involving all the sites of $\mathfrak{g}_j(\ell)$, but also contributions from proper subgraphs of $\mathfrak{g}_j(\ell)$. Let $\tilde{\varepsilon}_j(\ell)$ be the contribution to $\varepsilon_j(\ell)$ from $\mathfrak{g}_j(\ell)$ which would not appear in any subgraph. Then

$$\varepsilon_j(\ell) = \sum_{i=1}^j \sum_{\{\ell'\}} \vartheta_{ij}(\ell' | \ell) \tilde{\varepsilon}_i(\ell'). \quad (7.2.14)$$

Substituting into (7.2.12) and using (7.2.5) gives

$$\mathcal{E}(\{\pi_n(\ell)\}) = N \sum_{j=1}^n \omega_j \sum_{\{\ell\}} \tilde{\varepsilon}_j(\ell) \lambda_j(\ell) \pi_j(\ell). \quad (7.2.15)$$

Although (7.2.15) has the virtue of simplicity (7.2.12) is more useful for the development of the KHDeB cluster method.

7.2.3 The Hamiltonian and Free Energy

Let the contribution to the Hamiltonian from a graph g_j in occupation ℓ , including all contributions from proper subgraphs of g_j in configurations compatible with ℓ , be $\hat{h}_j(\ell)$. Since the Hamiltonian is additive over non-overlapping graphs it follows from (7.2.12), that it takes the form

$$\hat{H}(\{\pi_n(\ell)\}) := N \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \hat{h}_j(\ell) \lambda_j(\ell) \pi_j(\ell), \quad (7.2.16)$$

where an implicit linear dependence on *couplings* (not energy parameters) is assumed for the members of the set $\{\hat{h}_j(\ell)\}$, and hence for the whole Hamiltonian.

Let $W(\{\pi_n(\ell)\})$ be the number of microstate configurations σ consistent with the distribution numbers $\{\pi_n(\ell)\}$. Then the dimensionless free energy is

$$\Phi(\{\pi_n(\ell)\}) = -\ln \left[\sum_{\{\pi_n(\ell)\}} W(\{\pi_n(\ell)\}) \exp(-\hat{H}(\{\pi_n(\ell)\})) \right]. \quad (7.2.17)$$

Again, as in Sect. 6.1, the summation in (7.2.17) is approximated by the largest term, giving the cluster-variation free energy

$$\Phi(\{\pi_n(\ell)\}) = \hat{H}(\{\pi_n(\ell)\}) - \ln[W(\{\pi_n(\ell)\})]. \quad (7.2.18)$$

7.2.4 The Entropy

In Eq. (7.2.18)

$$S(\{\pi_n(\ell)\}) := \ln[W(\{\pi_n(\ell)\})], \quad (7.2.19)$$

is the cluster-variation entropy (cf. Eq. (1.1.10)). For brevity we denote $W(\{\pi_n(\ell)\})$ as W_n and, in a similar way, denote by W_j the number of microstate configurations consistent with the distribution numbers $\{\pi_j(\ell)\}$. Using the procedure of Hijmans and

de Boer (1955a), described in Sect. 7.2.2, let $G_k(j, \xi_j)$ be the number of microstate configurations on the j -assembly $\mathcal{N}_j(\xi_j)$ consistent with the distribution numbers $\{\pi_k(\ell)\}$. It is clear that $k \leq j$ and

$$G_j(j, \xi_j) = \Psi_j(N\xi_j), \quad \text{where} \quad \Psi_j(X) := \frac{X!}{\prod_{\{\ell\}} [\{X \pi_j(\ell)\}]^{\lambda_j(\ell)}}. \quad (7.2.20)$$

Using Stirling's formula and neglecting all terms which are negligible in the thermodynamic limit, $N \rightarrow \infty$,

$$\Psi_j(N\xi_j) = [\psi_j(N)]^{\xi_j}. \quad (7.2.21)$$

For brevity $\psi_j(N)$ is denoted by ψ_j . Consider now the quantity

$$C_k(j) := \frac{W_k}{G_k(j, \omega_j)}, \quad k \leq j. \quad (7.2.22)$$

The numerator is the number of microstate configurations on $\mathcal{N}$ consistent with the distribution numbers $\{\pi_k(\ell)\}$ and the denominator is the number of microstate configurations on $\mathcal{N}_j(\omega_j)$ consistent with the distribution numbers $\{\pi_k(\ell)\}$. It is clear that $C_k(j) < 1$. The assumption, originating in the work of Kikuchi (1951) and adopted in the derivation of the entropy in Hijmans and de Boer (1955a), is that $C_j(j) = C_{j-1}(j)$, for $j \leq n$. Thus, from (7.2.20)–(7.2.22),

$$\frac{W_j}{(\psi_j)^{\omega_j}} = \frac{W_{j-1}}{G_{j-1}(j, \omega_j)} \quad j = 2, 3, \dots, n. \quad (7.2.23)$$

Theorem 7.2.1 *With $\chi_i^{(j)}$, $i = 1, 2, \dots, k$ the components of the vector solution $\chi^{(j)}$ of (16.4.61) and ψ_i , $i = 1, 2, \dots, j$ given by (7.2.21), it follows from (7.2.23) that*

$$W_j = \prod_{i=1}^j (\psi_i)^{\chi_i^{(j)}}. \quad (7.2.24)$$

Proof As noted in Sect. 16.4.9, $\chi_j^{(j)} = \omega_j$ and (7.2.24) gives $W_1 = (\psi_1)^{\omega_1}$, which is true since this is the correct combinatorial factor for the single sites of the lattice. We now prove (7.2.24) by induction, assuming

$$W_{j-1} = \prod_{i=1}^{j-1} (\psi_i)^{\chi_i^{(j-1)}}. \quad (7.2.25)$$

This formula must also be valid for a j -figure assembly. This is equivalent to replacing, for all $i = 1, 2, \dots, j-1$ the lattice constant ω_i by $\Omega_i \omega_j$. To see the effect on the numbers $\chi_i^{(j-1)}$ refer to (16.4.67), which yields, from (7.2.25),

$$G_{j-1}(j, \omega_j) = \prod_{i=1}^{j-1} (\psi_i)^{\chi_i^{(j-1)} - \chi_i^{(j)}}. \quad (7.2.26)$$

Substituting from (7.2.25) and (7.2.26) into (7.2.23) and, using $\chi_j^{(j)} = \omega_j$, gives (7.2.24). $\square$

Using Stirling's formula, and neglecting all terms which are negligible in the thermodynamic limit, it follows from (7.2.19) and (7.2.24) that

$$S(\{\pi_n(\ell)\}) = -N \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \lambda_j(\ell) \pi_j(\ell) \ln[\pi_j(\ell)]. \quad (7.2.27)$$

7.2.5 Minimization

From (7.2.16), (7.2.18) and (7.2.27), the cluster-variation free energy per lattice site

$$\phi(\{\pi_n(\ell)\}) = \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \lambda_j(\ell) \pi_j(\ell) \{\hat{h}_j(\ell) + \ln[\pi_j(\ell)]\}, \quad (7.2.28)$$

in the limit $N \rightarrow \infty$, is a function of all the distribution numbers $\{\pi_j(\ell)\}$, for $j = 1, 2, \dots, n$ and it is with respect to variations of all of them that we wish to find an extremum. The distribution numbers $\{\pi_j(\ell)\}$ for $j < n$ are given in terms of $\{\pi_n(\ell)\}$ by the j - n consistency conditions (7.2.8), which yield

$$\frac{\partial \pi_j(\ell)}{\partial \pi_n(\ell')} = \frac{\vartheta_{jn}(\ell|\ell') \lambda_n(\ell')}{\Omega_{jn} \lambda_j(\ell)} \quad (7.2.29)$$

and using for the normalization condition (7.2.7) an undetermined multiplier ζ ,⁶ the extremum conditions are

$$\zeta = \sum_{j=1}^n \frac{\chi_j^{(n)}}{\Omega_{jn}} \sum_{\{\ell\}} \{\hat{h}_j(\ell) + \ln[\pi_j(\ell)]\} \vartheta_{jn}(\ell|\ell') \quad \forall \ell'. \quad (7.2.30)$$

⁶ Into which, from (7.2.6), a constant factor $\sum_{j=1}^n \chi_j^{(n)}$ has been absorbed.

Multiplying this equation by $\pi_n(\ell') \lambda_n(\ell')$ and summing over ℓ' gives, from (7.2.5),

$$\zeta = \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \pi_j(\ell) \lambda_j(\ell) \{\hat{h}_j(\ell) + \ln[\pi_j(\ell)]\} \quad (7.2.31)$$

and, from (7.2.28),

$$\phi(\{\pi_n(\ell)\}) = \zeta. \quad (7.2.32)$$

7.2.6 Labelling Configurations

From the point of view of the application of this analysis to particular examples the important equations are the normalization condition (7.2.7), the consistency conditions (7.2.8), the equilibrium conditions (7.2.31) and the equilibrium free-energy formula (7.2.32). In each of these the index ℓ is used to distinguish between non-isomorphic configurations. Suppose that ℓ is replaced by the pair of indices $(\ell; k)$, where ℓ now distinguishes configurations which make different contributions to the Hamiltonian and k (dependent on ℓ) distinguishes between non-isomorphic configurations within the set ℓ . Then $\hat{h}_j(\ell; k) = \hat{h}_j(\ell)$ for all j, ℓ and k . In Eqs. (7.2.7), (7.2.8) and (7.2.31) ℓ is replaced by $(\ell; k)$ and, it is assumed that:

(i)

$$\pi_j(\ell; k) = \pi_j(\ell), \quad \forall j, \ell \text{ and } k; \quad (7.2.33)$$

the probabilities of the configurations of $\mathfrak{g}_j$ are dependent only on $\hat{h}_j(\ell)$.

(ii)

$$\vartheta_{ij}(\ell' | \ell) = \sum_{\{k'\}} \vartheta_{ij}(\ell'; k' | \ell; k); \quad (7.2.34)$$

the total number of $g_i(\ell')$ (adding over all k') on $\mathfrak{g}_j(\ell; k)$ is independent of k .

Then, with

$$\lambda_j(\ell) := \sum_{\{k\}} \lambda_j(\ell; k), \quad (7.2.35)$$

the new versions of (7.2.7), (7.2.8) and (7.2.31) containing k can be summed over that index (and, in the case of (7.2.8), also over k') to restore them to their original form with the index ℓ for $\mathfrak{g}_j(\ell)$ now distinguishing energy levels with degeneracies $\lambda_j(\ell)$. Henceforth, it will be assumed that the index ℓ is used in this way.

7.3 The Bethe-Pair Approximation for the Ising Model

This method corresponds to the case where the partially-ordered sequence $\mathcal{S}$ consists of the section graphs $\mathfrak{g}_1$, a single site and $\mathfrak{g}_2$, a single edge (bond). Then with the coordination number of the lattice being z

$$\Omega = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad \omega = \begin{pmatrix} 1 \\ \frac{1}{2}z \end{pmatrix} \quad (7.3.1)$$

(Example 14.1) and, from (16.4.26), the weights are $\chi_1^{(2)} = 1 - z$ and $\chi_2^{(2)} = \frac{1}{2}z$. We now label the configurations on both $\mathfrak{g}_1$ and $\mathfrak{g}_2$ according to the aggregate spin and

$$\begin{aligned} \lambda_1(-1) &= 1, & \lambda_1(+1) &= 1, \\ \lambda_2(-2) &= 1, & \lambda_2(0) &= 2, & \lambda_2(+2) &= 1, \end{aligned} \quad (7.3.2)$$

$$\begin{aligned} \vartheta_{12}(-1|-2) &= 2, & \vartheta_{12}(-1|0) &= 1, & \vartheta_{12}(-1|+2) &= 0, \\ \vartheta_{12}(+1|-2) &= 0, & \vartheta_{12}(+1|0) &= 1, & \vartheta_{12}(+1|+2) &= 2, \end{aligned} \quad (7.3.3)$$

$$\begin{aligned} \hat{h}_1(-1) &= L, & \hat{h}_1(+1) &= -L, \\ \hat{h}_2(-2) &= -K + 2L, & \hat{h}_2(0) &= K, & \hat{h}_2(+2) &= -K - 2L, \end{aligned} \quad (7.3.4)$$

where $K := J/T$ and $L := \mathcal{H}/T$. The magnetization per lattice site is

$$m = \pi_1(+1) - \pi_1(-1) \quad (7.3.5)$$

and, from (7.2.8), the 1–2 consistency conditions are

$$\frac{1}{2}(1 \pm m) = \pi_2(0) + \pi_2(\pm 2). \quad (7.3.6)$$

From (7.2.28), the free energy per lattice site is

$$\begin{aligned} \phi(\{\pi_2(\ell)\}) &= mL - \frac{1}{2}Kz[\pi_2(-2) - 2\pi_2(0) + \pi_2(+2)] \\ &\quad + \frac{1}{2}(1 - z)[(1 + m)\ln(1 + m) + (1 - m)\ln(1 - m) - 2\ln(2)] \\ &\quad + \frac{1}{2}z[\pi_2(-2)\ln\pi_2(-2) + 2\pi_2(0)\ln\pi_2(0) + \pi_2(+2)\ln\pi_2(+2)] \end{aligned} \quad (7.3.7)$$

and from (7.2.30), the equilibrium conditions are

$$\left[\frac{\pi_2(+2)}{\pi_2(-2)} \right]^{\frac{z}{2}} \mathfrak{M} = \left(\frac{1 + m}{1 - m} \right)^{(z-1)}, \quad (7.3.8)$$

$$\frac{[\pi_2(0)]^2}{\pi_2(-2) \pi_2(+2)} = \mathfrak{X}^2, \quad (7.3.9)$$

where

$$\mathfrak{X} := \exp(-2K), \quad \mathfrak{M} := \exp(-2L). \quad (7.3.10)$$

Let

$$\tau := \sqrt{\frac{\pi_2(+2)}{\pi_2(-2)}}. \quad (7.3.11)$$

Then, from (7.3.6)–(7.3.9),

$$\mathfrak{M} = g(\tau), \quad \text{where} \quad g(\tau) = \tau^{-1} \left(\frac{\tau + \mathfrak{X}}{1 + \tau \mathfrak{X}} \right)^{(z-1)}, \quad (7.3.12)$$

and

$$m = \frac{\tau - \tau^{-1}}{\tau + 2\mathfrak{X} + \tau^{-1}}. \quad (7.3.13)$$

From (7.2.32), (7.3.8) and (7.3.9) the free-energy density is

$$\begin{aligned} \phi_1(K, L; \tau) &= \frac{1}{2}zK - \frac{1}{2}(z-1) \ln \left(\frac{1-m^2}{4} \right) + \frac{1}{2}z \ln[\pi_2(0)] \\ &= \frac{1}{2}(z-2) \ln(\tau + 2\mathfrak{X} + \tau^{-1}) \\ &\quad - \frac{1}{2}(z-1) \ln[1 + \mathfrak{X}(\tau + \tau^{-1}) + \mathfrak{X}^2] - \frac{1}{2}zK. \end{aligned} \quad (7.3.14)$$

In the mean-field analysis of the ferromagnetic Ising model in Sect. 6.1 the order parameter was the magnetization density m . Here it is more appropriate to regard τ as the order parameter. Once a solution $\tau = \tau(\mathfrak{M}, \mathfrak{X})$ has been obtained then the magnetization and free-energy density are given by (7.3.13) and (7.3.14). It is clear from (7.3.12) that $[g(\tau^{-1})]^{-1} = g(\tau)$. So, for fixed $\mathfrak{X}$, solutions to (7.3.12) occur in pairs $(\tau, \mathfrak{M})$ and $(\tau^{-1}, \mathfrak{M}^{-1})$.

When $\mathcal{H} = L = 0$ ($\mathfrak{M} = 1$) the equilibrium values of τ are solutions of $g(\tau) = 1$. One solution is $\tau = 1$, giving $m = 0$. To determine whether other pairs of solutions $\tau^{(\pm)} \neq 1$ exist consider

$$\frac{dg}{d\tau} = -g(\tau) \frac{\mathfrak{X}(\tau + \tau^{-1} - 2) + (\mathfrak{X} + 1)(2 - z + \mathfrak{X}z)}{(\tau + \mathfrak{X})(1 + \tau \mathfrak{X})}, \quad (7.3.15)$$

which is a maximum when $\tau = 1$. Ferromagnetic solution $\tau_s^{(\pm)}$ will therefore occur when $\mathfrak{X} < (z - 2)/z$ with corresponding magnetizations $\pm m_s$. This indicates that the critical temperature is given by

$$\mathfrak{X}_c = \frac{z-2}{z}, \quad K_c = \frac{1}{2} \ln[z/(z-2)], \quad T_c = \frac{2J}{\ln[z/(z-2)]}. \quad (7.3.16)$$

Critical temperatures derived from this formula are given in Table 3.1. However, to confirm that a transition occurs, it must be shown that

$$\phi(K; \tau_s) - \phi(K; 1) < 0, \quad K > K_c, \quad (7.3.17)$$

where it follows from (7.3.14) that $\phi(K; \tau_s^{-1}) = \phi(K; \tau_s)$.

For the square lattice, apart from the solutions $\tau = \pm 1$, $g(\tau) = 1$ yields

$$\tau^{(\pm)} = \frac{1 - 3\mathfrak{X}^2 \pm (1 - \mathfrak{X}^2)\sqrt{1 - 4\mathfrak{X}^2}}{2\mathfrak{X}^3}. \quad (7.3.18)$$

From (7.3.14), with $\tau_s = \tau^{(\pm)}$,

$$\phi(K; \tau_s) - \phi(K; 1) = \ln \left[\frac{1 - 2\mathfrak{X}^2}{2(1 - \mathfrak{X})^2} \right]. \quad (7.3.19)$$

For $\mathfrak{X} < \mathfrak{X}_c = \frac{1}{2}$ the argument of the logarithm in this equation is monotonically increasing to unity at the critical value, which establishes the inequality (7.3.17). The spontaneous magnetization derived from (7.3.13) and (7.3.18) is compared with that derived from (6.1.25) for the mean-field approximation in Fig. 3.1.

For the square lattice the next approximation in the hierarchy is to take $\mathcal{S}$ to consist of $\mathbf{g}_1 := [1, 1]$, a single site, $\mathbf{g}_2 := [2, 2]$, a single edge and $\mathbf{g}_3 := [4, 8]$, a single square (Table 16.1). The calculations for $\chi_j^{(3)}$, $j = 1, 2, 3$ are given in Example 16.4.4 and the critical temperature is $T_c = 2.426J$ (Kikuchi 1951). For the triangular lattice the next approximation in the hierarchy is used in Sect. 7.5 for the 3-state Potts model. For the spin- $\frac{1}{2}$ Ising model the corresponding calculations yield a critical temperature $T_c = 3.916J$ (Hijmans and de Boer 1955c).

7.4 Reduction to the Mean-Field Approximation

Consider, for simplicity, the case where $\mathbf{g}_1$ is the only single-site member of the sequence $\mathcal{S}$,⁷ which is chosen such that $\mathbf{g}_n$ is sufficiently large to contain any group of sites with an irreducible multi-site interaction. To approximate the Hamiltonian (7.2.15) and entropy (7.2.27) to forms expressible in terms of the single-site

⁷ The generalization to the case of a lattice divided into sublattices (Sect. 16.4.7) is straightforward.

probabilities $\{\pi_1(\ell)\}$ take

$$\pi_j(\ell) := \prod_{\{\ell'\}} [\pi_1(\ell')]^{\vartheta_{1j}(\ell'\ell)} \quad j = 2, 3, \dots, n. \quad (7.4.1)$$

This is a generalization to larger groups of sites of the assumption (6.1.4) that the states on a pair of sites are uncorrelated.

Let $\hat{u}_j(\ell)$ be the contribution to $\hat{h}_j(\ell)$ from $g_j(\ell)$ which would not appear in any subgraph. Then applying (7.2.15) to the Hamiltonian and substituting from (7.4.1) gives

$$\hat{H}(\{\pi_n(\ell)\}) = N \sum_{j=1}^n \omega_j \sum_{\{\ell\}} \lambda_j(\ell) \hat{u}_j(\ell) \prod_{\{\ell'\}} [\pi_1(\ell')]^{\vartheta_{1j}(\ell'\ell)}. \quad (7.4.2)$$

This expression will contain a term corresponding to any $\hat{u}_j(\ell) \neq 0$. For the entropy, substituting into (7.2.27) and using (7.2.5) gives

$$\begin{aligned} S(\{\pi_n(\ell)\}) &= N \sum_{j=1}^n \chi_j^{(n)} \sum_{\{\ell\}} \lambda_j(\ell) \pi_j(\ell) \sum_{\{\ell'\}} \vartheta_{1j}(\ell'\ell) \ln[\pi_1(\ell')] \\ &= N \sum_{j=1}^n \Omega_{1j} \chi_j^{(n)} \sum_{\{\ell\}} \lambda_1(\ell) \pi_1(\ell) \ln[\pi_1(\ell)] \\ &= N \omega_1 \sum_{\{\ell\}} \lambda_1(\ell) \pi_1(\ell) \ln[\pi_1(\ell)] = S_1(\{\pi_1(\ell)\}). \end{aligned} \quad (7.4.3)$$

So the approximation (7.4.1) reduces the entropy of the cluster method with sequence $\mathcal{S}$ to the mean-field (single-site) entropy.

As an example of this analysis consider the case of the Bethe-pair approximation for the Ising model discussed in Sect. 7.3. It is straightforward to show that substituting

$$\begin{aligned} \pi_2(+2) &= [\pi_1(+1)]^2 = \frac{1}{4}(1+m)^2, \\ \pi_2(0) &= \pi_1(+1) \pi_1(-1) = \frac{1}{4}(1-m^2), \\ \pi_2(-2) &= [\pi_1(-1)]^2 = \frac{1}{4}(1-m)^2, \end{aligned} \quad (7.4.4)$$

into (7.3.7) yields the mean-field free energy (6.1.9). However some questions remain. Whilst we do not expect the approximation (7.4.1) to give the equilibrium solution for the cluster approximation based on the sequence $\mathcal{S}$, we do expect it to satisfy the normalization and consistency conditions. Suppose, as we have indicated

in Sect. 7.2, that each microsystem on a lattice site can be in any one of v states. Then the single-site graph, with occupation k , is denoted by $\mathfrak{g}_1(k)$ with occupation number $\pi_1(k)$ and $\lambda_1(k) = 1$, for $k = 1, 2, \dots, v$. Let the graph $\mathfrak{g}_j$ have n_j sites. Its possible configurations are given by a set of numbers $\ell_j(k)$, $k = 1, 2, \dots, v$, where $\ell_j(k)$ denotes the number of sites with microsystem in state k , with

$$\sum_{k=1}^v \ell_j(k) = n_j. \quad (7.4.5)$$

Then

$$\lambda_j(\{\ell_j(k)\}) = \frac{n_j!}{\prod_{k=1}^v [\ell_j(k)]!}, \quad \vartheta_{1j}(k|\{\ell_j(k)\}) = \ell_j(k). \quad (7.4.6)$$

The mean-field assumption (7.4.1) now takes the form

$$\pi_j(\{\ell_j(k)\}) = \prod_{k=1}^v [\pi_1(k)]^{\ell_j(k)} \quad (7.4.7)$$

and it is straightforward to show that the normalization condition (7.2.2) and the consistency condition (7.2.5) (with $i = 1$ and $\Omega_{1j} = n_j$) are satisfied.

7.5 3-State Potts Model on a Triangular Lattice

Here it is convenient to use the Ising spin-1 variable $\sigma = -1, 0, 1$ to denote the three states on a lattice site. The Hamiltonian (3.6.53) is generalized so that the chemical potentials for states ± 1 are $\mu_{\pm}$ and there is a three-site energy $-R_3$ around every triangle, present only when the states of the sites are the same. We shall concentrate on the ‘ferromagnetic’ case $R \geq 0$, $R_3 \geq 0$.

Let the partially ordered sequence $\mathcal{S}$ be the section graphs $\mathfrak{g}_1$: a single site, $\mathfrak{g}_2$: a single edge and $\mathfrak{g}_3$: a single triangle. Then $\omega_1 = 1$, $\omega_2 = 3$ and $\omega_3 = 2$ and, as shown in Example 16.4.3, the weights are $\chi_1^{(3)} = 1$, $\chi_2^{(2)} = -3$ and $\chi_3^{(3)} = 2$. Labelling site, pair and triangle states in the obvious way⁸

$$\begin{aligned} \hat{h}_1(0) &= 0, & \hat{h}_1(\pm 1) &= -L_{\pm}, \\ \hat{h}_2(0, 0) &= -K, & \hat{h}_2(\pm 1, \pm 1) &= -2L_{\pm} - K, \end{aligned}$$

⁸ The ordering of the arguments in all the probabilities and variables defined here is arbitrary. In the interests of brevity a list of all the λ 's and ϑ 's has been omitted; it is easily calculated.

$$\begin{aligned}
\hat{h}_2(0, \pm 1) &= -L_{\pm 1}, & \hat{h}_2(\pm 1, \mp 1) &= -L_+ - L_-, \\
\hat{h}_3(0, 0, 0) &= -3K - K_3, & \hat{h}_3(\pm 1, \pm 1, \pm 1) &= -3L_{\pm} - 3K - K_3, \\
\hat{h}_3(0, \pm 1, \pm 1) &= -2L_{\pm 1} - K, & \hat{h}_3(\pm 1, \mp 1, \mp 1) &= -L_{\pm 1} - 2L_{\mp} - K, \\
\hat{h}_3(\pm 1, 0, 0) &= -L_{\pm} - K, & \hat{h}_3(0, +1, -1) &= -L_+ - L_-,
\end{aligned} \tag{7.5.1}$$

where $K := R/T$, $K_3 := R_3/T$ and $L := \mu/T$. In terms of the one- and two-site probabilities we now define the five variables

$$\begin{aligned}
\tau(\pm 1, \pm 1) &= \left[\frac{\pi_2(\pm 1, \pm 1)}{\pi_2(0, 0)} \right]^{\frac{1}{2}} \left[\frac{\pi_1(\pm 1)}{\pi_1(0)} \right]^{-\frac{1}{6}} \mathfrak{z}_{\pm}^{\frac{1}{6}}, \\
\tau(\pm 1, \mp 1) &= \left[\frac{\pi_2(\pm 1, \mp 1)}{\pi_2(0, 0)} \right]^{\frac{1}{2}} \\
&\quad \times \left[\frac{\pi_1(\pm 1)}{\pi_1(0)} \right]^{-\frac{1}{12}} \left[\frac{\pi_1(\mp 1)}{\pi_1(0)} \right]^{-\frac{1}{12}} (\mathfrak{z}_{\pm} \mathfrak{z}_{\mp})^{\frac{1}{12}}, \\
\tau(0, \pm 1) &= \left[\frac{\pi_2(0, \pm 1)}{\pi_2(0, 0)} \right]^{\frac{1}{2}} \left[\frac{\pi_1(\pm 1)}{\pi_1(0)} \right]^{-\frac{1}{12}} \mathfrak{z}_{\pm}^{\frac{1}{12}},
\end{aligned} \tag{7.5.2}$$

where

$$\mathfrak{z}_{\pm} := \exp(-L_{\pm}). \tag{7.5.3}$$

Then from the equilibrium conditions (7.2.30),

$$\begin{aligned}
\frac{\pi_3(\pm 1, \pm 1, \pm 1)}{\pi_3(0, 0, 0)} &= [\tau(\pm 1, \pm 1)]^3, \\
\frac{\pi_3(0, \pm 1, \pm 1)}{\pi_3(0, 0, 0)} &= [\tau(0, \pm 1)]^2 \tau(\pm 1, \pm 1) \mathfrak{x} \mathfrak{m}, \\
\frac{\pi_3(0, 0, \pm 1)}{\pi_3(0, 0, 0)} &= [\tau(0, \pm 1)]^2 \mathfrak{x} \mathfrak{m}, \\
\frac{\pi_3(\pm 1, \mp 1, \mp 1)}{\pi_3(0, 0, 0)} &= [\tau(\pm 1, \mp 1)]^2 \tau(\mp 1, \mp 1) \mathfrak{x} \mathfrak{m}, \\
\frac{\pi_3(-1, 0, +1)}{\pi_3(0, 0, 0)} &= \tau(-1, 0) \tau(0, +1) \tau(+1, -1) \mathfrak{x}^{\frac{3}{2}} \mathfrak{m},
\end{aligned} \tag{7.5.4}$$

where

$$\mathfrak{X} := \exp(-K), \quad \mathfrak{M} := \exp(-K_3). \quad (7.5.5)$$

Using the consistency conditions, five equations can be obtained relating the five τ 's. However, for simplicity we shall consider only the case where $\mathfrak{Z}_+ = \mathfrak{Z}_- = \mathfrak{Z}$ and the solution is in the subspace invariant under spin inversion. Now we have three variables

$$\begin{aligned} x &= \tau(0, +1) = \tau(0, -1), & y &= \tau(+1, -1), \\ z &= \tau(+1, +1) = \tau(-1, -1) \end{aligned} \quad (7.5.6)$$

and the 1–3 and 2–3 consistency conditions yield

$$\begin{aligned} z^3 \frac{f_2(x, y, z)}{f_1(x, y, z)} &= \mathfrak{Z} \left[\frac{g_2(x, y, z)}{g_1(x, y, z)} \right]^3, \\ \frac{f_2(x, y, z)}{f_1(x, y, z)} &= \mathfrak{Z}(\mathfrak{X}\mathfrak{M})^6 \left[\frac{g_3(x, y, z)}{g_1(x, y, z)} \right]^6, \\ y^3 \frac{f_2(x, y, z)}{f_1(x, y, z)} &= \mathfrak{Z}(\mathfrak{X}\mathfrak{M})^3 \left[\frac{g_2(x, y, z)}{g_1(x, y, z)} \right]^3, \end{aligned} \quad (7.5.7)$$

where

$$\begin{aligned} f_1(x, y, z) &:= 1 + 2x^2\mathfrak{X}\mathfrak{M}(2 + z + y\mathfrak{X}^{\frac{1}{2}}), \\ f_2(x, y, z) &:= z^3 + \mathfrak{X}\mathfrak{M}(x^2 + 2zx^2 + 3zy^2 + 2x^2y\mathfrak{X}^{\frac{1}{2}}), \\ g_1(x, y, z) &:= 1 + 2x^2\mathfrak{X}\mathfrak{M}, \\ g_2(x, y, z) &:= z^2 + \mathfrak{X}\mathfrak{M}(x^2 + zy^2), \\ g_3(x, y, z) &:= 1 + z + y\mathfrak{X}^{\frac{1}{2}}, \\ g_4(x, y, z) &:= 2yz + x^2\mathfrak{X}^{\frac{1}{2}}. \end{aligned} \quad (7.5.8)$$

It is not difficult to show that the Eq. (7.5.7) have the disordered solutions

$$x_{\text{D}}^{(\pm)} = y_{\text{D}}^{(\pm)} = \frac{\mathfrak{X}^{\frac{3}{2}}\mathfrak{M} \pm \sqrt{\mathfrak{X}^3\mathfrak{M}^2 - 8\mathfrak{X}\mathfrak{M}(1 - 2\mathfrak{X}\mathfrak{M})}}{4\mathfrak{X}\mathfrak{M}}, \quad z_{\text{D}} = 1. \quad (7.5.9)$$

At high temperatures these solutions are real with $x_{\text{D}}^{(-)} < 0$ when $T > (R + R_3)/\ln(2)$. Below a temperature given by $\mathfrak{X}^2\mathfrak{M} = 8(1 - 2\mathfrak{X}\mathfrak{M})$ the disordered

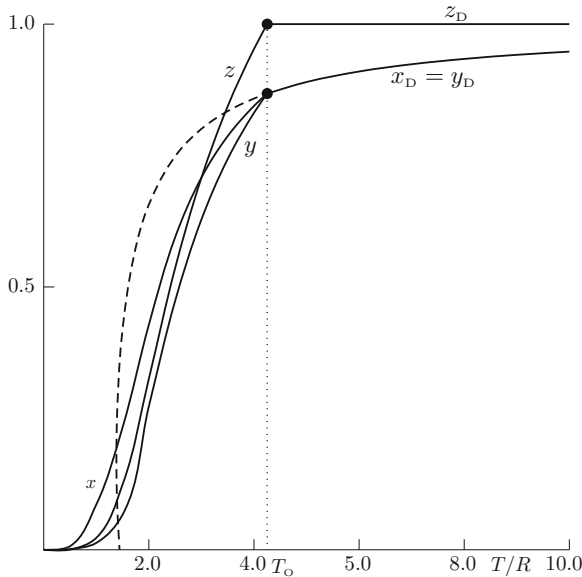


Fig. 7.1 The second-order transition in the 3-state Potts model for the case $R_3 = 0$. The continuation of the disordered solution $x_D = y_D$ below the transition is shown by a *broken line*

solutions no longer exist. A second-order transition to an ordered solution occurs at a point on the disordered solution curve, yielding a temperature T_o , which is a repeated root of the Eq. (7.5.7). Numerical calculations yield the results $T_o = 4.2520 R$, when $R_3 = 0$ and $T_o = 3.3761 R_3$, when $R = 0$ and graphs of the variables x , y and z for the case $R_3 = 0$ are shown in Fig. 7.1.

7.6 A Lattice Model for Fluid Water

The feature of the water molecule whose cooperative effect leads to the anomalous properties of liquid and supercritical water is the hydrogen bond.⁹ A number of statistical mechanical models, attempting to incorporate this feature have been developed. The simplest of these is the two-dimensional triangular lattice model of Bell and Lavis (1970) (see Sect. 3.7.2). In this section we consider a somewhat more realistic three-dimensional model, based on the body-centred cubic lattice (Fig. 16.8) and due to Bell (1972). This is a lattice fluid model of M molecules on $N(\geq M)$ sites with each molecule having four bonding arms pointing to the vertices of a tetrahedron. These arms are directed towards four of the eight first-neighbour sites and two arms have positive polarity and two negative, implying twelve distinguishable orientations

⁹ For an account of the role of hydrogen bonding in water see the footnote on p. 61.

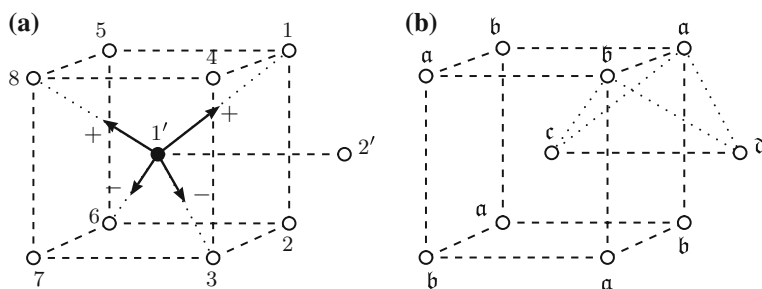


Fig. 7.2 **a** Sites on the body-centred cubic lattice with a water molecule placed on site 1'. (Reprinted from Bell (1972), by permission of the publisher IOP Publishing Ltd.). **b** The division of the body-centred cubic lattice into four sublattices

(Fig. 7.2a). Two first-neighbour molecules are connected by a hydrogen bond when a positive arm of one and a negative arm of the other lie along the segment connecting their sites. This occurs in 18 of the 144 possible orientational states of the two molecules.

The body-centred cubic lattice can be divided into two equivalent simple-cubic lattices. Let one be divided into two sublattices α and β as shown in Fig. 16.7 and the other into sublattices γ and δ, resulting in a division of the lattice into four equivalent sublattices (Fig. 7.2b). It can be seen that each of these sublattices forms a face-centred cubic lattice (Fig. 16.9) and the pairs α-γ, α-δ, β-γ, β-δ combine to form diamond lattices (Fig. 16.10). An open ice Ic structure, with one diamond sublattice of the body-centred cubic vacant and the other occupied by a bonded network, is thus possible. In the present fluid-state theory this arrangement occurs only as short-range order, but it is of crucial importance since it is the breakdown of the local open structure by increasing temperature which produces anomalous effects.¹⁰ Locally-ordered regions with the close-packed ice VII structure, with each diamond sublattice occupied by a fully-bonded structure, will exist at high pressures together with a range of other types of short-range order including the 'cage' structure in which unbonded molecules lie in the interstices of an open bonded network. However, the types of local order in real water are considerably more complex. The reduction of the ten or so observed ice structures to two indicates the approximations made in neglecting hydrogen bond distortion from the tetrahedral angles and the possibility of ordered hydrogen positions.

An interaction energy $-\varepsilon$ ($\varepsilon > 0$) is assumed for each unbonded first-neighbour pair of molecules which changes to $-\varepsilon - w$ ($w > 0$) when a hydrogen bond is formed. To ensure that the ice VII structure is stable only at high pressures, an energy penalty against close packing is needed. This is imposed by associating an energy $\frac{1}{3}u$ ($u > 0$) with the occupation by three molecules of a triad of sites where two second neighbours share a first neighbour (for example, the triads 121' or 1'2'1 in Fig. 7.2a).

¹⁰ For the nomenclature for the various phases of ice (I, Ic, II, etc.) see Eisenberg and Kauzmann (1969, Chap. 3). The difference between the locally similar ice Ic (diamond) and ice I (quartzite) structures is not likely to be significant in the fluid state.

The field conjugate to M is μ the chemical potential and to N is $-P$, where P is the pressure. In our calculations we choose μ to be the independent field, which is equivalent to using the grand-canonical distribution (Sect. 1.4.2) with free energy $-PN$. At zero temperature the free energy is equal to the enthalpy H and we compare its values in the open and closed-packed structures and also in an intermediate structure, which is similar to the open structure except that one of the remaining face-centred cubic sublattices is occupied by unbonded molecules. It is not difficult to show that

$$\begin{aligned} H_1(\text{open structure}) &= -N \left(\varepsilon + w + \frac{1}{2}\mu \right), \\ H_1(\text{intermediate structure}) &= -N \left(2\varepsilon + w - u + \frac{3}{4}\mu \right), \\ H_1(\text{close-packed structure}) &= -N (4\varepsilon + 2w - 4u + \mu). \end{aligned} \quad (7.6.1)$$

At high and low values of chemical potential the close-packed and open structures are respectively stable. By determining the boundaries of stability between the three phases it is easy to see that, if $2u > \varepsilon + w$, there will be a range of chemical potentials (and pressures) for which the intermediate phase is stable at $T = 0$. For simplicity we assume that this is not the case. Then there is a chemical potential μ_0 and pressure P_0 given by

$$\mu_0 = 2(4u - w - 3\varepsilon), \quad P_0 = 2(2u - \varepsilon), \quad (7.6.2)$$

such the open structure is stable at $T = 0$ for $\mu < \mu_0$ and the close-packed structure for $\mu > \mu_0$, with equivalent conditions applying to the pressure. If, on physical grounds, the pressure is restricted to positive values then, in order to ensure that both close-packed and open phases can be stable at T ,

$$P_0 > 0, \quad \text{which is equivalent to} \quad \mu_0 > -2(\varepsilon + w) \quad (7.6.3)$$

and thus

$$\varepsilon < 2u < \varepsilon + w. \quad (7.6.4)$$

To incorporate sublattice ordering we use a cluster approximation based on a tetrahedron of four sites, like $141'2'$ in Fig. 7.2a, one belonging to each of the four sublattices, as shown in Fig. 7.2b. The partially-ordered sequence of subgraphs $\mathfrak{g}_1, \mathfrak{g}_2, \dots, \mathfrak{g}_{15}$, where $\mathfrak{g}_{15}$ is the tetrahedron, are given in Example 16.4.7 and the parameters for the full KHDeB treatment are given in (16.4.51) and (16.4.52). However, this treatment is rather complicated and we shall present a version of the calculations of Bell (1972) using the Guggenheim and McGlashan (1951) method. For this the parameters $\tilde{\omega}_j^{(15)}, \tilde{\chi}_j^{(15)}$, $j = 1, 2, \dots, 15$ are given in (16.4.59). The states

of a single site ($g_1, \dots, g_4$) are denoted by an index $\ell = 0, 1$ for an unoccupied and an occupied site, respectively. Then¹¹

$$\begin{aligned} \lambda_j(0) &= 1, & \lambda_j(1) &= 12, \\ \hat{h}_j(0) &= 0, & \hat{h}_j(1) &= -\mu, \end{aligned} \quad j = 1, \dots, 4. \quad (7.6.5)$$

For the energy levels and parameters for g_{15} given in Table 7.1 we need to adjust the triad energy according to the prescription given by (16.4.51), (16.4.58) and (16.4.60). That is to say, $\frac{1}{3}u$ is replaced by u .

A number density ρ , equal to 1 for close packing, is

$$\rho := M/N, \quad (7.6.6)$$

in terms of which the single-site probabilities are given by

$$\pi_j(0) = 1 - \rho, \quad \pi_j(1) = \frac{1}{12}\rho, \quad j = 1, \dots, 4. \quad (7.6.7)$$

Using the parameters given here together with those contained in (16.4.59) we have, from (7.2.31), ten equilibrium conditions for $\pi_{15}(\ell)$, $\ell = 1, 2, \dots, 10$. With μ taken as the independent field, the free energy per lattice site is $-P$ and using the equilibrium condition for $\pi_{15}(4)$, from (7.2.31) and (7.2.32),

$$P = \frac{3}{2}T \left[\ln \left(\frac{\rho}{12} \right) + \ln(1 - \rho) \right] - T \ln[\pi_{15}(4)] - \varepsilon - w + \frac{1}{2}\mu. \quad (7.6.8)$$

A parameter τ is now defined by

$$\tau^2 := \frac{\pi_{15}(8)}{\pi_{15}(4)} = \exp \left(\frac{2\varepsilon - 4u + \mu}{2T} \right) \left[\frac{\rho}{12(1 - \rho)} \right]^{3/2}, \quad (7.6.9)$$

where the last expression follows from the equilibrium conditions. Reference to Table 7.1 shows that τ^2 expresses the ratio of ice VII-type to ice Ic-type configurations on the basic group of four sites. The equilibrium conditions can now be used to obtain all the ratios $\pi_{15}(\ell)/\pi_{15}(4)$ and substituting into the two 1–15 consistency conditions gives

$$\begin{aligned} \rho &= \mathfrak{X}^{-4} \mathfrak{M}^{-1} \tau^{-2} X(\tau) \pi_{15}(4), \\ 1 - \rho &= \mathfrak{X}^{-4} \mathfrak{M}^{-1} \tau^{-2} Y(\tau) \pi_{15}(4), \end{aligned} \quad (7.6.10)$$

¹¹ In this section it is convenient to use energies rather than couplings.

Table 7.1 Configurations and energy parameters for the tetrahedron $\mathfrak{g}_{15}$

Configuration index ℓ	Configuration	$\lambda_{15}(\ell)$	$\hat{h}_{15}(\ell)$
1		1	0
2		48	$-\mu$
3		288	-2μ
4		72	$-w - \varepsilon - 2\mu$
5		504	$-\varepsilon - 2\mu$
6		1728	$-w - 2\varepsilon + u - 3\mu$
7		5184	$-2\varepsilon + u - 3\mu$
8		648	$-2w - 4\varepsilon + 4u - 4\mu$
9		9072	$-w - 4\varepsilon + 4u - 4\mu$
10		11016	$-4\varepsilon + 4u - 4\mu$

where

$$\begin{aligned}
 X(\tau) := & 12\mathfrak{X}^3 \tau [\mathfrak{Y} \mathfrak{M}^3 + 3\mathfrak{X} \mathfrak{M} (1 + 7\mathfrak{Y}^2 + 4\mathfrak{X}^2 \mathfrak{Y}^2) \tau \\
 & + 108\mathfrak{X}^2 \mathfrak{Y} (1 + 3\mathfrak{Y}^2) \tau^2 + 54\mathfrak{X} \mathfrak{M} (1 + 14\mathfrak{Y}^2 + 17\mathfrak{Y}^4) \tau^3 \}, \quad (7.6.11)
 \end{aligned}$$

$$\begin{aligned}
 Y(\tau) := & \mathfrak{M}^5 + 36\mathfrak{X}^3 \mathfrak{Y} \mathfrak{M}^3 \tau + 36\mathfrak{X}^4 \mathfrak{M} (1 + 7\mathfrak{Y}^2 + 4\mathfrak{X}^2 \mathfrak{Y}^2) \tau^2 \\
 & + 432\mathfrak{X}^5 \mathfrak{Y} (1 + 3\mathfrak{Y}^2) \tau^3 \quad (7.6.12)
 \end{aligned}$$

and

$$\mathfrak{X} := \exp\left(-\frac{\varepsilon}{2T}\right), \quad \mathfrak{Y} := \exp\left(-\frac{w}{2T}\right), \quad \mathfrak{M} := \exp\left(-\frac{u}{T}\right). \quad (7.6.13)$$

Dividing the first relation of (7.6.10) by the second yields

$$\frac{\rho}{1-\rho} = \frac{X(\tau)}{Y(\tau)} \quad (7.6.14)$$

and, from (7.6.8) and (7.6.10),

$$\exp\left(\frac{P}{T}\right) = \frac{[X(\tau)]^3}{\mathfrak{M}^5 [X(\tau) + Y(\tau)]^2}. \quad (7.6.15)$$

Equations (7.6.14) and (7.6.15) constitute an implicit equation of state giving ρ in terms of P and T . Best results are found when u lies in the middle of the range given by (7.6.4). That is, when $2u = \frac{1}{2}w + \varepsilon$, which gives, from (7.6.2), $P_0 = w$. There is liquid–vapour phase separation, and density maxima occur on (ρ, T) isobars in the liquid state and in the supercritical state for $P < P_0$, as shown in Fig. 7.3. The temperature of maximum density varies very little with pressure. A compressibility minimum also occurs on isobars in the liquid state, and the mean number of first neighbours of a given molecule is between 4 and 5 at low pressures (Bell 1972). It can be seen that reasonable qualitative agreement with experiment is obtained in spite of the simplifications inherent in the model. The same model was used by Lavis and Christou (1977) to study dielectric properties and Wilson and Bell (1978) for an investigation of aqueous solutions. Bell and Salt (1976) used the mean-field approximation on a similar model to discuss transitions to ordered (ice-like) states.

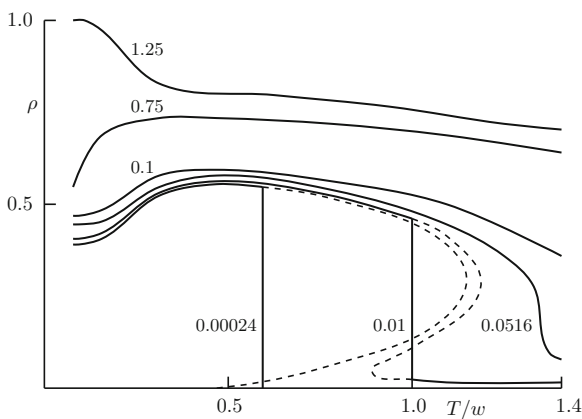


Fig. 7.3 Density–temperature isobars for $\varepsilon/w = 2$ and $u/w = 5/4$. Each isobar is labelled with the value of P/w . (Reprinted from Bell (1972), by permission of the publisher IOP Publishing Ltd.)

Weres and Rice (1972) used a model based on the body-centred cubic lattice, making rather different assumptions and were mainly concerned with local properties. These authors had reservations about Bell's triad term u which they considered made excessively large contributions to the energies of less dense structures. Other authors who used the body-centred cubic lattice with a different approach were O'Reilly (1973) and Fleming and Gibbs (1974) (see also the review by Stillinger 1975). A square lattice model with polarized bonds was discussed by Lavis and Christou (1979). Whitehouse et al. (1984) examined the Bell model using Monte Carlo methods and concluding that 'the cooperativity of the model is seriously underestimated by the first-order approximation, although the general behaviour of the model is correctly represented'. These authors consider that it would be more realistic to have $u = 0$, $\varepsilon < 0$, pointing out that the inequality (7.6.4) could still be satisfied and a positive P_0 obtained. Shinmi and Huckaby (1986) have used a similar model with non-polar bonding arms and repulsive bond energy to represent carbon tetrachloride.

Both the two-dimensional bonded lattice model of Bell and Lavis (1970) and the three-dimensional non-polar version of the Bell (1972) model, introduced by Shinmi and Huckaby (1986), can be expressed in Hamiltonian form as modified versions of the spin-1 Ising model. The modifications consist in the introduction of certain terms involving more than two sites. Based on this approach Southern and Lavis (1980) and Lavis and Southern (1984) used finite-size cluster real-space renormalization group methods to investigate the two models.

Part III

Exact Results

Introduction

For a one-component lattice fluid¹ on a finite lattice of N sites the grand partition function is an N -th degree polynomial in the fugacity $\mathfrak{z} := \exp(\mu/T)$. The only singularities of the free energy are the zeros of the partition function. However, none of these is real and positive since the coefficients of the partition-function polynomial are non-negative. Thus the lattice fluid on a finite lattice $\mathcal{N}$ can have no phase transitions. Since phase transitions do actually occur they must either be an artefact of the passage to the thermodynamic limit or be because the system is intrinsically infinite. The first of these possibilities is the one adopted in this book with the passage to the thermodynamic limit considered in detail in Sect. 8.1. However, it is of interest to consider briefly in Sect. 8.2 the formalism of statistical mechanics for infinite systems.

The *Peierls method* (Peierls 1936), which is described in Sect. 8.3, provides one of the few procedures for proving the existence of a phase transition without explicit calculation of the transition point. The original work of Peierls contains a proof that the Ising model on the square lattice has a spontaneous magnetic moment at sufficiently low temperatures. The argument was given a rigorous formulation by Griffiths (1964) and Dobrushin (1965). It was extended to lattice fluids with first-neighbour repulsion or exclusion by Dobrushin (1968a, b), for hypercubic lattices, and by Heilmann (1972), Abraham and Heilmann (1980) and Heilmann (1980) for triangular and face-centred cubic lattices. The Peierls method has been used to

¹ The equivalence between the spin- $\frac{1}{2}$ Ising model with first-neighbour interactions and a simple lattice fluid with first-neighbour pair interactions between particles was shown in Sect. 3.5.1.2 and, as we indicated there, this relationship holds between any one-component lattice fluid of particles of chemical potential μ and a spin- $\frac{1}{2}$ system on the same lattice in a magnetic field $\mathcal{H}$. It follows that the methods and results for the lattice fluid can be ‘translated’ into a spin- $\frac{1}{2}$ formulation and on that basis it is valid to describe the ideas developed in Chap. 8 in terms of a lattice fluid model.

established the existence of phase transitions at sufficiently low temperatures for the bonded lattice fluid models of Bell and Lavis (1970) and Bell (1972) by Heilmann and Huckaby (1979) and Lavis and Southern (1984) respectively.

The location and evolution of partition function zeros is clearly important for the understanding of the occurrence of phase transitions. This approach, which is the subject of Sect. 8.4, was initiated by the important insights contained in the papers of Yang and Lee, (Yang and Lee 1952; Lee and Yang 1952). Their approach is based on the idea of taking the fugacity $\mathfrak{z}$ to be a complex variable. They proved the well-known *circle theorem* that, for a spin- $\frac{1}{2}$ Ising ferromagnet with pairwise interactions of arbitrary range, the zeros of the grand partition function lie on the unit circle $|\mathfrak{z}| = 1$ in the complex plane of the variable $\mathfrak{z} := \exp(-2\mathcal{H}/T)$.² This result can be extended to a larger class of models (Asano 1968, 1970; Suzuki and Fisher 1971; Newman 1975; Lieb and Sokal 1981) but this does not include the v -state Potts model for $v > 2$ (Martin and Maillard 1986; O'Rourke et al. 1995). Lieb and Ruelle (1972) considered the distribution of zeros for the antiferromagnetic case and showed that, if the temperature is above the critical temperature, with an arc of the unit circle around $\mathfrak{z} = 1$ free of zeros, there will also be no zeros in the circle containing this arc and orthogonal to the unit circle. Heilmann and Lieb (1972) were able to derive the circle theorem for both Ising and Heisenberg models from an investigation of the partition function zeros of a monomer–dimer system in the monomer fugacity plane. Nishimori and Griffiths (1983) have investigated the motion of the zeros in the complex plane as the temperature is changed. Ruelle (1971) used the contraction method of Asano (1970) to prove a result which provides, in principle, a method for obtaining regions containing the grand partition function zeros of one-component lattice fluids with finite-range interactions whether or not they lie on $|\mathfrak{z}| = 1$.

The Yang–Lee circle theorem and Ruelle's theorem are both about the complex plane of the variable $\mathfrak{z}$, whether it is understood as the Boltzmann factor of the magnetic field $\mathcal{H}$ or of the chemical potential μ . The important point about $\mathfrak{z}$ is that it is the Boltzmann factor for a field associated with single-site occupations. Any partition function of a lattice system can, however, be expressed as a sum of powers of a number of Boltzmann factors, some of which will arise from particle interactions. These quantities can also, in the spirit of Yang and Lee, be treated as complex variables. Thus, for example, for the spin- $\frac{1}{2}$ ferromagnetic Ising model with Hamiltonian (Sect. 3.5.1) on the finite lattice $\mathcal{N}$ the grand partition function is a polynomial in $\mathfrak{X} := \exp(-2J/T)$ and $\mathfrak{z} := \exp(-2\mathcal{H}/T)$. As indicated above the distribution of zeros in the complex plane of $\mathfrak{z}$ satisfies the circle theorem. The distribution of zeros in the complex plane of $\mathfrak{X}$ has also been investigated (Fisher 1965; Jones 1966; Brascamp and Kunz 1974) and it has been shown that in the limit

² There is, of course, an equivalent lattice fluid formulation of this result (Sect. 8.4.2), but since it relies on the Ising model hole–particle symmetry condition $E(\mathcal{M}) = E(\mathcal{N} - \mathcal{M})$, where $\mathcal{N} - \mathcal{M}$ is the complement of $\mathcal{M}$ on the lattice $\mathcal{N}$, it has the rather artificial property that two empty sites interact with the same energy as two occupied sites.

of an infinite system, when $\mathcal{H} = 0$, the asymptotic locus of zeros is given by the *Fisher circles*

$$\mathfrak{X} = \sqrt{2} \exp(i\theta) \pm 1.$$

For the square-lattice Ising model on a semi-infinite lattice Wood (1985) showed that the Fisher circles could be computed as curves on which transfer matrix eigenvalues are degenerate in modulus and thus that these latter yield the asymptotic locus of partition function zeros.

As we saw in Sect. 3.6.4, the 2-state Potts model is isomorphic to the zero-field spin- $\frac{1}{2}$ Ising model with $R = 2J$. The Fisher circle with the plus sign has been generalized to the two-dimensional v -state Potts model (Kenna 1998) to show that, in the complex plane of $\mathfrak{X} := \exp(-R/T)$, the asymptotic locus of zeros lies on the Fisher circle

$$\mathfrak{X} = G(v; \theta) := \frac{\sqrt{v} \exp(i\theta) - 1}{v - 1}.$$

The treatment of duality for the v -state Potts model in Sect. 9.2.1 yields the self-dual circle

$$\mathfrak{X} = F(v; \theta) := 1 + \sqrt{v} \exp(i\theta).$$

Taking into account the fact that the variable $\mathfrak{X}$ used here is the reciprocal of that used in Sect. 9.2.1, it is not difficult to show that $\mathfrak{X} = G(v; \theta)$ is simply the self-dual circle $\mathfrak{X} = 1/F(v; \varphi)$, where $\exp(i\varphi) = [\sqrt{v} - \exp(i\theta)]/[\sqrt{v} \exp(i\theta) - 1]$. The ‘extra’ Fisher circle with the minus sign is due to the ferromagnetic–antiferromagnetic symmetry present in the spin- $\frac{1}{2}$ Ising model but not in the Potts model when $v \neq 2$.

One of the most useful techniques in statistical mechanics is the formulation of a transformation leading to an equivalence between two different models. We have already seen in Chap. 3 that such equivalences exist between different interpretations of the Ising model (ferromagnetic, antiferromagnetic, lattice gas and binary mixture) and between Ising, Potts and vertex models. Underlying the transformation methods described in Chaps. 9 and 10 is a *lattice transformation* of the type described in Sect. 16.2.2. Briefly summarized, for a given lattice the first step is to add a lattice site to some or all of either its faces or edges. The second step is then to either remove the original lattice, leaving only the new sites, or to consider the ‘decorated’ lattice of new and old sites. The Wegner transformation, incorporating the duality and weak-graph transformations and treated in Sect. 9.2, is a case where sites are added to the faces of the lattice and the original lattice is removed. The star-triangle transformation of Sect. 9.4 is an example of *face decoration*, where in this case sites are added to half the faces of a triangular lattice with the originally lattice retained to give a honeycomb lattice. Chapter 10 considers various models obtained by edge-decorating an Ising model. The aim for all these transformation methods is to obtain relationships between the couplings for the systems on the original and transformed lattices which yield a relationship between their partition functions. This means that if one model is solved then so is the other.

Given a finite lattice $\mathcal{N}$ of dimension d the thermodynamic limit in which $\mathcal{N}$ tends to the infinite lattice $\mathcal{L}$ is normally understood to be a growth in lattice size in

all d lattice dimensions (Sect. 8.1) and the critical exponents of a phase transition in the system will be functions of d (Sects. 4.5.6 and 4.7.2). In Sect. 4.10 we considered the modified situation where the thermodynamic limit was taken in only $\mathfrak{d} < d$ dimensions. The resulting lattice $\mathcal{L}_{\mathfrak{d}}$ then has the critical properties of a $\mathfrak{d}$ -dimensional system, with no non-analytic behaviour associated with phase transitions unless $\mathfrak{d} > 1$.

Transfer matrices in statistical mechanics are used to grow a finite lattice step-by-step to the infinite limit in one lattice dimension. In principle, this method can be applied to a lattice of any dimension and it has been successfully applied to some three-dimensional systems (Huang et al. 1997; Popkov and Nienhuis 1997). However, most applications have been to one- and two-dimensional lattices and in Chaps. 11 and 12 attention is restricted to these cases.

The first use of transfer matrices is due to Kramers and Wannier (1941a, b). As a preparation for a study of the two-dimensional square-lattice Ising model they first re-derived in transfer matrix form the solution of the one-dimensional Ising model, which was first obtained by Ising (1925) by combinatorial means. They then proceeded to investigate lattices with increasing values of N_2 , the number of lines of sites. Although, of course, these models effectively represent increasingly complicated one-dimensional models without true phase transitions, they were able to show that an increasingly steep maximum in the heat capacity gave progressively more accurate estimates of the critical temperature of the two-dimensional system, which they also derived using a duality transformation. The use of systems with finite values of N_2 and *incipient phase transitions* to estimate the critical parameters of truly two-dimensionally infinite systems has proved a valuable tool in statistical mechanics (see, for example, Lavis 1976) and, after an introduction to transfer matrices in Sect. 11.1 this method is developed and used in Sect. 11.2.

An interesting approach due to the late David Wood and his coworkers (Wood 1987; Wood et al. 1987; Wood and Turnbull 1988; Wood et al. 1989; Wood and Ball 1990) combines the investigation of partition function zeros contained in the work of Yang and Lee and Ruelle with an exploitation of transfer matrices. This method, described in Sect. 8.4, is able, by tracing the evolution of partition function zeros with increasing N_2 to obtain estimates for critical points and exponents.

In the context of transfer matrices the ‘solution’ of a two-dimensional model means the determination of the partition function in the limiting case where $N_2 \rightarrow \infty$. Thus the problem is to find some way of handling the limit where the dimension of the transfer matrix becomes infinite. This must involve a calculation of some, or all, of the transfer matrix eigenvalues and ideally also the eigenvectors, as functions of N_2 .

Although there are now many different published solutions of the two-dimensional Ising model,³ the first solution of the square-lattice case by Onsager (1944) was an important landmark in the history of statistical mechanics. By applying

³ To emphasize this Baxter and Enting (1978) called their method (of which an account is also given in Lavis and Bell 1999a, Chap. 8) the ‘339th’ presentation.

group representation theory he was able to determine the eigenvalues of the transfer matrix as functions of N_2 . Subsequent simplification was achieved by Kaufman (1949) who showed that the transfer matrix belongs to the group of spinor operators, with Kaufman and Onsager (1949) deriving the pair correlation function. Although the work of Onsager and Kaufman is restricted to the square lattice, their work was soon generalized to the triangular and honeycomb lattices (Houtappel 1950; Husimi and Syôzi 1950; Newell 1950; Syôzi 1950; Temperley 1950; Wannier 1950).

A general introduction to vertex models has been given in Sect. 3.8. As was noted there a major breakthrough for vertex models was the derivation of the entropy of the ice-rule (six-vertex) model by Lieb (1967a, b), followed by solutions for the F model (Lieb 1967c) and the KDP model (Lieb 1967d).⁴ Unlike Onsager's solution of the zero-field Ising model, with which they are comparable in importance, Lieb's solutions of the six-vertex model include an external field. They also proved fruitful in opening up the search for exact solutions of other models particularly the eight-vertex model which was solved by Baxter (1972b, 1973a, b, c). In Chap. 12 we find it convenient to model our presentation on the account by Baxter (1978) of the eight-vertex model on a rapidity lattice (Sect. 16.3). This yields the *checkerboard Ising model* of Baxter (1986) and the six-vertex model as special cases, although for the latter this approach has the disadvantage that it cannot include an external field.⁵

Section 3.9 contains a brief account of dimer models. In Chap. 13 a method is described for evaluating exactly the dimer partition function. This was first presented for the square-lattice case by Temperley and Fisher (1961), Fisher (1961) and Kasteleyn (1961), but it can be applied to any regular two-dimensional lattice. Vertex and Ising model equivalences to the dimer model are derived in Sect. 13.3. It is of particular interest to note that, although the work of Chap. 12 is restricted to the square-lattice eight-vertex model, the dimer equivalence of Sect. 13.3.2 allows the eight-vertex free-energy densities for the honeycomb and triangular lattice models to be derived.

⁴ Elliott Lieb (private communication) was alerted to the problem of calculating the entropy of the ice-rule model by reading the paper of Nagle (1966a) in which he used a series method to obtain an approximate value of $\ln(1.543)$ for the entropy per site; this being compared to the exact result of $\frac{3}{2}\ln(\frac{4}{3}) \simeq \ln(1.5396)$, subsequently obtained by Lieb (Sect. 12.3.10.4).

⁵ For this the reader is referred to the original papers or to Lieb and Wu (1972).

Chapter 8

Algebraic Methods

8.1 The Thermodynamic Limit

For a one-component lattice fluid on a finite lattice $\mathcal{N}$ the state of the system can be described by specifying the subset $\mathcal{M} \subseteq \mathcal{N}$ of sites occupied by particles, and the Hamiltonian in the *grand-canonical distribution* (Sects. 1.4.2 and 2.1) is

$$\hat{H}(\mathcal{N}, \mu; \mathcal{M}) := E(\mathcal{M}) - \mu|\mathcal{M}|, \quad (8.1.1)$$

$E(\mathcal{M})$ being the interaction energy between the particles of $\mathcal{M}$. The equilibrium (Gibbs) distribution is

$$p(\mathcal{N}, T, \mu; \mathcal{M}) = \frac{\exp[-\hat{H}(\mathcal{N}, \mu; \mathcal{M})/T]}{Z_1(\mathcal{N}, T, \mu)}, \quad (8.1.2)$$

and

$$Z_1(\mathcal{N}, T, \mu) := \sum_{\{\mathcal{M}\}} \exp[-\hat{H}(\mathcal{N}, \mu; \mathcal{M})/T] \quad (8.1.3)$$

is the grand partition function, with the pressure and density given by

$$P(\mathcal{N}, T, \mu) = \frac{T}{N} \ln Z_1(\mathcal{N}, T, \mu) \quad (8.1.4)$$

and

$$\rho = \frac{1}{N} \sum_{\{\mathcal{M}\}} |\mathcal{M}| p(\mathcal{N}, T, \mu; \mathcal{M}) = \frac{\partial P}{\partial \mu}, \quad (8.1.5)$$

respectively. If the number M of particles on $\mathcal{N}$ were fixed, the distribution would be the *canonical distribution* with independent extensive variables $N := |\mathcal{N}|$ and M , and partition function

$$Z_0(\mathcal{N}, M, T) := \sum_{\{|\mathcal{M}|=M\}} \exp[-E(\mathcal{M})/T], \quad (8.1.6)$$

where the sum is over all subsets of $\mathcal{N}$ with M sites. The grand partition function is now given by

$$Z_1(\mathcal{N}, T, \mathfrak{z}) = \sum_{\hat{M}=0}^N Z_0(\mathcal{N}, \hat{M}, T) \mathfrak{z}^{\hat{M}}, \quad (8.1.7)$$

where $\mathfrak{z} := \exp(\mu/T)$ is the *fugacity* or *activity*.

We consider the thermodynamic limit of this fluid for an infinite sequence of lattices $\mathcal{N}$ growing so that $|\mathcal{N}| \rightarrow \infty$ and $\mathcal{N}$ tends to an infinite lattice $\mathcal{L}$. To achieve a sensible outcome in this procedure it is necessary that

$$\lim_{\mathcal{N} \rightarrow \mathcal{L}} P(\mathcal{N}, T, \mathfrak{z}) = P(\mathcal{L}, T, \mathfrak{z}), \quad (8.1.8)$$

where $P(\mathcal{L}, T, \mathfrak{z})$ is the finite pressure in the infinite lattice $\mathcal{L}$. Yang and Lee (1952) established the existence of the thermodynamic limit for the grand-canonical distribution of a one-component fluid of particles with hard cores and finite-range interactions in a volume V . This work was extended by Witten (1954), who relaxed the condition of hard cores. The corresponding theorem for lattice systems is given by Ruelle (1969) (see also Griffiths 1972). Three ingredients are necessary for the existence of the thermodynamic limit:

- (i) translational invariance of the interparticle interactions,
- (ii) some restriction on the range of the interactions,
- (iii) a limitation on the proportional number of boundary sites as the lattice grows.

These conditions can be stated more precisely in the following way:

Given any subset $\mathcal{X}$ of the lattice sites, let $\mathcal{E}(\mathcal{X})$ be the contribution to the energy which would arise if every site of $\mathcal{X}$ were occupied, but which would be absent if one or more of its sites were empty. This means that $\mathcal{E}(\mathcal{X})$ cannot be decomposed in a sum of contributions from proper subsets of $\mathcal{X}$. If $|\mathcal{X}| = X > 1$ and $\mathcal{E}(\mathcal{X}) \neq 0$, then $\mathcal{X}$ is called an *interacting set*. The system is *translationally invariant* if, for all $\mathcal{X}$,

$$\mathcal{E}(\mathcal{X}) \text{ is dependent only on the geometric form of } \mathcal{X} \text{ and not on its location on the lattice.} \quad (8.1.9)$$

Now choose a site $\mathbf{r}_0$ belonging to every lattice $\mathcal{N}$ in the sequence and let $\{\mathcal{X}\}_0$ be all the subsets, of all possible $\mathcal{N}$, which contain $\mathbf{r}_0$. The necessary condition on the range of interactions is that

$$\sum_{\{\mathcal{X}\}_0} \frac{|\mathcal{E}(\mathcal{X})|}{|\mathcal{X}|} < \infty. \quad (8.1.10)$$

Finally suppose that $\mathcal{N}_\ell$ is the set of sites of $\mathcal{N}$ which are within some fixed distance ℓ of the boundary. A limiting procedure $\mathcal{N} \rightarrow \mathcal{L}$ is said to be taken *in the sense of van Hove* (Ruelle 1969, p. 14) if

$$\lim_{\mathcal{N} \rightarrow \mathcal{L}} \frac{|\mathcal{N}_\ell|}{|\mathcal{N}|} = 0, \quad \forall \ell. \quad (8.1.11)$$

The following theorem can now be proved (ibid p. 22):

Theorem 8.1.1 *The limit (8.1.8) exists for $\beta > 0$ with $P(\mathcal{L}, T, \beta)$ a continuous non-decreasing function of β if (8.1.9)–(8.1.11) are satisfied.*

Because of the equivalence between the one-component lattice fluid and spin- $\frac{1}{2}$ lattice models this result can be translated into an equivalent result for the latter, establishing the continuity and monotonicity of the free-energy density as a function of the appropriate variable.

8.2 The Infinite-System Approach

Here, instead of considering a sequence of finite lattices, a system on the infinite lattice $\mathcal{L}$ is adopted at the start, eliminating the need to consider boundary effects.¹ In such a system equilibrium can no longer be described using the Gibbs distribution (2.1.2). There are a number of alternatives available in the literature (Penrose 1979, Sect. 2.2). We shall give a brief summary of the first steps of one of these, based on the work of Dobrushin (1968a) and Lanford and Ruelle (1969), in which the Gibbs distribution is replaced by the more general *Gibbs measure*.

The validity of the distribution (2.1.2), defined for $\sigma \in \Sigma$, is dependent on its numerator being summable (or for a continuous system integrable) over Σ to give the partition function. This breaks down for an infinite system, although as we saw in Sect. 8.1 we can under certain specific conditions obtain convergence for the partition function for a sequence of increasing large systems. The aim here is to replace $p(\sigma)$ by a probability (that is to say, normalized) measure $m(\sigma)$ on Σ which coincides with $p(\sigma)$ for a finite system but retains its meaning for an infinite system.²

To be specific consider the lattice fluid of Sect. 8.1 with the set $\mathcal{M}$ of occupied sites replaced by the equivalent designation of σ standing for the set of states $\sigma(\mathbf{r}) = 0, 1$,

¹ Our discussion will be concerned solely with classical systems. However, it should be noted that another advantage of using the infinite-system approach for quantum systems is that the quasi-periodicity of finite quantum systems arising from the discrete spectrum of the Hamiltonian is absent in the infinite system.

² For convenience we drop reference to η and the independent intensive variables.

for all $\mathbf{r} \in \mathcal{N}$, where 0 and 1 denote respectively the presence or absence of a fluid particle. Then $|\mathcal{N}| = N$, $|\Sigma| = 2^N$ and we choose a pair-interaction of arbitrary range depending on the distance between the pair of sites. Thus

$$E(\varepsilon; \sigma) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}} \varepsilon(|\mathbf{r} - \mathbf{r}'|) \sigma(\mathbf{r}) \sigma(\mathbf{r}'). \quad (8.2.1)$$

This is the simple lattice fluid of Sect. 3.5.1.2, except that the interaction is not restricted to first neighbours. Let $\mathcal{N}_A (\neq \emptyset) \subset \mathcal{N}$ with $\mathcal{N}_B := \mathcal{N} \setminus \mathcal{N}_A \neq \emptyset$. The configurations σ_A and σ_B of spins on $\mathcal{N}_A$ and $\mathcal{N}_B$ can be regarded as events occurring on the respective sublattices with $\sigma_A \wedge \sigma_B$ an event on $\mathcal{N}$. The probability that the event σ_A occurs, given that σ_B has occurred is

$$\begin{aligned} m(\sigma_A | \sigma_B) &= \frac{m(\sigma_A \wedge \sigma_B)}{m(\sigma_B)} \\ &= \frac{\exp[-\hat{H}(\sigma_A | \sigma_B)/T]}{\sum_{\{\sigma_A\}} \exp[-\hat{H}(\sigma_A | \sigma_B)/T]} = \frac{\exp[-\hat{H}_A(\sigma_A | \sigma_B)/T]}{Z_A(\sigma_B)}, \end{aligned} \quad (8.2.2)$$

where

$$\begin{aligned} \hat{H}_A(\sigma_A | \sigma_B) &:= - \sum_{\{\mathbf{r}, \mathbf{r}' \in \mathcal{N}_A\}} \varepsilon(|\mathbf{r} - \mathbf{r}'|) \sigma(\mathbf{r}) \sigma(\mathbf{r}') \\ &\quad - \sum_{\{\mathbf{r} \in \mathcal{N}_A\}} \sigma(\mathbf{r}) \left[\mu + \sum_{\{\mathbf{r}' \in \mathcal{N}_B\}} \varepsilon(|\mathbf{r} - \mathbf{r}'|) \sigma(\mathbf{r}') \right], \end{aligned} \quad (8.2.3)$$

$$Z_A(\sigma_B) := \sum_{\{\sigma_A\}} \exp[-\hat{H}_A(\sigma_A | \sigma_B)/T]. \quad (8.2.4)$$

The Hamiltonian $\hat{H}_A(\sigma_A | \sigma_B)$ is that of the system on $\mathcal{N}_A$ with ‘boundary conditions’ given by the event σ_B . In fact with any finite-range condition $\varepsilon(|\mathbf{r} - \mathbf{r}'|) = 0$, if $|\mathbf{r} - \mathbf{r}'| > R$, imposed on the interaction, the boundary conditions will be specified by only part of σ_B ; so it will remain valid when $\mathcal{N}_B$ and hence $\mathcal{N}$ becomes infinite.

This example encapsulates the essential feature of this approach, which is that, given a system on a infinite lattice $\mathcal{L}$, equilibrium is defined locally in terms of a measure on finite subsets of $\mathcal{L}$, which is conditional on boundary conditions provided by the state on the lattice exterior to the sets in question. To generalize:

Definition 8.2.1 Let $\Sigma := \mathfrak{E}^{\mathcal{S}}$ be a product space where $\mathcal{S}$ is countably infinite and $\mathfrak{E}$ is any measurable space. Then m is a *Gibbs measure* if, for every finite subset $\mathcal{A}$ of $\mathcal{S}$ and m -almost every configuration σ_B outside $\mathcal{A}$, the conditional distribution of the configuration in $\mathcal{A}$ given σ_B is a Gibbs distribution (2.1.2).

The ‘ m -almost’ caveat in the definition makes provision for exceptional boundary conditions, but apart from that it will be seen that, for any $\mathcal{A}$, the Gibbs measure corresponds to a family $m(\cdot|\sigma_b)$ of conditional Gibbs distributions.

With the interpretation that $\mathfrak{E} := \{0, 1\}$, the two-point space of states at a single site, $\mathcal{S} := \mathcal{L}$ the infinite lattice and $\mathcal{A} := \mathcal{N}_\Lambda$ we have, by construction, obtained the Gibbs measure for the lattice fluid. The distribution (8.2.2) is called the *Dobrushin–Lanford–Ruelle (DLR) equilibrium condition* and a distribution on $\mathcal{L}$ which satisfies it is said to be in a *Gibbs state*. For the consequences of this development particularly in relation to the treatment of phase transitions and long-range correlation functions the reader is referred to Georgii (1988), Sewell (1989) and Simon (1993), which are concerned wholly or partly with this approach.

8.3 Lower Bounds for Phase Transitions: The Peierls Method

Phase transitions in statistical mechanical models correspond to the onset of some kind of order. Except in exceptional circumstances,³ immediately below the phase transition this ordering is imperfect, with the perfectly-ordered state occurring only at zero temperature. An ordered phase can thus be understood by reference to its perfect realization at zero temperature and the types of ordered phases can in turn be understood by defining the concept of a *ground state*⁴ and, as indicated in Sect. 4.2.1, the *order parameter* of the phase is the expectation value of a quantity which measures the extent to which the site occupations mirror those of the ground state. For the Ising ferromagnet on the square lattice, with Hamiltonian (3.5.1), the magnetization density is

$$m(\mathcal{N}, K, L) = 2 \left[\frac{\langle \widehat{N}(\uparrow) \rangle}{N} - \frac{1}{2} \right], \quad (8.3.1)$$

where $K := J/T$, $L := \mathcal{H}/T$ and $\widehat{N}(\uparrow)$ is the number from the N sites of $\mathcal{N}$ which are in the spin-up state. The order parameter for the all spin-up ferromagnetic phase in zero field on the infinite lattice $\mathcal{L}$ is

$$m(\mathcal{L}, K, 0) = \lim_{L \rightarrow 0+} \lim_{\mathcal{N} \rightarrow \mathcal{L}} m(\mathcal{N}, K, L). \quad (8.3.2)$$

³ Of which the transition to the frozen ferroelectric state in the six-vertex model is an example (Sect. 3.8.2).

⁴ We have already made informal use of the idea of a ground state, beginning with the antiferromagnet in Sect. 3.5.1.1.

Let

$$m_{\uparrow}(\mathcal{N}, K, L) = 2 \left[\frac{\langle \hat{N}(\uparrow) \rangle_{\uparrow}}{N} - \frac{1}{2} \right], \quad (8.3.3)$$

where $\langle \cdot \rangle_{\uparrow}$ denotes the constrained expectation value with all the boundary sites containing up spins. There are two steps in the Peierls argument for the square-lattice Ising ferromagnet (Peierls 1936):

(a) To show that, if there exists an $\alpha > 0$, independent of N , such that

$$m_{\uparrow}(\mathcal{N}, K, 0) \geq \alpha, \quad \forall K > K' \text{ and } \mathcal{N}, \quad (8.3.4)$$

then

$$m(\mathcal{L}, K, 0) \geq \alpha, \quad \forall K > K'. \quad (8.3.5)$$

(b) To establish the existence of α and K' .

There are many presentations of details of this argument (Thompson 1988). Rather than giving another we shall adopt the more general approach of Heilmann and Huckaby (1979) and concentrate on the derivation of a lower temperature bound for the existence of a phase transition. Although Heilmann and Huckaby (1979) present their development for some bonded lattice fluid models it is not in essence specific to those models and can be applied to any one-component lattice fluid satisfying conditions (8.1.9)–(8.1.11).

Suppose that $E(\mathcal{M})$, defined in Sect. 8.1 is, for all $\mathcal{M}$, a linear function of parameters $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n_e}$. Then the Hamiltonian is a linear function of this set and the chemical potential μ . A *ground state* G is defined to be a configuration of the system for which the set $\mathcal{G} \subseteq \mathcal{N}$ of sites is occupied by particles and for which:

- (i) $\mathcal{G}$ is a periodic arrangement of particles on the lattice.
- (ii) There is a set of values of $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n_e}, \mu$ for which

$$\hat{H}(\mathcal{G}) \leq \hat{H}(\mathcal{M}), \quad \forall \mathcal{M} \subseteq \mathcal{N}, \quad (8.3.6)$$

with equality holding, if at all, only when $\mathcal{M}$ is also periodic and thus itself a ground state.

G is said to be a *stable* ground state in the region of parameter space in which it satisfies (8.3.6). The number $\omega(G)$ of ground states which are stable in the region in which G is stable is called the *degeneracy* of G .

Consider now a ground state G_0 . For any occupation $\mathcal{M}$, a site of the lattice is said to *belong to* G_0 if it is occupied in the way it would be if the system were in the ground state G_0 . For any occupation of the lattice let $N_0(\mathcal{M})$ be the number of sites which belong to G_0 and

$$\rho_0(\mathcal{N}, K_1, \dots, K_{n_e}, L) := \frac{\langle N_0(\mathcal{M}) \rangle_0}{N} - \frac{1}{\omega(G_0)}, \quad (8.3.7)$$

where now $K_\ell := \varepsilon_\ell/T$, $L := \mu/T$ and $\langle \cdot \rangle_0$ denotes the constrained expectation value taken over all $\mathcal{M}$ for which the boundary sites belong to G_0 . When the parameters $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n_e}, \mu$ take values for which G_0 is a stable ground state the limit $\mathcal{N} \rightarrow \mathcal{L}$ of $\rho_0(\mathcal{N}, K_1, \dots, K_{n_e}, L)$ is the order parameter for G_0 . The replacement for (8.3.4) is then

$$\rho_0(\mathcal{N}, K_1, \dots, K_{n_e}, L) \geq \alpha > 0, \quad (8.3.8)$$

(Heilmann and Huckaby 1979; Dobrushin 1968a, b). It is clear that (8.3.8) could never be satisfied if $\omega(G_0) = 1$; non-degenerate ground states cannot lead to ordered phases. This would allow us, if we chose, to normalize $\rho_0(\mathcal{N}, K_1, \dots, K_{n_e}, L)$ at $T = 0$ by dividing by $1 - 1/\omega(G_0)$. It is in fact this normalized quantity which is equivalent to $m_\uparrow(\mathcal{N}, K, 0)$ in the zero-field Ising ferromagnet. To obtain a useful lower bound on phase transitions we must now use (8.3.8) to derive an inequality involving the temperature.

Let the lattice $\mathcal{N}$ be divided into *units* $\mathcal{C}_i$, $i = 1, 2, \dots, \eta$ which consist of the smallest subsets of the sites with the following properties:

- (a) All the units have the same number of sites $C = |\mathcal{C}_i|$ and the same geometrical structure. That is, they can be transformed into each other by rotations and translations.
- (b) For any ground state and for some labelling of the sites of the units each unit is identically occupied both with respect to the number of particles and the labelling.
- (c) Every interacting set is contained within at least one unit.

It is not difficult to see that it is always possible to divide the lattice into units in the way described here as long as the boundaries of the lattice are constructed in a suitable way. For each unit $\mathcal{C}_i$,

$$\mathcal{C}_i := \left\{ \mathbf{r} : \sum_{j=1}^C \lambda_j \mathbf{r}_j^{(i)} = \mathbf{r}, \forall \lambda_j \geq 0 \text{ such that } \sum_{j=1}^C \lambda_j = 1 \right\}, \quad (8.3.9)$$

is a *cell*, where $\mathbf{r}_1^{(i)}, \mathbf{r}_2^{(i)}, \dots, \mathbf{r}_C^{(i)}$ are the sites of the unit $\mathcal{C}_i$, which are called the *vertices* of $\mathcal{C}_i$. The cells of the lattice are closed convex sets of points in the space occupied by the lattice and spanned by their vertices.⁵ In a d -dimensional lattice the dimension $d_{\mathcal{C}}$ of the cells is less than or equal to d . The elements of the surfaces of the cells spanned by sets of $d_{\mathcal{C}} - 1$ vertices are called *faces*.⁶ We shall confine

⁵ If the points of the unit are linearly independent, as is the case for the models treated by Heilmann and Huckaby (1979), then (8.3.9) is the usual definition of a *simplex* (Munkres 1984).

⁶ For a simplex, all such linear combinations will give points on the surface, but this is not generally true for the definition of a cell given here.

our attention to the case where the only overlaps between cells are the points on the faces and where each cell face lies in two cells which will be called *neighbours*. A simply-connected set of cells will be called a *cluster*. The *surface* of a cluster consists of those faces of the cells of the cluster which belong to only one cell of the cluster. The *boundary* of the cluster consists of those cells which have faces on the surface. An *interior site* of a cluster is one which does not lie on its surface. An *interior cell* of the cluster is one which is not part of the boundary. Suppose that all sites of the lattice not lying on the boundary belong to $q_0 \geq 2$ cells. Then a cluster with an interior site must contain at least q_0 cells.

Each cell of the lattice can be occupied in 2^C ways denote by $\mathcal{O}_k, k = 1, 2, \dots, 2^C$. Among the particle arrangements when each cell is identically occupied will be, by definition, all the ground states of the system. However, it is not necessarily the case that all such configurations are ground states, since there may be some corresponding to particle occupations which do not satisfy the minimization condition (8.3.6) for any set of values of the energy parameters. Any particle occupation $\mathcal{M}$ of the lattice implies the occupation $\mathcal{C}_i \cap \mathcal{M}$ for the cell $\mathcal{C}_i$. Each $\mathcal{C}_i \cap \mathcal{M}$ is one of the occupations $\mathcal{O}_k$ and if, for occupation $\mathcal{M}$, $v_k(\mathcal{M})$ of the cells have occupation $\mathcal{O}_k$ the Hamiltonian can be expressed in the form

$$\hat{H}(\mathcal{M}) = \sum_{i=1}^{\eta} \hat{H}_{\mathcal{C}_i}(\mathcal{C}_i \cap \mathcal{M}) = \sum_{k=1}^{2^C} v_k(\mathcal{M}) \hat{H}_{\mathcal{C}}(\mathcal{O}_k), \quad (8.3.10)$$

where $\hat{H}_{\mathcal{C}}(\cdot)$ is the contribution to the Hamiltonian from a single cell $\mathcal{C}$. This contribution will contain all factors of the form $\mathcal{E}(\mathcal{X})$, for every fully occupied interacting set $\mathcal{X} \subseteq \mathcal{C}$. If $\mathcal{X}$ lies in more than one unit the contribution is divided equally between that subset of units. For any occupation $\mathcal{M}$ in which each of the cells is identically occupied each of the terms in the first sum in (8.3.10) will be equal and the second sum will have only one term. For the ground state $\mathcal{G}_0$ in which each cell has occupation $\mathcal{C} \cap \mathcal{G}_0$

$$\hat{H}(\mathcal{G}_0) = \eta \hat{H}_{\mathcal{C}}(\mathcal{C} \cap \mathcal{G}_0). \quad (8.3.11)$$

A face of a cell is said to be $\mathcal{G}_0$ -ordered (or more briefly *ordered*, since we are concentrating on a fixed ground state) if all its sites belong to $\mathcal{G}_0$. Otherwise it is said to be *disordered*. Given any occupation $\mathcal{M} \neq \mathcal{G}_0$, two cells of the lattice are said to be *linked* if they share a disordered face. Given the specific occupation $\mathcal{M}$ of the lattice a cell $\mathcal{C}_i$ with some occupation $\mathcal{O}_k \neq \mathcal{C}_i \cap \mathcal{G}_0$ is called a *contour segment*. The boundary of a cluster is called a *contour* if:

- (i) Each of its cells is a contour segment. (Some, but not necessarily all, of the interior cells of the cluster will also be contour segments.)
- (ii) The cluster contains at least one interior site not belonging to $\mathcal{G}_0$. (A contour must, therefore, consist of at least q_0 cells.)

- (iii) The cells of the contour are not linked to any cell not in the cluster. (The surface of the cluster must consist of ordered faces.)
- (iv) The cluster cannot be decomposed into two disjoint clusters, each bounded by a contour.

We denote a particular contour by $[q; j]$, where q is the number of segments in the contour and $j = 1, 2, \dots, m(q)$ will be the variable which differentiates between different contours which can be formed on the lattice with q segments. Let $n_j(q)$ be the number of interior sites of the cluster bounded by the contour $[q; j]$.

Suppose that the occupation $\mathcal{M}$ has all the boundary sites belonging to G_0 . We are interested in obtaining an upper bound on the number of sites of $\mathcal{M}$ which do not belong to G_0 . This will clearly be given by the total number of interior sites contained within the contours present in $\mathcal{M}$. Thus

$$N - N_0(\mathcal{M}) \leq \sum_{\{q\}} \sum_{j=1}^{m(q)} n_j(q) \delta[q; j], \quad (8.3.12)$$

where the outer summation is over all possible contour sizes and

$$\delta[q; j] := \begin{cases} 1, & \text{if } [q; j] \text{ is present in } \mathcal{M}, \\ 0, & \text{otherwise.} \end{cases} \quad (8.3.13)$$

Let $\{\mathcal{M}\}_0$ be the set of lattice occupations for which the boundary sites of $\mathcal{M}$ belong to G_0 and let $\{\mathcal{M}\}_0^{[q; j]}$ be the set which satisfy the additional requirement that the contour $[q; j]$ is present. Then

$$\langle \delta[q; j] \rangle_0 = \frac{\sum_{\{\mathcal{M}\}_0^{[q; j]} \exp[-\hat{H}(\mathcal{M})/T]}{\sum_{\{\mathcal{M}\}_0 \exp[-\hat{H}(\mathcal{M})/T]}. \quad (8.3.14)$$

For every $\mathcal{M} \in \{\mathcal{M}\}_0^{[q; j]}$ there is an $\mathcal{M}^* \in \{\mathcal{M}\}_0$, but $\notin \{\mathcal{M}\}_0^{[q; j]}$, which differs from $\mathcal{M}$ only in that all the interior points of $[q; j]$ not belonging to G_0 have had their occupations changed to put them into G_0 . We denote this new set of occupations by $\{\mathcal{M}^*\}_0^{[q; j]}$. Thus

$$\langle \delta[q; j] \rangle_0 < \frac{\sum_{\{\mathcal{M}\}_0^{[q; j]} \exp[-\hat{H}(\mathcal{M})/T]}{\sum_{\{\mathcal{M}^*\}_0^{[q; j]} \exp[-\hat{H}(\mathcal{M}^*)/T]}$$

$$= \exp \left\{ - \sum_{k=1}^{2^c} \theta_k[q; j] [\widehat{H}_{\mathcal{C}}(\mathcal{O}_k) - \widehat{H}_{\mathcal{C}}(\mathcal{C} \cap \mathcal{G}_0)]/T \right\}, \quad (8.3.15)$$

where $\theta_k[q; j]$ is the number of cells with occupation $\mathcal{O}_k$ in contour $[q; j]$. To find an upper bound to the final term in (8.3.15) it is necessary to replace $\widehat{H}_{\mathcal{C}}(\mathcal{O}_k)$ by its least possible value for each cell of the cluster bounded by $[q; j]$. For the interior cells of the cluster it is possible for $\widehat{H}_{\mathcal{C}}(\mathcal{O}_k) = \widehat{H}_{\mathcal{C}}(\mathcal{C} \cap \mathcal{G}_0)$, either because $\mathcal{O}_k = \mathcal{C} \cap \mathcal{G}_0$ or because $\mathcal{O}_k$ is the occupation appropriate to some other ground-state degenerate with $\mathcal{G}_0$. The cells of the contour itself, however, are contour segments with occupations different from $\mathcal{C} \cap \mathcal{G}_0$. They must also be different from the cell occupation of some other ground state degenerate with $\mathcal{G}_0$.⁷ It follows that

$$\langle \delta[q; j] \rangle_0 < \exp[-\varepsilon_0 q/T], \quad (8.3.16)$$

where

$$\varepsilon_0 := \min_{\widehat{H}_{\mathcal{C}}(\mathcal{O}_k) \neq \widehat{H}_{\mathcal{C}}(\mathcal{C} \cap \mathcal{G}_0)} \{ \widehat{H}_{\mathcal{C}}(\mathcal{O}_k) \} - \widehat{H}_{\mathcal{C}}(\mathcal{C} \cap \mathcal{G}_0). \quad (8.3.17)$$

From (8.3.7), (8.3.12) and (8.3.16),

$$1 - \sum_{\{q\}} Q(q) \exp(-\varepsilon_0 q/T) - \frac{1}{\omega(\mathcal{G}_0)} < \rho_0(\mathcal{N}, K_1, \dots, K_{n_e}, L), \quad (8.3.18)$$

where $Q(q)$ is some function of q such that

$$\frac{1}{N} \sum_{j=1}^{m(q)} n_j(q) \leq Q(q). \quad (8.3.19)$$

If such a $Q(q)$ can be found such that the series in (8.3.18) is convergent for $T \in [0, \varepsilon_0 t(\mathcal{G}_0)]$, where $t(\mathcal{G}_0)$ is the least positive solution of the equation

$$\sum_{\{q\}} Q(q) \exp \left[-\frac{q}{t(\mathcal{G}_0)} \right] = 1 - \frac{1}{\omega(\mathcal{G}_0)}, \quad (8.3.20)$$

then (8.3.8) is satisfied if

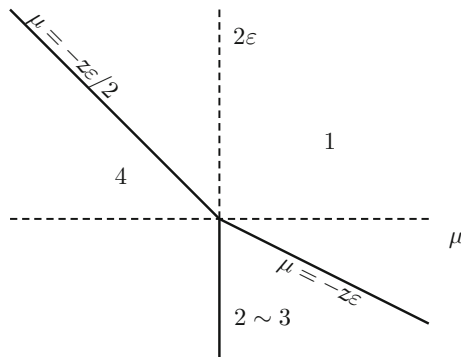
$$T < \varepsilon_0 t(\mathcal{G}_0), \quad (8.3.21)$$

and, in the thermodynamic limit, the system, with parameter values appropriate to the ground state $\mathcal{G}_0$, will be in the corresponding ordered state.

⁷ If this were not the case it would be possible to ‘mix’ these two occupations to achieve a ground-state degeneracy of $O(N)$.

Table 8.1 Cell occupations for first-neighbour pairs for the simple lattice fluid

ι	Occupation	$\hat{H}_{\varepsilon}(\mathcal{C} \cap \mathcal{G}_{\iota})$
1	●—●	$-\varepsilon - 2\mu/z$
2	●—○	$-\mu/z$
3	○—●	$-\mu/z$
4	○—○	0

**Fig. 8.1** The ground-state diagram for the simple lattice fluid. Regions of stability are labelled by ι , $\iota = 1, \dots, 4$. The region $2 \sim 3$ is that in which the degenerate ground states 2 and 3 are both stable

8.3.1 The Simple Lattice Fluid

As shown in Sect. 3.5.1.2, the Hamiltonian (3.5.1) for the spin- $\frac{1}{2}$ Ising model can be transformed⁸ into the form (3.5.20), for the simple lattice fluid where the lattice-fluid variables ε and μ are related to the Ising variables J and $\mathcal{H}$ by (3.5.22). The cells of the lattice can be taken to be the first-neighbour pairs, of which there are the four possible occupations listed in Table 8.1. They are denoted by $\mathcal{C} \cap \mathcal{G}_{\iota}$, $\iota = 1, \dots, 4$, since for each there is a region of parameter space for which it corresponds to a stable ground state. These regions are shown in Fig. 8.1. In the region $2 \sim 3$ the two ground states 2 and 3 are degenerate so it is possible for an ordered phase to occur. Otherwise ordered states can occur only on the phase boundaries. In particular on the boundary $\mu = -\frac{1}{2}z\varepsilon$, $\varepsilon > 0$ the ground states 1 and 4 are degenerate.⁹ The method described above is now applied to the model along this phase boundary to investigate the existence of the ordered phase corresponding to G_1 . From (8.3.17)

⁸ Apart from a constant C , given by (3.5.21), which we can for the present purposes neglect.

⁹ This line corresponds to the zero-field axis $\mathcal{H} = 0$ in the spin representation and the model is equivalent to the zero-field Ising ferromagnet.

$\varepsilon_0 = -\mu/z = \varepsilon/2$ and, of course, $\omega(G_1) = 2$, $q_0 = z$. It is now necessary to determine a quantity $Q(q)$ satisfying (8.3.19). This will depend on the lattice structure under consideration.

Example 8.3.1 (The Linear Lattice) Although it is well-known that there is no phase transition in this case (Sect. 11.2) it is of interest to see how the theory can be applied. Consider a linear lattice of N sites with the two boundary sites occupied by particles. The only possible contours are ones for which $q = 2$. Such contours can bound clusters with n interior sites for $n = 1, 2, \dots, N - 2$ and a cluster with n interior sites can be placed on the lattice in $N - n - 1$ ways. So (8.3.19) is satisfied by $Q(2) = \frac{1}{6}N^2$ and, substituting into (8.3.20), gives $t(G_1) = 1/\ln(N/\sqrt{3})$. Since $t(G_1) \rightarrow 0$ as $N \rightarrow \infty$ the method, as is to be expected, fails to establish the existence of an ordered phase.

Example 8.3.2 (The Square Lattice) For a particular occupation, with all the boundary sites of the lattice occupied by particles, a contour can be represented by drawing a line through the centre of each boundary segment (each first-neighbour pair with occupation 2 or 3), between the centres of the two lattice squares of which the segment is a side. It is clear that an even number of lines will meet at the centre of each square and that the lines form closed polygons. Whenever four lines meet at the centre of a square the lines are cut to disconnect the polygon into two simply-connected parts. The effect of this procedure is to enclose all the interior sites of the clusters bounded by the contours within polygons (Fig. 8.2). Now a contour of q segments is equivalent to a polygon consisting of q line elements drawn between the centres of lattice squares and the possible values of q are 4, 6, 8, $\dots$. A polygon encloses the maximum number of sites when it is a square. It follows that $n_j(q) \leq (q/4)^2$. We now need an upper bound on the total number $m(q)$ of closed polygons of q line elements. Ignoring the boundaries of the lattice we chose the centre of one of the N squares and begin to draw a polygon. The first line element can be chosen in four ways and each subsequent line element in three ways.

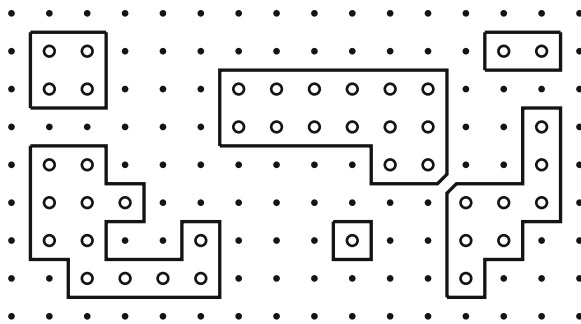


Fig. 8.2 The equivalence between contours and polygons on the square lattice. The maximum number of vacant sites corresponds to the case, as shown here, where all the sites enclosed by the polygons are empty

There are, therefore, $4N3^{q-1}$ ways of completing this construction. Among these will be all closed polygons, including those which cross themselves, and all open arcs. Each simply-connected closed polygon will have been counted q times and thus $m(q) < 4N3^{q-1}/q$. It follows that

$$\frac{1}{N} \sum_{j=1}^{m(q)} n_j(q) < \frac{1}{4} q 3^{q-1} = Q(q). \quad (8.3.22)$$

Substituting into (8.3.20) gives

$$x^3 + x^2 - 6x + 3 = 0, \quad (8.3.23)$$

as long as $x = 9 \exp[-2/t(G_1)] < 1$. This equation has the single root $x = 0.5936$ in $(0, 1)$ and, from (8.3.21), this establishes the existence of an ordered phase when $T < 0.3678\varepsilon$. This can be compared with the exact result for the critical temperature given, from (9.2.36), by $T_c = \varepsilon/[2 \ln(1 + \sqrt{2})] = 0.5673\varepsilon$. The difference between these two values illustrates the point that the importance of the Peierls method consists not in its ability to yield good values for critical points, for which series methods are much better (Chap. 14), but rather in its ability to establish in a rigorous way the existence of an ordered phase.

8.4 Grand Partition Function Zeros and Phase Transitions

Suppose that $\mathbb{O}_{\mathcal{N}}(T) := \{\mathfrak{z}_1, \mathfrak{z}_2, \dots, \mathfrak{z}_N\}$ is the set of zeros of the grand partition function (8.1.7) in the complex plane $\mathbb{C}_3$. The connection between the roots of the grand partition function and phase transitions in the system is established by the following theorem, due to Yang and Lee (1952), for which we shall provide a proof:

Theorem 8.4.1 *Let $\mathbb{W}(T)$ be a region in $\mathbb{C}_3$ which contains a segment of the positive real axis $\mathbb{R}^{(+)}$, but none of the members of $\mathbb{O}_{\mathcal{N}}(T)$, for any of the sequence of lattices involved in the limiting process described in Sect. 8.1. Then for all $\mathfrak{z} \in \mathbb{W}(T)$ the limit (8.1.8) is uniform and $P(\mathcal{L}, T, \mathfrak{z})$ is an analytic function¹⁰ of $\mathfrak{z}$ in $\mathbb{W}(T)$.*

Proof Let $\mathbb{V}(T) \subseteq \mathbb{W}(T)$ be a closed disc of finite radius with its centre on $\mathbb{R}^{(+)}$. We first prove the theorem for $\mathbb{V}(T)$ using the *Vitali convergence theorem*.¹¹ Consider the sequence of functions $P(\mathcal{N}, T, \mathfrak{z})$ of $\mathfrak{z}$ given by the set of lattices leading to the

¹⁰ See the footnote on p. 161 for the usage of the term ‘analytic’ for complex-valued functions.

¹¹ The statement of the theorem is as follows (for a proof see Titchmarsh 1939): Let $f_n(\mathfrak{z})$ be a sequence of functions which are analytic in a region $\mathbb{D} \subset \mathbb{C}_3$ and have the following properties: (i) $|f_n(\mathfrak{z})| \leq B$ for all n and all $\mathfrak{z} \in \mathbb{D}$. (ii) $f_n(\mathfrak{z})$ tends to a limit as $n \rightarrow \infty$ at a set of points having a limit point inside $\mathbb{D}$. Then $f_n(\mathfrak{z})$ tends uniformly to a limit in any region bounded by a contour interior to $\mathbb{D}$, the limit being an analytic function of $\mathfrak{z}$.

limit in Theorem 8.1.1. These are analytic functions in $\mathbb{V}(T)$ for the reasons given above. According to Theorem 8.1.1 they tend to a limit on $\mathbb{V}(T) \cap \mathbb{R}^{(+)}$. Now let $\mathfrak{z}^*$ be the right-hand extremity of the diameter of $\mathbb{V}(T)$. The sequence of finite real numbers $P(\mathcal{N}, T, \mathfrak{z}^*)$ tends to the limit $P(\mathcal{L}, T, \mathfrak{z}^*)$. It is therefore bounded above by some number B and since $|\mathfrak{z}| \leq \mathfrak{z}^*$ on $\mathbb{V}(T)$ it follows that $P(\mathcal{N}, T, \mathfrak{z}) \leq B$ for all lattices $\mathcal{N}$ in the sequence and all $\mathfrak{z} \in \mathbb{V}(T)$. This establishes the theorem for $\mathbb{V}(T)$. Now take any disc $\mathbb{V}'(T) \subseteq \mathbb{W}(T)$ with centre in $\mathbb{V}(T)$. The conditions of the Vitali convergence theorem now hold in $\mathbb{V}'(T)$ with $\mathfrak{z}^*$ replaced by the real number equal to the maximum modulus of $\mathfrak{z}$ in $\mathbb{V}'(T)$. By repeating the process with a sequence of discs the whole of $\mathbb{W}(T)$ can be covered. $\square$

Suppose that $\mathbb{Q}(T)$ is a closed region in $\mathbb{C}_3$ which contains the set $\mathbb{O}_{\mathcal{N}}(T)$ of zeros of $\Xi(\mathcal{N}, T, \mathfrak{z})$ for all of the sequence of finite lattices. It follows from Theorem 8.4.1 that $P(\mathcal{L}, T, \mathfrak{z})$ is an analytic function of $\mathfrak{z}$ in any $\mathbb{W}(T) \subseteq \mathbb{C}_3$ for which $\mathbb{Q}(T) \cap \mathbb{W}(T) = \emptyset$. Any phase transition of the system at temperature T must therefore occur at a chemical potential μ for which $\exp(\mu/T) \in \mathbb{Q}(T) \cap \mathbb{R}^{(+)}$.

We have now defined the main problem in this approach to the theory of phase transitions, namely, for any particular model leading to a set of canonical partition functions $Z_0(\mathcal{N}, T, M)$, to determine the set $\mathbb{Q}(T)$. This set will not, of course, be unique and different choices may have different advantages according to circumstances. Under rather general conditions a lattice fluid will not have any phase transitions at sufficiently high temperatures. If a set $\mathbb{Q}(T)$ can be found which evolves on to $\mathbb{R}^{(+)}$ as late as possible as the temperature is decreased then a good upper bound on T at any fixed μ will be obtained. On the other hand, if a $\mathbb{Q}(T)$ can be found, for which the range of μ lying in $\mathbb{Q}(T) \cap \mathbb{R}^{(+)}$ is as narrow as possible, then good bounds of μ at constant T will be obtained. These differing aims in the construction of $\mathbb{Q}(T)$ will not, as we shall see below, necessarily lead to the same set, although we can always form the intersection of all known sets of type $\mathbb{Q}(T)$ in order to improve the bounds.

There is one class of one-component lattice fluids for which it is not difficult to obtain some information about $\mathbb{Q}(T) \cap \mathbb{R}^{(+)}$ without any calculation. If

$$Z_0(\mathcal{N}, N - M, T) = Z_0(\mathcal{N}, M, T), \quad M = 1, 2, \dots, N - 1, \quad (8.4.1)$$

then, if $\mathfrak{z}_0 \in \mathbb{O}_{\mathcal{N}}(T)$, so does its conjugate complex $\bar{\mathfrak{z}}_0$ and $1/\mathfrak{z}_0$ and $1/\bar{\mathfrak{z}}_0$. This set of roots is invariant under inversion and conjugation and it follows that $\mathbb{O}_{\mathcal{N}}(T)$ must also have this property. Thus, if $\mathbb{O}_{\mathcal{N}}(T)$ is exterior to the disc $|\mathfrak{z}| < R < 1$ it is contained in the annulus $R \leq |\mathfrak{z}| \leq 1/R$. If condition (8.4.1) is replaced by

$$Z_0(\mathcal{N}, N - M, T) = Z_0(\mathcal{N}, M, T)[Z_0(\mathcal{N}, N, T)]^{1 - \frac{2M}{N}}, \quad M = 1, 2, \dots, N - 1, \quad (8.4.2)$$

it can be shown that the grand partition function can be symmetrized by a change of variable. If the lattice fluid is expressed in terms of Ising spin variables then condition (8.4.2) is equivalent to the restriction that the only odd degree terms in the Hamiltonian are linear. This condition is satisfied for a lattice fluid with only pairwise

interactions satisfying conditions (8.1.9) and (8.1.10). If condition (8.4.1) holds then, for there to be a phase transition at temperature T , the condition $1 \in \mathbb{Q}(T) \cap \mathbb{R}^{(+)}$ must be satisfied. In this case the region of phase transitions can be reduced to the largest closed interval $\mathbb{I}(T) \subset \mathbb{Q}(T) \cap \mathbb{R}^{(+)}$ which maps into itself under inversion. If (8.4.1) holds and the interactions are pairwise attractive the circle theorem of Lee and Yang (1952) proves that the set $\mathbb{Q}_{\mathcal{N}}(T)$ lies on the unit circle $|\mathfrak{z}| = 1$. This result is a special case of the theorem of Ruelle (1971), which will now be discussed.

8.4.1 Ruelle's Theorem

Given the set of variables $\{\mathfrak{z}(\mathbf{r})\}$ defined on the sites of $\mathcal{N}$, let

$$\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r})\}) := \sum_{\{\mathcal{M}\}} \left\{ \exp[-E(\mathcal{M})/T] \prod_{\{\mathbf{r} \in \mathcal{M}\}} \mathfrak{z}(\mathbf{r}) \right\}. \quad (8.4.3)$$

It is clear, from (8.1.6) and (8.1.7), that

$$\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r}) = \mathfrak{z}\}) = Z_1(\mathcal{N}, T, \mathfrak{z}). \quad (8.4.4)$$

Suppose now that a set of closed subsets $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$ of $\mathbb{C}_{\mathfrak{z}}$ can be found such that

$$\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r})\}) \neq 0, \quad \text{if } \mathfrak{z}(\mathbf{r}) \notin \mathbb{S}_{\mathcal{N}}(T; \mathbf{r}), \quad \forall \mathbf{r} \in \mathcal{N}. \quad (8.4.5)$$

The zeros of $Z_1(\mathcal{N}, T, \mathfrak{z})$ must lie in the closed set

$$\mathbb{U}_{\mathcal{N}}(T) := \bigcup_{\{\mathbf{r} \in \mathcal{N}\}} \mathbb{S}_{\mathcal{N}}(T; \mathbf{r}) \quad (8.4.6)$$

and a possible set $\mathbb{Q}(T)$ would be one which satisfied the condition that

$$\mathbb{U}_{\mathcal{N}}(T) \subseteq \mathbb{Q}(T), \quad \forall \mathcal{N} \text{ in the sequence } \mathcal{N} \rightarrow \mathcal{L}. \quad (8.4.7)$$

In the cases discussed in Sects. 8.4.2 and 8.4.3, below, the determination of $\mathbb{Q}(T)$ from $\mathbb{U}_{\mathcal{N}}(T)$ is trivial, since the application of Ruelle's theorem leads to a set $\mathbb{U}_{\mathcal{N}}(T)$, which is independent of $\mathcal{N}$.

We now define a *covering* $\mathcal{Y}_{\mathcal{N}}$ of the lattice $\mathcal{N}$ to be the set $\mathcal{Y}_{\kappa}$, $\kappa = 1, 2, \dots, \nu$ of subsets of $\mathcal{N}$ such that:

- (i) Every site of $\mathcal{N}$ belongs to at least one member of $\mathcal{Y}_{\mathcal{N}}$.
- (ii) Every interacting set is contained within at least one member of $\mathcal{Y}_{\mathcal{N}}$.

It is clear that the set of units, defined in Sect. 8.3, is one possible covering of the lattice, but for present purposes it is convenient to be less specific. For any occupation $\mathcal{M}$ of the lattice we wish to be able to write

$$E(\mathcal{M}) = \sum_{\kappa=1}^{\nu} E_{\kappa}(\mathcal{Y}_{\kappa} \cap \mathcal{M}), \quad (8.4.8)$$

where $E_{\kappa}(\mathcal{Y}_{\kappa} \cap \mathcal{M})$ is the contribution to the energy from $\mathcal{Y}_{\kappa}$. To do this we must avoid multiple counting of energies from interacting sets of particles which occur in more than one $\mathcal{Y}_{\kappa}$. One way to do this is to divide the energy of such a set equally between all $\mathcal{Y}_{\kappa}$ in which it is contained. Another way would be to define the covering so that every interacting set is a member of $\mathcal{Y}_{\mathcal{N}}$ and, in the energy $E_{\kappa}(\mathcal{Y}_{\kappa} \cap \mathcal{M})$, to exclude contributions from proper subsets of $\mathcal{Y}_{\kappa}$. For each $\mathcal{Y}_{\kappa} \in \mathcal{Y}_{\mathcal{N}}$, let

$$\xi(\mathcal{Y}_{\kappa}, T, \{\mathfrak{z}_{\kappa}(\mathbf{r})\}) := \sum_{\{\mathcal{X} \subseteq \mathcal{Y}_{\kappa}\}} \left\{ \exp[-E_{\kappa}(\mathcal{X})/T] \prod_{\{\mathbf{r} \in \mathcal{X}\}} \mathfrak{z}_{\kappa}(\mathbf{r}) \right\}, \quad (8.4.9)$$

where $\{\mathfrak{z}_{\kappa}(\mathbf{r})\}$ is a set of variables defined for $\mathbf{r} \in \mathcal{Y}_{\kappa}$. For later reference we note, from (8.4.8), that, when the members of the covering are disjoint sets,

$$\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r})\}) = \prod_{\kappa=1}^{\nu} \xi(\mathcal{Y}_{\kappa}, T, \{\mathfrak{z}_{\kappa}(\mathbf{r}) = \mathfrak{z}(\mathbf{r})\}). \quad (8.4.10)$$

Ruelle's theorem can now be expressed in the following way:

Theorem 8.4.2 *If, for all $\mathcal{Y}_{\kappa} \in \mathcal{Y}_{\mathcal{N}}$ and all $\mathbf{r} \in \mathcal{Y}_{\kappa}$, there exists a closed set $\mathbb{X}_{\kappa}(T; \mathbf{r}) \subset \mathbb{C}_{\mathfrak{z}}$ such that $0 \notin \mathbb{X}_{\kappa}(T; \mathbf{r})$ and $\xi(\mathcal{Y}_{\kappa}, T, \{\mathfrak{z}_{\kappa}(\mathbf{r})\}) \neq 0$ when $\mathfrak{z}_{\kappa}(\mathbf{r}) \notin \mathbb{X}_{\kappa}(T; \mathbf{r})$, for all $\mathbf{r} \in \mathcal{Y}_{\kappa}$, then $\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r})\}) \neq 0$ when¹²*

$$\mathfrak{z}(\mathbf{r}) \notin \mathbb{S}_{\mathcal{N}}(T; \mathbf{r}) = - \prod_{\{\kappa\}} [-\mathbb{X}_{\kappa}(T; \mathbf{r})], \quad \forall \mathbf{r} \in \mathcal{N}, \quad (8.4.11)$$

where the product in (8.4.11) is over those values of κ for which $\mathbf{r} \in \mathcal{Y}_{\kappa}$.

Proof An important step in the proof of this theorem is the derivation of a formula for $\Xi(\mathcal{N}, T, \{\mathfrak{z}(\mathbf{r})\})$ in terms of the set $\xi(\mathcal{Y}_{\kappa}, T, \{\mathfrak{z}_{\kappa}(\mathbf{r})\})$, $\kappa = 1, 2, \dots, \nu$. When the covering $\mathcal{Y}_{\mathcal{N}}$ consists of disjoint sets, this formula is (8.4.10) and the theorem is trivial. This is because each lattice site belongs to exactly one member of the covering, the product in (8.4.11) has one term with $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r}) = \mathbb{X}_{\kappa}(T; \mathbf{r})$ and $\mathfrak{z}(\mathbf{r}) = \mathfrak{z}_{\kappa}(\mathbf{r})$. The result follows from the assumption $\xi(\mathcal{Y}_{\kappa}, T, \{\mathfrak{z}_{\kappa}(\mathbf{r})\}) \neq 0$ and (8.4.10).

When the covering contains overlapping sets (8.4.10) is no longer valid. Its replaced is given by the *contraction procedure* of Asano (1970):

Lemma 8.4.1 *Let*

$$\phi_{\alpha} = \phi_{\alpha-1} \circ \xi(\mathcal{Y}_{\alpha}, T, \{\mathfrak{z}_{\alpha}(\mathbf{r})\}), \quad \alpha = 2, 3, \dots, \nu, \quad (8.4.12)$$

¹² Notation: For subsets $\mathbb{A}$ and $\mathbb{B}$ of $\mathbb{C}_{\mathfrak{z}}$, $-\mathbb{A}$ is the set of all points $-\mathfrak{z}_A$, for $\mathfrak{z}_A \in \mathbb{A}$, and $\mathbb{A}\mathbb{B}$ is the set of all points $\mathfrak{z}_A \mathfrak{z}_B$, for $\mathfrak{z}_A \in \mathbb{A}$ and $\mathfrak{z}_B \in \mathbb{B}$.

where

$$\phi_1 = \xi(\mathcal{Y}_1, T, \{\mathfrak{Z}_1(\mathbf{r})\}), \quad (8.4.13)$$

and the contraction symbol $\circ$ means that, when the product is taken, terms containing factors of the form $\prod_{[\kappa]} \mathfrak{Z}_\kappa(\mathbf{r})$ are deleted unless the product contains all $\mathfrak{Z}_\kappa(\mathbf{r})$ for which $\mathbf{r} \in \mathcal{Y}_\kappa$ for $\kappa = 1, 2, \dots, \alpha$. Then with the substitution

$$\mathfrak{Z}(\mathbf{r}) = \prod_{\{\kappa\}} \mathfrak{Z}_\kappa(\mathbf{r}), \quad (8.4.14)$$

$$\Xi(\mathcal{N}, T, \{\mathfrak{Z}(\mathbf{r})\}) = \phi_v. \quad (8.4.15)$$

Proof (of Lemma) From (8.4.9), each of the terms in ϕ_v is of the form

$$\prod_{\kappa=1}^v \left\{ \exp[-E_\kappa(\mathcal{X}_\kappa)/T] \prod_{\{\mathbf{r} \in \mathcal{X}_\kappa\}} \mathfrak{Z}_\kappa(\mathbf{r}) \right\}, \quad (8.4.16)$$

where $\mathcal{X}_\kappa \subseteq \mathcal{Y}_\kappa$. The sum is over certain subsets $\mathcal{X}_\kappa$ of the covering sets $\mathcal{Y}_\kappa$ with a constraint imposed by the contraction procedure of (8.4.12). This can be expressed by the condition that if, for some subsets $\mathcal{Y}_i, \mathcal{Y}_j, \mathcal{Y}_k, \dots$ of the covering $\mathcal{Y}_{\mathcal{N}}$,

$$\mathcal{Y}_i \cap \mathcal{Y}_j \cap \mathcal{Y}_k \cap \dots \neq \emptyset, \quad (8.4.17)$$

then

$$\mathcal{Y}_i \cap \mathcal{X}_i = \mathcal{Y}_j \cap \mathcal{X}_j = \mathcal{Y}_k \cap \mathcal{X}_k = \dots. \quad (8.4.18)$$

With this condition we can define

$$\mathcal{M} = \bigcup_{\kappa=1}^v \mathcal{X}_\kappa, \quad (8.4.19)$$

with

$$\mathcal{X}_\kappa = \mathcal{M} \cap \mathcal{Y}_\kappa, \quad \kappa = 1, 2, \dots, v, \quad (8.4.20)$$

and the lemma follows from (8.4.3) and (8.4.8). $\square$

The formula (8.4.11) for $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$ can now be obtained by induction, using the contraction procedure of Lemma 8.4.1 and the following result, proved by Ruelle (1971):

Lemma 8.4.2 *Let $\mathbb{A}$ and $\mathbb{B}$ be closed subsets of $\mathbb{C}_{\mathfrak{Z}}$ which do not contain 0. Suppose that the complex polynomial*

$$c + a\mathfrak{Z}_A + b\mathfrak{Z}_B + g\mathfrak{Z}_A\mathfrak{Z}_B$$

can vanish only when $\mathfrak{Z}_A \in \mathbb{A}$ and $\mathfrak{Z}_B \in \mathbb{B}$. Then

$$c + g\mathfrak{Z}$$

can vanish only when $\mathfrak{Z} \in -\mathbb{A}\mathbb{B}$.

To complete the theorem using induction on the contraction procedure we first define $\mathbb{S}_{\mathcal{N}}^{(\alpha)}(T; \mathbf{r})$ by a formula like (8.4.11), except that the product is restricted to values of κ in the range $[1, \alpha]$ for which $\mathbf{r} \in \mathcal{Y}_\kappa$. The statement of the theorem, applied to the lattice $\mathcal{N}_\alpha = \cup_{\kappa=1}^\alpha \mathcal{Y}_\kappa$, is that $\Xi(\mathcal{N}_\alpha, T, \{\mathfrak{Z}(\mathbf{r})\}) \neq 0$ when $\mathfrak{Z}(\mathbf{r}) \notin \mathbb{S}_{\mathcal{N}}^{(\alpha)}(T; \mathbf{r})$, for all $\mathbf{r} \in \mathcal{N}_\alpha$. This is true for $\alpha = 1$, when $\Xi(\mathcal{N}_1, T, \{\mathfrak{Z}(\mathbf{r})\}) = \xi(\mathcal{Y}_1, T, \{\mathfrak{Z}(\mathbf{r})\})$, $\mathbb{S}_{\mathcal{N}}^{(1)}(T; \mathbf{r}) = \mathbb{X}_1(T; \mathbf{r})$. Let it be assumed for $\alpha - 1$. It is now important to note that the substitution (8.4.14) can be made step by step in the contraction procedure with $\phi_{\alpha-1}$ having a single variable $\mathfrak{Z}(\mathbf{r})$ defined for each site $\mathbf{r} \in \mathcal{N}_{\alpha-1}$. Consider now the formation of ϕ_α according to the formula (8.4.12). For a site $\mathbf{r}$ in $\mathcal{Y}_\alpha$ but not in $\mathcal{N}_{\alpha-1}$, $\mathbb{S}_{\mathcal{N}}^{(\alpha)}(T; \mathbf{r}) = \mathbb{X}_\alpha(T; \mathbf{r})$ and for a site $\mathbf{r}$ in $\mathcal{N}_{\alpha-1}$ but not in $\mathcal{Y}_\alpha$, $\mathbb{S}_{\mathcal{N}}^{(\alpha)}(T; \mathbf{r}) = \mathbb{S}_{\mathcal{N}}^{(\alpha-1)}(T; \mathbf{r})$. In both these cases no contraction is necessary for the terms in $\mathfrak{Z}(\mathbf{r})$ and $\mathfrak{Z}_\alpha(\mathbf{r})$. For a site which belongs to both $\mathcal{Y}_\alpha$ and $\mathcal{N}_{\alpha-1}$ the contraction procedure is applied. Linear terms in $\mathfrak{Z}(\mathbf{r})$ and $\mathfrak{Z}_\alpha(\mathbf{r})$ are deleted and $\mathfrak{Z}(\mathbf{r})\mathfrak{Z}_\alpha(\mathbf{r})$ is replaced by $\mathfrak{Z}(\mathbf{r})$. According to Lemma 8.4.2, $\Xi(\mathcal{N}_\alpha, T, \{\mathfrak{Z}(\mathbf{r})\})$ vanishes as a function of $\mathfrak{Z}(\mathbf{r})$ only when $\mathfrak{Z}(\mathbf{r}) \in -\mathbb{S}_{\mathcal{N}}^{(\alpha-1)}(T; \mathbf{r})\mathbb{X}_\alpha(T; \mathbf{r}) = \mathbb{S}_{\mathcal{N}}^{(\alpha)}(T; \mathbf{r})$. This completes the proof of Ruelle's theorem. $\square$

We now summarize the 'practical' procedure for obtaining the set $\mathbb{U}_{\mathcal{N}}(T)$:

- (i) For the lattice $\mathcal{N}$ choose the covering $\mathcal{Y}_{\mathcal{N}}$ of subsets $\mathcal{Y}_\kappa$, $\kappa = 1, 2, \dots, \nu$.
- (ii) For each $\mathcal{Y}_\kappa$ form the polynomial $\xi(\mathcal{Y}_\kappa, T, \{\mathfrak{Z}_\kappa(\mathbf{r})\})$.
- (iii) For each $\mathbf{r} \in \mathcal{Y}_\kappa$ find a subset $\mathbb{X}_\kappa(T; \mathbf{r}) \subset \mathbb{C}_\mathfrak{Z}$ and not containing 0 such that $\xi(\mathcal{Y}_\kappa, T, \{\mathfrak{Z}_\kappa(\mathbf{r})\}) \neq 0$ when $\mathfrak{Z}_\kappa(\mathbf{r}) \notin \mathbb{X}_\kappa(T; \mathbf{r})$ for every $\mathbf{r} \in \mathcal{Y}_\kappa$.
- (iv) For each $\mathbf{r} \in \mathcal{N}$ construct $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$ according to the formula (8.4.11).
- (v) Obtain $\mathbb{U}_{\mathcal{N}}(T)$ by taking the union, as in (8.4.6).

It is clear that the tractability of this procedure depends very much on the covering chosen. The number of geometrically different covering sets should be very small. In practice it is often convenient to use the lattice units, defined in Sect. 8.1, as the elements of the covering.

8.4.2 The Yang–Lee Circle Theorem

Consider the generalization

$$\widehat{H}(\{J(\mathbf{r}, \mathbf{r}')\}, \mathcal{H}; \boldsymbol{\sigma}) := - \sum_{\{\mathbf{r}, \mathbf{r}'\}} J(\mathbf{r}, \mathbf{r}') \sigma(\mathbf{r}) \sigma(\mathbf{r}') - \mathcal{H} \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}) \quad (8.4.21)$$

of the spin- $\frac{1}{2}$ Ising model Hamiltonian (3.5.1), where the first summation is over all pairs of sites, with $J(\mathbf{r}, \mathbf{r}') = J(\mathbf{r}', \mathbf{r})$ and $J(\mathbf{r}, \mathbf{r}) = 0$. Using the lattice fluid variables $\tau(\mathbf{r})$ of Sect. 3.5.1.2 and omitting the constant term, (8.4.21) can be re-expressed in the form

$$\hat{H}(\{\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}')\}, \tilde{\mu}; \tau) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}} \tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') [2\tau(\mathbf{r})\tau(\mathbf{r}') - \tau(\mathbf{r}) - \tau(\mathbf{r}')] - \tilde{\mu} \sum_{\{\mathbf{r}\}} \tau(\mathbf{r}), \quad (8.4.22)$$

where $\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') := 2J(\mathbf{r}, \mathbf{r}')$ and $\tilde{\mu} := 2\mathcal{H}$. The terms in the first summation in (8.4.22) are the energies of a pair of sites. The linear terms are retained in this summation, rather than being absorbed into the chemical potential, in order to maintain the hole-particle symmetry. This implies the rather artificial situation where the occupations $\circ-\circ$ and $\bullet-\bullet$ both have zero interaction energy, with non-zero energies occurring for $\circ-\bullet$ and $\bullet-\circ$. This can be regarded as a mathematical device for simplifying the application of Ruelle's theorem, with a transformation to more natural variables being applied at the end. The interacting sets for this system consist of all pairs of sites for which $\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') \neq 0$ and thus, from (8.1.10), we must have

$$\frac{1}{2} \sum_{\{\mathbf{r}'\}} |\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}')| < \infty \quad \forall \mathbf{r} \in \mathcal{N}. \quad (8.4.23)$$

This is achieved by imposing the condition

$$\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') = 0 \quad \text{when } |\mathbf{r} - \mathbf{r}'| > R, \quad (8.4.24)$$

where R is the *range of interaction*. A particular pair of lattice sites is denoted by $[\mathbf{r}, \mathbf{r}']$. The covering $\mathcal{Y}_{\mathcal{N}}$ can now be taken to be all the pairs $[\mathbf{r}, \mathbf{r}']$ for which $\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') \neq 0$ and the polynomials (8.4.9) take the form

$$\begin{aligned} \xi([\mathbf{r}, \mathbf{r}'], T, \mathfrak{Z}(\mathbf{r}'; \mathbf{r}), \mathfrak{Z}(\mathbf{r}; \mathbf{r}')) &= 1 + \tilde{\mathfrak{X}}(T; \mathbf{r}, \mathbf{r}') [\mathfrak{Z}(\mathbf{r}'; \mathbf{r}) + \mathfrak{Z}(\mathbf{r}; \mathbf{r}')] \\ &\quad + \mathfrak{Z}(\mathbf{r}'; \mathbf{r}) \mathfrak{Z}(\mathbf{r}; \mathbf{r}'), \end{aligned} \quad (8.4.25)$$

where $\tilde{\mathfrak{X}}(T; \mathbf{r}, \mathbf{r}') := \exp[-\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}')/T]$ and $\mathfrak{Z}(\mathbf{r}'; \mathbf{r})$ is the variable for the site $\mathbf{r}$ and the covering set $[\mathbf{r}, \mathbf{r}']$. The analysis of this problem now depends of whether the terms $\tilde{\mathfrak{X}}(T; \mathbf{r}, \mathbf{r}')$ are greater or less than unity. The case to which the Yang-Lee circle theorem applies is when

$$\tilde{\mathfrak{X}}(T; \mathbf{r}, \mathbf{r}') < 1, \quad \forall [\mathbf{r}, \mathbf{r}'], \quad (8.4.26)$$

which in spin formulation corresponds to all interactions being ferromagnetic. $\xi([\mathbf{r}, \mathbf{r}'], T, \mathfrak{Z}(\mathbf{r}'; \mathbf{r}), \mathfrak{Z}(\mathbf{r}; \mathbf{r}'))$ is of the form given by Eq. (17.5.2). The regions $\mathbb{X}_{\kappa}(T; \mathbf{r})$ are then given by (17.5.9) for all covering sets and all lattice sites for any choice of $\kappa \neq 1$. As will be seen in the discussions in the next section, the difficult

step in the application of Ruelle's theorem is (iv), the construction of $\mathbb{S}_{\mathcal{N}}(T)$ according to the formula (8.4.11). There is, however, one choice of x for which this process is straightforward, namely $x = 0$, when

$$\mathbb{X}_{\kappa}(T; \mathbf{r}) = \mathbb{D}^{(+)} := \{\mathfrak{Z} : |\mathfrak{Z}| \geq 1\}. \quad (8.4.27)$$

To obtain $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$ (8.4.11) must be used. In general this can be difficult, particularly since boundary sites are members of fewer covering sets than interior sites and must be considered separately. However, in this case the form of $\mathbb{D}^{(+)}$ means that the product procedure has no effect and $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r}) = \mathbb{D}^{(+)}$ for every site of the lattice. It then follows from (8.4.6) that $\mathbb{U}_{\mathcal{N}}(T) = \mathbb{D}^{(+)}$, which is independent of $\mathcal{N}$. Thus $\mathbb{D}^{(+)}$ is a possible choice of $\mathbb{Q}(T)$ for all T . However, in this case the symmetry condition (8.4.1) is satisfied. If some $\mathfrak{Z}_0 \in \mathbb{O}_{\mathcal{N}}(T)$, with $|\mathfrak{Z}_0| > 1$, then $1/\mathfrak{Z}_0 \in \mathbb{O}_{\mathcal{N}}(T)$. This contradicts the result of Ruelle's theorem and it follows that

$$\mathbb{Q}(T) = \{\mathfrak{Z} : |\mathfrak{Z}| = 1\}, \quad \forall T. \quad (8.4.28)$$

This is the Yang–Lee circle theorem.

For this formulation in lattice fluid terms, for which condition (8.4.26) is equivalent to an attractive potential between pair of particles (and pairs of holes), we have shown that, for an arbitrary but finite range of interaction on any lattice, phase transitions can occur only for $\tilde{\mu} = 0$. In the magnetic spin formulation this is equivalent to the result that the Ising ferromagnet with arbitrary range of interaction can have a phase transition only on the zero-field axis. Equation (8.4.22) can be re-expressed in the more usual lattice fluid form (Eq. (3.5.20))

$$\hat{H}(\{\varepsilon(\mathbf{r}, \mathbf{r}')\}, \mu; \boldsymbol{\tau}) = - \sum_{\{\mathbf{r}, \mathbf{r}'\}} \varepsilon(\mathbf{r}, \mathbf{r}') \tau(\mathbf{r}) \tau(\mathbf{r}') - \mu \sum_{\{\mathbf{r}\}} \tau(\mathbf{r}), \quad (8.4.29)$$

where

$$\begin{aligned} \varepsilon(\mathbf{r}, \mathbf{r}') &:= 2\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') = 4J(\mathbf{r}, \mathbf{r}'), \\ \mu &:= \tilde{\mu} - \sum_{\{\mathbf{r}\}} \varepsilon(\mathbf{r}, \mathbf{r}') = 2\mathcal{H} - 2 \sum_{\{\mathbf{r}\}} J(\mathbf{r}, \mathbf{r}'). \end{aligned} \quad (8.4.30)$$

In this case it follows that, as long as the summations in (8.4.30) have the same value for each lattice site, a phase transition can occur only on the line

$$\mu = -\frac{1}{2} \sum_{\{\mathbf{r}\}} \varepsilon(\mathbf{r}, \mathbf{r}'). \quad (8.4.31)$$

8.4.3 Systems with Pair Interactions

For the application of Ruelle's theorem, the choice of the sets $\mathbb{X}_\kappa(T; \mathbf{r})$ is, as we have seen above, not unique. For covering sets consisting of pairs of sites this is exemplified by the different allowed values of the parameters x and y of Sect. 17.5. In particular cases one choice may have particular advantages, in either narrowing the intersection of $\mathbb{Q}(T)$ with $\mathbb{R}^{(+)}$ or obtaining a good estimate of the critical temperature by delaying as much as possible, as T is decreased, the value of T at which $\mathbb{Q}(T)$ intersects $\mathbb{R}^{(+)}$. The circle theorem, obtained by the application of Ruelle's theorem and symmetry, narrows $\mathbb{Q}(T) \cap \mathbb{R}^{(+)}$ to the single point $\mathfrak{z} = 1$. This is an extreme case, where the value $\tilde{\mu} = 0$ for the phase transition (if it occurs) is predicted with complete accuracy. This result is, however, achieved at the expense of a total inability to obtain an upper bound for the critical temperature. In this section we consider other possible choices for $\mathbb{X}_\kappa(T; \mathbf{r})$ for systems with pair interactions. We first examine the situation where

$$\tilde{\varepsilon}(\mathbf{r}, \mathbf{r}') := \begin{cases} \tilde{\varepsilon} > 0, & \text{for every first-neighbour pair } [\mathbf{r}, \mathbf{r}'], \\ 0, & \text{otherwise,} \end{cases} \quad (8.4.32)$$

which gives

$$\tilde{\mathfrak{X}}(T; \mathbf{r}, \mathbf{r}') = \begin{cases} \tilde{\mathfrak{X}}(T) = \exp(-\tilde{\varepsilon}/T) < 1, & \text{for every first-} \\ & \text{neighbour pair } [\mathbf{r}, \mathbf{r}'], \\ 1, & \text{otherwise.} \end{cases} \quad (8.4.33)$$

This is a special case of the situation to which the circle theorem applies. We can however, from (17.5.9), choose

$$\mathbb{X}_\kappa(T; \mathbf{r}) = \mathbb{X}_x(T) := \{\mathfrak{z} : |2\tilde{\mathfrak{X}}(T)\mathfrak{z} + x| \leq \sqrt{4\tilde{\mathfrak{X}}^2(T)(1-x) + x^2}\}, \quad (8.4.34)$$

for every lattice site, every covering set and any $x > 1$. The region $\mathbb{X}_x(T)$ is a disc not including $\mathfrak{z} = 0$ and with centre on $\mathbb{R}^{(-)}$. From (8.4.11) and (8.4.34),

$$\mathbb{S}_\mathcal{N}(T; \mathbf{r}) = -[-\mathbb{X}_x(T)]^z, \quad (8.4.35)$$

for every interior lattice site, where z is the coordination number of the lattice. For boundary sites the formula is the same as (8.4.35) except that z is replaced by some $z' < z$, this being the number of first-neighbour sites of the particular boundary site. Since, however, $-1 \in \mathbb{X}_x(T)$, for all $x > 1$, it follows that

$$-[-\mathbb{X}_x(T)]^{z'} \subset -[-\mathbb{X}_x(T)]^z \quad (8.4.36)$$

and thus, from (8.4.6),

$$\mathbb{U}_{\mathcal{N}}(T) = -[\mathbb{X}_x(T)]^2. \quad (8.4.37)$$

Since this $\mathbb{U}_{\mathcal{N}}(T)$ does not involve $\mathcal{N}$ it can be taken to be $\mathbb{Q}(T)$. From (17.5.7) the points $\mathfrak{Z}^{(\pm)} = -\tilde{\mathfrak{X}}(T) \pm i\sqrt{1 - \tilde{\mathfrak{X}}^2(T)}$ lie on the boundary of $\mathbb{X}_x(T)$ for all $x > 1$. These points subtend an angle $\arccos[\tilde{\mathfrak{X}}(T)]$ with $\mathbb{R}^{(-)}$. It follows that the arc $\text{Arg}(\mathfrak{Z}) \in [z\{\pi - \arccos[\tilde{\mathfrak{X}}(T)\}], z\{\pi + \arccos[\tilde{\mathfrak{X}}(T)\}]]$ of the unit circle $|\mathfrak{Z}| = 1$ is contained in $\mathbb{Q}(T)$. Since the zeros of the partition function lie on $|\mathfrak{Z}| = 1$ an upper bound $T = T_u$ for the occurrence of a phase transition is given by

$$\tilde{\mathfrak{X}}(T_u) = \cos\left(\frac{\pi}{z}\right), \quad T_u := \frac{\tilde{\varepsilon}}{-\ln(\pi/z)} = \frac{\varepsilon}{-2\ln(\pi/z)}, \quad (8.4.38)$$

where $\varepsilon := 2\tilde{\varepsilon}$ is the usually lattice-fluid interaction energy given by the first-neighbour case of (8.4.30). For the square lattice $T_u = 1.4427\varepsilon$, which is consistent with the known critical temperature $T_c = 0.5673\varepsilon$ (formula (9.2.36) with $\varepsilon = 4J$).

Consider now the homogeneous first-neighbour lattice fluid with Hamiltonian (3.5.20). In this case

$$\xi([\mathbf{r}, \mathbf{r}'], T, \mathfrak{Z}, \mathfrak{Z}') = 1 + \mathfrak{Z} + \mathfrak{Z}' + \mathfrak{X}(T)\mathfrak{Z}\mathfrak{Z}', \quad (8.4.39)$$

where $\mathfrak{X}(T) := \exp(\varepsilon/T)$, for every covering set, consisting of a pair of first-neighbour sites. The expression (8.4.39) is of the form given by (17.5.12) and the sets $\mathbb{X}_\kappa(T; \mathbf{r})$ can be taken to be all the same and given by (17.5.14) or (17.5.15) according as $\mathfrak{X}(T) <, > 1$. These cases correspond respectively to repulsive and attractive first-neighbour interactions between particles.

When the interparticle interaction is attractive the possible covering sets $\mathbb{X}_\kappa(T, \mathbf{r})$ are given by (17.5.15). The useful choices are given by $x > 1$ and consist of discs not containing $\mathfrak{Z} = 0$. The best upper bound for the critical temperature is given by (8.4.35) with $\mathbb{X}$ replaced by $\mathbb{X}'$ when x is chosen so that $\mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$ reaches $\mathbb{R}^{(+)}$ as late as possible as the temperature is decreased. This involves the choice of the disc for which the tangent from the origin to the boundary subtends the smallest angle with $\mathbb{R}^{(-)}$. As indicated in Sect. 17.5 this is given by choosing $x = 2$ with

$$\mathbb{X}_\kappa(T; \mathbf{r}) = \mathbb{X}'_2(T) := \left\{ \mathfrak{Z} : |\mathfrak{Z} + 1| \leq \sqrt{1 - \mathfrak{X}^{-1}(T)} \right\}, \quad (8.4.40)$$

for every lattice site and every covering set. The tangent from the origin to the boundary of this disc subtends angles $\pm \arccos[1/\sqrt{\mathfrak{X}(T)}]$ with $\mathbb{R}^{(-)}$. Again, since $-1 \in \mathbb{X}'_2(T)$, we can make the identification $\mathbb{Q}(T) = \mathbb{U}_{\mathcal{N}}(T) = \mathbb{S}_{\mathcal{N}}(T; \mathbf{r})$, where $\mathbf{r}$ is any interior site of the lattice. Using (8.4.35) the upper bound temperature T_u is then given by

$$T_u = -\frac{|\varepsilon|}{2\ln(\pi/z)}, \quad (8.4.41)$$

which, as is to be expected, is the same as the form given by (8.4.38).

In a similar way, when the interparticle interaction is repulsive, the possible sets $\mathbb{X}_\kappa(T; \mathbf{r})$ are given by (17.5.14) with the best choice being given by $y = 0$. Thus

$$\mathbb{X}_\kappa(T; \mathbf{r}) = \mathbb{Y}'_0(T) := \left\{ \mathfrak{Z} : |\mathfrak{X}(T)\mathfrak{Z} + 1| \leq \sqrt{1 - \mathfrak{X}(T)} \right\}, \quad (8.4.42)$$

for every lattice site and every covering set. The tangent from the origin to the boundary of the disc subtends angles $\pm \arccos[\sqrt{\mathfrak{X}(T)}]$ with $\mathbb{R}^{(-)}$ and the upper bound temperature is again given by (8.4.42), except that ε is now positive.

A particular case of this analysis, which was investigated by Runnels and Hubbard (1972), is that of the lattice fluid with first-neighbour exclusion. This corresponds to the limit $\mathfrak{X}(T) \rightarrow 0$, for all T , when Eq. (8.4.42) takes the form

$$\mathbb{X}_\kappa(T; \mathbf{r}) = \left\{ \mathfrak{Z} : \Re(\mathfrak{Z}) \leq -\frac{1}{2} \right\}. \quad (8.4.43)$$

The points on the boundary of this region are given by

$$\mathfrak{Z} = -\frac{\exp(i\theta)}{2 \cos(\theta)}, \quad -\pi \leq \theta \leq \pi. \quad (8.4.44)$$

The set $\mathbb{Q}$, containing the zeros of the grand partition function, will consist of the whole of the plane $\mathbb{C}_\mathfrak{Z}$ except a region containing the origin. The boundary of this region will be obtained from products of z points given by (8.4.44). However, not all such products lie on the boundary. Suppose that $R(\psi) \exp(i\psi)$ is a point on the boundary. It is clear, from (8.4.35), that

$$R(\psi) = -\frac{1}{2^z} \min \prod_{j=1}^z \sec(\theta_j), \quad (8.4.45)$$

where the minimization is subject to the constraint

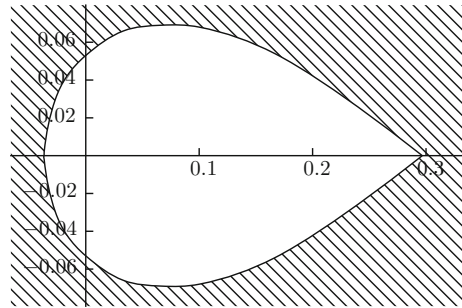
$$\sum_{j=1}^z \theta_j = \psi. \quad (8.4.46)$$

Using a Lagrange undetermined multiplier, it is not difficult to show that the required minimum is achieved when $\theta_j = \psi/z$, for all j , and the boundary curve is given by

$$\mathfrak{Z} = -\frac{\exp(i\psi)}{[2 \cos(\psi/z)]^z}, \quad -\pi \leq \psi \leq \pi. \quad (8.4.47)$$

The point where the curve (8.4.47) meets $\mathbb{R}^{(+)}$ gives a lower bound $\mathfrak{Z}_L$ for the critical fugacity $\mathfrak{Z}_c$. The square lattice fluid model with first-neighbour exclusion is known as the *hard-square model*. From (8.4.47), setting $\psi = \pi$ and $z = 4$, $\mathfrak{Z}_L = \frac{1}{4}$. The

Fig. 8.3 The result given by Eq. (8.4.48) for the hard-hexagon model. The zeros of the grand partition function lie in the shaded region



corresponding model on the triangular lattice is the *hard-hexagon model*. In this case $z = 6$ and $\mathfrak{Z}_L = \frac{1}{27}$. The hard-square model has been shown to be equivalent to a sixteen-vertex model on the square lattice (Wood and Goldfinch 1980). Estimates of the critical fugacity were been obtained by transfer matrix (Runnels and Combs 1971) and series (Gaunt and Fisher 1965) methods. Baxter et al. (1980) used corner transfer matrices to obtain the high-density series of the order parameter to twenty-four terms giving an estimate 3.7962 ± 0.0001 for the critical fugacity. The hard-hexagon model is one of the few examples of a model with a known exact solution (Baxter 1980, 1981, 1982b; Joyce 1988a, b), the critical fugacity being $\frac{1}{2}(5\sqrt{5} + 11) = 11.09$. As Runnels and Hubbard (1972) observed for the hard-square model, the lower bounds derived by the procedure described here, although of course valid, differ substantially from the critical values. These authors obtained the improvement $\mathfrak{Z}_L = \frac{1}{2}$ for the hard-square model by choosing covering sets consisting of half the elementary squares of the lattice with each site belonging to two squares. The same procedure can be applied to the hard-hexagon model. In that case Eq. (8.4.47) is replaced by

$$\mathfrak{Z} = -\frac{\exp(i\psi)}{[3 \cos(\psi/3)]^3}, \quad -\pi \leq \psi \leq \pi, \quad (8.4.48)$$

which gives $\mathfrak{Z}_L = \frac{8}{27}$. The region free from zeros obtained from (8.4.48) is shown in Fig. 8.3.

The analysis of this section has given no unexpected information, showing simply that for first-neighbour lattice fluids with either repulsive or attractive interactions there is no ordered phase at high temperatures and giving consistency with known critical fugacities for models with first-neighbour exclusion. In all but the last example we have chosen covering sets consisting only of pairs of sites. The last example gives some indication that improvements can be gained by increasing the number of sites within a covering.

Chapter 9

Transformation Methods

9.1 Related Systems

We consider a transformation from a model on lattice $\mathcal{N}$ with couplings $\mathbf{K}$ and partition function Z , denoted by the triplet $\{\mathcal{N}, \mathbf{K}, Z\}$, to a model denoted by $\{\mathring{\mathcal{N}}, \mathring{\mathbf{K}}, \mathring{Z}\}$ with all or some of its lattice, couplings and partition function different from those of $\{\mathcal{N}, \mathbf{K}, Z\}$. It is supposed that the numbers of lattice sites $N := |\mathcal{N}|$ and $\mathring{N} := |\mathring{\mathcal{N}}|$ are related by $\mathring{N} = n_0 N$, for some rational number n_0 and the relationships between couplings and partition functions are of the form

$$\mathring{\mathbf{K}} = \mathcal{T}(\mathbf{K}), \quad (9.1.1)$$

$$\mathring{Z}(\mathring{\mathcal{N}}, \mathring{\mathbf{K}}) = \mathcal{Z}(\mathring{\mathcal{N}}, \mathcal{N}, \mathring{\mathbf{K}}, \mathbf{K}) Z(\mathcal{N}, \mathbf{K}). \quad (9.1.2)$$

It is useful to distinguish two different situations in which these equations can arise and be used¹:

- (a) If no singularities are contributed by $\mathcal{Z}(\mathring{\mathcal{N}}, \mathcal{N}, \mathring{\mathbf{K}}, \mathbf{K})$ to (9.1.2) in the limit $N \rightarrow \infty$, then (9.1.1) represents a mapping between critical couplings of $\{\mathcal{N}, \mathbf{K}, Z\}$ and $\{\mathring{\mathcal{N}}, \mathring{\mathbf{K}}, \mathring{Z}\}$. In particular if the two systems differ only in their coupling values, the fixed points of (9.1.1) will give the critical couplings of the system. Duality for the square lattice ν -state Potts model (Sect. 9.2.2.1) and a combination of duality and the star-triangle transformation for the triangular and honeycomb 2-state Potts model (Sects. 9.4.2.2 and 9.4.2.3) are examples of this, although the partition function of the general ν -state Potts model is unknown. For the square-lattice Ising model Kramers and Wannier (1941a) derived the critical temperature by duality in advance of Onsager's derivation of the partition function in 1944.

¹ They also feature in the real-space renormalization group method of Chap. 15. There the aim is to obtain an exact, or approximate, transformation for which $\mathcal{N}$ and $\mathring{\mathcal{N}}$ are lattices of the same type with $n_0 < 1$. Repeated application of this transformation preserves the critical properties associated with the couplings and leads to trajectories in the space of couplings for which the fixed points can be used to determine critical exponents.

- (b) If the model $\{\mathcal{N}, \mathbf{K}, Z\}$ is solved (in the sense that its partition function is known), then (9.1.1) and (9.1.2) yield the solution for $\{\mathring{\mathcal{N}}, \mathring{\mathbf{K}}, \mathring{Z}\}$. Examples of this are provided by the transformation between the eight-vertex and Ising models (Sect. 3.8.1.3)² and the staggered six-vertex model and the ν -state Potts model (Sect. 3.8.2.1). In the latter case the staggered six-vertex model is solved only for particular parameter values which under the transformation entail the Potts model being self-dual, or equivalently at its critical point.

Except where otherwise stated the lattices $\mathcal{N}$ and $\mathring{\mathcal{N}}$ are two-dimensional and of one of the regular types described in Sect. 16.2.1.³ This dimensionality restriction is not necessary for the edge-decoration transformations discussed in Chap. 10.

9.2 The Wegner Transformation

There are many ways of deriving duality transformations. Here we adopt the general formulation of Wegner (1973) which has the advantage that it incorporates both duality and the weak-graph transformation (Nagle and Temperley 1968), which is needed for the eight-vertex model in Sect. 9.3.1.

Consider a system on the lattice $\mathcal{N}$ with the microstate at $\mathbf{r} \in \mathcal{N}$ specified by $\sigma(\mathbf{r})$ and the complete microstate of $\mathcal{N}$ denoted by σ . Suppose that the lattice is divided into n identical (possibly overlapping) subsets of sites,⁴ denoted by $\mathcal{Y}_k$, $k = 1, 2, \dots, n$ and that the Hamiltonian for the system can be expressed in the form

$$\hat{H}(\mathcal{N}, \mathbf{K}; \sigma) := \sum_{k=1}^n \hat{h}(\mathcal{Y}_k, \mathbf{K}; \sigma_k), \quad (9.2.1)$$

where σ_k denotes the subset of the variables σ which determine the state of $\mathcal{Y}_k$ and $\mathbf{K}$ denotes whatever set of couplings is involved.

With $X(\mathcal{Y}_k, \mathbf{K}; \sigma_k) := \exp(-\hat{h}(\mathcal{Y}_k, \mathbf{K}; \sigma_k))$ the partition function takes the form

$$Z(\mathcal{N}, \mathbf{K}) = \sum_{\{\sigma\}} \prod_{k=1}^n X(\mathcal{Y}_k, \mathbf{K}; \sigma_k). \quad (9.2.2)$$

² Although these models were solved independently, it is an interesting and by no means trivial problem (Lavis 1996) to derive the Ising model partition function from that of the eight-vertex model.

³ We shall also use the graph theory terminology of Sect. 16.1.

⁴ This condition can be easily relaxed for specific lattices to include a small number of different kinds of subsets of sites.

Although the components of σ can be chosen independently the components of σ_k , for all k cannot, since in general the subsets $\mathcal{Y}_k$, and in consequence their state variables, overlap. The steps of the Wegner transformation, if it can be applied,⁵ are:

- (i) To find an alternative state vector τ_k for $\mathcal{Y}_k$ so that all the distinct values of $\hat{h}(\mathcal{Y}_k, \mathbf{K}; \sigma_k)$ for different choices of σ are reproduced by a transformed Hamiltonian $\hat{h}'(\mathcal{Y}_k, \mathbf{K}; \tau_k)$ for different choices of τ_k . *But*, whereas different σ_k may share components, all the components of all the τ_k are distinct and the state of the whole system, denoted by τ is just the union of τ_k , for all k . This means that each τ variable is associated with a unique $\mathcal{Y}_k$; a relationship denoted as $\tau \in \mathcal{Y}_k$. Of course, this change of variables comes with a cost. The components of τ will now be constrained. That is to say, certain τ will be impossible and ones which are possible may correspond to more than one term in the partition function (9.2.2) which can be re-expressed in the form

$$Z(\mathcal{N}, \mathbf{K}) = \sum_{\{\tau\}} \eta(\tau) \prod_{k=1}^n X'(\mathcal{Y}_k, \mathbf{K}; \tau_k), \quad (9.2.3)$$

where $X'(\mathcal{Y}_k, \mathbf{K}; \tau_k) := \exp(-\hat{h}'(\mathcal{Y}_k, \mathbf{K}; \tau_k))$ and $\eta(\tau)$ is equal to the number of σ configurations corresponding to the configuration τ , including, of course, zero if the τ configuration is impossible.

- (ii) For a lattice $\hat{\mathcal{N}}$, to associate each constraint conditions with a site $\hat{r} \in \hat{\mathcal{N}}$ and denote the fact that a particular variable τ belongs to the constraint condition associated with $\hat{r}$ by writing $\tau \in \hat{r}$. It should be emphasized that a particular variable τ can in general be associated with more than one constraint condition and thus with more than one site $\hat{r}$.
- (iii) Suppose that on the sites $\hat{r} \in \hat{\mathcal{N}}$ variables $\sigma(\hat{r})$ exist with the same range of values as those on $\mathcal{N}$.
- (iv) To find a function ν so that the constraints can be expressed in the form

$$\Upsilon(\hat{r}; \{\tau \in \hat{r}\}) := \sum_{\{\sigma(\hat{r})\}} \prod_{\tau \in \hat{r}} \nu(\sigma(\hat{r}); \tau) \begin{cases} \neq 0, & \text{if the constraint} \\ & \text{is satisfied,} \\ = 0, & \text{otherwise.} \end{cases} \quad (9.2.4)$$

- (v) To check that Ω exists, independent of τ , so that

$$\eta(\tau) = \Omega \prod_{\{\hat{r}\}} \Upsilon(\hat{r}; \{\tau \in \hat{r}\}). \quad (9.2.5)$$

If each of the steps (i)–(v) can be carried out then, substituting from (9.2.4) and (9.2.5) into (9.2.3), gives

⁵ As Wegner says (op.cit. p. 571), “the total class of systems to which the transformation can be applied with meaningful results is not known”.

$$Z(\mathcal{N}, \mathbf{K}) = \Omega \sum_{\{\tau\}} \prod_{\{\hat{r}\}} \sum_{\{\sigma(\hat{r})\}} \left[\prod_{\{\tau \in \hat{r}\}} v(\sigma(\hat{r}); \tau) \right] \left[\prod_{k=1}^n X'(\mathcal{Y}_k, \mathbf{K}; \tau_k) \right]. \quad (9.2.6)$$

The aim, of course, is to sum over the τ variables in (9.2.6) leaving an expression like (9.2.2), but with a sum over the state $\hat{\sigma}$, representing the set of state variables on all the sites of $\hat{\mathcal{N}}$. This means that we need to determine the equivalent groups $\hat{\mathcal{Y}}_k$ of sites of $\hat{\mathcal{N}}$. Through the relationship $\tau \in \hat{r}$ a certain number of the τ variables will be associated with a particular $\hat{\mathcal{Y}}_k$; a relationship denoted by $\tau \in \hat{\mathcal{Y}}_k$. But, of course, we already have, from (i) above, the relationship $\tau \in \mathcal{Y}_k$, associating the τ variables with the subsets of $\mathcal{N}$. Thus to complete the procedure we need the final condition that:

- (vi) Given a subset $\mathcal{Y}_k$, there exists a unique subset $\hat{\mathcal{Y}}_j$ such that, if $\tau \in \mathcal{Y}_k$, then $\tau \in \hat{\mathcal{Y}}_j$. Of course, this does not necessarily establish a one-to-one relationship between the subsets of $\mathcal{N}$ and $\hat{\mathcal{N}}$. The τ variables in $\hat{\mathcal{Y}}_j$ may divide into subsets belonging to a number of different $\mathcal{Y}_k$. However, in the cases we consider in Sects. 9.2.1–9.2.3 the relationship is one-to-one and for simplicity this relationship will be assumed. Thus we have $\mathcal{Y}_k \leftrightarrow \hat{\mathcal{Y}}_k$ with a particular subset of the τ variables associated with this pair of subsets.

Given this final condition, let

$$X(\hat{\mathcal{Y}}_k, \hat{\mathbf{K}}; \hat{\sigma}_k) \zeta(\mathbf{K}, \hat{\mathbf{K}}) := \sum_{\{\tau_k\}} \left[\prod_{\{\hat{r} \in \hat{\mathcal{Y}}_k\}} \prod_{\{\tau \in \hat{r}\}} v(\sigma(\hat{r}); \tau) \right] X'(\mathcal{Y}_k, \mathbf{K}; \tau_k), \quad (9.2.7)$$

for some function $\zeta(\mathbf{K}, \hat{\mathbf{K}})$. Then, from (9.2.6),

$$Z(\mathcal{N}, \mathbf{K}) = \Omega [\zeta(\mathbf{K}, \hat{\mathbf{K}})]^n Z(\hat{\mathcal{N}}, \hat{\mathbf{K}}), \quad (9.2.8)$$

where $Z(\hat{\mathbf{K}}, \hat{\mathcal{N}})$ is given by (9.2.2) with all variables overscripted with ‘ $\circ$ ’.

9.2.1 Duality for the v -State Potts Model

Consider the case where there is just a first-neighbour interaction, no chemical potential term and the Hamiltonian is given by (3.6.49). The subsets of sites $\mathcal{Y}_k$ are first-neighbour pairs which are referenced by their two constituent sites (viz. $[\mathbf{r}, \mathbf{r}']$). Then, in terms of the general notation given, above

$$X([\mathbf{r}, \mathbf{r}'], K; \{\sigma(\mathbf{r}), \sigma(\mathbf{r}')\}) = \mathfrak{X}^{\delta^{\mathbf{K}\tau}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))}, \quad (9.2.9)$$

where $K := R/T$, $\mathfrak{X} := \exp(K)$ and $\sigma(\mathbf{r}) = 0, 1, \dots, \nu - 1$. For each $[\mathbf{r}, \mathbf{r}']$, the two pair variables

$$\tau(\mathbf{r}, \mathbf{r}') := \sigma(\mathbf{r}) - \sigma(\mathbf{r}') \pmod{\nu}, \quad (9.2.10)$$

are defined and distinguished by the ordering of the pair of sites, with, of course, the consistency conditions

$$\tau(\mathbf{r}, \mathbf{r}') + \tau(\mathbf{r}', \mathbf{r}) = 0 \pmod{\nu}. \quad (9.2.11)$$

We now consider the possible boundary conditions for the lattice when, as is our aim here, the Wegner transformation is applied from a lattice to its dual with the groups of subsets being first-neighbour pairs. For this it is necessary to establish a one-to-one relationship between a first-neighbour pair in the lattice and a first-neighbour pair in the dual lattice and obviously there needs to be an equal number of each. Of the four boundary conditions (I)–(IV) listed in Sect. 16.2.2 the obvious choice is the periodic case (I); and indeed duality is often presented in this way. However, for our purposes and for its application in Sect. 14.2.2.1 it is more convenient to use (IV) with duality applied to the lattice $\mathcal{N}^{(+)}$ which includes the linking edges and the terminating sites connecting $\mathcal{N}$ to the boundary, for each of which we set $\sigma = 0$. So, from (16.2.13), $n = N_{\text{E}}^{(+)} = N_{\text{E}}^*$ and the one-to-one mapping between pairs in $\mathcal{N}^{(+)}$ and $\mathcal{N}^*$ is that of orthogonal bisection.

Suppose that $\mathbf{r}_0$ is an arbitrarily chosen site of $\mathcal{N}$ and $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_m$ gives a simply connected graph of lattice edges between $\mathbf{r}_0$ and $\mathbf{r}_m$. Then,

$$\sigma(\mathbf{r}_m) = \tau(\mathbf{r}_m, \mathbf{r}_{m-1}) + \dots + \tau(\mathbf{r}_1, \mathbf{r}_0) + \sigma(\mathbf{r}_0), \pmod{\nu}. \quad (9.2.12)$$

If $\mathbf{r}_0$ is a boundary site, $\sigma(\mathbf{r}_0) = 0$ and all the site states can be derived from the pair variables. There are N state variables and $2N_{\text{E}}^{(+)}$ pair variables, although this number is effectively reduced to $N_{\text{E}}^{(+)}$ by the consistency conditions (9.2.11). Thus we need $N_{\text{E}}^{(+)} - N$ independent constraint conditions. These are provided by the special case of (9.2.12), where $\mathbf{r}_0 = \mathbf{r}_m$ and the sequence of sites lies on the edges of one of the set of elementary cycles around the faces of $\mathcal{N}$.⁶ From (16.2.13) there are $N^* = N_{\text{E}}^{(+)} - N + 1$ elementary cycles; we seem to have one too many constraints, however, as we shall see they are not all independent. Let

$$\alpha(\mathbf{r}^*) := \tau(\mathbf{r}_{z^*}, \mathbf{r}_{z^*-1}) + \dots + \tau(\mathbf{r}_2, \mathbf{r}_1) \pmod{\nu}, \quad (9.2.13)$$

where $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{z^*}$ are the sites around the face containing $\mathbf{r}^*$. Then the constraint condition

$$\alpha(\mathbf{r}^*) = 0, \quad (9.2.14)$$

⁶ We showed in Sect. 16.1.2 that the number of independent cycles for a section graph is the cyclomatic number, which for a planar graph is the number of faces; in this case it is the number of dual lattice sites.

is associated with the dual-lattice site $\mathbf{r}^*$. As indicated above, there are N^* conditions, but their independence is reduced by

$$\sum_{\{\mathbf{r}^*\}} \alpha(\mathbf{r}^*) = 0. \quad (9.2.15)$$

The lattice $\mathcal{N}$, used in the general description of the Wegner transformation in Sect. 9.2, is now the dual lattice $\mathcal{N}^*$ (with $\mathcal{N}^{(+)}$ replacing $\mathcal{N}$) and taking

$$\nu(\sigma(\mathbf{r}^*); \tau(\mathbf{r}, \mathbf{r}')) := \exp\left(i \frac{2\pi}{\nu} \sigma(\mathbf{r}^*) \tau(\mathbf{r}, \mathbf{r}')\right), \quad (9.2.16)$$

gives, from (9.2.4), (9.2.13), (9.2.14), (18.1.10) and (18.1.11),

$$\Upsilon(\mathbf{r}^*; \{\tau(\mathbf{r}, \mathbf{r}')\}) = \sum_{\sigma=0}^{\nu} \exp\left(i \frac{2\pi}{\nu} \sigma \alpha(\mathbf{r}^*)\right) = \nu \delta^{\text{Kr}}(\alpha(\mathbf{r}^*)). \quad (9.2.17)$$

We now use Ω as it appears in (9.2.5) to compensate for the ν , appearing in (9.2.17) for each constraint condition around the faces of $\mathcal{N}^*$; thus

$$\Omega = 1/\nu^{N^*}. \quad (9.2.18)$$

Every $\tau(\mathbf{r}, \mathbf{r}')$, for an edge of $\mathcal{N}^{(+)}$ will belong to two constraint conditions, in one of which it will appear as $\tau(\mathbf{r}', \mathbf{r})$, and these will be related to two sites $\mathbf{r}^*$ and $\mathbf{r}^{*'}$. Thus a pair of sites $(\mathbf{r}, \mathbf{r}') \in \mathcal{N}^{(+)}$ is in a one-to-one relationship to the pair $(\mathbf{r}^*, \mathbf{r}^{*'}) \in \mathcal{N}^*$; being, of course, pairs of intersecting edges of $\mathcal{N}^{(+)}$ and $\mathcal{N}^*$. From (9.2.7), (9.2.11), (9.2.16) and (18.1.11)

$$\begin{aligned} (\mathfrak{X}^*)^{\delta^{\text{Kr}}(\sigma(\mathbf{r}^*) - \sigma(\mathbf{r}^{*'}))} \zeta(\mathfrak{X}, \mathfrak{X}^*) &= \sum_{\tau=0}^{\nu-1} \exp\left(i \frac{2\pi}{\nu} [\sigma(\mathbf{r}^*) - \sigma(\mathbf{r}^{*'})] \tau\right) \mathfrak{X}^{\delta^{\text{Kr}}(\tau)} \\ &= \{\mathfrak{X} - 1 + \nu \delta^{\text{Kr}}(\sigma(\mathbf{r}^*) - \sigma(\mathbf{r}^{*'}))\}. \end{aligned} \quad (9.2.19)$$

Making the two choices $\sigma(\mathbf{r}^*) = \sigma(\mathbf{r}^{*'})$ and $\sigma(\mathbf{r}^*) \neq \sigma(\mathbf{r}^{*'})$ gives

$$\zeta(\mathfrak{X}, \mathfrak{X}^*) = (\mathfrak{X} - 1), \quad (9.2.20)$$

$$(\mathfrak{X} - 1)(\mathfrak{X}^* - 1) = \nu. \quad (9.2.21)$$

This last relationship is symmetric under $\mathfrak{X} \leftrightarrow \mathfrak{X}^*$ and yields both

$$\mathfrak{X}^{\delta^{\text{Kr}}(x)} = 1 + \frac{\nu \delta^{\text{Kr}}(x)}{\mathfrak{X}^* - 1} \quad \text{and} \quad (\mathfrak{X}^*)^{\delta^{\text{Kr}}(x)} = 1 + \frac{\nu \delta^{\text{Kr}}(x)}{\mathfrak{X} - 1}. \quad (9.2.22)$$

Substituting from (9.2.18) and (9.2.20) into (9.2.8) gives⁷

$$Z(\mathcal{N}^{(+)}, K, 0) = \nu^{-N^*} (\mathfrak{X} - 1)^{N^*} Z^*(\mathcal{N}^*, K^*, 0), \quad (9.2.23)$$

which, by applying (16.2.13) and (9.2.21), becomes

$$\frac{Z(\mathcal{N}^{(+)}, K, 0)}{(\mathfrak{X} - 1)^N} = \frac{Z^*(\mathcal{N}^*, K^*, 0)}{(\mathfrak{X}^* - 1)^{N^*}}. \quad (9.2.24)$$

This relationship has an appearance of symmetric, but it must be remembered that this is not really the case since we have supposed a fixed boundary state on $\mathcal{N}^{(+)}$ and $\mathcal{N}^*$ with free boundary conditions. However, for large lattices, the boundary conditions become unimportant; N and N^* are large and also large with respect to N_L , the number of boundary sites.⁸ From (9.2.24), the free energies per lattice site are related by the symmetric equation

$$z \{ \ln(\mathfrak{X}^* - 1) + \phi_1(\mathcal{N}^*, K^*) \} = z^* \{ \ln(\mathfrak{X} - 1) + \phi_1(\mathcal{N}, K) \}. \quad (9.2.25)$$

Equations (9.2.21) and (9.2.25) are the relationships between couplings and free energies for a pair of dual lattices. In the complex- $\mathfrak{X}$ plane (9.2.21) is satisfied by the points $\mathfrak{X} = F(\nu; \theta)$, $\mathfrak{X} = F(\nu; -\theta)$, which lie on the unit circle

$$\mathfrak{X} = F(\nu; \theta) := 1 + \sqrt{\nu} \exp(i\theta), \quad (9.2.26)$$

which is invariant under the duality transformation.

For the self-dual square lattice if $\mathfrak{X} = \mathfrak{X}^* = \mathfrak{X}_c$ then, from (9.2.21),

$$\mathfrak{X}_c = F(\nu, 0) = 1 + \sqrt{\nu}. \quad (9.2.27)$$

In terms of the coupling $K := \ln(\mathfrak{X})$, the transformation (9.2.21) connects any value of K in the high-temperature interval $[0, K_c]$ with one in the low-temperature interval $[K_c, \infty]$, and vice versa. A critical point arises from a singularity in $Z(\mathcal{N}, K)$, and thus from (9.2.21) a critical point in $[0, K_c]$ implies a critical point in $[K_c, \infty]$. If there is a unique critical point, it must given by

$$K = K_c, \quad T_c = R / \ln(1 + \sqrt{\nu}). \quad (9.2.28)$$

⁷ It should be noted that $Z(\mathcal{N}^{(+)}, K, 0)$ is the partition function for the case where the boundary states have a particular specified value (without loss of generality $\sigma = 0$). If we had simply chosen the condition that all the boundary states were the same the factor ν^{-N^*} in (9.2.23) would be replaced by ν^{1-N^*} .

⁸ The limit of a large lattice is in the sense of van Hove; see Sect. 8.1.

9.2.2 Duality for the Spin- $\frac{1}{2}$ Ising Model

As described in Sect. 3.6.4, the $\mu = 0$ 2-state Potts model is equivalent to the zero-field spin- $\frac{1}{2}$ Ising model, with the Ising energy parameter $J = \frac{1}{2}R$. So with $\nu = 2$, $K := J/T$, $\mathfrak{K} := \exp(2K/T)$ and $\mathfrak{K}^* := \exp(2K^*/T)$, (9.2.21) gives

$$\sinh(2K) \sinh(2K^*) = 1 \quad (9.2.29)$$

and applying (3.6.57) to (9.2.25) gives

$$z^* \left\{ \frac{1}{2} \ln[2 \sinh(2K)] + \phi_1(\mathcal{N}, K, 0) \right\} = z \left\{ \frac{1}{2} \ln[2 \sinh(2K^*)] + \phi_1^*(\mathcal{N}^*, K^*, 0) \right\}. \quad (9.2.30)$$

Then, from (2.3.27),

$$\frac{u(\mathcal{N}, K, 0)}{z \coth(2K)} + \frac{u^*(\mathcal{N}^*, K^*, 0)}{z^* \coth(2K^*)} = -\frac{1}{2}. \quad (9.2.31)$$

This duality relationship between internal energy densities is also, from (3.5.8), a relation between first-neighbour correlations.

9.2.2.1 The Square Lattice

This lattice is self-dual and (9.2.29) is a transformation $K_s \leftrightarrow K_s^*$ between two values of the square-lattice coupling with, from (9.2.29)–(9.2.31),

$$\sinh(2K_s^*) \sinh(2K_s) = 1, \quad (9.2.32)$$

$$\frac{Z(\mathcal{N}_s, K_s, 0)}{[\sinh(2K_s)]^{\frac{1}{2}N}} = \frac{Z(\mathcal{N}_s^*, K_s^*, 0)}{[\sinh(2K_s^*)]^{\frac{1}{2}N}}, \quad (9.2.33)$$

$$\phi_1(\mathcal{N}_s, K_s^*, 0) + \frac{1}{2} \ln[\sinh(2K_s^*)] = \frac{1}{2} \ln[\sinh(2K_s)] + \phi_1(\mathcal{N}_s, K_s, 0), \quad (9.2.34)$$

$$\frac{u(\mathcal{N}_s, K_s, 0)}{\coth(2K_s)} + \frac{u(\mathcal{N}_s, K_s^*, 0)}{\coth(2K_s^*)} = -2. \quad (9.2.35)$$

From (9.2.32), when $K_s \rightarrow \infty$, $K_s^* \rightarrow 0$ and vice-versa. The transformation is, therefore, between high and low temperatures and it will map between critical points with singular thermodynamic characteristics. Intuitively a single critical temperature seems likely for the Ising ferromagnet since there is only one simple type of long-range order, corresponding to the zero temperature ground state.⁹ This must be given by the fixed point $K_s = K_s^* = K_{sc}$ of the transformation. Thus we have

$$\sinh^2(2K_{sc}) = 1, \quad T_c = J/K_{sc} = 2J/\ln(1 + \sqrt{2}). \quad (9.2.36)$$

⁹ That this is the case is evident from Onsager's solution of the model (see Sect. 12.2.7.3).

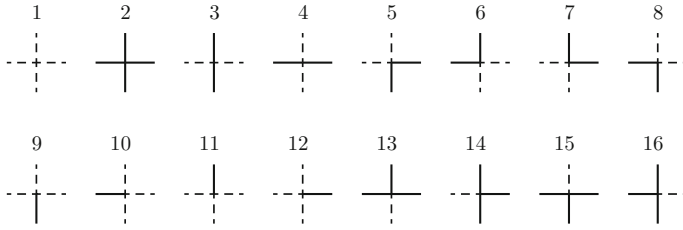


Fig. 9.1 The sixteen vertex types on the square lattice

Substituting $K = K^* = K_{sc}$ in (9.2.31) or (9.2.35), the critical point values of the first-neighbour correlation (see (3.5.8)) and the internal energy per site for the square lattice are given by

$$u_c = -2\langle\sigma(\mathbf{r})\sigma(\mathbf{r}')\rangle_{1c} = -\coth(2K_{sc}) = -\sqrt{2}. \quad (9.2.37)$$

Numerical values for T_c and u_c are given in Table 3.1. It is important to note that the duality form (9.2.34) is not sufficient to determine the critical value $\phi_1(\mathcal{N}_s, K_c, 0)$ for the free-energy density, nor since, from (2.3.28),

$$s = uK - \phi_1(\mathcal{N}_s, K, 0), \quad (9.2.38)$$

is it possible from duality to determine the critical entropy density. For this we need the solution of the two-dimensional Ising model of which a version is given in Sect. 12.2.

9.2.3 The Weak-Graph Transformation

Here attention is restricted to a square lattice $\mathcal{N}_s$ with toroidal boundary conditions. Consider all the 2^{2N} graphs obtain by drawing, or not drawing, a graph edge along an edge of $\mathcal{N}_s$. These graphs $\mathbf{g}$ are weakly embedded in $\mathcal{N}_s$ (Sect. 16.4)¹⁰ and there are 16 vertex types according to whether a particular edge at the vertex contains, or does not contain, an edge of $\mathbf{g}$. They are shown in Fig. 9.1 with the first eight being those of the eight-vertex model shown in Fig. 3.2. The Boltzmann weights for the different vertices are denoted by $\mathfrak{X}(k)$, $k = 1, 2, \dots, 16$ and the partition function is

$$Z(\mathcal{N}_s, \mathfrak{X}) = \sum_{\{\mathbf{g}\}} \prod_{\{\mathbf{r}\}} X(\mathbf{r}; \mathbf{g}), \quad (9.2.39)$$

¹⁰ Hence the name of this transformation.

where the sum is over all graphs $\mathbf{g}$ and $X(\mathbf{r}; \mathbf{g})$ is the particular $\mathfrak{X}(k)$ at site $\mathbf{r}$ when the lattice occupation is $\mathbf{g}$.

In order to apply the Wegner transformation the edges of $\mathcal{N}_s$ are decorated with sites with the location of the site on the first-neighbour pair $[\mathbf{r}, \mathbf{r}'] \in \mathcal{N}_s$ given by $\hat{\mathbf{r}} := (\mathbf{r} + \mathbf{r}')/2$. Suppose that at each edge site $\hat{\mathbf{r}}$, there is a variable $\sigma(\hat{\mathbf{r}})$, which can take the value 1 or 0, according as the graph $\mathbf{g}$ does or does not have a edge through $\hat{\mathbf{r}}$. Then the set of all weakly embedded graphs on $\mathcal{N}_s$ is given by choosing all values of the vector $\boldsymbol{\sigma}$ whose components are the states $\sigma(\hat{\mathbf{r}})$ of all the edge sites.

Now it is necessary to re-express the Boltzman weights in terms of these edge variables. Let $\mathfrak{X}(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, be the Boltzmann weight at a site, where the neighbouring edge variables are labelled in clockwise order starting at the bottom with σ_1 ,¹¹ (e.g. $\mathfrak{X}(1) \rightarrow \mathfrak{X}(0, 0, 0, 0)$, $\mathfrak{X}(13) \rightarrow \mathfrak{X}(0, 1, 1, 1)$), with the partition function re-expressed in the form

$$Z(\mathcal{N}_s, \mathfrak{X}) = \sum_{\{\boldsymbol{\sigma}\}} \prod_{\{\mathbf{r}\}} \mathfrak{X}(\sigma_1(\mathbf{r}), \sigma_2(\mathbf{r}), \sigma_3(\mathbf{r}), \sigma_4(\mathbf{r})). \quad (9.2.40)$$

Of course these variables are not all independent: with a lattice spacing of unity and unit horizontal and vertical vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, $\sigma_1(\mathbf{r}) = \sigma_3(\mathbf{r} + \hat{\mathbf{y}})$, $\sigma_2(\mathbf{r}) = \sigma_4(\mathbf{r} + \hat{\mathbf{x}})$, $\sigma_3(\mathbf{r}) = \sigma_1(\mathbf{r} - \hat{\mathbf{y}})$ and $\sigma_4(\mathbf{r}) = \sigma_2(\mathbf{r} - \hat{\mathbf{x}})$. These constraints are made explicit by replacing, for all $\mathbf{r}$, the variables $\sigma_\ell(\mathbf{r})$ with the *independent* variables $\tau_\ell(\mathbf{r})$, for $\ell = 1, \dots, 4$, with (9.2.40) becoming

$$\begin{aligned} Z(\mathcal{N}_s, \mathfrak{X}) &= \sum_{\{\boldsymbol{\tau}\}} \prod_{\{\mathbf{r}\}} \mathfrak{X}(\tau_1(\mathbf{r}), \tau_2(\mathbf{r}), \tau_3(\mathbf{r}), \tau_4(\mathbf{r})) \\ &\quad \times \delta^{\text{Kr}}(\tau_1(\mathbf{r}) - \tau_3(\mathbf{r} + \hat{\mathbf{y}})) \delta^{\text{Kr}}(\tau_2(\mathbf{r}) - \tau_4(\mathbf{r} + \hat{\mathbf{x}})). \end{aligned} \quad (9.2.41)$$

The next step is to factorize the delta functions into a product of contributions from each of the sites in the argument. For an arbitrary real variable q , let

$$V(q) := \frac{1}{\sqrt{1+q^2}} \begin{pmatrix} +1 & q \\ q & -1 \end{pmatrix}, \quad (9.2.42)$$

where, for convenience, we shall index the rows and columns by 0 and 1, rather than 1 and 2. Now suppose that τ and τ' are two variables on one edge and introduce a new edge variable $\hat{\sigma}$ between them at the corresponding decoration site. Then, since $V(q)$ is a symmetric unitary matrix,

$$\delta^{\text{Kr}}(\tau - \tau') = \sum_{\hat{\sigma}=0}^1 V(\hat{\sigma}, \tau; q) V(\hat{\sigma}, \tau'; q). \quad (9.2.43)$$

¹¹ See the footnote concerning the relationship between toroidal boundary conditions and lattice orientation on p. 60. This function is essentially the same as that defined in Sect. 3.8.1 except that here it is more convenient to take the argument values 0, 1 rather than $-1, +1$.

With

$$\mathring{\mathfrak{X}}(\mathring{\sigma}_T, \mathring{\sigma}_R, \mathring{\sigma}_B, \mathring{\sigma}_L) := \sum_{\tau_1=0}^1 \sum_{\tau_2=0}^1 \sum_{\tau_3=0}^1 \sum_{\tau_4=0}^1 \mathfrak{X}(\tau_1, \tau_2, \tau_3, \tau_4) \prod_{\ell=1}^4 V(\mathring{\sigma}_\ell, \tau_\ell; q), \quad (9.2.44)$$

$$Z(\mathcal{N}_s, \mathfrak{X}) = Z(\mathcal{N}_s, \mathring{\mathfrak{X}}), \quad (9.2.45)$$

where

$$Z(\mathcal{N}_s, \mathring{\mathfrak{X}}) = \sum_{\{\mathring{\sigma}\}} \prod_{\{r\}} \mathring{\mathfrak{X}}(\mathring{\sigma}_1(r), \mathring{\sigma}_2(r), \mathring{\sigma}_3(r), \mathring{\sigma}_4(r)). \quad (9.2.46)$$

Using (9.2.42), (9.2.44) can be re-expressed in the simpler form

$$\mathring{\mathfrak{X}}(k) = \sum_{j=1}^{16} \mathfrak{C}(k, j; q) \mathfrak{X}(j), \quad (9.2.47)$$

where

$$\mathfrak{C}(k, j; q) := \frac{(-1)^{v_{kj}} q^{\gamma_k + \gamma_j - 2v_{kj}}}{(1 + q^2)^2}, \quad (9.2.48)$$

γ_k is the valency of a vertex k and v_{kj} is the number of graph edges which coincide for two superimposed vertices k and j . It is not difficult to show that the matrix with elements $\mathfrak{C}(k, j; q)$ is symmetric and unitary and thus, as in the case of duality, that the weak-graph relationship between two sets of weights is symmetric. It can also be shown that the transformation is satisfied by vertex weights which are functions only of the valencies.

For the eight-vertex model (Sect. 3.8.1) $\mathfrak{X}(k) = \mathfrak{X}(k + 1)$, $k = 1, 3, 5, 7$ and $\mathfrak{X}(k) = 0$, $k = 9, 10, \dots, 16$. Since, from Fig. 9.1 and (9.2.48), $\mathfrak{C}(k, j; 1) = -\mathfrak{C}(k, j + 1; 1)$, $j = 1, 3, 5, 7$, this condition holds for the weights $\mathring{\mathfrak{X}}(k)$, when $q = 1$ and, as we shall now see, the weak-graph transformation gives a symmetry transformation $\mathcal{T}_{\text{wg}}$ for the eight-vertex model.

9.3 The Regular Square-Lattice Eight-Vertex Model

As in Sect. 3.8.1 the lattice carrying the vertices is denoted by $\mathcal{N}^{\otimes}$ with the spins on the sites of $\mathcal{N} + \mathcal{N}^*$. The vertex weights are expressible in spin and edge-variable forms by (3.8.9) and (3.8.10), where now the couplings are site-independent, as are the eight-vertex Boltzmann weights a, b, c and d , given in (3.8.11). In terms of these

$$\Delta := \frac{a^2 + b^2 - c^2 - d^2}{2(ab + cd)}, \quad \Gamma := \frac{ab - cd}{ab + cd}, \quad (9.3.1)$$

were introduced in (3.8.12) and (3.8.13).

It will be shown, in an investigation of the symmetry properties of the eight-vertex partition function $Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c, d)$, that the weak-graph transformation plays a fundamental role and that, as in the case of duality for the Ising model, this leads to a determination of the critical region.

9.3.1 Symmetry Properties and Transformations

Symmetry properties of the partition function are obtained by considering one-to-one transformations between arrow configurations. This can be done using the direction rule and the list of vertex configurations in Fig. 3.2 (see, for example Lavis and Bell 1999b, Sect. 5.4). Here, however, we shall use the shading rule and Fig. 3.3. Then

- Rotation of the lattice through any multiple of 90° represents a translation of the edge variables, while leaving their values unchanged. A clockwise rotation is given by $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (\varsigma_4, \varsigma_1, \varsigma_2, \varsigma_3)$.¹² Thus we have the transformation

$$\begin{aligned} \mathcal{T}_0 : \quad (a, b, c, d) &\leftrightarrow (a, b, d, c), \\ \Delta &\leftrightarrow \Delta, \quad \Gamma \leftrightarrow \Gamma, \end{aligned} \quad (9.3.2)$$

$$Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c, d) = Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, d, c).$$

It is evident from Figs. 3.2 and 3.3, that an alternative way to produce the effect of $\mathcal{T}_0$ on Boltzmann weights is to interchange the pair of edge variable on either of the lines through the vertex. That is, for example, $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (\varsigma_3, \varsigma_2, \varsigma_1, \varsigma_4)$. It is clear that this procedure cannot be carried out over the whole lattice simultaneously. However, it is important for the transfer matrix, which is mapped into its transpose, an element of which is obtained by interchanging c and d .

- Reversing all the arrows on horizontal lines corresponds to $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (\varsigma_1, -\varsigma_2, \varsigma_3, -\varsigma_4)$, giving

$$\begin{aligned} \mathcal{T}_1 : \quad (a, b, c, d) &\leftrightarrow (b, a, d, c), \\ \Delta &\leftrightarrow \Delta, \quad \Gamma \leftrightarrow \Gamma, \end{aligned} \quad (9.3.3)$$

$$Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c, d) = Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, b, a, d, c).$$

This transformation can, of course, also be derived by considering reversal of vertical arrows.

¹² If the direction rule were used the transformation would correspond to $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (-\varsigma_4, \varsigma_1, -\varsigma_2, \varsigma_3)$ giving $(a, b, c, d) \rightarrow (b, a, c, d)$.

- Suppose, on an arbitrary arrow configuration, a graph in the form of one of the two complete sets of bottom-left-to-top-right (BL-TR) staircases, with step length equal to that of one edge (Fig. 3.4a), is superposed. One way to do this is to choose, for the graph edges, 2 and 3 at sites of $\mathcal{N}_E^\oplus$ and 4 and 1 at sites of $\mathcal{N}_O^\oplus$. A transformation which reverses the arrows on the edges of the staircases has the effect $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (\varsigma_1, -\varsigma_2, -\varsigma_3, \varsigma_4)$, on $\mathcal{N}_E^\oplus$ and $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (-\varsigma_1, \varsigma_2, \varsigma_3, -\varsigma_4)$, on $\mathcal{N}_O^\oplus$, giving the BL-TR *staircase transformation*

$$\begin{aligned} \mathcal{T}_2 : \quad & (a, b, c, d) \leftrightarrow (c, d, a, b), \\ & \Delta \leftrightarrow -\Delta, \quad \Gamma \leftrightarrow -\Gamma, \end{aligned} \tag{9.3.4}$$

$$Z^{(8_{\text{VM}})}(\mathcal{N}^\oplus, a, b, c, d) = Z^{(8_{\text{VM}})}(\mathcal{N}^\oplus, c, d, a, b).$$

- Referring to Fig. 3.4b, the corresponding bottom-right-to-top-left (BR-TL) *staircase transformation* is $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (\varsigma_1, \varsigma_2, -\varsigma_3, -\varsigma_4)$, on $\mathcal{N}_E^\oplus$ and $(\varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) \rightarrow (-\varsigma_1, -\varsigma_2, \varsigma_3, \varsigma_4)$, on $\mathcal{N}_O^\oplus$, giving

$$\begin{aligned} \mathcal{T}_3 : \quad & (a, b, c, d) \leftrightarrow (d, c, b, a), \\ & \Delta \leftrightarrow -\Delta, \quad \Gamma \leftrightarrow -\Gamma, \end{aligned} \tag{9.3.5}$$

$$Z^{(8_{\text{VM}})}(\mathcal{N}^\oplus, a, b, c, d) = Z^{(8_{\text{VM}})}(\mathcal{N}^\oplus, d, c, b, a).$$

It is clear that the transformations $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ are self-inverse with $\mathcal{T}_3 = \mathcal{T}_1\mathcal{T}_2$. In fact, together with the identity, they form a commutative subgroup of the symmetric group $\mathbb{S}_4$.

- Comparing Figs. 3.2 and 3.3 it will be seen that the changes of signs of the edge variables in $\mathcal{T}_2$ is equivalent to the mapping needed to change between the arrow assignments made using the direction and shading rules; meaning that both sublattice arrow configurations, for vertices $p = 1, 2, \dots, 8$ in Fig. 3.3, map into the corresponding arrow configurations for these vertices in Fig. 3.2. This can thus be used to count the numbers of vertices of each type on the lattice. Let these numbers be $n_p, p = 1, 2, \dots, 8$. Assuming periodic boundary conditions in both directions, the number of changes of arrow directions left-to-right along a horizontal line must be the same as the number of right-to-left changes. The same must be the case for up-to-down changes on vertical lines. This gives

$$n_5 + n_7 = n_6 + n_8, \quad n_6 + n_7 = n_5 + n_8, \tag{9.3.6}$$

from which it follows that $n_5 = n_6, n_7 = n_8$. The weights c and d appear to an even power in $Z^{(8_{\text{VM}})}(\mathcal{N}^\oplus, a, b, c, d)$, and, because of the transformation $\mathcal{T}_2$ so must a and b . Thus we have the (unphysical) transformations which independently negate the Boltzmann weights

$$\begin{aligned}
\mathcal{T}_a : a &\leftrightarrow -a, & \mathcal{T}_b : b &\leftrightarrow -b, & \mathcal{T}_c : c &\leftrightarrow -c, & \mathcal{T}_d : d &\leftrightarrow -d, \\
\Delta &\leftrightarrow \Delta, & \Gamma &\leftrightarrow 1/\Gamma, \\
Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c, d) &= Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, \pm a, \pm b, \pm c, \pm d),
\end{aligned} \tag{9.3.7}$$

for all choices of signs.

- The weak-graph transformation of Sect. 9.2.3 is applicable to the eight-vertex model. From (9.2.47) and (9.2.48) with $q = 1$

$$\begin{aligned}
\mathcal{T}_{\text{WG}} : (a, b, c, d) &\leftrightarrow (\overset{\circ}{a}, \overset{\circ}{b}, \overset{\circ}{c}, \overset{\circ}{d}), \\
\Delta &\leftrightarrow 1/\Delta, & \Gamma &\leftrightarrow \Gamma/\Delta, \\
Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c, d) &= Z^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, \overset{\circ}{a}, \overset{\circ}{b}, \overset{\circ}{c}, \overset{\circ}{d}),
\end{aligned} \tag{9.3.8}$$

where

$$\begin{aligned}
\overset{\circ}{a} &:= \frac{1}{2}(a + b + c + d), & \overset{\circ}{b} &:= \frac{1}{2}(a + b - c - d), \\
\overset{\circ}{c} &:= \frac{1}{2}(a - b + c - d), & \overset{\circ}{d} &:= \frac{1}{2}(a - b - c + d).
\end{aligned} \tag{9.3.9}$$

Referring to the ground-state regions defined in Sect. 3.8.1.4, the application of $\mathcal{T}_{\text{WG}}$ in region I, is replaced in regions II, III and IV by

$$\begin{aligned}
\mathcal{T}_1 \mathcal{T}_{\text{WG}} \mathcal{T}_1 : (a, b, c, d) &\rightarrow (\overset{\circ}{b}, \overset{\circ}{a}, -\overset{\circ}{d}, -\overset{\circ}{c}), \\
\Delta &\leftrightarrow 1/\Delta, & \Gamma &\leftrightarrow \Gamma/\Delta, \\
\mathcal{T}_2 \mathcal{T}_{\text{WG}} \mathcal{T}_2 : (a, b, c, d) &\rightarrow (\overset{\circ}{c}, -\overset{\circ}{d}, \overset{\circ}{a}, -\overset{\circ}{b}), \\
\Delta &\leftrightarrow 1/\Delta, & -\Gamma &\leftrightarrow \Gamma/\Delta, \\
\mathcal{T}_3 \mathcal{T}_{\text{WG}} \mathcal{T}_3 : (a, b, c, d) &\rightarrow (\overset{\circ}{d}, -\overset{\circ}{c}, -\overset{\circ}{b}, \overset{\circ}{a}), \\
\Delta &\leftrightarrow 1/\Delta, & -\Gamma &\leftrightarrow \Gamma/\Delta,
\end{aligned} \tag{9.3.10}$$

respectively, where $(\overset{\circ}{a}, \overset{\circ}{b}, \overset{\circ}{c}, \overset{\circ}{d})$ are given by (9.3.9). It follows that the points of the surface:

$$\mathcal{C}(\text{I}) : a = b + c + d \quad \text{are the fixed points of } \mathcal{T}_{\text{WG}} \text{ in I,} \tag{9.3.11}$$

$$\mathcal{C}(\text{II}) : b = c + d + a \quad \text{are the fixed points of } \mathcal{T}_1 \mathcal{T}_{\text{WG}} \mathcal{T}_1 \text{ in II,} \tag{9.3.12}$$

$$\mathcal{C}(\text{III}) : c = d + a + b \quad \text{are the fixed points of } \mathcal{T}_2 \mathcal{T}_{\text{WG}} \mathcal{T}_2 \text{ in III,} \tag{9.3.13}$$

$$\mathcal{C}(\text{IV}) : d = a + b + c \quad \text{are the fixed points of } \mathcal{T}_3 \mathcal{T}_{\text{WG}} \mathcal{T}_3 \text{ in IV.} \tag{9.3.14}$$

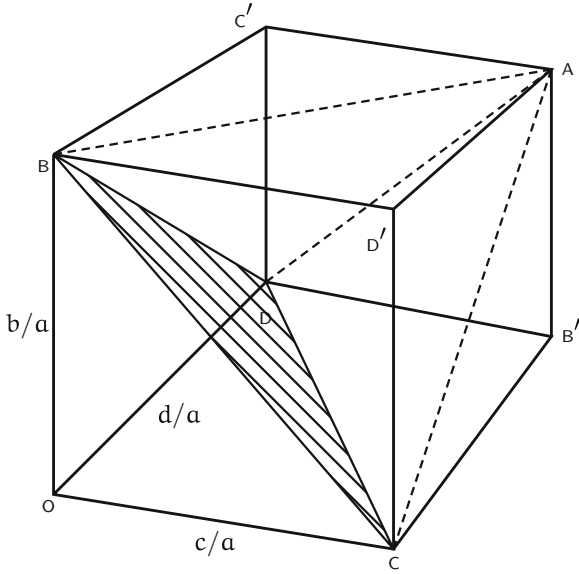


Fig. 9.2 The unit cube corresponding to region I in the space of $(b/a, c/a, d/a)$. The surface $\mathcal{C}(I)$ is hatched

9.3.2 The Case of Region I

This region, in which a is the largest Boltzmann weight, can be represented by the interior of the unit cube $OBCDB'C'D'A$ shown in Fig. 9.2.

9.3.2.1 Transformations and Zones

The cube is divided into two zones¹³

$$\mathcal{R}_L(I) : a > b + c + d, \quad \mathcal{R}_H(I) : a < b + c + d, \quad (9.3.15)$$

corresponding respectively to the tetrahedron $OBCD$ and the remaining part of the cube and separated by the shaded planar triangle BCD , which corresponds to the surface $\mathcal{C}(I)$ defined in (9.3.11).

As in (3.8.19) we expose the temperature dependence of the Boltzmann weights by letting $q =: \exp(-\varepsilon_q/T)$ for $q = a, b, c, d$ and consider the trajectory followed, as temperature varies, by the point (a, b, c, d) . With the vertex energy satisfying the conditions for region I, the trajectory remains in the unit cube and, as the temperature

¹³ For convenience we refer to the different significant subsets of the regions I, II, III and IV as ‘zones’ and ‘subzones’.

is decreased, it passes from **A**, where $T = \infty$ and $a = b = c = d = 1$ to **O**, where $T = 0$ and $b/a = c/a = d/a = 0$. Thus, as the temperature decreases from infinity to zero, the trajectory passes from the high-temperature zone $\mathcal{R}_H(I)$ to the low-temperature zone $\mathcal{R}_L(I)$. It must, therefore, intersect the surface $\mathcal{C}(I)$.

Take a point in $\mathcal{R}_L(I)$ and perform the weak-graph transformation. From (9.3.9) and (9.3.15), $\mathring{a} > \mathring{b} > 0$, $\mathring{a} > \mathring{c} > 0$ and $\mathring{a} > \mathring{d} > 0$ with

$$(\mathring{b} + \mathring{c} + \mathring{d}) - \mathring{a} = a - (b + c + d). \quad (9.3.16)$$

It follows that $\mathcal{T}_{WG}$ maps a point in $\mathcal{R}_L(I)$ into a point in $\mathcal{R}_H(I)$. In general the points are not on the same trajectory (they do not correspond to the same values of the ε_q). An exception to this is the case where, in the vertex energies (3.8.9), $K^{(f)} = K^{(b)}$, $K^{(d)} = 0$. Then the weak-graph transformation becomes the duality transformation of Sect. 9.2.2.1.

The image of $\mathcal{R}_L(I)$ under $\mathcal{T}_{WG}$ is not the whole of $\mathcal{R}_H(I)$. Since a , b , c and d are positive,¹⁴ from (9.3.9),

$$\mathring{a} + \mathring{b} > \mathring{c} + \mathring{d}, \quad \mathring{a} + \mathring{c} > \mathring{b} + \mathring{d}, \quad \mathring{a} + \mathring{d} > \mathring{b} + \mathring{c}. \quad (9.3.17)$$

These inequalities define the tetrahedron **ABCD** of Fig. 9.2, denoted by $\mathcal{R}_O(I)$, and an application of $\mathcal{T}_{WG}$ to a point $(\mathring{a}, \mathring{b}, \mathring{c}, \mathring{d})$ in $\mathcal{R}_O(I)$ takes it back to a point (a, b, c, d) in $\mathcal{R}_L(I)$.

However, this is the case only for points in $\mathcal{R}_O(I)$. Since $\mathring{a} > \mathring{b}$, $\mathring{a} > \mathring{c}$ and $\mathring{a} > \mathring{d}$, not more than one of the three inequalities (9.3.17) can be violated at any one point. The zone $\mathcal{R}_H(I)$ can thus be divided into four subzones; referring to Fig. 9.2, they are the tetrahedra:

ABCD, denoted by $\mathcal{R}_O(I)$, where (9.3.1.7) applies,

AB'CD, denoted by $\mathcal{R}_b(I)$, where $\mathring{a} + \mathring{b} < \mathring{c} + \mathring{d}$,

ABC'D, denoted by $\mathcal{R}_c(I)$, where $\mathring{a} + \mathring{c} < \mathring{b} + \mathring{d}$,

ABCD', denoted by $\mathcal{R}_d(I)$, where $\mathring{a} + \mathring{d} < \mathring{b} + \mathring{c}$.

Although it can easily be verified that the entire high-temperature part of the trajectory of the isotropic Ising model¹⁵ lies in $\mathcal{R}_O(I)$, this is not the case for every trajectory.

Example 9.3.1 Suppose that $\varepsilon_a + \varepsilon_d > \varepsilon_b + \varepsilon_c$, ($ad < bc$). By expanding a , b , c and d in inverse powers of T , it can be shown that, after leaving the $T = \infty$ point **A**, it follows a trajectory which traverses $\mathcal{R}_d(I)$ before entering $\mathcal{R}_O(I)$ and crossing the surface $\mathcal{C}(I)$.

¹⁴ The special case of the six-vertex model, where $d = 0$, is discussed below.

¹⁵ For which $K^{(f)} = K^{(b)} = K_s$, $K^{(d)} = 0$, yielding, from (3.8.11), $\frac{1}{2}(\varepsilon_a + \varepsilon_b) = \varepsilon_c = \varepsilon_d$.

A point in $\mathcal{R}_d(I)$ is mapped by $\mathcal{T}_{wG}$ into a ‘non-physical’ point (a, b, c, d') , with a, b and c positive, but d' negative. The transformation $\mathcal{T}_d$ defined in (9.3.7) maps this point into (a, b, c, d) , with $d = -d'$, so that all the coordinates are positive. The transformation $\mathcal{T}_d \mathcal{T}_{wG}$ takes a point in $\mathcal{R}_d(I)$ into a point in $\mathcal{R}_l(I)$ where, from (9.3.9) and $\mathring{a} > \mathring{b}$, $\mathring{a} > \mathring{c}$, $\mathring{a} > \mathring{d}$, it follows that $d < b$ and $d < c$. The inverse process is effected by the transformation $\mathcal{T}_{wG} \mathcal{T}_d$. More specifically it can be shown that:

- (i) points in the (low-temperature) triangle OBC of Fig. 9.2 are mapped by $\mathcal{T}_{wG}$ ($\mathcal{T}_d$ is, of course, unnecessary in this case) into points for which

$$\mathring{a} + \mathring{d} = \mathring{b} + \mathring{c}, \quad \mathring{d} > 0. \quad (9.3.18)$$

This is the triangle ABC, which is the boundary between the (high-temperature) zones $\mathcal{R}_o(I)$ and $\mathcal{R}_d(I)$.

- (ii) points in the (high-temperature) triangle D'BC are mapped by $\mathcal{T}_d \mathcal{T}_{wG}$ into DBC (the surface $\mathcal{C}(I)$ given by (9.3.11)).

The invariant points for the transformation $\mathcal{T}_{wG} \mathcal{T}_d$ are

$$\mathring{a} = \mathring{b} + \mathring{c}, \quad \mathring{d} = 0. \quad (9.3.19)$$

This is a mapping between the six-vertex model and special cases of the eight-vertex model. The former corresponds, in Fig. 9.2 for region I, to the front face of the cube in the plane $d/a = 0$ and to the square region in Fig. 3.5.

The high-temperature subzones $\mathcal{R}_b(I)$ and $\mathcal{R}_c(I)$ are similarly connected to $\mathcal{R}_l(I)$ by the transformations $\mathcal{T}_b \mathcal{T}_{wG}$ and $\mathcal{T}_c \mathcal{T}_{wG}$ respectively. It follows that the partition function $Z^{(8vM)}(\mathcal{N}^{(8)}, a, b, c, d)$ can be deduced for the entire region I if it is known for the low-temperature zone.

9.3.2.2 Phase Transitions

Suppose that, on some trajectory, a transition occurs at a point in $\mathcal{R}_l(I)$. Making reasonable continuity assumptions, variation of the trajectory (produced by changing the ε_q) generates a transition surface in $\mathcal{R}_l(I)$. Since this is a surface of singularities in the free energy, it follows by applying $\mathcal{T}_{wG}$, that there is a corresponding surface of singularities in $\mathcal{R}_o(I)$. However, the boundary surface $\mathcal{C}(I)$ between $\mathcal{R}_l(I)$ and $\mathcal{R}_h(I)$ is the invariant surface of $\mathcal{T}_{wG}$. So, if it is assumed that there is a unique transition surface in region I (Sutherland 1970) then this surface *may* coincide with $\mathcal{C}(I)$.

However, we know that the transformations $\mathcal{T}_{wG} \mathcal{T}_b$, $\mathcal{T}_{wG} \mathcal{T}_c$ and $\mathcal{T}_{wG} \mathcal{T}_d$ also take points from $\mathcal{R}_l(I)$ to $\mathcal{R}_h(I)$ and it is necessary to examine whether the corresponding invariant surfaces are possible surfaces of transitions. As we have seen the invariant points for $\mathcal{T}_{wG} \mathcal{T}_d$ are given by (9.3.19) which specifies the line BC in Fig. 9.2, which is the intersection of $\mathcal{C}(I)$ and the invariant surface $d/a = 0$ of $\mathcal{T}_d$. Now a trajectory cannot cross the plane $d = 0$, since this would involve entering a region with $d < 0$. In fact, for $T > 0$, no trajectory can touch this plane except in the limiting case of

the six-vertex model, $\varepsilon_d = \infty$, when it lies entirely in the plane. It can similarly be shown that the transformations $\mathcal{T}_{wg}\mathcal{T}_b$ and $\mathcal{T}_{wg}\mathcal{T}_c$ do not generate any possible transition surfaces. Hence, if there is only one transition surface it must be the plane surface $\mathcal{C}(I)$, which implies that there is exactly one transition point on any trajectory, occurring at its intersection with $\mathcal{C}(I)$. It also implies the existence of ferroelectric type I order throughout $\mathcal{R}_l(I)$, while $\mathcal{R}_h(I)$ is a zone of disorder. The assumption that there is only one transition surface is confirmed by the analysis of the exact expression for the zero-field free energy of the eight-vertex model in Sect. 12.3.8.

In the high-temperature (disordered) zone $\mathcal{R}_h(I)$ it is of interest to look at the properties of the model on the boundaries between $\mathcal{R}_o(I)$, where all the conditions of (9.3.17) apply, and the other subzones. For the subzone $\mathcal{R}_d(I)$ the boundary in question is the triangle ABC, given by (9.3.18). As we have seen $\mathcal{T}_{wg}$ maps a point on ABC into a point of the triangle OBC, for which $d = 0$ and $a > b + c$. This represents a subcritical temperature where the state is one of perfect order in terms of spins or arrows with

$$Z^{(8vm)}(\mathcal{N}^{\otimes}, a, b, c, 0) = Z^{(6vm)}(\mathcal{N}^{\otimes}, a, b, c) = a^{N^{\otimes}}. \quad (9.3.20)$$

Applying $\mathcal{T}_{wg}$ gives, for a point on ABC,

$$Z^{(8vm)}(\mathcal{N}^{\otimes}, \mathring{a}, \mathring{b}, \mathring{c}, \mathring{d}) = \frac{1}{4N^{\otimes}} (2\mathring{a} + 2\mathring{b} + 2\mathring{c} + 2\mathring{d})^{N^{\otimes}}. \quad (9.3.21)$$

Application of the transformation $\mathcal{T}_0\mathcal{T}_1$ shows that this formula also applies on the triangle ABD where

$$\mathring{a} + \mathring{c} = \mathring{b} + \mathring{d}. \quad (9.3.22)$$

Comparison with the eight-vertex ice result (3.8.6) shows that (9.3.21) corresponds to an equilibrium state in which the vertex types are distributed as randomly as possible without breaking the compatibility rules for nearest-neighbour vertices. Hence (9.3.18) and (9.3.22) appear to be surfaces of *disorder points*.¹⁶ That this is indeed the case in the circumstances described here was confirmed by correlation function calculations of Baxter (1982b). Thus trajectories starting in $\mathcal{R}_c(I)$ or $\mathcal{R}_d(I)$ pass through a point of disorder in the range between $T = \infty$ and the critical temperature.

From (3.8.11), the Ising model (with $K^{(4)} = 0$) is compatible with (9.3.18) only when $K^{(b)} = 0$ and the entire trajectory then lies in that surface. Since $K^{(i)} \neq 0$ there are correlations between spins on any line of sites in the forward direction, but $K^{(b)} = 0$ means that the spins on different forward lines are uncorrelated. The system is thus essentially one-dimensional and, from (9.3.18) and (9.3.21) with $K^{(0)} = K^{(b)} = K^{(4)} = 0$,

¹⁶ A disorder point in a system is defined to be one, for finite temperature, at which all correlation functions between the states on pairs of sites are zero, as they would be at infinite temperature.

$$\phi(K^{(i)}) = -\ln[2 \cosh(K^{(i)})], \quad (9.3.23)$$

which is equivalent to the formula for the free energy of the one-dimensional zero-field Ising model (Domb 1960, Sect. 3.2.1). There is no transition for $T > 0$, which is to be expected since (9.3.18) has no points in common with (9.3.11). The situation on ABD, given by (9.3.22), is similar, with $K^{(i)} = 0$ in the Ising model case. Points on ACD, where $\mathring{a} + \mathring{b} = \mathring{c} + \mathring{d}$, have different properties from those on ABC and ABD since $\mathcal{T}_{\text{wg}}$ maps them into points (a, b, c, d) where $b = 0$ and $a > c + d$. From (9.3.4) the free energy is equal to that of a six-vertex antiferroelectric below its critical temperature. However the six-vertex antiferroelectric state is not one of perfect order and no equation like (9.3.21) can be deduced.

9.3.3 Regions and Variables

Region I is mapped into regions II, III and IV by $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$, respectively so the analysis of Sect. 9.3.2 carries over to these cases with transition surfaces, $\mathcal{C}(\text{II})$, $\mathcal{C}(\text{III})$ and $\mathcal{C}(\text{IV})$, which become, in the six-vertex $d \rightarrow 0$ limit, the dotted lines in Fig. 3.5.

When the least two values in the set $\{\varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d\}$ are degenerate the trajectory of the system phase point for varying temperature lies on the boundary between two regions. For example, suppose that $\varepsilon_a = \varepsilon_b$. Then $a = b$, $c < a$ and $d < a$ at all points on the trajectory, which lies in the boundary between regions I and II. In such cases none of the transition conditions can be satisfied and hence no transition occurs for $T > 0$. This behaviour is similar to that of the IKDP and IF cases of the six-vertex model (Sect. 3.8.2).

Linear pair combinations of the Boltzmann weights, denoted by $w_j, j = 1, 2, 3, 4$, and satisfying

$$w_1 \geq w_2 \geq w_3 \geq |w_4|, \quad (9.3.24)$$

play an important role in the solution of the eight-vertex model. The definitions for these variables in the different regions and their subzones must necessarily be different. However, we now have all the information needed to generate them. One way to do this is to start with the definitions for one zone and then generate the definitions for the rest of the phase space by applying transformations.

Since, in $\mathcal{R}_L(\text{I})$, $a > b + c + d$, the definitions of the w_j in the first line of Table 9.1 satisfy (9.3.24). The definitions for the remaining zones of region I and for the zones of regions II, III and IV are then derived from the first line by the transformations listed in column 1. In each of the regions

$$\mathcal{T}_{\text{wg}} : (w_1, w_2, w_3, w_4) \leftrightarrow (w_1, w_3, w_2, w_4). \quad (9.3.25)$$

Table 9.1 The definitions of the variables (w_1, w_2, w_3, w_4) in the different regions of the phase space of (a, b, c, d)

Zone	Transformation	w_1	w_2	w_3	w_4	W_Δ	W_Γ	Δ^*	Γ^*
$\mathcal{R}_l(I)$		$\frac{1}{2}(a+b)$	$\frac{1}{2}(a-b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c-d)$	$W(2, 3 1, 4)$	$W(3, 4 2, 1)$	Δ	Γ
$\mathcal{R}_o(I)$	$\mathcal{T}_{WG}$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a-b)$	$\frac{1}{2}(c-d)$	$W(2, 3 4, 1)$	$W(2, 4 3, 1)$	$1/\Delta$	Γ/Δ
$\mathcal{R}_b(I)$	$\mathcal{T}_{WG}\mathcal{T}_b$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a-b)$	$\frac{1}{2}(c-d)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ
$\mathcal{R}_d(I)$	$\mathcal{T}_{WG}\mathcal{T}_d$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c-d)$	$\frac{1}{2}(a-b)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_c(I)$	$\mathcal{T}_{WG}\mathcal{T}_c$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(d-c)$	$\frac{1}{2}(b-a)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_l(II)$	$\mathcal{T}_1$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(b-a)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(d-c)$	$W(2, 3 1, 4)$	$W(3, 4 2, 1)$	Δ	Γ
$\mathcal{R}_o(II)$	$\mathcal{T}_1\mathcal{T}_{WG}$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(b-a)$	$\frac{1}{2}(d-c)$	$W(2, 3 4, 1)$	$W(2, 4 3, 1)$	$1/\Delta$	Γ/Δ
$\mathcal{R}_a(II)$	$\mathcal{T}_1\mathcal{T}_{WG}\mathcal{T}_1\mathcal{T}_a\mathcal{T}_1$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(b-a)$	$\frac{1}{2}(d-c)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ
$\mathcal{R}_c(II)$	$\mathcal{T}_1\mathcal{T}_{WG}\mathcal{T}_1\mathcal{T}_c\mathcal{T}_1$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(d-c)$	$\frac{1}{2}(b-a)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_d(II)$	$\mathcal{T}_1\mathcal{T}_{WG}\mathcal{T}_1\mathcal{T}_d\mathcal{T}_1$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c-d)$	$\frac{1}{2}(a-b)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_l(III)$	$\mathcal{T}_2$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c-d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a-b)$	$W(1, 4 2, 3)$	$W(1, 2 4, 3)$	$-\Delta$	$-\Gamma$
$\mathcal{R}_o(III)$	$\mathcal{T}_2\mathcal{T}_{WG}$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c-d)$	$\frac{1}{2}(a-b)$	$W(1, 4 3, 2)$	$W(1, 3 4, 2)$	$-1/\Delta$	Γ/Δ
$\mathcal{R}_d(III)$	$\mathcal{T}_2\mathcal{T}_{WG}\mathcal{T}_2\mathcal{T}_d\mathcal{T}_2$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(c-d)$	$\frac{1}{2}(a-b)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_b(III)$	$\mathcal{T}_2\mathcal{T}_{WG}\mathcal{T}_2\mathcal{T}_b\mathcal{T}_2$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a-b)$	$\frac{1}{2}(c-d)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ/Γ
$\mathcal{R}_a(III)$	$\mathcal{T}_2\mathcal{T}_{WG}\mathcal{T}_2\mathcal{T}_a\mathcal{T}_2$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(b-a)$	$\frac{1}{2}(d-c)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ/Γ
$\mathcal{R}_l(IV)$	$\mathcal{T}_3$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(d-c)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(b-a)$	$W(1, 4 2, 3)$	$W(1, 2 4, 3)$	$-\Delta$	$-\Gamma$
$\mathcal{R}_o(IV)$	$\mathcal{T}_3\mathcal{T}_{WG}$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(d-c)$	$\frac{1}{2}(b-a)$	$W(1, 4 3, 2)$	$W(1, 3 4, 2)$	$-1/\Delta$	Γ/Δ
$\mathcal{R}_c(IV)$	$\mathcal{T}_3\mathcal{T}_{WG}\mathcal{T}_3\mathcal{T}_c\mathcal{T}_3$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(d-c)$	$\frac{1}{2}(b-a)$	$W(2, 4 3, 1)$	$W(2, 3 4, 1)$	$1/\Gamma$	Δ/Γ
$\mathcal{R}_a(IV)$	$\mathcal{T}_3\mathcal{T}_{WG}\mathcal{T}_3\mathcal{T}_a\mathcal{T}_3$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(b-a)$	$\frac{1}{2}(d-c)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ/Γ
$\mathcal{R}_b(IV)$	$\mathcal{T}_3\mathcal{T}_{WG}\mathcal{T}_3\mathcal{T}_b\mathcal{T}_3$	$\frac{1}{2}(c+d)$	$\frac{1}{2}(a+b)$	$\frac{1}{2}(a-b)$	$\frac{1}{2}(c-d)$	$W(1, 3 4, 2)$	$W(1, 4 3, 2)$	$-1/\Gamma$	Δ/Γ

So the transition surfaces are given by

$$w_2 = w_3. \quad (9.3.26)$$

Any transformation

$$(w_1, w_2, w_3, w_4) \rightarrow (\pm w_i, \pm w_j, \pm w_k, \pm w_\ell), \quad (9.3.27)$$

where (w_i, w_j, w_k, w_ℓ) is a permutation of (w_1, w_2, w_3, w_4) , and the choices of sign can be made independently, corresponds to some product of the transformations defined above. Let

$$W(i, j|k, \ell) := \frac{w_i^2 - w_j^2}{w_k^2 - w_\ell^2}. \quad (9.3.28)$$

It is clear that, for these ratio functions, it can be assumed that $i < j$. Then $W(i, j|k, \ell) \geq, \leq 0$ according as $k <, > \ell$. Further if $j - i = 1$, $|k - \ell| = 3$, $|W(i, j|k, \ell)| \leq 1$ or, conversely, if $j - i = 3$, $|k - \ell| = 1$, $|W(i, j|k, \ell)| \geq 1$. There are only two independent ratio functions.¹⁷ which can be chosen by selecting two sets of indices $(i, j|k, \ell)$ and $(i', j'|k', \ell')$ so that

¹⁷ Every ratio function is either the negative or reciprocal of one of $W(1, 2|3, 4)$, $W(1, 4|2, 3)$ and $W(1, 3|2, 4)$, with

$$W(1, 3|2, 4) = [1 + W(1, 2|3, 4)W(1, 4|2, 3)]/[W(1, 2|3, 4) + W(1, 4|2, 3)].$$

$$W(i, j|k, \ell) = W_{\Delta} := \frac{\Delta - 1}{\Delta + 1} = \frac{(a - b)^2 - (c + d)^2}{(a + b)^2 - (c - d)^2}, \quad (9.3.29)$$

$$W(i', j'|k', \ell') = W_{\Gamma} := \frac{\Gamma - 1}{\Gamma + 1} = \frac{(c + d)^2 - (c - d)^2}{(a - b)^2 - (a + b)^2}.$$

The choices of indices, which vary of course between zones, are shown in Table 9.1. In $\mathcal{R}_L(I)$, $(i, j|k, \ell) = (2, 3|1, 4)$ and $(i', j'|k', \ell') = (3, 4|2, 1)$ and using these ratio functions we define

$$\Delta^* := \frac{1 + W(2, 3|1, 4)}{1 - W(2, 3|1, 4)}, \quad \Gamma^* := \frac{1 + W(3, 4|2, 1)}{1 - W(3, 4|2, 1)}, \quad (9.3.30)$$

so that in that zone $\Delta^* = \Delta$ and $\Gamma^* = \Gamma$. Since $0 < W(2, 3|1, 4) < 1$ and $W(3, 4|2, 1) < 0$,

$$\Delta^* \geq 1, \quad -1 \leq \Gamma^* \leq 1. \quad (9.3.31)$$

From this and the last two columns of Table 9.1 conditions on the values of Δ and Γ in all the zones can be derived. In particular,

$$\begin{aligned} \Delta \geq 1, \quad -1 \leq \Gamma \leq 1, \quad & \text{in } \mathcal{R}_L(J), \\ 0 \leq \Delta \leq 1, \quad -\Delta \leq \Gamma \leq \Delta, \quad & \text{in } \mathcal{R}_O(J), \end{aligned} \quad \text{for } J = I \text{ and } II, \quad (9.3.32)$$

and

$$\begin{aligned} \Delta \leq -1, \quad -1 \leq \Gamma \leq 1, \quad & \text{in } \mathcal{R}_L(J), \\ -1 \leq \Delta \leq 0, \quad \Delta \leq \Gamma \leq -\Delta, \quad & \text{in } \mathcal{R}_O(J), \end{aligned} \quad \text{for } J = III \text{ and } IV. \quad (9.3.33)$$

The transition surfaces $\mathcal{C}(I)$ and $\mathcal{C}(II)$ corresponds to $\Delta = \Delta_c := 1$, with Δ decreasing as temperature is increased across the surfaces. The transition surfaces $\mathcal{C}(III)$ and $\mathcal{C}(IV)$ corresponds to $\Delta = \Delta_c := -1$, with Δ increasing as temperature is increased across the surfaces. From (3.8.12), the coupling forms for the critical surfaces are

$$\sinh(2K^{(f)}) \sinh(2K^{(b)}) + \tanh(2K^{(4)}) \cosh(2K^{(f)}) \cosh(2K^{(b)}) = \Delta_c. \quad (9.3.34)$$

When $K^{(4)} = 0$, this gives

$$\sinh(2K^{(f)}) \sinh(2K^{(b)}) = \Delta_c. \quad (9.3.35)$$

In the spin representation this is the formula for the critical temperature of two identical, anisotropic, non-interacting Ising models, with, by reference to the ground-states in Sect. 3.8.1.4, both $K^{(f)}$ and $K^{(b)}$ positive in region I and negative in region II, and with opposite signs in regions III and IV.

9.4 The Star-Triangle Transformation

As we saw in Sects. 9.2.1 and 9.2.2, duality alone can be used to determine the critical temperature for the self-dual square lattice, but is not sufficient for the triangle/honeycomb dual pair.¹⁸ We need the face-decoration star-triangle transformation.

Consider a honeycomb lattice $\mathcal{N}_H$ of $2N$ sites as shown in Fig. 16.14.¹⁹ The lattice is loose-packed and divides into two triangular sublattices, each of N sites, represented by white and black circles. Suppose that there are three unequal couplings $K_H^{(1)}$, $K_H^{(2)}$ and $K_H^{(3)}$ as shown in Fig. 9.3. With the state at $\mathbf{r} \in \mathcal{N}_H$ given by $\sigma(\mathbf{r})$, the Hamiltonian can be expressed in the form

$$\hat{H}(\mathcal{N}_H, K_H^{(1)}, K_H^{(2)}, K_H^{(3)}; \sigma) = \sum_{\{\mathbf{r}_0\}} \left\{ K_H^{(1)} \epsilon(\sigma(\mathbf{r}_1), \sigma(\mathbf{r}_0)) + K_H^{(2)} \epsilon(\sigma(\mathbf{r}_2), \sigma(\mathbf{r}_0)) + K_H^{(3)} \epsilon(\sigma(\mathbf{r}_3), \sigma(\mathbf{r}_0)) \right\}, \quad (9.4.1)$$

where the sum is over all downward pointing triangles identified by the locations $\{\mathbf{r}_0\}$ of their white sites²⁰ and the partition function is

$$Z(\mathcal{N}_H, K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) = \sum_{\{\sigma\}} \prod_{\{\mathbf{r}_0\}} \left\{ \prod_{x=1}^3 X(K_H^{(x)}; \sigma(\mathbf{r}_x), \sigma(\mathbf{r}_0)) \right\}, \quad (9.4.2)$$

where

$$X(K; \sigma, \sigma') := \exp(-K \epsilon(\sigma, \sigma')). \quad (9.4.3)$$

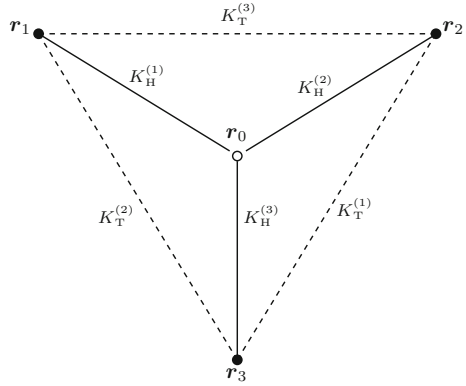
The state σ can be divided into two parts, the state on the white sites and the state on the black sites and the summation can be performed in two parts, first over the states $\{\sigma(\mathbf{r}_0)\}$ on the white sites and then over the states on the black sites. After the first of these has been performed the partition function will be that for the triangular lattice $\mathcal{N}_T$ of N black sites. However, in general it may be of a complicated form. The star-triangle can be carried out if new coupling $K_T^{(1)}$, $K_T^{(2)}$ and $K_T^{(3)}$ (as shown in Fig. 9.3) can be found so that the partition function can, apart from a multiplicative function of the couplings $K_H^{(x)}$, $x = 1, 2, 3$, be expressed as a sum over first-neighbour pairs on $\mathcal{N}_T$ with the same state variable $\sigma(\mathbf{r})$, the same interaction function $\epsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'))$ and the couplings $K_T^{(1)}$, $K_T^{(2)}$ and $K_T^{(3)}$. This means, as in the case of duality, that three equations can be found relating each triangle coupling to the three honeycomb

¹⁸ Although Wegner (1973) has shown that a version of his transformation can be used to transform the triangular lattice directly into itself.

¹⁹ Toroidal boundary conditions are assumed, although the analysis applies equally well to a bounded lattice with an equal number of white and black sites.

²⁰ The labels $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ in Fig. 9.3 for the neighbours of site $\mathbf{r}_0$ are relative to $\mathbf{r}_0$ and are not sublattice labels which can be attached with the same orientation to the black sites in Fig. 16.14. However the coupling designations are applicable to all the downward pointing triangles in Fig. 16.14.

Fig. 9.3 Unequal couplings for the star-triangle transformation



couplings. However, this turns out to be rather more restrictive than was the case for duality.

In order to effect the star-triangle transformation we need to extract the contribution to the partition function from one site r_0 and its neighbours (r_1, r_2, r_3) and set it equal (apart from a multiplicative factor) to interactions around the triangle (r_1, r_2, r_3). Using the abbreviated notation $\sigma_0 := \sigma(r_0)$ $\sigma_x := \sigma(r_x)$ for $x = 1, 2, 3$, this is given by

$$\sum_{\{\sigma_0\}} X(K_H^{(1)}; \sigma_1, \sigma_0) X(K_H^{(2)}; \sigma_2, \sigma_0) X(K_H^{(3)}; \sigma_3, \sigma_0) \\ = \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) X(K_T^{(1)}; \sigma_2, \sigma_3) X(K_T^{(2)}; \sigma_3, \sigma_1) X(K_T^{(3)}; \sigma_1, \sigma_2), \quad (9.4.4)$$

and then

$$Z(\mathcal{N}_H, K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) = [\zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)})]^N Z(\mathcal{N}_T, K_T^{(1)}, K_T^{(2)}, K_T^{(3)}). \quad (9.4.5)$$

As we shall see in Chap. 12, equations of the form (9.4.4) have a remarkable ubiquity in the analysis of statistical mechanical lattice systems. Now often called *Yang-Baxter equations*, they are a key to the solution of many problems quite unrelated to the relationship between the honeycomb and triangular lattice.

A set of transformation equations can be obtained from (9.4.4) by making different choices for the states $(\sigma_1, \sigma_2, \sigma_3)$. Given the need to obtain a form for $\zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)})$, the set must contain four independent equations when the couplings are different and two when the couplings are equal. With this in mind, it is of interest to consider some special cases.

9.4.1 The ν -State Potts Model

Now $X(K_w^{(x)}; \sigma, \sigma') = [\mathfrak{X}_w^{(x)}]^{\delta^{K_r}(\sigma - \sigma')}$ and $\mathfrak{X}_w^{(x)} := \exp(-K_w^{(x)})$, $K_w^{(x)} := R_w^{(x)}/T$, $W = T, H$, $x = 1, 2, 3$. There are three distinguishable cases of (9.4.4) for the states $(\sigma_1, \sigma_2, \sigma_3)$:

(a) All states the same giving

$$\mathfrak{X}_H^{(1)} \mathfrak{X}_H^{(2)} \mathfrak{X}_H^{(3)} + \nu - 1 = \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) \mathfrak{X}_T^{(1)} \mathfrak{X}_T^{(2)} \mathfrak{X}_T^{(3)}. \quad (9.4.6)$$

(b) Two the same and one different, giving (when the couplings are not equal) three equations

$$\begin{aligned} \mathfrak{X}_H^{(x)} + \mathfrak{X}_H^{(y)} \mathfrak{X}_H^{(z)} + \nu - 2 &= \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) \mathfrak{X}_T^{(x)}, & x = 1, 2, 3, \\ y \neq z \neq x. & \end{aligned} \quad (9.4.7)$$

(c) Only when $\nu > 2$; all different giving

$$\mathfrak{X}_H^{(1)} + \mathfrak{X}_H^{(2)} + \mathfrak{X}_H^{(3)} + \nu - 3 = \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}). \quad (9.4.8)$$

There are, therefore, five equations when the couplings are different and three when the couplings are equal; one too many unless $\nu = 2$ when the numbers reduce to four and two respectively. Thus we conclude that the transformation works, in the sense of yielding the triangular couplings as functions of the honeycomb couplings, only for the 2-state Potts model. For $\nu > 2$ Eqs. (9.4.6)–(9.4.8) can be solved to give

$$\begin{aligned} &\mathfrak{X}_T^{(1)} \mathfrak{X}_T^{(2)} \mathfrak{X}_T^{(3)} - \mathfrak{X}_T^{(1)} \mathfrak{X}_T^{(2)} - \mathfrak{X}_T^{(2)} \mathfrak{X}_T^{(3)} - \mathfrak{X}_T^{(3)} \mathfrak{X}_T^{(1)} \\ &\quad - (\nu - 1)[\mathfrak{X}_T^{(1)} + \mathfrak{X}_T^{(2)} + \mathfrak{X}_T^{(3)}] - \nu^2 + 3\nu - 1 = 0, \\ &\mathfrak{X}_H^{(1)} \mathfrak{X}_H^{(2)} \mathfrak{X}_H^{(3)} - \mathfrak{X}_H^{(1)} - \mathfrak{X}_H^{(2)} - \mathfrak{X}_H^{(3)} - \nu + 2 = 0. \end{aligned} \quad (9.4.9)$$

with their joint solutions satisfying the duality relationship (9.2.21) (with $\mathfrak{X} = \mathfrak{X}_H$ and $\mathfrak{X}^* = \mathfrak{X}_T$, or vice-versa). Equations (9.4.9) are the critical conditions derived by Hindermann et al. (1978).²¹ It must be emphasised that the calculation given here is *not a derivation* of the critical conditions. It simply shows that the star-triangle transformation is valid only at criticality. This is consonant with the fact that the ν -state Potts model is solved only at criticality; this solution arising from that of a special case of the staggered six-vertex model (see Sect. 3.8.2.1). In view of Eq. (3.8.47) this is also referred to as the solution of the ‘self-dual Potts model’.

²¹ The derivation given by Hindermann et al. is for $\nu \geq 4$, but, as remarked by Wu (1982), since it agrees with the $\nu = 2$ (Ising) result (Onsager 1944, see also the comment at the end of Sect. 9.4.2.1), “it is expected the [it] is also exact for $\nu = 3$ ”.

9.4.2 The Spin- $\frac{1}{2}$ Ising Model

Since the two-state Potts model is equivalent to the zero-field spin- $\frac{1}{2}$ Ising model we consider that model, for which $X(K_w^{(x)}; \sigma, \sigma') = \exp(-K_w^{(x)} \sigma \sigma')$, $K_w^{(x)} := J_w^{(x)}/T$, $w = T, H$, $x = 1, 2, 3$. Then, from (9.4.4), for the two cases (a) and (b),

$$\begin{aligned} 2 \cosh(K_H^{(1)} + K_H^{(2)} + K_H^{(3)}) &= \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) \\ &\times \exp(K_T^{(1)} + K_T^{(2)} + K_T^{(3)}), \\ 2 \cosh(2K_H^{(x)} - K_H^{(1)} - K_H^{(2)} - K_H^{(3)}) &= \zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)}) \\ &\times \exp(2K_T^{(x)} - K_T^{(1)} - K_T^{(2)} - K_T^{(3)}) \quad x = 1, 2, 3. \end{aligned} \quad (9.4.10)$$

It is convenient at this point to introduce a compact notation for the hyperbolic functions of the honeycomb and triangle couplings²²

$$\begin{aligned} \text{ch}_w^{(x)} &:= \cosh(2K_w^{(x)}), & \text{sh}_w^{(x)} &:= \sinh(2K_w^{(x)}), \\ \text{th}_w^{(x)} &:= \tanh(2K_w^{(x)}), & \text{cth}_w^{(x)} &:= \coth(2K_w^{(x)}), \\ \text{sch}_w^{(x)} &:= \text{sech}(2K_w^{(x)}), & x &= 1, 2, 3, \quad w = T, H. \end{aligned} \quad (9.4.11)$$

Then it can be shown from (9.4.10) that

$$\text{sh}_T^{(x)} = \frac{\text{sh}_H^{(y)} \text{sh}_H^{(z)}}{q_H}, \quad \text{ch}_T^{(x)} = \frac{\text{ch}_H^{(x)} + \text{ch}_H^{(y)} \text{ch}_H^{(z)}}{q_H}, \quad (9.4.12)$$

where in (9.4.12) and subsequent equations (x, y, z) is any permutation of $(1, 2, 3)$ and

$$q_H := \frac{1}{2} [\zeta(K_H^{(1)}, K_H^{(2)}, K_H^{(3)})]^2. \quad (9.4.13)$$

From (9.4.12),

$$q_H^2 = 2\text{ch}_H^{(a)} \text{ch}_H^{(b)} \text{ch}_H^{(c)} + [\text{ch}_H^{(a)}]^2 + [\text{ch}_H^{(b)}]^2 + [\text{ch}_H^{(c)}]^2 - 1. \quad (9.4.14)$$

It can now be shown that

$$\text{ch}_H^{(x)} = \frac{\text{sh}_T^{(x)} \text{ch}_T^{(y)} \text{ch}_T^{(z)} + \text{ch}_T^{(x)} \text{sh}_T^{(y)} \text{sh}_T^{(z)}}{\text{sh}_T^{(x)}}. \quad (9.4.15)$$

²² This notation, not all of which is needed immediately, will be used throughout this and the following section.

This is the inverse relation giving honeycomb couplings in terms of triangle couplings. From the first relation of (9.4.12),

$$\text{sh}_T^{(x)} \text{sh}_H^{(x)} = k^{-1} \quad x = 1, 2, 3, \quad (9.4.16)$$

where

$$k := \frac{q_H}{\text{sh}_H^{(1)} \text{sh}_H^{(2)} \text{sh}_H^{(3)}}. \quad (9.4.17)$$

This is an expression for k in terms of the honeycomb couplings. From (9.4.15),

$$k = \frac{[\text{sch}_T^{(1)} \text{sch}_T^{(2)} \text{sch}_T^{(3)}]^2}{4q_T}, \quad (9.4.18)$$

is an expression in terms of triangle couplings, where

$$q_T^2 := [\text{th}_T^{(1)} + \text{th}_T^{(2)} \text{th}_T^{(3)}][\text{th}_T^{(2)} + \text{th}_T^{(3)} \text{th}_T^{(1)}][\text{th}_T^{(3)} + \text{th}_T^{(1)} \text{th}_T^{(2)}] \\ \times [1 + \text{th}_T^{(1)} \text{th}_T^{(2)} \text{th}_T^{(3)}]. \quad (9.4.19)$$

9.4.2.1 Equal Couplings

We now take the case $K_w^{(1)} = K_w^{(2)} = K_w^{(3)} = K_w$, $w = H, T$ to complete the transformation between the honeycomb and triangular lattices. From (9.4.5) and (9.4.15),

$$\exp(4K_T) = 2 \cosh(2K_H) - 1, \quad (9.4.20)$$

$$\phi_1(\mathcal{N}_T, K_T, 0) = 2\phi_1(\mathcal{N}_H, K_H, 0) + \ln[A(K_T, K_H)], \quad (9.4.21)$$

$$\frac{u(\mathcal{N}_T, K_T)}{\coth(2K_T)} + 1 = B(K_T, K_H) \left[\frac{u(\mathcal{N}_H, K_H)}{\coth(2K_H)} + 1 \right], \quad (9.4.22)$$

where, from (9.4.13), (9.4.14) and (9.4.20),

$$A(K_T, K_H) := \exp(K_T) [\exp(4K_T) + 3]^{\frac{1}{2}} = 2[\cosh(3K_H) \cosh^3(K_H)]^{\frac{1}{4}}, \\ B(K_T, K_H) := \frac{3 \exp(4K_T)}{\exp(4K_T) + 3} = \frac{2[2 \cosh(2K_H) + 1]}{\cosh(2K_H) + 1}. \quad (9.4.23)$$

Suppose that the duality transformation is performed on a triangular lattice Ising model with N sites and coupling K'_T , mapping it into a honeycomb lattice Ising model with coupling K'_H . As seen in Sect. 16.2.2.1 this gives $N^* = 2N$ and, from (9.2.29)–(9.2.31),

$$\sinh(2K'_T) \sinh(2K'_H) = 1, \quad (9.4.24)$$

$$\begin{aligned} \frac{1}{2} \ln[2 \sinh(2K'_T)] + \phi_1(\mathcal{N}_T, K'_T, 0) &= \ln[2 \sinh(2K'_H)] \\ &+ 2\phi_1(\mathcal{N}_H, K'_H, 0), \end{aligned} \quad (9.4.25)$$

$$\frac{u(\mathcal{N}_T, K'_T)}{\coth(2K'_T)} + \frac{2u(\mathcal{N}_H, K'_H)}{\coth(2K'_H)} = -3. \quad (9.4.26)$$

We now have two strategies.

9.4.2.2 The Triangular Lattice

A duality transformation is applied to the triangular lattice, then the equal-couplings star-triangle transformation, given by (9.4.20), is performed on the resulting honeycomb lattice. Setting $K'_H = K_H$, and eliminating this quantity between (9.4.20) and (9.4.24) gives

$$[\exp(4K_T) - 1][\exp(4K'_T) - 1] = 4, \quad (9.4.27)$$

which relates the triangular lattice to itself at two different coupling values. In a similar way, from (9.4.21)–(9.4.26),

$$\phi_1(\mathcal{N}_T, K'_T, 0) + \frac{1}{2} \ln[\sinh(2K'_T)] = \frac{1}{2} \ln[\sinh(2K_T)] + \phi_1(\mathcal{N}_T, K_T, 0), \quad (9.4.28)$$

$$\frac{u(\mathcal{N}_T, K_T)}{\coth(2K_T)} + 1 = -\frac{1}{2} Y(K_T) \left[\frac{u(\mathcal{N}_T, K'_T)}{\coth(2K'_T)} + 1 \right]. \quad (9.4.29)$$

The first of these equations is identical in form to the square-lattice duality transformation relation (9.2.34), but the two K_T values are now connected by (9.4.27) instead of (9.2.29). Like (9.2.29), Eq. (9.4.27) relates values of K_T in a high-temperature interval $[0, K_{TC}]$ with those in a low-temperature interval $[K_{TC}, \infty]$ where K_{TC} is now given by

$$[\exp(4K_{TC}) - 1]^2 = 4. \quad (9.4.30)$$

By similar reasoning to that used in the square case, if there is a unique critical point it must be at $K_T = K_{TC}$ where the two intervals meet, and the critical temperature is then

$$T_c = J/K_{TC} = 4J/\ln(3). \quad (9.4.31)$$

It follows from (9.4.29) that u_c is given by (9.2.37).

9.4.2.3 The Honeycomb Lattice

The star-triangle transformation is followed by a duality transformation. Now we set $K'_T = K_T$ in (9.4.24) and (9.4.25) and after a similar manipulation to that described for the triangular lattice we have

$$[\cosh(2K_H) - 1][\cosh(2K'_H) - 1] = 1, \quad (9.4.32)$$

$$\phi_1(\mathcal{N}_H, K'_H, 0) + \frac{1}{2} \ln[\sinh(2K'_H)] = \frac{1}{2} \ln[\sinh(2K_H)] + \phi_1(\mathcal{N}_H, K_H, 0), \quad (9.4.33)$$

$$\frac{u(\mathcal{N}_H, K'_H)}{\coth(2K'_H)} + 1 = -\frac{1}{2} Y(K_H) \left[\frac{u(\mathcal{N}_H, K_H)}{\coth(2K_H)} + 1 \right], \quad (9.4.34)$$

the second equation again being identical in form to the square-lattice duality transformation relation (9.2.34) with the two K_H values now being connected by (9.4.32). The meeting point of the high- and low-temperature intervals connected by (9.4.32) is at $K_H = K_{HC}$, where

$$[\cosh(2K_{HC}) - 1]^2 = 1 \quad (9.4.35)$$

and, if there is a unique critical point, it occurs at temperature

$$T_c = J/K_{HC} = 2J / \ln(2 + \sqrt{3}). \quad (9.4.36)$$

Again it follows from (9.4.34) that u_c is given by (9.2.37). Numerical values for T_c and u_c are given in Table 3.1.

Finally we note that, translating the critical point Eqs. (9.4.31) and (9.4.36) into the notation for the 2-state Potts model gives $\mathfrak{X}_{TC} = \sqrt{3}$, $\mathfrak{X}_{HC} = 2 + \sqrt{3}$, which satisfy (9.4.9).

Chapter 10

Edge-Decorated Ising Models

10.1 Primary and Secondary Sites

Consider a lattice $\mathcal{N}$ of dimension d , coordination number z , N sites, periodic boundary conditions and an Ising variable $\sigma(\mathbf{r}) = \pm 1$ at every $\mathbf{r} \in \mathcal{N}$. Let the lattice be edge-decorated by taking $\mathring{\mathcal{N}} = \mathcal{N} + \mathcal{N}^{\otimes}$, where $\mathcal{N}^{\otimes}$ is the *medial lattice* (Sect. 16.2.2) with $|\mathring{\mathcal{N}}| = (1 + \frac{1}{2}z)$. The sites of $\mathcal{N}$ are referred to as *primary sites*, labelling them $\mathbf{p}$, and of $\mathcal{N}^{\otimes}$ as *secondary sites*, labelling them $\mathbf{s}$. Suppose that at the secondary site on the edge connecting the first-neighbour pair of sites $[\mathbf{r}, \mathbf{r}']$ there is a microsystem whose state is denoted by $\tau(\mathbf{r}, \mathbf{r}')$. Given that this microsystem interacts only with the microsystems on its neighbouring primary sites and that the primary lattice microsystems form a first-neighbour spin- $\frac{1}{2}$ Ising model the Hamiltonian (3.5.1) is modified to

$$\begin{aligned} \hat{H}(\mathring{\mathcal{N}}, K, \mathring{K}, L; \boldsymbol{\sigma}, \boldsymbol{\tau}) = & \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \left\{ \mathring{K} \mathfrak{e}(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \tau(\mathbf{r}, \mathbf{r}')) - K \sigma(\mathbf{r}) \sigma(\mathbf{r}') \right\} \\ & - L \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}), \end{aligned} \quad (10.1.1)$$

where $K := J/T$, $L := \mathcal{H}/T$ and $\mathring{K}$ is the primary-secondary coupling, for which the interaction has the general form $\mathfrak{e}(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \tau(\mathbf{r}, \mathbf{r}'))$. The partition function is now

$$Z(\mathring{\mathcal{N}}, K, \mathring{K}, L) = \sum_{\{\boldsymbol{\sigma}, \boldsymbol{\tau}\}} \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \left\{ \mathfrak{X}^{\sigma(\mathbf{r})\sigma(\mathbf{r}')} \mathring{\mathfrak{X}}^{-\mathfrak{e}(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \tau(\mathbf{r}, \mathbf{r}'))} \mathfrak{Z}^{\sigma(\mathbf{r})/z} \mathfrak{Z}^{\sigma(\mathbf{r}')/z} \right\}, \quad (10.1.2)$$

where $\mathfrak{X} := \exp(K)$, $\mathring{\mathfrak{X}} := \exp(\mathring{K})$, $\mathfrak{Z} := \exp(L)$ and where the single-site magnetic field terms have been divided between each of their z first-neighbour pairs. The next stage of the procedure is similar to that used for the star-triangle transformation in Sect. 9.4; we sum over the states of the secondary sites and demand that the

resulting partition function be equal, within a multiplicative constant, to that of an equivalent (undecorated) Ising model. The effective Ising couplings are $K_e := J_e/T$ and $L_e := \mathcal{H}_e/T$, with $\mathfrak{X}_e := \exp(K_e)$, $\mathfrak{Z}_e := \exp(L_e)$ and, concentrating on the single first-neighbour pair $[\mathbf{r}, \mathbf{r}']$,

$$\mathfrak{X}^{\sigma(\mathbf{r})\sigma(\mathbf{r}')}\mathfrak{Z}^{\sigma(\mathbf{r})/z}\mathfrak{Z}^{\sigma(\mathbf{r}')/z}\phi(\hat{K}; \sigma(\mathbf{r}), \sigma(\mathbf{r}')) = \zeta(\hat{K})\mathfrak{X}_e^{\sigma(\mathbf{r})\sigma(\mathbf{r}')}\mathfrak{Z}_e^{\sigma(\mathbf{r})/z}\mathfrak{Z}_e^{\sigma(\mathbf{r}')/z}, \quad (10.1.3)$$

where

$$\phi(\hat{K}; \sigma(\mathbf{r}), \sigma(\mathbf{r}')) := \sum_{\{\tau\}} \hat{\mathfrak{X}}^{-\epsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \tau)}. \quad (10.1.4)$$

By making the three choices $\sigma(\mathbf{r}) = \sigma(\mathbf{r}') = \pm 1$, $\sigma(\mathbf{r}) = -\sigma(\mathbf{r}')$, this gives

$$\zeta(\hat{K}) = \left\{ \phi(\hat{K}; +1, +1) \phi(\hat{K}; -1, -1) [\phi(\hat{K}; +1, -1)]^2 \right\}^{\frac{1}{4}}, \quad (10.1.5)$$

$$\mathfrak{X}_e^4 = \mathfrak{X}^4 \frac{\phi(\hat{K}; +1, +1) \phi(\hat{K}; -1, -1)}{[\phi(\hat{K}; +1, -1)]^2}, \quad (10.1.6)$$

$$\mathfrak{Z}_e^4 = \mathfrak{Z}^4 \left[\frac{\phi(\hat{K}; +1, +1)}{\phi(\hat{K}; -1, -1)} \right]^z, \quad (10.1.7)$$

and, from (10.1.2) and (10.1.3),

$$Z(\hat{\mathcal{N}}, K, \hat{K}, L) = [\zeta(\hat{K})]^{\frac{1}{2}Nz} Z^{(\text{im})}(\mathcal{N}, K_e, L_e), \quad (10.1.8)$$

where the latter partition function is that of the Ising model on $\mathcal{N}$. The general form of these relations is due to Fisher (1959). The equivalent primary lattice Ising model with couplings K_e and L_e is a ferromagnet or antiferromagnet according as K_e is positive or negative. For $K_e > 0$, any critical point or curve in the space of the decorated model thermodynamic variables maps into the point $K_e = K_c$, $L_e = 0$ in the K_e - L_e plane of the Ising model while any first-order transition line or surface maps into the segment $L_e = 0$ $K_e > K_c$.

10.2 Super-Exchange or Bond-Dilution

Consider now the special case of the edge-decorated Ising model where $K = 0$ and the one coupling $\hat{K}$ is replaced by two couplings $K_p := J_p/T$ and $L_s := \mu/T$ with

$$\hat{K}\epsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \tau(\mathbf{r}, \mathbf{r}')) = -\tau(\mathbf{r}, \mathbf{r}')[L_s + K_p\sigma(\mathbf{r})\sigma(\mathbf{r}')] \quad (10.2.1)$$

and $\tau(\mathbf{r}, \mathbf{r}')$ able to take two values: zero, meaning that there is no interaction between the Ising spins on $\mathbf{r}$ and $\mathbf{r}'$ and one, meaning that there is a solute atom, with a chemical potential μ , on the secondary s site which mediates an Ising interaction J_p between the two spins on primary p sites. This model can be regarded as one of *super-exchange* mediated by secondary atoms or of *bond-dilution* (Rapaport 1972). Then, from (10.1.5)–(10.1.8),

$$\exp(2K_e) = \frac{\exp(L_s + K_p) + 1}{\exp(L_s - K_p) + 1}, \quad L_e = L, \quad (10.2.2)$$

$$Z(\mathring{\mathcal{N}}, K_p, L, L_s) = [(\exp(L_s + K_p) + 1)(\exp(L_s - K_p) + 1)]^{\frac{1}{4}Nz} \\ \times Z^{(\text{IM})}(\mathcal{N}, K_e, L). \quad (10.2.3)$$

From (3.5.7), (3.5.8) and (10.2.1)–(10.2.3), the magnetization per primary lattice site is

$$m := \langle \sigma(\mathbf{r}) \rangle = - \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, L, L_s)}{\partial L} \right)_{K_p, L_s} = m^{(\text{IM})}(K_e, L), \quad (10.2.4)$$

the density of solute atoms is

$$\rho := \langle \tau(\mathbf{r}, \mathbf{r}') \rangle = \frac{2}{z} \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, L, L_s)}{\partial L_s} \right)_{K_p, L} = f(K_p, K_e, L), \quad (10.2.5)$$

and the internal energy per primary site (made dimensionless by dividing by J_p) is¹

$$u := -\frac{1}{2}z \langle \tau(\mathbf{r}, \mathbf{r}') \sigma(\mathbf{r}) \sigma(\mathbf{r}') \rangle_1 = \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, L, L_s)}{\partial K_p} \right)_{L, L_s} \\ = \frac{1}{2}zg(K_p, K_e, L), \quad (10.2.6)$$

where

$$f(K_p, K_e, L) := C(K_p, K_e) - 2u^{(\text{IM})}(K_e, L)S(K_p, K_e)/z, \quad (10.2.7)$$

$$g(K_p, K_e, L) := S(K_p, K_e) - 2u^{(\text{IM})}(K_e, L)C(K_p, K_e)/z, \quad (10.2.8)$$

$$C(K_p, K_e) := \frac{\sinh(K_e) \cosh(K_p - K_e)}{\sinh(K_p)}, \\ S(K_p, K_e) := \frac{\sinh(K_e) \sinh(K_p - K_e)}{\sinh(K_p)} \quad (10.2.9)$$

¹ Where the subscript '1' is now a reminder that $\mathbf{r}$ and $\mathbf{r}'$ are first-neighbour *primary* sites.

and $m^{(\text{IM})}(K_e, L)$ and $u^{(\text{IM})}(K_e, L)$ are respectively the magnetization and internal energy per primary lattice site on the equivalent Ising model (the latter being made dimensionless by dividing by J_e). From (2.3.29), (10.2.5) and (10.2.6), the heat capacity at constant density per primary site is

$$\begin{aligned} c(K_p, K_e, L, \rho^*) &:= -k_B K_p^2 \left(\frac{\partial u}{\partial K_p} \right)_\rho \\ &= - \frac{k_B z K_p^2 \{ f_e(K_p, K_e, L) g_p(K_p, K_e, L) - f_p(K_p, K_e, L) g_e(K_p, K_e, L) \}}{2 f_e(K_p, K_e, L)}, \end{aligned} \quad (10.2.10)$$

where the subscripts ‘p’ and ‘e’ attached to f and g denote partial differentiation with respect to K_p and K_e respectively.

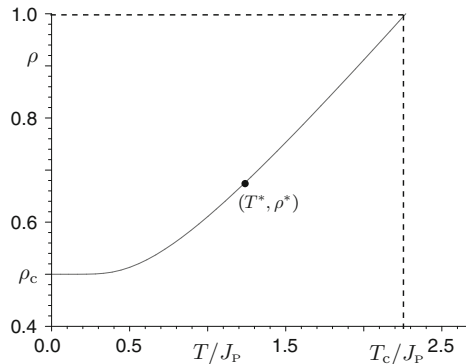
10.2.1 Critical Properties and Exponent Renormalization

Equation (10.2.5) is an implicit relation giving K_e in terms of ρ , K_p and L . In the equivalent Ising model there is a second-order transition when K_e passes through the critical value K_c at $L_e = 0$. A curve of ρ as a function of T/J_p , with $\mathcal{H} = L = 0$ and $K_e = K_c$, can be computed from (10.2.5) and is shown in Fig. 10.1. The bond-dilute model is equivalent to the standard Ising model when $\rho = 1$, and it can be verified from (10.2.5) that the critical curve meets the line $\rho = 1$ at $T = T_c$, the Ising critical temperature. As T decreases from T_c along the critical curve, ρ decreases steadily and $d\rho/dT = 0$ when the curve meets the zero-temperature axis at $\rho = \rho_c$, given by

$$\rho_c = \frac{1 + u^{(\text{IM})}(K_c, 0)}{1 + \coth(K_c)} . \quad (10.2.11)$$

For $\rho < \rho_c$ there is no transition at any value of T .

Fig. 10.1 The critical curve for the bond-dilute model. The curve was calculated for the square lattice, using (9.2.36) and (9.2.37), to give $\rho = \frac{1}{2} + \frac{\sqrt{2}}{4}[\coth(K_p) - 1]$, and for which $\rho_c = \frac{1}{2}$



We now investigate the critical properties of the system in a neighbourhood of the critical curve. Consider the point (T^*, ρ^*) on this curve approached along the line $\rho = \rho^*$. The critical exponents for the equivalent Ising model are denoted in the usual way (Sect. 4.4.1) as α, β, γ and δ , with primes used to distinguish between approaches from below or above the critical temperature.² The corresponding exponents for the approach to $T = T^*$ along the line $\rho = \rho^*$ are denoted by $\alpha_r, \beta_r, \gamma_r$ and δ_r . Since $L_e = L$ and, from (10.2.4), the magnetization is the same as that in the equivalent Ising model, $\delta_r = \delta$. It is tempting to suppose that the same reasoning would apply to γ_r and β_r , but, of course, the asymptotic property of the equivalent Ising model is in terms of $\Delta K_e = K_c - K_e$ and that of the full model is in terms of $\Delta K_p = K_p^* - K_p$. So the first task is to relate these two small quantities to each other. This will involve α the exponent for the heat capacity and we must distinguish two cases: $\alpha = 0$, which is the case for $d = 2$, where the heat capacity has a logarithmic singularity (Sect. 12.2.7.3) and $\alpha \neq 0$. Setting aside the case $\alpha = 0$, which we shall consider at the end of the section, from (4.4.12),

$$u^{(\text{im})}(K_e, 0) \simeq \begin{cases} u_0 + u_1 |\Delta K_e|^{1-\alpha}, & K_e < K_c, \\ u_0 + u_1' |\Delta K_e|^{1-\alpha'}, & K_e > K_c, \end{cases} \quad (10.2.12)$$

for some constants u_0, u_1 and u_1' . In order that the internal energy is finite at the critical point and the heat capacity is singular it is necessary that α lies in the interval $(0, 1]$. In fact we shall impose the more restrictive condition that $0 < \alpha < \frac{1}{2}$, which is certainly true for $d = 3$ (Domb and Hunter 1965) and which has been assumed in the derivation of the asymptotic form (10.2.16). From (10.2.5) and (10.2.7), for small $\Delta K_e > 0$ ($T > T_c$),

$$\Delta K_p \simeq p_1(K_p^*, K_c) |\Delta K_e|^{1-\alpha} + p_2(K_p^*, K_c) \Delta K_e, \quad (10.2.13)$$

where

$$\begin{aligned} p_1(K_p^*, K_c) &:= \frac{-2u_1 S(K_p^*, K_c)}{zC_p(K_p^*, K_c) - 2u_0 S_p(K_p^*, K_c)}, \\ p_2(K_p^*, K_c) &:= \frac{2u_0 S_e(K_p^*, K_c) - zC_e(K_p^*, K_c)}{zC_p(K_p^*, K_c) - 2u_0 S_p(K_p^*, K_c)}, \end{aligned} \quad (10.2.14)$$

with again the subscripts 'p' and 'e' denoting partial differentiation with respect to K_p and K_e respectively. Since $S(K_c, K_c) = 0$,

$$\Delta K_p \simeq \begin{cases} p_1(K_p^*, K_c) |\Delta K_e|^{1-\alpha}, & \text{when } K_p^* \neq K_c, \\ p_2(K_c, K_c) \Delta K_e, & \text{when } K_p^* = K_c. \end{cases} \quad (10.2.15)$$

² In the Ising model $\sigma = \alpha$ (Sect. 4.5.6).

Thus we have, when $K_p^* \neq K_c$, $\gamma_r = \gamma/(1 - \alpha)$. Similar calculations for the case $\Delta K_c < 0$ ($T < T_c$), yield $\gamma'_r = \gamma'/(1 - \alpha')$ and $\beta_r = \beta/(1 - \alpha')$.

To obtain the exponent α_r the asymptotic form for the heat capacity must be determined from (10.2.10). Again consider the two cases where K_p^* is or is not equal to K_c and, retaining only those terms which play a role in the asymptotic form for one of these cases

$$c(K_p, K_c, 0, \rho^*) \simeq -k_B K_p^2 \frac{(1 - \alpha)B_1(K_p^*, K_c)|\Delta K_c|^{-\alpha} + B_2(K_p^*, K_c)}{(1 - \alpha)B_3(K_p^*, K_c)|\Delta K_c|^{-\alpha} + B_4(K_p^*, K_c)}, \quad (10.2.16)$$

where

$$\begin{aligned} B_1(K_p^*, K_c) &:= u_1 \left\{ \frac{1}{2} z [C(K_p^*, K_c)C_p(K_p^*, K_c) - S(K_p^*, K_c)S_p(K_p^*, K_c)] \right. \\ &\quad \left. + u_0 [S(K_p^*, K_c)C_p(K_p^*, K_c) - C(K_p^*, K_c)S_p(K_p^*, K_c)] \right\}, \\ B_2(K_p^*, K_c) &:= \left(\frac{1}{4} z^2 - u_0^2 \right) \{ S_p(K_p^*, K_c)C_e(K_p^*, K_c) \\ &\quad - S_e(K_p^*, K_c)C_p(K_p^*, K_c) \}, \\ B_3(K_p^*, K_c) &:= -u_1 S(K_p^*, K_c), \\ B_4(K_p^*, K_c) &:= \frac{1}{2} z C_e(K_p^*, K_c) - u_0 S_e(K_p^*, K_c). \end{aligned} \quad (10.2.17)$$

For $K_p^* \neq K_c$ (and $\Delta K_c > 0$) it follows from (10.2.15) that $p_1(K_p^*, K_c) > 0$ and, with

$$\theta := \frac{\Delta K_p}{p_1(K_p^*, K_c)} > 0, \quad (10.2.18)$$

$$\begin{aligned} \Delta c(\theta) &:= c(K_p, K_c, 0, \rho^*) - c(K_p^*, K_c, 0, \rho^*) \\ &\simeq -\frac{k_B K_p^2 D(K_p^*, K_c)}{\theta^{\alpha_r}(1 - \alpha)}, \end{aligned} \quad (10.2.19)$$

$$\frac{d\Delta c(\theta)}{d\theta} \simeq \frac{\alpha_r k_B K_p^2 D(K_p^*, K_c)}{\theta^{\alpha_r+1}(1 - \alpha)},$$

where $-1 < \alpha_r := -\alpha/(1 - \alpha) < 0$ and

$$D(K_p^*, K_c) := \frac{[B_2(K_p^*, K_c)B_3(K_p^*, K_c) - B_1(K_p^*, K_c)B_4(K_p^*, K_c)]}{B_3^2(K_p^*, K_c)}. \quad (10.2.20)$$

In the case $\Delta K_c < 0$, α' and $p'_1(K_p^*, K_c)$, obtained by replacing u_1 by u'_1 in (10.2.15), replace α and $p_1(K_p^*, K_c)$ respectively. Now it must be the case that $p'_1(K_p^*, K_c) < 0$ and we obtain the expressions equivalent to (10.2.19) but with primed exponents. Collecting together all the exponents

$$\begin{aligned} \alpha_r &= -\frac{\alpha}{1-\alpha}, & \gamma_r &= \frac{\gamma}{1-\alpha}, & \delta_r &= \delta, \\ \alpha'_r &= -\frac{\alpha'}{1-\alpha'}, & \gamma'_r &= \frac{\gamma'}{1-\alpha'}, & \beta_r &= \frac{\beta}{1-\alpha'}. \end{aligned} \quad (10.2.21)$$

These transformations are called *exponent renormalization*. From (10.2.21) it is easily verified that, if the Ising exponents satisfy the Essam–Fisher scaling law (4.5.44) and the Widom scaling law (4.5.45), then

$$\begin{aligned} \alpha'_r + 2\beta_r + \gamma'_r &= 2 \\ \gamma'_r &= \beta_r(\delta_r - 1). \end{aligned} \quad (10.2.22)$$

These exponents also satisfy the scaling laws with $\alpha'_r = \alpha_r$ and $\gamma'_r = \gamma_r$ if (4.5.47) holds.

For $\rho = 1$, $K_p^* = K_c$ and it follows from (10.2.15) and (10.2.16) that the exponents revert to their Ising values. There are thus finite discontinuities in the exponents, since their limits as $\rho \rightarrow 1 - 0$ are not equal to their values at $\rho = 1$. It is remarkable that for any non-zero value of $1 - \rho$, however small, there is a qualitative change in the properties of the model, the infinite discontinuity in the heat capacity being replaced by a cusp with an infinite slope at the critical point. However, the smaller the value of $1 - \rho$, the narrower becomes the range around the critical point where the exponent changes have a significant numerical effect (Fisher and Scesney 1970).³

The exponent renormalization results from taking ρ as an independent variable which can change at constant T or remain constant when T changes. It is not simply a consequence of the existence of broken bonds. If the latter are regarded as defects with an activation energy $\Delta\varepsilon$, then $L_s = \Delta\varepsilon/T$ and K_c , as given by (10.2.2), depends only on T . Thus $\Delta K_c \sim \Delta K_p$ and no renormalization occurs. The concentration ρ in zero field becomes a function of T . In fact, it is more realistic physically to take ρ as an independent variable in the interpretation of this model as a mixture, where secondary sites are occupied/unoccupied by solute atoms than it is when the two states of a secondary site correspond to a bond which is unbroken/broken. In the mixture case, the mole fraction of solute is $\rho z/(2 + z)$ which is a legitimate thermodynamic density.

³ These authors used a different model, but the same general principles must apply here.

As we shall see in Sect. 12.2.7.3, in two-dimensional lattices the Ising model has $\alpha = 0$ with (10.2.12) replaced by⁴

$$u^{(\text{IM})}(K_c, 0) = u_0 + u_1 \Delta K_c \ln |\Delta K_c|. \quad (10.2.23)$$

Then, when $K_p^* \neq K_c$, (10.2.15), (10.2.18) and (10.2.19) are replaced by

$$\theta \simeq -\Delta K_c \ln |\Delta K_c|, \quad (10.2.24)$$

and

$$\Delta c(\theta) \simeq -k_B K_p^2 \frac{D(K_p^*, K_c)}{\ln |\Delta K_c|}. \quad (10.2.25)$$

The appropriate solution to (10.2.24) is

$$\Delta K_c = -\frac{\theta}{\mathcal{W}(-\theta)}, \quad \ln |\Delta K_c| = \mathcal{W}(-\theta), \quad (10.2.26)$$

where $\mathcal{W}$ is the branch of the Lambert W function chosen so that $\lim_{\theta \rightarrow 0} \mathcal{W}(-\theta) = -\infty$ and for which it is the case that $\lim_{\theta \rightarrow 0} \sqrt{\theta} \mathcal{W}(-\theta) = 0$.⁵ Then

$$\left. \begin{aligned} \Delta c(\theta) &\simeq -k_B K_p^2 \frac{D(K_p^*, K_c)}{\mathcal{W}(-\theta)} \rightarrow 0, \\ \frac{d\Delta c(\theta)}{d\theta} &\simeq k_B K_p^2 \frac{D(K_p^*, K_c)}{\theta [\mathcal{W}(-\theta)]^2} \rightarrow \infty, \end{aligned} \right\} \quad \text{as } \theta \rightarrow 0. \quad (10.2.27)$$

There is thus a cusp with vertical tangents at the transition point, as in the $\alpha > 0$ case.

Renormalization relations like (10.2.21) were derived by Essam and Garelick (1967) for the *Syozu dilution model* (Syozu 1972; Essam 1972), which is slightly more elaborate than the one treated here. Widom (1967) derived the expression $\gamma/(1 - \alpha)$ for the compressibility exponent in a decorated mixture model with the density of one component fixed. Fisher (1968) used thermodynamic arguments, in conjunction with examples from statistical mechanics, to establish general conditions for exponent renormalization. These were the existence of a ‘hidden variable’ subject to a constraint, for instance the presence of an equilibrium distribution of a fixed

⁴ The result in Sect. 12.2.7.3 is just for the square lattice, but it was shown by Houtappel (1950) that (10.2.23) also applied to the triangular and honeycomb lattices.

⁵ The Lambert W function (see e.g. Corless et al. 1996) is the multi-valued inverse of the function $z \exp(z)$.

number of impurity atoms. ‘Ideal’ exponents are obtained when the largest possible number of fields are taken as independent variables, that is $\eta = n - 1$, and are renormalized when a field is replaced as an independent variable by a density.

10.3 A Ferrimagnet

In this type of model, which has been discussed by Syozi and Nakono (1955), Syozi (1972) and Bell (1974b), both the primary p and secondary s sites are occupied by Ising spins but they have different magnet moments νm_0 and m_0 respectively. This is equivalent to taking the magnetic field as $\nu\mathcal{H}$ on primary sites and $\mathcal{H}$ on secondary sites. The other particular feature of this model is that the primary–primary interaction between first-neighbour pairs is ferromagnetic, but the interaction between a secondary spin and its two neighbouring primary spins is antiferromagnetic. The Hamiltonian is (10.1.1) with⁶

$$\mathbb{K}\epsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \sigma(\mathbf{r}, \mathbf{r}')) = K_s \sigma(\mathbf{r}, \mathbf{r}') [\sigma(\mathbf{r}) + \sigma(\mathbf{r}')] - L_s \sigma(\mathbf{r}, \mathbf{r}'), \quad (10.3.1)$$

and $K_s := J_s/T$, $L_s := \mathcal{H}/T$, where $J_s > 0$ is the p – s interaction energy. We shall also, in the interests of conformity, denote the direct p – p Ising interaction, as it appears in (10.1.1), by J_p rather than J with $K_p := J_p/T$, replacing L by $L_p := \nu\mathcal{H}/T$.

We can now carry out a ground-state analysis similar to that for the antiferromagnet in Sect. 3.5.1.1. The number of p sites is N and the number of s sites is $\frac{1}{2}zN$. In the paramagnetic ground state, which will give the lowest Hamiltonian value for large $\mathcal{H}$, all spins are aligned with the field. The other ground state will be one in which the p spins and s spins have an antiparallel alignment. Since the number of p and s spins differ there will be, unlike in the case of antiferromagnetism, a residual magnetic moment and this phenomenon is called *ferrimagnetism*. The ferrimagnetic ground state has a degeneracy of two for zero field and for small $\mathcal{H}$ the spins on s sites will be aligned with the field if $\frac{1}{2}z > \nu$, which we shall assume unless otherwise stated. Thus we have

$$\begin{aligned} \hat{H}_p &= \frac{1}{2}zN J_p - zN J_s - N \left(\frac{1}{2}z + \nu \right) |\mathcal{H}|, \\ \hat{H}_s &= \frac{1}{2}zN J_p + zN J_s - N \left(\frac{1}{2}z - \nu \right) |\mathcal{H}| \end{aligned} \quad (10.3.2)$$

and $\hat{H}_s < \hat{H}_p$, making the ferrimagnetic state stable if

$$-\mathcal{H}_c < \mathcal{H} < \mathcal{H}_c := zJ_s/\nu. \quad (10.3.3)$$

⁶ We use $\sigma(\mathbf{r}, \mathbf{r}')$ rather than $\tau(\mathbf{r}, \mathbf{r}')$ for the secondary site variable since it is an Ising spin variable.

Equations (10.1.5)–(10.1.8) now apply with $\mathring{K}$ replaced by the pair (K_s, L_s) and

$$\exp(4K_e) = \frac{\exp(4K_p) \cosh(2K_s + L_s) \cosh(2K_s - L_s)}{\cosh^2(L_s)}, \quad (10.3.4)$$

$$\exp(4L_e) = \exp(4L_p) \left[\frac{\cosh(2K_s - L_s)}{\cosh(2K_s + L_s)} \right]^z, \quad (10.3.5)$$

$$Z(\mathring{\mathcal{N}}, K_p, K_s, L_p, L_s) = [16 \cosh(2K_s + L_s) \cosh(2K_s - L_s) \cosh^2(L_s)]^{\frac{1}{8}Nz} \times Z^{(\text{im})}(\mathcal{N}, K_e, L_e). \quad (10.3.6)$$

Then the magnetization densities m_p and m_s on the p and s sites are

$$\begin{aligned} m_p &= - \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, K_s, L_p, L_s)}{\partial L} \right)_{K_p, K_s, L_s} \\ &= - \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, K_s, L, L_s)}{\partial L_e} \right)_{K_p, K_s, L_s} = m^{(\text{im})}(K_e, L), \end{aligned} \quad (10.3.7)$$

$$m_s = - \frac{2}{z} \left(\frac{\partial \phi(\mathring{\mathcal{N}}, K_p, K_s, L, L_s)}{\partial L_s} \right)_{K_p, K_s, L} = f(K_e, K_s, L_e, L_s), \quad (10.3.8)$$

where

$$\begin{aligned} f(K_e, K_s, L_e, L_s) &= \tanh(L_s) \\ &+ \frac{1}{4} \left[1 - \frac{2}{z} u^{(\text{im})}(K_e, L_e) \right] [\tanh(2K_s + L_s) - \tanh(2K_s - L_s) - 2 \tanh(L_s)] \\ &- \frac{1}{2} m^{(\text{im})}(K_e, L_e) [\tanh(2K_s + L_s) + \tanh(2K_s - L_s)]. \end{aligned} \quad (10.3.9)$$

The overall magnetization density is given by

$$m = \frac{2vm_p + zm_s}{2v + z}. \quad (10.3.10)$$

We need dimensionless variables to measure temperature and magnetic field. These we take to be $T/J_s = 1/K_s$ and $\mathcal{H}/J_s = L_s/K_s$. We already have the relationship $L_p = vL_p$ and now we take $\xi := J_p/J_s$, so that $K_p = \xi K_s$. It is clear that this model retains the $\mathcal{H} \leftrightarrow -\mathcal{H}$ symmetry of the Ising model, so attention is restricted to the region $\mathcal{H} \geq 0$. As $T \rightarrow 0$, $m \rightarrow (z-2v)/(z+2v)$ when $\mathcal{H} < \mathcal{H}_c$ and $m \rightarrow 1$ when $\mathcal{H} > \mathcal{H}_c$.

10.3.1 The Zero-Field Axis

From (10.3.5), $L_p = L_s = 0$ implies $L_e = 0$ and, from (10.3.4),

$$\exp(2K_e) = \exp(2\xi K_s) \cosh(2K_s), \quad (10.3.11)$$

with the spontaneous magnetization given by

$$m_s = \pm m_s^{(\text{IM})}(K_e) \left(\frac{2\nu - z \tanh(2K_s)}{2\nu + z} \right), \quad (10.3.12)$$

where $m_s^{(\text{IM})}(K_e)$ (>0) is the spontaneous magnetization of the Ising ferromagnet. Since $\mathcal{H} \rightarrow +0$ the $\pm$ sign must be chosen to make $m_s > 0$. The critical temperature, above which the spontaneous magnetization is zero, is given by solving (10.3.11) with $K_e = K_c$, the critical parameter value for the Ising model. In particular, for the triangular lattice, from (9.4.31), $K_c = \ln(3)/4 = 0.2747$. Then, when $\xi = 1$ ($J_p = J_s$), $T_c = 4.4353 J_s$ and, when $\xi = 0$ ($J_p = 0$), $T_c = 1.7449$. When $T > T_c$, $K_e < K_c$ so that $m_s^{(\text{IM})}(K_e) = 0$ while, as T decreases from T_c to 0, $\tanh(2K_s)$ increases from $\tanh(2J_s/T_c)$ to 1. The following cases can be distinguished:

- (i) $\nu < \nu_0$, (ii) $\nu = \nu_0$, (iii) $\frac{1}{2}z > \nu > \nu_0$,
 (iv) $\nu = \frac{1}{2}z$, (v) $\nu > \frac{1}{2}z$,

where

$$\nu_0 := \frac{1}{2}z \tanh(2J_s/T_c). \quad (10.3.13)$$

For (i)–(iii), $\nu < \frac{1}{2}z$ so that, in accordance with our discussion of the ground state,

$$m_p = -m_s^{(\text{IM})} = -1, \quad m_s = \frac{z - 2\nu}{z + 2\nu} > 0. \quad (10.3.14)$$

For (iv), $m_s = 0$ at $T = 0$ and, for (v),

$$m_p = +m_s^{(\text{IM})} = +1, \quad m_s = \frac{2\nu - z}{2\nu + z} > 0. \quad (10.3.15)$$

The most interesting case is (iii), where there is a *compensation temperature* $T_{\text{comp}} < T_c$, given by

$$\nu = \frac{1}{2}z \tanh(2J_s/T_{\text{comp}}), \quad (10.3.16)$$

where $m_s = 0$, because the first factor on the right-hand side of (10.3.12) is zero. For $T_{\text{comp}} < T < T_c$,

$$m_p = +m_s^{(fIM)}(K_e) > 0, \quad m_s < 0, \quad (10.3.17)$$

but, for $0 < T < T_{\text{comp}}$,

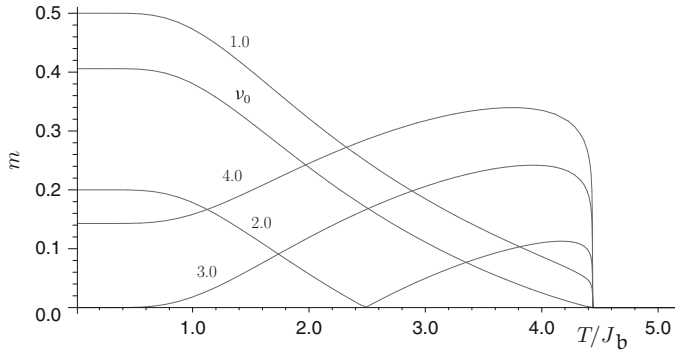


Fig. 10.2 Curves of magnetization density as a function of temperature along the zero-field axis for the triangular lattice with $\xi = 1$, parameterized by the value of ν . From (10.3.13) $\nu_0 = 1.26710$ and for the curve $\nu = 2$, from (10.3.16), $T_{\text{comp}} = 2.4853 J_s$

$$m_p = -m_s^{(\text{im})}(K_e) < 0, \quad m_s > 0. \quad (10.3.18)$$

Both m_p and m_s change sign discontinuously at $T = T_{\text{comp}}$ although $|m_p|$ and $|m_s|$ are continuous. The magnetization density (10.3.12) can be calculated for the square, honeycomb and triangular lattices using the formula for the Ising model spontaneous magnetization given in Sect. 12.3.9. Such curves for the triangular lattice with $\xi = 1$ are given in Fig. 10.2.

Since $m_s^{(\text{im})}(K_e) \sim |\Delta K_e|^\beta$, as $\Delta K_e = K_e - K_c \rightarrow 0$, β being the Ising exponent, (10.3.12) gives $m_s \sim (T_c - T)^\beta$, as $T \rightarrow T_c - 0$, except in case (ii) when $r - \frac{1}{2}z \tanh(2K_s) \sim T_c - T$, as $T \rightarrow T_c - 0$, yielding $m_s \sim (T_c - T)^{\beta+1}$. Since $0 < \beta < 1$ for both $d = 2$ and $d = 3$ lattices it follows that the plot of m_s against T at $\mathcal{H} = 0$ has a vertical tangent at $T = T_c$ except for case (ii) where the tangent is horizontal (see Fig. 10.2). The zero-field susceptibility can be deduced from (10.3.9).

10.3.2 Non-Zero Field

In addition to the segment $[0, T_c]$ of the zero-field axis, where m undergoes a discontinuous change of sign, there are first-order transition lines on which $|\mathcal{H}| \neq 0$ and for which exact relations can be derived (Bell 1974b). This arises from the fact that the mapping $(T, \mathcal{H}) \rightarrow (K_e^{-1}, L_e)$ from the plane of this ferrimagnet into the plane of the Ising model is not one-to-one. Equation (10.3.5), with $L_e = 0$, is satisfied by $L_s = 0$, but after some rearrangement it also yields

$$\exp(4K_s) = \frac{\sinh[(1 + 2\nu/z)L_s]}{\sinh[(1 - 2\nu/z)L_s]}, \quad (10.3.19)$$

which has a solution which gives a curve $\mathcal{L}(\nu)$ in the T - $\mathcal{H}$ plane when $\nu < \frac{1}{2}z$. This curve is clearly symmetrical about $\mathcal{H} = 0$ and meets the zero-temperature axis at $\pm\mathcal{H}_c$, where $\mathcal{H}_c$ is the critical field given by (10.3.3). It cuts the zero-field axis orthogonally at a temperature T_i , which, by taking the limit of the right-hand side of (10.3.19) as $L_s \rightarrow 0$, is given as the solution of

$$\nu = \frac{1}{2}z \tanh(2K_s). \quad (10.3.20)$$

For case (i), $T_i > T_c$ and, as can be seen from (10.3.16), for case (ii), $T_i = T_{\text{comp}} = T_c$ and, for case (iii), $T_i = T_{\text{comp}} < T_c$. From (10.3.4) the curve $\mathcal{C}(\xi)$ of points in the plane T - $\mathcal{H}$ which map into the Ising critical point $(K_c^{-1}, 0)$ is given by

$$\exp[4(K_c - \xi K_s)] - 1 = \frac{\sinh^2(2K_s)}{\cosh^2(L_s)}. \quad (10.3.21)$$

This curve meets the zero-field axis at $T = T_c$ and, when $\xi \neq 0$, has the infinite-field asymptote $T = J_s/K_c$. When $\xi = 0$ the curve meets the zero-temperature axis $\mathcal{H} = J_s$. The curves $\mathcal{L}(\nu)$ are independent of ξ and the curves $\mathcal{C}(\xi)$ are independent of ν . The intersection of an $\mathcal{L}(\nu)$ and a $\mathcal{C}(\xi)$, if it exists for particular values of ν and ξ , is the mapping of the Ising critical point $(K_c^{-1}, 0)$ and it corresponds to the end-point of a line of first-order transitions along $\mathcal{L}(\nu)$. The case of the triangular lattice with $\xi = 1$, shown in Fig. 10.3a, is typical of the situation where $\xi > 0$ and Fig. 10.3b is the special case $\xi = 0$, where there is no Ising interaction between primary p sites. The first-order transition lines in the T - $\mathcal{H}$ plane are the images of $(K_c^{-1}, 0)$ for $K_c^{-1} < K_c^{-1}$, so lie below the curve $\mathcal{C}(\xi)$.

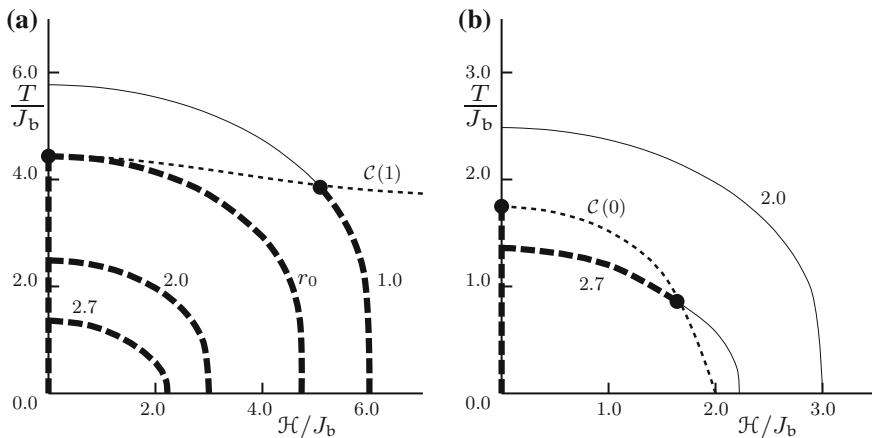


Fig. 10.3 Curves $\mathcal{L}(\nu)$, parameterized by ν for the triangular lattice. First-order transitions are indicated by *heavy broken lines*. No transition occurs along the continuous lines. Curves $\mathcal{C}(\xi)$ for **a** $\xi = 1$ and **b** $\xi = 0$ are given by *light broken lines*

For $\xi \neq 0$ (Fig. 10.3a) the curve $\nu = 1$ is an example of case (i), where there are two first-order transition lines which are mirror images of each other in the zero-field axis. The curves $\nu = 2$ and $\nu = 2.7$ are examples of case (iii). They lie entirely in the $K_c > K_c$ region and yield first-order transition lines crossing the $[0, T_c]$ segment on the zero-field at $T = T_{\text{comp}}$. The compensation point is thus the intersection of two lines of first-order transitions. The configurational energy is continuous over the segment $[0, T_c]$ of the zero-field axis but, together with the magnetization m , it is discontinuous at first-order transition points for which $\mathcal{H} \neq 0$.

For $\xi = 0$ case (i) curves like $\nu = 1$ contain no transitions except on the zero-field axis. Only case (iii) curves like $\nu = 2$ and $\nu = 2.7$ (Fig. 10.3b) intersect $\mathcal{C}(0)$. The first-order transition segment $[0, T_c]$ on the zero-field axis is crossed at the compensation point by another first-order transition line on $\mathcal{L}(\nu)$. This transition line terminates at two critical points in the regions $\mathcal{H} \geq 0$, which are reached by *reducing* the temperature from T_{comp} . They are examples of *lower critical points*, which will be encountered and described in more detail in the next section.

10.4 A Competing-Interaction Magnetic Model

Here we consider only the case $\mathcal{H} = 0$. The primary sites form a spin- $\frac{1}{2}$ *antiferromagnetic* Ising model and there is a ferromagnetic primary-secondary interaction with an Ising spin of magnitude $\frac{1}{2}q$ on the secondary sites, where q can take any positive integer value. The secondary-site variable $\sigma(\mathbf{r}, \mathbf{r}') = q, q-2, \dots, -q$ and the secondary-site interaction is $(J/q) > 0$, where the scaling factor q is introduced so that the energies of maximum alignment are $\pm J$. The Hamiltonian is (10.1.1) with

$$\hat{K} \epsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \sigma(\mathbf{r}, \mathbf{r}')) = -(K/q)\sigma(\mathbf{r}, \mathbf{r}')[\sigma(\mathbf{r}) + \sigma(\mathbf{r}')], \quad (10.4.1)$$

and $K := J/T$ ($J > 0$), with L , as it appears in (10.1.1), set to zero and K replaced by $-\nu K$ ($1 > \nu > 0$). Although the primary-secondary interaction term in (10.4.1) is taken to be ferromagnetic, an interaction of either sign tends to align the spins on the primary sites $\mathbf{r}$ and $\mathbf{r}'$. The primary-secondary interaction is thus in conflict with the antiferromagnetic primary-primary interaction, which tends to make the spins on the primary sites $\mathbf{r}$ and $\mathbf{r}'$ antiparallel.

The energy of a perfect ferromagnetic state with all primary site $\sigma(\mathbf{r}) = 1$ and all secondary site $\sigma(\mathbf{r}, \mathbf{r}') = q$ (or all $\sigma(\mathbf{r}) = -1$ and all $\sigma(\mathbf{r}, \mathbf{r}') = -q$) is $-\frac{1}{2}zN(2-\nu)J$. It is also possible to have a state of perfect antiferromagnetic order on a loose-packed primary lattice in which all primary first-neighbour pairs have $\sigma(\mathbf{r}) + \sigma(\mathbf{r}') = 0$ and the energy is $-\frac{1}{2}zN\nu J$. Since we assume $\nu < 1$, the zero-temperature ground state is ferromagnetic. We thus expect ferromagnetic long-range order to be stable in a range $(0, T_c)$, where T_c is a critical temperature. However, all the $q+1$ values of $\sigma(\mathbf{r}, \mathbf{r}')$ are equally probable when the spins on primary sites $\mathbf{r}$ and $\mathbf{r}'$ are antiparallel, contributing a term $k_B \ln(q+1)$ to the entropy. States

with primary spins antiferromagnetically ordered have a high entropy. Ordered states with this unusual property can be destabilized by either an increase or a decrease in temperature but may be stable in a range (T_{lc}, T_{uc}) , where T_{uc} and T_{lc} are respectively termed the *upper* and *lower critical temperatures*. For appropriate parameter values, this situation exists for the present model with the antiferromagnetic lower critical temperature $T_{lc} > T_{uc}$. From (10.1.4) and (10.4.1),

$$\begin{aligned}\phi(K; \pm 1, \pm 1) &= \exp(-\nu K) \sum_{\ell=0}^q \exp[(q - 2\ell)2K/q] \\ &= \exp(-\nu K) \frac{\sinh[2(q + 1)K/q]}{\sinh(2G/q)},\end{aligned}\quad (10.4.2)$$

$$\phi(K; +1, -1) = (q + 1) \exp(\nu K)$$

and, substituting into (10.1.6) and (10.1.7),

$$\exp(2K_e) = \exp(-2\nu K) \frac{\sinh[2(q + 1)K/q]}{(q + 1) \sinh(2K/q)}, \quad (10.4.3)$$

$$L_e = 0.$$

From (10.4.3),

$$\exp(2K_e) \simeq \frac{\exp[2(1 - \nu)K]}{q + 1}, \quad \text{as } K \rightarrow \infty. \quad (10.4.4)$$

Hence $K_e \rightarrow \infty$, as $T \rightarrow 0$, in accordance with the previous deduction of a ferromagnetic ground state. However, K_e does not increase monotonically with K since it follows, from (10.4.3), that

$$K_e \simeq -\nu K + \frac{q + 2}{3q} K^2, \quad \text{as } K \rightarrow 0. \quad (10.4.5)$$

For small K , $K_e < 0$ but, for large K , $K_e > 0$. It is not difficult to show, from (10.4.3), that $d^2 K_e / dK^2 > 0$ for all $K > 0$ and it follows that K_e decreases from 0 as K increases from 0 to reach a unique minimum $K_e = K_{\min} < 0$ at $K = J / T_{\min}$. Then K_e increases monotonically, passing through 0 at $K = J / T_0$ and through K_c (the critical value for the primary lattice Ising ferromagnet) at $K = J / T_c$ ($T_c < T_0$) to approach infinity with K as $K \rightarrow \infty$.

From the above it follows that $K_e > K_c$ for the temperature interval $(0, T_c)$, giving long-range ferromagnetic order when $T < T_c$. If $K_{\min} < -K_c$ then $K_e = -K_c$ at temperatures T_{uc} and T_{lc} , with $T_0 < T_{lc} < T_{\min} < T_{uc}$. Thus $K_e < -K_c$ in the interval (T_{lc}, T_{uc}) . Since we have specified a loose-packed primary lattice, there is

long-range antiferromagnetic order when $T_{lc} < T < T_{uc}$. It follows that T_{uc} and T_{lc} are respectively upper and lower critical points. The value of K_{min} , which depends on q and ν and which is given by

$$\frac{dK_c}{dK} = -\nu + \frac{q+1}{q} \coth \left[\frac{2(q+1)K}{q} \right] - \frac{1}{q} \coth \left(\frac{2K}{q} \right) = 0, \quad (10.4.6)$$

is crucial. It can be shown from (10.4.6) that, with a fixed value of q , K_{min} has a greatest lower bound equal to $-\frac{1}{2} \ln(q+1)$ for the range $0 < \nu < 1$. This bound is approached as $\nu \rightarrow 1 - 0$. Hence antiferromagnetic critical points occur over some range of ν if $\frac{1}{2} \ln(q+1) > K_c$, where K_c , of course, depends only on the primary lattice. With $q = 1$ this condition is satisfied for the $d = 3$ simple cubic and body-centred cubic lattices. With $q = 2$ it is also satisfied for the square and diamond lattices, and with $q = 3$ for the honeycomb lattice as well (Table 3.1). As an example, with a square primary lattice, $q = 3$ and $\nu = 0.8$ we have $T_{uc} = 1.0150 J$, $T_{lc} = 0.4387 J$, $T_0 = 0.1443 J$ and $T_c = 0.0882 J$. For the general case, since $dK_c/dK \neq 0$ when $K_c = \pm K_c$, $\Delta K \sim \Delta K_c$ and the heat capacity singularity at each critical point is of similar type to that at the critical point of the primary lattice Ising model. The $q = 1$ form of the present model was investigated by Nakono (1968) and Syozi (1972) discusses related models.

The temperature $T = T_0$, where K_c changes sign, is of some interest. Since $K_c = 0$ at $T = T_0$ the primary lattice spins are completely disordered, as they are at $T = \infty$. Suppose that $\langle \sigma(\mathbf{0})\sigma(\mathbf{r}) \rangle$ is the correlation between primary lattice spins whose relative position is represented by $\mathbf{r}$. As the temperature increases through T_0 , K_c changes from positive to negative and $\langle \sigma(\mathbf{0})\sigma(\mathbf{r}) \rangle$ changes from a positive quantity which decreases with $|\mathbf{r}|$ to one which alternates in sign, though its magnitude still decreases with $|\mathbf{r}|$. At $T = T_0$ itself $\langle \sigma(\mathbf{0})\sigma(\mathbf{r}) \rangle = 0$ for all $\mathbf{r}$. This is an example of a *disorder point*. Early work on these is discussed by Stephenson (1970), who cites the $d = 1$, $q = 1$ case of the present model as one example (see also Georges et al. 1986). As in the present model, disorder points in other cases are due to competing interactions which promote different kinds of short-range order. Disorder points in the eight-vertex model are discussed Sect. 9.3.2.2.

10.5 Decoration with Orientable Molecules

This section is concerned with a binary mixture which has closed-loop coexistence curves in the composition–temperature plane and thus both upper and lower critical solution temperatures. Phenomena of this type occur in real mixtures, for example glycerol-ethylbenzylamine and water-nicotine.

A primary or secondary site can be occupied either by a molecule of component A with chemical potential μ_A or one of component B with chemical potential μ_B . The Ising variable $\sigma(\mathbf{r})$ is used with $\sigma(\mathbf{r}) = +1$ indicating an A occupation and $\sigma(\mathbf{r}) = -1$ indicating a B occupation. We use the binary mixture variables given in

Sect. 3.5.1.3 and the Ising/binary mixture equivalence given by (3.5.27) and (3.5.28). However, we wish to impose symmetry between the two components and this can be achieved, without loss of generality, by taking $\varepsilon_{AA} = \varepsilon_{BB} = 0$ and $\varepsilon_{AB} = \varepsilon$. Thus, from (3.5.28), for the couplings of (10.1.1) we have $K := \varepsilon/(2T)$ and $L := (\mu_A - \mu_B)/(2T)$. The constant coupling $C := [\varepsilon - (2/z)(\mu_A + \mu_B)]/(2T)$, given by (3.5.27), features only in the calculation of the partition function. The existence of a lower critical point is likely to be an entropic phenomenon, as we saw in Sect. 10.4. To achieve the necessary high entropy it is postulated that each A or B molecule on a secondary site has ω distinguishable orientation states, labelled $1, 2, \dots, \omega$. In state 1 an active *arm* is directed towards the molecule on one of the adjacent primary sites and $\varepsilon'_{AA}, \varepsilon'_{AB}$ or ε'_{BB} is added to $\varepsilon_{AA}, \varepsilon_{AB}$ or ε_{BB} according to the A or B occupations of the secondary and the respective primary site. In state 2 the same thing is the case with respect to the other primary site. In the other $\omega - 2$ states the secondary site molecule is said to be *disoriented* and there is no additional energy contribution. Let $\varepsilon'_{AA} = \varepsilon'_{BB} = -\xi\varepsilon$ and $\varepsilon'_{AB} = -\nu\varepsilon$. We again use the Ising equivalence of Sect. 3.5.1.3, but now must include the constant contribution since the disoriented states set the energy zero. Thus, from (3.5.27) and (3.5.28), an addition factor $(\xi + \nu)K$ must be added to the Ising term together with a constant $(\xi - \nu)K$ (unaffected by the spin orientations) for a primary–secondary pair with an active arm. We attach to the secondary site $(\mathbf{r}, \mathbf{r}')$ two variables $\sigma(\mathbf{r}, \mathbf{r}') = \pm 1$ signifying, as for the primary sites, A or B occupation and $\tau(\mathbf{r}, \mathbf{r}') = 1, 2, \dots, \omega$, which specifies the orientational state. The orientations of the molecules on the primary sites are taken as fixed so that these sites have only two distinguishable states, occupation by A or B respectively. The Hamiltonian is (10.1.1) with

$$\begin{aligned} \hat{K}\varepsilon(\sigma(\mathbf{r}), \sigma(\mathbf{r}'); \sigma(\mathbf{r}, \mathbf{r}'), \tau(\mathbf{r}, \mathbf{r}')) = & -L\sigma(\mathbf{r}, \mathbf{r}') \\ & -K\{1 + (\xi - \nu)\delta^{Kr}(\tau(\mathbf{r}, \mathbf{r}') - 1)\}\sigma(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}) \\ & -K\{1 + (\xi - \nu)\delta^{Kr}(\tau(\mathbf{r}, \mathbf{r}') - 2)\}\sigma(\mathbf{r}, \mathbf{r}')\sigma(\mathbf{r}') \\ & -K(\xi + \nu)\{\delta^{Kr}(\tau(\mathbf{r}, \mathbf{r}') - 1) + \delta^{Kr}(\tau(\mathbf{r}, \mathbf{r}') - 2)\}. \end{aligned} \quad (10.5.1)$$

From (10.1.4), summing over the values of both $\sigma(\mathbf{r}, \mathbf{r}')$ and $\tau(\mathbf{r}, \mathbf{r}')$ gives

$$\begin{aligned} \phi(K, L; \pm 1, \pm 1) = & \exp(2K + L)[\omega - 2 + 2\exp(2K\xi)] \\ & + \exp(-2K - L)[\omega - 2 + 2\exp(2K\nu)], \end{aligned} \quad (10.5.2)$$

$$\phi(K, L; +1, -1) = 2 \cosh(L)[\omega - 2 + \exp(2K\xi) + \exp(2K\nu)].$$

We are interested in phase equilibrium which corresponds in the equivalent Ising model to $L_c = 0$. From (10.5.2) and (10.1.7) this is the case if $L = 0$, that is $\mu_A = \mu_B$. However, we also need the mole fractions x_A and x_B . To obtain these we must compute the partition function for general μ_A and μ_B , using (10.1.5) and (10.1.8), where in the later $g(K, L)$ is, as indicated above, replaced by $\exp(C)g(K, L)$. Now differentiating the resulting free energy with respect to μ_A and μ_B and setting $\mu_A = \mu_B$ yields

$$x_A = \frac{1}{2} \pm m_s^{(\text{im})}(K_e) X(K), \quad x_B = \frac{1}{2} \mp m_s^{(\text{im})}(K_e) X(K), \quad (10.5.3)$$

where

$$X(K) = \frac{1}{z+2} \left\{ 1 + \frac{z}{2} \frac{(\omega-2) \sinh(2K) + \exp[2(\xi+1)K] - \exp[2(\nu-1)K]}{(\omega-2) \cosh(2K) + \exp[2(\xi+1)K] + \exp[2(\nu-1)K]} \right\} \quad (10.5.4)$$

and, by substituting from (10.5.2) into (10.1.6),

$$\exp(2K_e) = \frac{(\omega-2) \cosh(2K) + \exp[2(\xi+1)K] + \exp[2(\nu-1)K]}{\omega-2 + \exp(2\xi K) + \exp(2\nu K)}. \quad (10.5.5)$$

We now assume that $\varepsilon > 0$, $\xi > 0$ and $\nu > 0$. Then the possibly stable ground states are coexistence between phases all with primary–secondary–primary triplets with one of the occupations:

- I AA~B or A~BB with energy $-(\nu-1)\varepsilon$,
- II AA~A or B~BB with energy $-\xi\varepsilon$,

where ‘~’ represents the presence of an active arm. We now explore the properties of the curves, derived from (10.5.5), of K_e as a function of T . Their limiting properties are given by

$$K_e \simeq \begin{cases} \frac{1}{2} \ln \left[\frac{\exp[\varepsilon(\xi+1)/T] + \exp[\varepsilon(\nu-1)/T]}{\exp[\varepsilon\xi/T] + \exp[\varepsilon\nu/T]} \right], & \text{as } T \rightarrow 0, \\ \frac{\varepsilon^2[\omega-2(\nu-\xi)]}{4T^2\omega}, & \text{as } T \rightarrow \infty, \end{cases} \quad (10.5.6)$$

from which it can be seen that the behaviour in the low-temperature limit is one of three types:

When $\xi < \nu - 1$, $K_e \rightarrow -\infty$ as $T \rightarrow 0$:

The stable ground state is I and on the primary lattice AB first-neighbour pairs are favoured at low temperatures, corresponding to antiferromagnetic local order in the equivalent Ising model. For a loose-packed primary lattice the ground state is one of long-range sublattice order with As occupying one sublattice, Bs the other and ground-state energy $-\frac{1}{2}zN(\nu-1)\varepsilon$. However for a primary–secondary–primary triad of sites there are also occupation states AAA or BBB, with the molecule on the secondary site disoriented, which have zero energy but an entropy $k_B \ln(\omega-2)$. Thus for high enough ω and temperature, AA or BB first-neighbour pairs on the primary lattices will be favoured and separation into A-rich and B-rich phases will occur, corresponding to ferromagnetic order in the equivalent Ising model. This is borne out by the behaviour of K_e . From (10.5.6), $K_e < 0$ at low temperature, but at high temperatures, if $\omega > 2(\nu-\xi)$, $K_e > 0$. Plots of K_e against temperature are

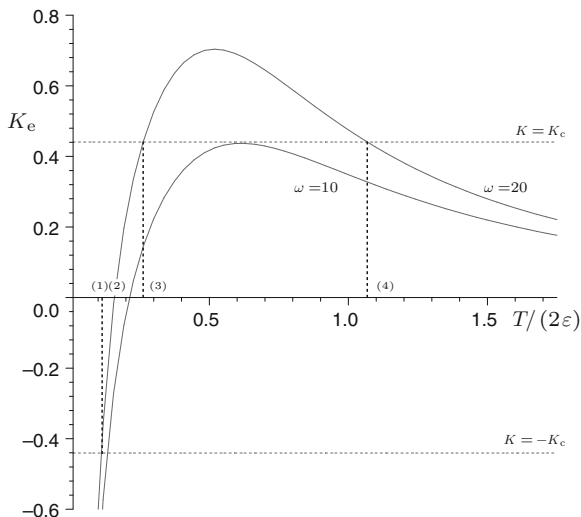


Fig. 10.4 Curves of K_e against temperature derived from (10.5.6) for $\xi = 0.1$, $\nu = 1.2$. The curves are labelled with their value of ω and the lines $K_e = \pm K_c$ are for the square lattice where, from (9.2.36), $K_c = \frac{1}{2} \ln(1 + \sqrt{2})$. The labelled points on the axis are (1) $T = T_{sc} = 0.1289(2\varepsilon)$, (2) $T = T_{do} = 0.2029(2\varepsilon)$, (3) $T = T_{lc} = 0.2618(2\varepsilon)$, (4) $T = T_{uc} = 1.0684(2\varepsilon)$

shown in Fig. 10.4 for a case where $\xi < \nu - 1$ and the high-temperature limit of K_e is positive if $\omega \geq 3$. For sufficiently large ω , the maximum value of $K_e > K_c$, the critical Ising parameter value for the primary lattice, as shown in Fig. 10.4, where for the square lattice ω must be greater than 10. The values of T at which $K_e = K_c$ are lower and upper critical temperatures T_{lc} and T_{uc} . For temperatures in the range (T_{lc}, T_{uc}) , $m_s^{(im)}(K_e) > 0$ and (10.5.3) gives the compositions of A-rich and B-rich conjugate phases. Outside this range of temperature $m_s^{(im)}(K_e) = 0$ and (10.5.3) gives $x_A = \frac{1}{2}$. Thus in the temperature–mole fraction plane there is a *closed-loop coexistence curve* (similar to the upper part of curve (c) in Fig. 10.6) symmetrical about the line $x_A = \frac{1}{2}$ with lower and upper critical temperatures for phase separation at $T = T_{lc}$ and $T = T_{uc}$ respectively. There is a *disorder point*, with $K_e = 0$, at some temperature $T_{do} < T_{lc}$ (Fig. 10.4) and, for a loose-packed primary lattice, there is a transition at a temperature T_{sc} to a sublattice ordered state when K_e passes through the value $-K_c$.

When $\xi = \nu - 1$, $K_e \rightarrow 0$ as $T \rightarrow 0$:

The stable ground state is a coexistence of I and II. The (T, K_e) plot assumes the form of curve (b) in Fig. 10.5.

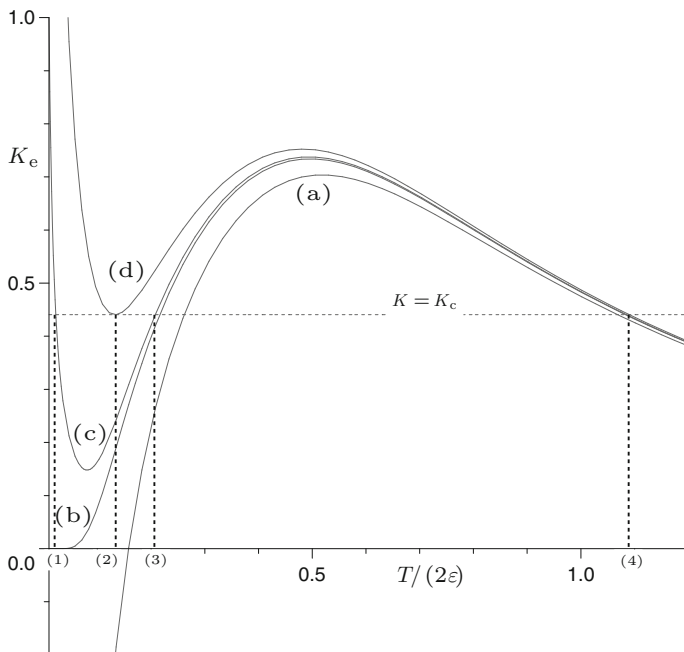


Fig. 10.5 Curves as in Fig. 10.4 for $\nu = 1.2$, with (a) $\xi = 0.1 < \nu - 1$, (b) $\xi = 0.2 = \nu - 1$, (c) $\xi = 0.21 > \nu - 1$, (d) $\xi = 0.2451$. The labelled points on the axis are for curve (c): (1) $T = T_{\text{auc}} = 0.0227(2\varepsilon)$, (3) $T = T_{\text{lc}} = 0.2079(2\varepsilon)$ and (4) $T = T_{\text{uc}} = 1.0845(2\varepsilon)$ and for curve (d): (2) $T = T_{\text{dc}} = 0.1329(2\varepsilon)$

When $\xi > \nu - 1$, $K_e \rightarrow \infty$ as $T \rightarrow 0$:

A curve where ξ is slightly greater than $\nu - 1$ is shown as (c) in Fig. 10.5. It remains close to (b) except for low temperatures where $K_e \rightarrow \infty$. There is thus a low temperature interval $(0, T_{\text{auc}})$ where $K_e > K_c$, so that phase separation occurs. The coexistence diagram, calculated from (10.5.3) and (10.5.4), is shown with a broken line in Fig. 10.6.⁷ There is a closed-loop coexistence curve with upper critical temperature T_{uc} and lower critical temperature T_{lc} . Separated from the closed-loop curve by the temperature interval $(T_{\text{auc}}, T_{\text{lc}})$, with $x_A = \frac{1}{2}$, is a low-temperature coexistence curve with (another) upper critical temperature T_{auc} . For both intervals $(T_{\text{lc}}, T_{\text{uc}})$ and $(0, T_{\text{auc}})$ there is separation into A-rich and B-rich phases, but for the upper temperature interval the phase separated state is entropically stabilized, whereas for the lower temperature interval it is energetically stabilized.

As ξ is increased again relative to $\nu - 1$, the minimum of the curve rises until it occurs on the line $K_e = K_c$, as shown in curve (d) of Fig. 10.5. The lower critical point of the upper branch of the coexistence curve and the upper critical point of

⁷ For the sake of clarity we have used $\xi = 0.235$, in Fig. 10.6, rather than $\xi = 0.21$ for the (c) curve.

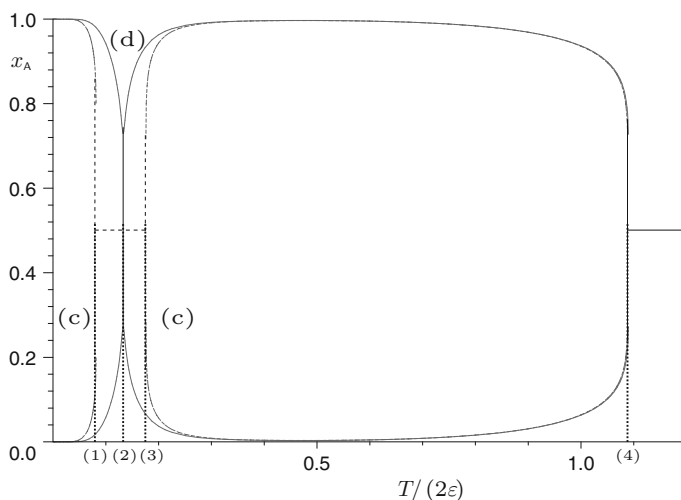


Fig. 10.6 Coexistence curves in the temperature–mole fraction plane for the curves of type (c) (broken line) and (d) (continuous line) of Fig. 10.5. Labels on the temperature axis are as in Fig. 10.5

the lower branch have coalesced into a single point at $T = T_{dc}$. This is a *critical double-point*. In Fig. 10.6 the two branches of the coexistence curve are shown (with a continuous line) as meeting with a common tangent at the critical double-point. That this is the case can be shown in the following way. Let Δx_A denote the difference between the values of x_A in the conjugate phases. From (10.5.3), $\Delta x_A \sim m_s^{(im)}(K_e)$, as $\Delta K_e = K_e - K_c \rightarrow +0$, and $m_s^{(im)}(K_e) \sim (\Delta K_e)^\beta$ where β is the Ising model critical exponent. Now K_e has a minimum at $K_e = K_c$, $T = T_d$ so that $\Delta K_e \sim (T - T_d)^2$. This applies for both positive and negative $T - T_d$, and hence on both sides of the critical double-point we have $\Delta x_A \sim |T - T_d|^{\beta_d}$ where $\beta_d = 2\beta$.⁸ Since $\beta < \frac{1}{2}$ for both $d = 2$ and $d = 3$, $\beta_d < 1$ and the two coexistence branches meet with a common tangent parallel to the x_A -axis.⁹ If ξ is again increased relative to $\nu - 1$ the minimum in the curve of K_e moves up to a value greater than K_c . The two branches of the coexistence curve are joined to form a rather bizarre ‘neck’ shape and there is now only one critical temperature.

The $\xi = 0$, $\nu > 1$ case of the above theory was developed by Wheeler (1975, 1977) and Anderson and Wheeler (1978a) fitted theoretical results to the experimental closed-loop curves for some organic mixtures, for example glycerol–ethylbenzylamine. They required the very large value $\omega = 5000$ for the number of secondary site orientation states. Although the model is highly schematic, this value of ω may reflect the large number of configurations available to fairly

⁸ This result was derived for a somewhat different decorated solution model by Bartis and Hall (1974).

⁹ If it were true that $\beta \geq \frac{1}{2}$, the coexistence curves would form a double cusp at their meeting point.

Table 10.1 The contribution to the Hamiltonian from a primary–secondary–primary triplet of sites

Phase Name	Triplet Occupation	Hamiltonian Contribution
Vapour	○—○—○	0
Intermediate-Primary	●—○—○	$-\mu/z$
Intermediate-Secondary	○—●—○	$-\mu$
Open	●—○—●	$-\varepsilon - 2\mu/z$
Partially-Packed	●—●—○	$-\nu\varepsilon - (1+z)\mu/z$
Close-Packed	●—●—●	$(\xi - 2\nu - 1)\varepsilon - (2+z)\mu/z$

The different configurations are labelled with their phase names

complex organic molecules. Barker and Fock (1953) had earlier obtained a closed-loop coexistence curve from a regular lattice model with orientation dependent interaction energies. This is not equivalent to an Ising model, and a first-order approximation was used. Closed-loop coexistence curves are observed experimentally for some aqueous mixtures (in, for example, water-nicotine) but are highly unsymmetrical with respect to the two components. Anderson and Wheeler (1978b) fitted theoretical results obtained by taking $\varepsilon_{AA} \neq \varepsilon_{BB}$ and $\varepsilon'_{AA} \neq \varepsilon'_{BB}$.

10.6 A Decorated Lattice Fluid

In this section we consider liquid/vapour phase phenomena by using a decorated lattice fluid model. Both primary and secondary lattice sites can either be unoccupied (○) or occupied (●) by a particle with chemical potential μ . The occupation states of a primary–secondary–primary triad of sites, together with their contributions to the Hamiltonian are shown in Table 10.1. We assume throughout that $\varepsilon > 0$.¹⁰ The special case of this model with $\xi = 0$ was introduced by Mermin (1971a) for the square lattice and, under the name of *Mermin's decorated lattice gas*, has been investigated and generalized to other lattices by Zollweg and Mulholland (1972) and Mulholland and Rehr (1974) (see also Wheeler 1977). For additivity of pair energies $\xi = 0$, but it is useful to include the non-additive case, particularly in order to incorporate as a special case the water-like model of Bell and Sallouta (1975) described in Sect. 10.6.2.

Each of the triplet configuration states represent a possible ground state of the system, when every triplet of the lattice is identically occupied, with the caveat that the intermediate-primary and partially-packed ground states can be perfectly realized only for loose-packed lattices. At zero temperature the Hamiltonian is the enthalpy,

¹⁰ Meaning that the equivalent Ising model on the primary sites is ferromagnetic.

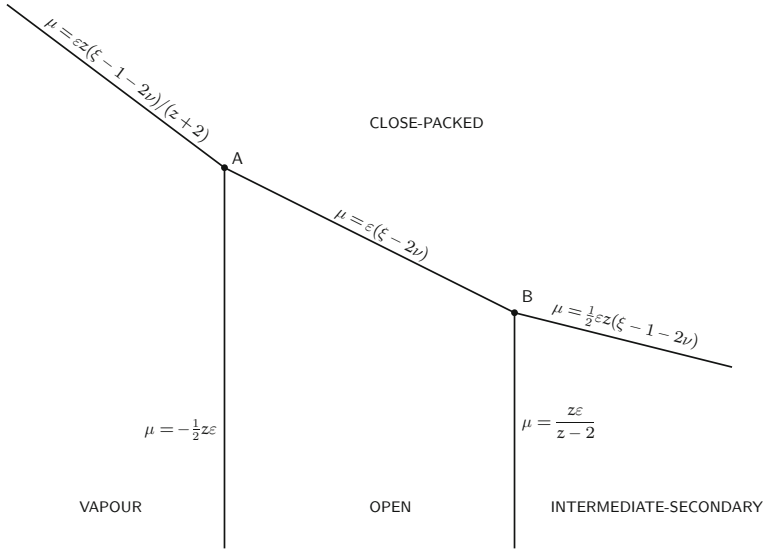


Fig. 10.7 Ground-state diagram in the μ - ν plane for the decorated lattice fluid. The coordinates of the point A of vapour/open/close-packed coexistence and of the point B of open/intermediate-secondary/close-packed coexistence are respectively (μ_A, ν_A) and (μ_B, ν_B)

which from Sect. 1.4.2 is equal to $-PN$. Thus the contributions to the Hamiltonian listed in Table 10.1 are equal to the values of $-2P/z$ in the different phases.¹¹

The stable ground state is that with least enthalpy. It is clear that the intermediate-primary state can never be stable, since if $\mu < 0$ the vapour state is more stable and if $\mu > 0$ the intermediate-secondary state is more stable. It is also not difficult to show that if $0 \leq \xi \leq 1$ (a condition which we shall henceforth assume) then the partially-packed state can also never be stable. Figure 10.7 is the ground-state diagram in the μ - ν plane where the coordinates of A and B are respectively

$$\begin{aligned} \mu_A &= -\frac{1}{2}z\varepsilon, & \nu_A &= \frac{1}{4}(z + 2\xi), \\ \mu_B &= -\frac{z\varepsilon}{z-2}, & \nu_B &= \frac{\xi(z-2) - z}{2(z-2)}. \end{aligned} \quad (10.6.1)$$

It can be seen that there are three ranges of ν giving three different types of zero-temperature behaviour:

- (I) When $\nu > \nu_A$ there is a vapour/close-packed transition at $P = 0$.
- (II) When $\nu_A > \nu > \nu_B$ there is a vapour/open transition at $P = 0$ and an open/close-packed transition at

$$P = P_0 := \frac{1}{2}\varepsilon(z + 2\xi - 4\nu). \quad (10.6.2)$$

¹¹ We are using the grand-canonical distribution and scaling with respect to the number of primary-secondary-primary triplets rather than the number of primary sites.

- (III) When $\nu_b > \nu$ there is a vapour/open transition at $P = 0$, an open/intermediate-secondary transition at $P = P_1$ and an intermediate-secondary/close-packed transition at P_2 , where

$$P_1 := \frac{\varepsilon z^2}{2(z-2)}, \quad P_2 = \frac{1}{4}\varepsilon z^2(\xi - 1 - 2\nu). \quad (10.6.3)$$

Representing the occupied and vacant states by $\sigma(\mathbf{r}) = +1$ and $\sigma(\mathbf{r}) = -1$ respectively we use, for the primary lattice, the Ising/lattice fluid equivalence given by (3.5.22). However it is convenient to define the couplings $K := \varepsilon/T$ and $L := \mu/T$, which means that in (10.1.1) K is replaced by $\frac{1}{4}K$ and L by $\frac{1}{4}(2L + zK)$. The constant coupling, given by (3.5.21), which now becomes $C := -\frac{1}{4}(4L + zK)/z$, features only in the calculation of the partition function. The Hamiltonian is (10.1.1) with

$$\begin{aligned} & \frac{1}{8} \{ \sigma(\mathbf{r}), \sigma(\mathbf{r}'); \sigma(\mathbf{r}, \mathbf{r}'), \tau(\mathbf{r}, \mathbf{r}') \} = \frac{1}{8} [1 + \sigma(\mathbf{r}, \mathbf{r}')] \\ & \times \{ 4L + K(2\nu - \xi)[2 + \sigma(\mathbf{r}) + \sigma(\mathbf{r}')] + K\xi[1 - \sigma(\mathbf{r})\sigma(\mathbf{r}')] \} \end{aligned} \quad (10.6.4)$$

and, from (10.2.10),

$$\begin{aligned} \phi(K, L; +1, +1) &= 1 + 3\mathfrak{X}^{(2\nu-\xi)}, & \phi(K, L; -1, -1) &= 1 + 3, \\ \phi(K, L; +1, -1) &= 1 + 3\mathfrak{X}^\nu, \end{aligned} \quad (10.6.5)$$

where $\mathfrak{Z} := \exp(L)$ and $\mathfrak{X} := \exp(K)$. Then, from (10.1.6) and (10.1.7), (remembering that the prefactor $\mathfrak{X}^4$ in (10.1.6) must be replaced by $\mathfrak{X}$ and $\mathfrak{Z}^4$ in (10.1.7) by $\mathfrak{Z}^2\mathfrak{X}^z$)

$$\exp(4K_c) = \frac{\mathfrak{X}(1 + \mathfrak{Z})(1 + 3\mathfrak{X}^{(2\nu-\xi)})}{(1 + 3\mathfrak{X}^\nu)^2}, \quad (10.6.6)$$

$$\exp(4L_c) = \mathfrak{Z}^2 \left\{ \frac{\mathfrak{X}(1 + 3\mathfrak{X}^{(2\nu-\xi)})}{1 + \mathfrak{Z}} \right\}^z. \quad (10.6.7)$$

The density ρ (equal to 1 when all primary and secondary sites are occupied) is given by

$$\rho = -\frac{2\mathfrak{Z}}{(2+z)} \left(\frac{\partial \phi(K, L, \tilde{\mathcal{N}})}{\partial \mathfrak{Z}} \right)_{\mathfrak{X}}, \quad (10.6.8)$$

where ϕ is the dimensionless free energy per primary lattice site of the decorated model, given in terms of the partition function (10.1.8) by (2.3.6), where, from (10.1.5) (and incorporating the contribution from C)

$$g(\mathfrak{X}, \mathfrak{Z}) = \mathfrak{Z}^{1/z} \{ \mathfrak{X}(1 + \mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}) (1 + \mathfrak{Z})(1 + \mathfrak{Z}\mathfrak{X}^\nu)^2 \}^{1/4}. \quad (10.6.9)$$

Thus, from (10.6.6)–(10.6.9)

$$\rho = p(\mathfrak{X}, \mathfrak{Z}) + B(\mathfrak{X}, \mathfrak{Z})m^{(\text{im})}(K_e, L_e) + C(\mathfrak{X}, \mathfrak{Z})u^{(\text{im})}(K_e, L_e), \quad (10.6.10)$$

where

$$\begin{aligned} p(\mathfrak{X}, \mathfrak{Z}) &= \frac{1}{4(2+z)} \left\{ 4 + \frac{z\mathfrak{Z}}{1+\mathfrak{Z}} + \frac{z\mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}}{1+\mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}} + \frac{2z\mathfrak{Z}\mathfrak{X}^\nu}{1+\mathfrak{Z}\mathfrak{X}^\nu} \right\}, \\ B(\mathfrak{X}, \mathfrak{Z}) &= \frac{1}{2(2+z)} \left\{ 2 + \frac{z\mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}}{1+\mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}} - \frac{z\mathfrak{Z}}{1+\mathfrak{Z}} \right\}, \\ C(\mathfrak{X}, \mathfrak{Z}) &= -\frac{\mathfrak{Z}}{2(2+z)} \left\{ \frac{1}{1+\mathfrak{Z}} + \frac{\mathfrak{X}^{(2\nu-\xi)}}{1+\mathfrak{Z}\mathfrak{X}^{(2\nu-\xi)}} - \frac{2\mathfrak{X}^\nu}{1+\mathfrak{Z}\mathfrak{X}^\nu} \right\} \end{aligned} \quad (10.6.11)$$

and $m^{(\text{im})}(K_e, L_e)$ and $u^{(\text{im})}(K_e, L_e)$ are respectively the magnetization and internal energy per primary lattice site on the equivalent Ising model (the latter being made dimensionless by dividing by J_e). We denote the curve in the μ – T plane derived from (10.6.6) by setting $K_e = K_c$, with specific values of z , ν and ξ , as $\mathcal{C}(z, \nu, \xi)$ and the corresponding curve, derived from (10.6.7) by setting $L_e = 0$, as $\mathcal{L}(z, \nu, \xi)$. The first-order transition lines in the μ – T plane are the parts of $\mathcal{L}(z, \nu, \xi)$ lying in the region where $K_e > K_c$, that is below $\mathcal{C}(z, \nu, \xi)$. This is always a simply-connected curve with asymptotes $T \rightarrow 1/(4K_c)$ as $\mu \rightarrow \infty$ and $T \rightarrow (1 - \xi)/(4K_c)$ as $\mu \rightarrow -\infty$. Critical temperatures occur at the intersection of $\mathcal{L}(z, \nu, \xi)$ with $\mathcal{C}(z, \nu, \xi)$. We now consider the three cases I–III.

10.6.1 Case I: A Single Vapour/Liquid Transition

Here there is a first-order transition line from the vapour phase to a close-packed liquid phase (Fig. 10.8). The conjugate vapour and liquid densities, ρ_v and ρ_ℓ respectively, are given by taking L_e in (10.6.8) and setting $m^{(\text{im})}(K_e, 0) = -m_s^{(\text{im})}(K_e)$ and $m^{(\text{im})}(K_e, 0) = +m_s^{(\text{im})}(K_e)$. Since $u^{(\text{im})}(K_e, K_e)$ is continuous across $\mathcal{L}(z, \nu, \xi)$, the coexistence width

$$\rho_w := \rho_\ell - \rho_v = 2B(\mathfrak{X}, \mathfrak{Z})m_s^{(\text{im})}(K_e). \quad (10.6.12)$$

Since $m_s^{(\text{im})}(K_e) \sim (\Delta K_e)^\beta$, as $K_e \rightarrow K_c + 0$, and, since K_e and L_e are regular functions of μ and T ,

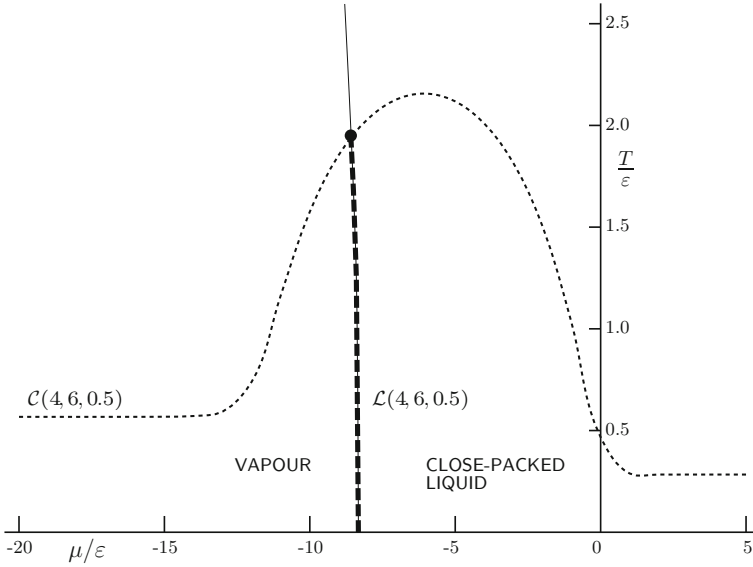


Fig. 10.8 Curves $K_e = K_c$ and $L_e = 0$ for the square lattice with $\nu = 6$ and $\xi = 0.5$

$$\rho_w \sim (\Delta T)^\beta, \quad \text{as } T \rightarrow T_c - 0. \quad (10.6.13)$$

This is similar to the behaviour of the simple lattice fluid, but there are important differences, both globally and in the neighbourhood of the critical points. The hole–particle symmetry of the simple lattice fluid, described in Sect. 3.5.1.2, is lost (except for a special case to be discussed in Sect. 10.6.3) and the coexistence curves in the ρ – T plane are no longer symmetrical about $\rho = \frac{1}{2}$. The coexistence diameter

$$\rho_d := \frac{1}{2}(\rho_\ell + \rho_v) \quad (10.6.14)$$

is, from (3.5.24) and (3.5.25), equal to $\frac{1}{2}$ for the simple lattice fluid, but for the decorated model, from (10.6.10) and (10.6.11),

$$\rho_d = p(\mathfrak{X}, \mathfrak{Z}) + C(\mathfrak{X}, \mathfrak{Z})u^{(\text{im})}(K_e, 0). \quad (10.6.15)$$

The empirical *law of the rectilinear diameter* (Sect. 3.5.1.2) implies that $d\rho_d/dT$ remains finite as $T \rightarrow T_c - 0$. This law is satisfied by the simple lattice fluid, since $d\rho_d/dT = 0$ for all $T < T_c$. However, for the decorated model, the right-hand side of (10.6.15) is not a linear function of T near the critical point. Unless $C(\mathfrak{X}, \mathfrak{Z}) = 0$, the derivative $d\rho_d/dT$ contains a term proportional to $du^{(\text{im})}(K_e, 0)/dK_e$ and, from (4.4.12) and (10.2.23),

$$\frac{du^{(IM)}(K_c, 0)}{dK_c} \sim \begin{cases} (\Delta K_c)^{-\alpha'}, & \text{for } d \neq 2, \\ -\ln |\Delta K_c|, & \text{for } d = 2. \end{cases} \quad (10.6.16)$$

Hence

$$\frac{d\rho_d}{dT} \sim \begin{cases} (\Delta T)^{-\alpha'}, & \text{for } d \neq 2, \\ -\ln |\Delta T|, & \text{for } d = 2, \end{cases} \quad (10.6.17)$$

where, for $d = 3$, $0 < \alpha < \frac{1}{2}$. So for $d = 2$ and 3 , $d\rho_d/dT \rightarrow \infty$ as $T \rightarrow T_c - 0$. The special case of this result for the square lattice, with no primary–primary interaction and $\xi = 0$, was obtained by Mermin (1971a) who also showed (Mermin 1971b) that a similar result occurs for a model in which the secondary sites are occupied by bars. This indicates that a non-singular diameter is an artifact of the simple lattice fluid. As pointed out in Sect. 3.5.1.2, more general arguments for the existence of the singularity have been developed. However, experimentally the situation is still not completely resolved (Zollweg and Mulholland 1972; Rowlinson and Swinton 1982).

10.6.2 Case II: A Water-Like Model

This case was investigated by Bell and Sallouta (1975) as a model for water, by taking the open state, with all primary sites occupied and all secondary sites empty, to represent a completely hydrogen-bonded ice-like structure. The interaction energy $-\varepsilon$ of a primary–primary pair of molecules with the intervening secondary site empty is attributed to the formation of a hydrogen bond, whereas the primary–secondary pair energy $-\nu\varepsilon$ is due to the non-bonding interaction. It is assumed that no hydrogen bond can form between a first-neighbour primary pair of molecules with the intervening secondary site occupied so that $\xi = 1$ and, from (10.6.1), case II would correspond to $(z + 2)/4 > \nu > -1/(z - 2)$. However, we shall follow Bell and Sallouta and take $\nu > 0$, meaning that the secondary–secondary interaction is attractive. The close-packed state is taken to represent the liquid phase.

This model is less realistic structurally than the water model treated in Sect. 7.6 and less satisfactory for fluid phases in that the open structure is imposed *ab initio* rather than arising as short-range order due to bonding between appropriately placed and oriented molecules. However, it does have the advantage that, because of the equivalence to the Ising model, some accurate results can be obtained for comparison with those derived from standard approximations.¹²

¹² This decorated model can be looked on as one of the class of *interstitial models* (see, for instance, Perram 1971) in which the open structure is disordered by displacement of molecules from ‘framework’ onto interstitial sites.

Bearing in mind, as indicated above, that the Hamiltonian contributions in Table 10.1, at $T = 0$, are equal to $-2P/z$, the chemical potentials in the open and close-packed ground states can be expressed in the form

$$\mu^{(o)} = P - \frac{1}{2}z\varepsilon \quad \mu^{(cp)} = \frac{2}{z+2}(P - 2z\nu\varepsilon). \quad (10.6.18)$$

From (10.6.2) and Table 10.7 an open/close-packed transition occurs at $T = 0$ when

$$P = P_0 := \frac{1}{2}(z + 2 - 4\nu) > 0, \quad \mu = \mu_0 := \varepsilon(1 - 2\nu) \quad (10.6.19)$$

and a vapour/open transition at $T = 0$ when

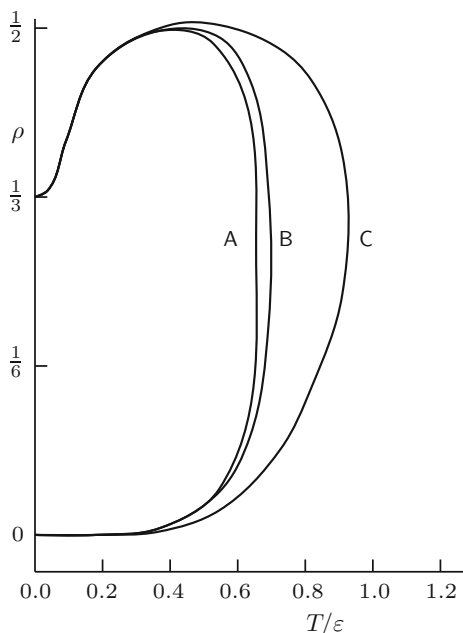
$$P = 0, \quad \mu = -\frac{1}{2}z\varepsilon. \quad (10.6.20)$$

It can be shown from (10.6.7) that the $L_e = 0$ locus $\mathcal{L}(z, \nu, 1)$ in the T - μ plane meets the zero-temperature axis at $\mu = -\frac{1}{2}z\varepsilon$. However the open/close-packed transition point $\mu = \mu_0$ on the zero-temperature axis is an isolated singularity. It cannot be the terminus of any branch of the $\mathcal{L}(z, \nu, 1)$ because, at low temperatures, the primary lattice is nearly full in both the open and close-packed states and so both correspond to $L_e > 0$.

Bell and Sallouta (1975) used the method described for case I to determine the vapour/open first-order transition curve and critical point and then used (10.6.10) to compute the densities of the coexistent phases. They obtained results using the exact values of the zero-field magnetization and internal energy for the square and honeycomb lattices and values derived from the series work of Essam and Sykes (1963) and Baker and Gaunt (1967) for the diamond lattice. Their results for this last case are shown as curve A in Fig. 10.9.¹³ The terminal points $\rho = 0$ and $\rho = \frac{1}{3}$ correspond to the vapour and open structure phases respectively. The most striking feature, apart from the marked lack of symmetry about $\rho = \frac{1}{2}$, is the pronounced maximum on the liquid density branch. Unfortunately the calculation of (ρ, T) isobaric curves requires Ising model results away from the $\mathcal{H} = 0$ axis. However, near the maximum on the liquid branch of the coexistence curve, the first-order plot is quite close to the accurate one. First-order calculations show maxima on the adjacent isobaric (ρ, T) curves. By considering perturbations from the perfect open structure, Bell and Sallouta (1975) showed that near $T = 0$ the density ρ increases with T on isobars for $0 < P < P_0$, provided that $\nu > \frac{1}{2}$. This implies a density maximum on any continuous isobar in this pressure range. Since P_0 is considerably greater than the critical pressure P_c , density anomalies occur well into the supercritical region, in accordance with experimental results. Compressibility minima are also found.

¹³ Curves B and C respectively, given for comparison, show the results of using formulae from first-order and mean-field approximations for $m_s^{(im)}(K_c)$ and $u^{(im)}(K_c, 0)$.

Fig. 10.9 Coexistence curves in the temperature–density plane for the water model for the diamond lattice with $\nu = 1.8$. (Reprinted from Bell and Sallouta 1975, by permission of the publisher Taylor and Francis Ltd.)



There is thus good qualitative agreement between the results for this model and those for the model of Sect. 7.6, where first-order approximation methods had to be used exclusively.

10.6.3 Case III: Maxithermal, Critical Double and Cuspoidal Points

When $\mathcal{C}(z, \nu, \xi)$ and $\mathcal{L}(z, \nu, \xi)$ are both single curves with a unique intersection in the μ – T plane, the phase diagram in this plane is topologically similar to that of a simple lattice fluid, though as we have shown for case I, there are important differences in properties. However, for case III there are three phase transitions at zero temperature and the mapping from the μ – T plane of the decorated model to the K_c^{-1} – L_c plane of the Ising model is no longer one-to-one and $\mathcal{L}(z, \nu, \xi)$ has several distinct branches. Consider now the case $\xi = 0$, when the pair energies are additive.¹⁴ Defining

$$x := \mathfrak{X}^\nu = \exp(\nu\varepsilon/T) \qquad y := 3\mathfrak{X}^\nu = \exp[(\mu + \nu\varepsilon)/T], \qquad (10.6.21)$$

¹⁴ This has been treated extensively by Mulholland and Rehr (1974).

(10.6.6) and (10.6.7) can be re-expressed in the form

$$\exp(4K_e - \varepsilon/T) - 1 = X(x, y), \quad (10.6.22)$$

$$\exp[4L_e - \varepsilon z(v - v^*)/(v^*T)] = Y(x, y), \quad (10.6.23)$$

where

$$X(x, y) := \frac{x + x^{-1} - 2}{y + y^{-1} + 2} = X(x, y^{-1}), \quad (10.6.24)$$

$$Y(x, y) := y^2 \left(\frac{1 + xy}{x + y} \right)^z = \frac{1}{Y(x, y^{-1})} \quad (10.6.25)$$

and

$$v = v^* := -\frac{z}{z - 2}. \quad (10.6.26)$$

It follows from the symmetry property of $X(x, y)$ that the curve $\mathcal{C}(z, v, 0)$

- (i) is symmetrical about the line $\mu = -v\varepsilon$ ($y = 1$),
- (ii) has the line $T = \varepsilon/(4K_e)$ as asymptote when $\mu \rightarrow \pm\infty$,
- (iii) lies entirely above this asymptote,
- (iv) has a maximum at $\mu = -v\varepsilon$,

as is shown by the example in Fig. 10.10. Although $Y(x, y)$ also has a symmetry property it will be reflected in the form of the curve $\mathcal{L}(z, v, 0)$ derived from (10.6.23) with $L_e = 0$ only when $v = v^*$ giving

$$L_e(x, y) = -L_e(x, y^{-1}). \quad (10.6.27)$$

The case $v = v^*$:

The curve $\mathcal{L}(z, v^*, 0)$ is given by the intersections of $Y = Y(x, y)$ plotted as a function of y for fixed x with the line $Y = 1$.¹⁵ For high temperatures ($x = \exp(v^*\varepsilon/T)$ close to unity) Y is monotonically increasing with only one solution $y = 1$ for $Y(x, y) = 1$. This solution exists for all x and so one branch of $\mathcal{L}(z, v^*, 0)$ is the entire line $\mu = -v^*\varepsilon$ in the μ - T plane. As T (and x) decrease towards zero the curve of $Y(x, y)$ develops a loop with two additional solutions $y(x)$ and $1/y(x)$, giving two other branches of $\mathcal{L}(z, v^*, 0)$ symmetrically placed about $\mu = -v^*\varepsilon$ in the μ - T plane. This is shown in Fig. 10.10 for the square lattice, although the topology and

¹⁵ There are interesting similarities between this analysis and the discussion of solution of the Bethe-pair approximation for the Ising model using the function $g(\tau)$ in Sect. 7.3.

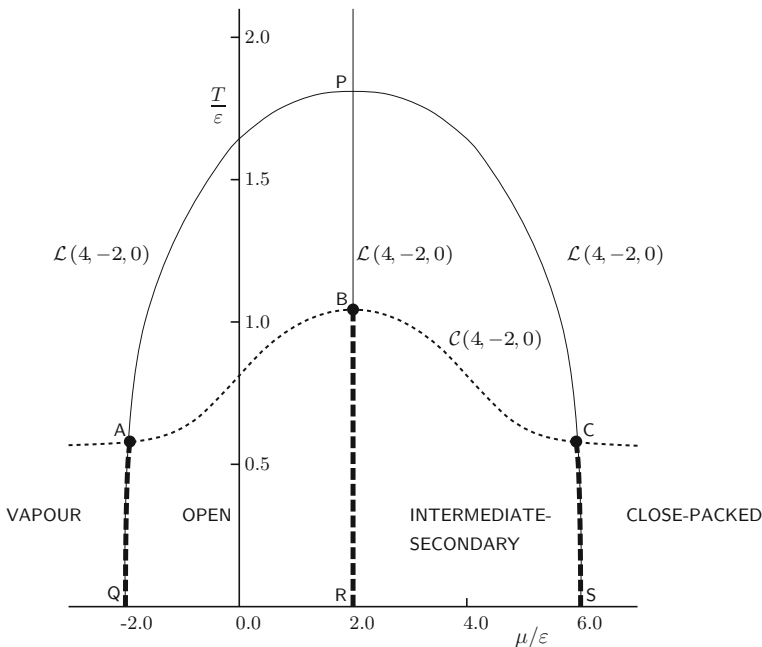


Fig. 10.10 Curves $K_e = K_c$ and $L_e = 0$ for the square lattice with $v = v^* = -2.0$ and $\xi = 0$

symmetry of the curves are not confined to that lattice. The onset of the symmetric branches of $\mathcal{L}(z, v^*, 0)$ at **P** corresponds to a point inflexion in $Y(x, y)$ at $y = Y = 1$ when

$$x = x_0 := \frac{z-2}{z+2}, \quad \text{giving} \quad T = T_0 := v\varepsilon \left[\ln \left(\frac{z-2}{z+2} \right) \right]^{-1}. \quad (10.6.28)$$

When K_e is calculated at **P** it proves to be less than K_c for all common lattices. So, as is shown for the square lattice in Fig. 10.10, there are three distinct first-order transition curves, **RB** on the line $\mu = -v^*\varepsilon$ with **QA** and **SC** placed symmetrically on either side. Taking the Ising model magnetization and internal energy densities are functions of x and y it follows from (10.6.27) that $m^{(\text{im})}(K_e, L_e)(x, y^{-1}) = -m^{(\text{im})}(K_e, L_e)(x, y)$ and $u^{(\text{im})}(K_e, L_e)(x, y^{-1}) = u^{(\text{im})}(K_e, L_e)(x, y)$. In a similar way, from (10.6.11), $p(x, y^{-1}) = 1 - p(x, y)$, $B(x, y^{-1}) = B(x, y)$ and $C(x, y^{-1}) = -C(x, y)$. It follows from (10.6.10) that

$$\rho(x, y^{-1}) = 1 - \rho(x, y). \quad (10.6.29)$$

Thus there is particle-hole symmetry and the phase pattern in the ρ - T plane is symmetrical about $\rho = \frac{1}{2}$. Since $P = 0$ at **Q**, **QA** can be regarded as the vapour/liquid transition line. On **RB** the conjugate densities are the limits of $\rho(x, y)$ and $\rho(x, y^{-1})$

as $y \rightarrow 1$, giving, from (3.5.25) and (10.6.29), $\rho_d = \frac{1}{2}$. There is no singularity in the diameter which is attributable to the fact that, for $\xi = 0$, $C(x, 1) = 0$. However, $C(x, y) \neq 0$ on the lines QA and SC, so that in both cases there is a singular diameter.

The case $\nu \neq \nu^*$:

To find the curve $\mathcal{L}(z, \nu, 0)$ we again use the function $Y(x, y)$, but, from (10.6.23), we are now looking for solutions to

$$Y(x, y) = \exp[\varepsilon z(\nu - \nu^*)/(\nu^* T)]. \quad (10.6.30)$$

In case III, from (10.6.1), $\nu < \frac{1}{2}\nu^* < 0$. Although the hole-particle symmetry is no longer present, it can be seen from the definitions of x and y in (10.6.21) and the defining equations (10.6.22) and (10.6.30) for $\mathcal{C}(z, \nu, 0)$ and $\mathcal{L}(z, \nu, 0)$ that there is a ‘hole-particle mapping’ $(\varepsilon, \nu) \leftrightarrow (\varepsilon', \nu')$ given by

$$\nu\varepsilon = \nu'\varepsilon', \quad (\nu - \nu^*)/\nu = -(\nu' - \nu^*)/\nu', \quad (10.6.31)$$

which reflects μ about the line $\mu = -\nu\varepsilon$ and makes the hole-particle phase interchanges vapour $\leftrightarrow$ close-packed and open $\leftrightarrow$ intermediate-secondary. With $\nu \in [\nu^*, -\infty)$, $\nu' \in (\frac{1}{2}\nu^*, \nu^*]$. So we can concentrate on the case $\nu < \nu^*$, knowing that our conclusions with respect to the topology of the phase diagram map into a case with $\nu^* < \nu < \frac{1}{2}\nu^*$.

As ν decreases from ν^* the right-hand side of (10.6.30) increases from unity and $\mathcal{L}(z, \nu, 0)$ breaks into a closed loop from Q to R with a second branch beginning at S and asymptotic to $\mu = -\varepsilon\nu$. Initially, as shown in Fig. 10.11a,¹⁶ the only effect on the first-order transition lines is to remove the symmetry and RB will like QA and SC now have a singular diameter. A further reduction of ν yields the situation shown in Fig. 10.11b, where $\mathcal{C}(z, \nu, 0)$ passes through the maximum of the Q–R loop of $\mathcal{L}(z, \nu, 0)$. The point B becomes a special type of critical point called a *cuspid point*, where both branches of the coexistence curve in the ρ – T plane approach the critical point from densities greater than the critical value. As ν is further reduced $\mathcal{C}(z, \nu, 0)$ cuts the Q–R loop of $\mathcal{L}(z, \nu, 0)$ at two points A and A' to the left of the maximum point B (Fig. 10.11c). These are both critical points and B becomes a *maxithermal point*, which is like two coinciding cusp points with coexistent regions in the ρ – T plane approaching from above and below the critical density. The maxithermal point persists as ν is further reduced. Figure 10.11d shows the case where $\mathcal{C}(z, \nu, 0)$ touches the Q–R loop of $\mathcal{L}(z, \nu, 0)$ at a single point A, which is the termination of two

¹⁶ The diagrams in Fig. 10.11 are, unlike Fig. 10.10, qualitative rather than quantitative. Accurate calculations for the simple cubic lattice are described by Mulholland and Rehr (1974) who use them to plot diagrams for all these cases in the ρ – T plane. They also investigate the ‘antiferromagnetic’ case $\varepsilon < 0$, $\xi = 0$.

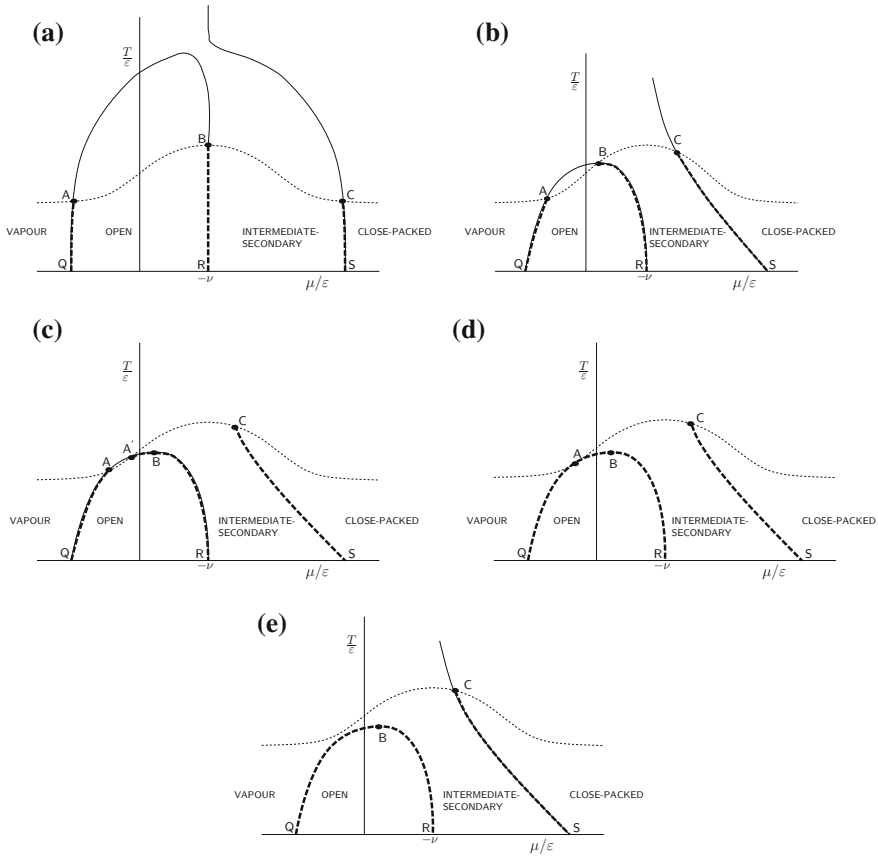


Fig. 10.11 Figures a–e show the evolution away from the symmetric phase diagram in Fig. 10.10, as ν decreases from $\nu = \nu^*$

coexistence regions and is called a *critical double-point*. Finally $\mathcal{C}(z, \nu, 0)$ no longer intersects the Q–R loop of $\mathcal{L}(z, \nu, 0)$ and apart from the critical point on SC the only critical point is the maxithermal point at B. However, if $\mathcal{C}(z, \nu, 0)$ passes sufficiently close to $\mathcal{L}(z, \nu, 0)$ it has the effect of producing a ‘neck’ in the coexistence region for Q–B (Mulholland and Rehr 1974, Fig. 10).

Chapter 11

Transfer Matrices: Incipient Phase Transitions

11.1 The Transfer Matrix Formulation

Let $\mathcal{N}$ be a two-dimensional lattice of N sites with unit vectors $\mathbf{i}$ and $\mathbf{j}$ in the two lattice directions. It is pictured as consisting of N_1 identical non-overlapping *slices* $\mathcal{N}_i$, $i = 0, 1, \dots, N_1 - 1$ parallel to $\mathbf{j}$ and consisting of N_2 sites, with $N_1 N_2 = N$. In the simplest case of a square $N_1 \times N_2$ lattice the slices can consist simply of rows of N_2 sites and a one-dimensional lattice is included as the case $N_2 = 1$. The lattice growth achieved by the use of transfer matrices is to the infinite limit for N_1 in the direction of $\mathbf{i}$.¹ Let $\tau(i)$ denote the microstate of slice i with interactions restricted to contributions from within slices and between neighbouring slices. Then the Hamiltonian can be expressed in the form

$$\hat{H}(\tau) := \hat{b}(\tau(N_1 - 1), \tau(0)) + \sum_{i=0}^{N_1-2} \hat{h}(\tau(i), \tau(i+1)), \quad (11.1.1)$$

where $\hat{h}(\tau(i), \tau(i+1))$ contains the contribution in the bulk of the system from interactions between $\mathcal{N}_i$ and $\mathcal{N}_{i+1}$ and *half* the contributions from within these two slices. The term $\hat{b}(\tau(N_1 - 1), \tau(0))$ contains any boundary contributions at $\mathcal{N}_0$ and $\mathcal{N}_{N_1-1}$, including half the contributions from interactions within these slices. By varying the form of $\hat{b}(\tau(N_1 - 1), \tau(0))$ the model can incorporate the cases where $\tau(0)$ and $\tau(N_1 - 1)$ are fixed or are allowed to vary freely (open boundary condition) or where periodic boundary conditions are applied so that $\hat{b}(\tau(N_1 - 1), \tau(0)) = \hat{h}(\tau(N_1 - 1), \tau(0))$. Let

$$\begin{aligned} V(\tau(i), \tau(i')) &:= \exp[-\hat{h}(\tau(i), \tau(i'))], \\ B(\tau(i), \tau(i')) &:= \exp[-\hat{b}(\tau(i), \tau(i'))], \end{aligned} \quad (11.1.2)$$

¹ Here and in Chap. 12 we picture the lattice oriented so the growth is upwards with the rows of N_2 sites horizontal. In Chap. 14 it is more convenient to picture these rows as vertical columns with growth to the right.

and M be the number of states of a slice, which are denoted by $\tau_m, m = 1, 2, \dots, M$. Using the notation of Sect. 17.2, the $M \times M$ dimensional *transfer matrix* V and *boundary matrix* B are given by

$$\langle m|V|m' \rangle := V(\tau_m, \tau_{m'}), \quad \langle m|B|m' \rangle := B(\tau_m, \tau_{m'}). \quad (11.1.3)$$

11.1.1 The Eigen Problem

It is not in general the case that V is symmetric. However its elements are real and thus, if it commutes with its transpose, it is normal, and this condition, which we assume (and which needs to be checked in particular applications), is sufficient for it to be diagonalizable (see Sect. 17.2) in the form

$$X^\dagger V X = \Lambda, \quad (11.1.4)$$

where X is unitary and has columns which are the right eigenvectors $\{x_m\}$ of V and the diagonal matrix Λ has elements $\{\Lambda_m\}$ which are the eigenvalues. The rows of $Y := X^\dagger$ are the left eigenvectors $\{y_m^t\}$ of V . In general some of the eigenvalues of V will be complex. Since, however, V has only real elements, the eigenvectors corresponding to the conjugate $\overline{\Lambda_m}$ of some complex eigenvalue Λ_m will be $\overline{y_m}$ and $\overline{x_m}$ and the elements of the eigenvectors corresponding to a real eigenvalue will be real. We suppose that the eigenvalues are labelled in order of decreasing magnitude.

If V is a positive matrix it has, according to Perron's theorem (Sect. 17.2.4), a (real) positive eigenvalue larger in magnitude than any other eigenvalue with left and right eigenvectors with positive elements.² The transfer matrix of the Ising model satisfies this condition. However, in the case of the eight-vertex model discussed in Chap. 12, the transfer matrix is merely non-negative and also reducible (Sect. 12.3.5.2) so Frobenius generalization of Perron's theorem is inapplicable and we might expect, and in fact find (Sect. 12.3.6.6) cases where the largest eigenvalue is degenerate. In general we shall assume that the transfer matrices of systems of interest are simple and non-negative. For such a matrix a useful result (Sect. 17.2.4) is that it has an eigenvalue equal to its spectral radius. We now introduce a partial ordering of eigenvalues with $|\Lambda_m| \geq |\Lambda_{m'}|$, for $m < m'$, and *principal eigenvalue* $\Lambda_1 = \mathcal{SR}(V)$.

² It is important to emphasize that this non-degeneracy of the largest eigenvalue is for *finite* M . Degeneracy in the limit $N_2 \rightarrow \infty$ (and hence $M \rightarrow \infty$) is crucial for the occurrence of phase transitions as we shall see in the discussion in Sect. 11.2 and in the treatment of the two-dimensional Ising model in Sect. 12.12.

11.1.2 The Partition Function

The partition function is³

$$\begin{aligned} Z(\mathcal{N}) &:= \sum_{\{\boldsymbol{\tau}\}} B(\boldsymbol{\tau}(N_1 - 1), \boldsymbol{\tau}(0)) \prod_{i=0}^{N_1-2} V(\boldsymbol{\tau}(i), \boldsymbol{\tau}(i+1)), \\ &= \text{Trace}\{\mathbf{B}\mathbf{V}^{N_1-1}\}. \end{aligned} \quad (11.1.5)$$

We shall for the sake of simplicity throughout this chapter assume the application of periodic boundary conditions, so that $\mathbf{B} = \mathbf{V}$ and

$$Z(\mathcal{N}) = \text{Trace}\{\mathbf{V}^{N_1}\} = \sum_{m=1}^M [\Lambda_m]^{N_1}. \quad (11.1.6)$$

However, it is necessary for development of the finite-lattice series method in Sect. 14.5, to note the case where the lattice has lines of boundary sites at both ends (First and Last) so that

$$\hat{b}(\boldsymbol{\tau}_m, \boldsymbol{\tau}_{m'}) := \hat{b}^{(\text{F})}(\boldsymbol{\tau}_{m'}) + \hat{b}^{(\text{L})}(\boldsymbol{\tau}_m). \quad (11.1.7)$$

With the vectors $\mathbf{b}^{(\text{F})}$ and $\mathbf{b}^{(\text{L})}$ having elements $\exp[-\hat{b}^{(\text{F})}(\boldsymbol{\tau}_m)/T]$ and $\exp[-\hat{b}^{(\text{L})}(\boldsymbol{\tau}_m)/T]$ respectively

$$\mathbf{B} = \mathbf{b}^{(\text{L})} \otimes [\mathbf{b}^{(\text{F})}]^{(\text{T})} \quad (11.1.8)$$

and, from (11.1.5),

$$Z(\mathcal{N}) = [\mathbf{b}^{(\text{F})}]^{(\text{T})} \mathbf{V}^{N_1-1} \mathbf{b}^{(\text{L})}. \quad (11.1.9)$$

This expression can also be re-expressed in terms of powers of the eigenvalues of $\mathbf{V}$, so in both cases the partition function is dependent on N_1 only through the powers of the eigenvalues, although, of course, the number and form of the eigenvalues are dependent on N_2 .

We return to the case of periodic boundary conditions where the partition function is given by (11.1.6) and consider three examples:

Case A: $\mathbf{V}$ is irreducible and primitive ($p = 1$):

Λ_1 is the only eigenvalue on the spectral circle $SC(\mathbf{V})$ and, defining scaled eigenvalues

$$\Omega_m := \Lambda_m / \Lambda_1, \quad (11.1.10)$$

³ The displayed arguments of functions are treated rather flexibly, with $\mathcal{N}$ sometimes replaced by N_1 , N_2 or, in the case where the limit $N_1 \rightarrow \infty$ has been taken, $\mathcal{L}_1$, N_2 .

in the limit $N_1 \rightarrow \infty$, $|\Omega_m|^{N_1} \rightarrow 0$, for $m > 1$, giving

$$Z(\mathcal{N}) \simeq [\Lambda_1]^{N_1}. \quad (11.1.11)$$

Case B: V is irreducible period $p > 1$:

There are p eigenvalues $\Lambda_1 \exp(2\pi i k/p)$, $k = 0, 1, \dots, p-1$, on $\mathcal{SC}(V)$ and, with M divisible by p , the remaining eigenvalues also fall into sets of p eigenvalues of equal magnitude separated by the same phase factor. V^{N_1} is traceless unless N_1 is divisible by p and restricting N_1 to such values and using (11.1.10), in the limit $N_1 \rightarrow \infty$, $|\Omega_m|^{N_1} \rightarrow 0$, for $m > p$, giving

$$Z(\mathcal{N}) \simeq p[\Lambda_1]^{N_1}. \quad (11.1.12)$$

Case C: Λ_1 is of degeneracy ω :

A necessary but not sufficient condition for this to be the case is that V is reducible to a block-wise upper triangular matrix with q non-reducible square blocks on the diagonal, where $q \geq \omega$. Since the diagonal blocks are irreducible they each have a non-degenerate largest eigenvalue, of which ω are equal to Λ_1 . For simplicity we suppose these blocks are primitive. Using (11.1.10), in the limit $N_1 \rightarrow \infty$, $|\Omega_m|^{N_1} \rightarrow 0$, for $m > \omega$, giving

$$Z(\mathcal{N}) \simeq \omega[\Lambda_1]^{N_1}. \quad (11.1.13)$$

The factors p in (11.1.12) and ω in (11.1.13) have a negligible effect on the free-energy density, which is given in each case by

$$\phi(\mathcal{L}_1, N_2) = - \lim_{N_1 \rightarrow \infty} \frac{\ln[Z(\mathcal{N})]}{N_1 N_2} = - \frac{1}{N_2} \ln[\Lambda_1], \quad (11.1.14)$$

where $\mathcal{L}_1$ indicates that the thermodynamic limit has been taken to a lattice which is infinite in the direction of the interaction between slices.

11.1.3 Correlation Functions and Lengths

The probability that the state of the slice $\mathcal{N}_i$ of the lattice is τ_m is given by

$$\text{Prob}[\tau(i) = \tau_m] = \frac{\text{Trace}\{V^{i-1}|m\rangle\langle m|V^{N_1-i+1}\}}{\text{Trace}\{V^{N_1}\}} = \frac{\text{Trace}\{\Theta^{(m)}\Lambda^{N_1}\}}{\text{Trace}\{\Lambda^{N_1}\}}, \quad (11.1.15)$$

where

$$\Theta^{(m)} := Y|m\rangle\langle m|X. \quad (11.1.16)$$

As is to be expected this probability is independent of i . Similarly, for $i \geq i'$,

$$\text{Prob} \begin{bmatrix} \tau(i') = \tau_{m'} \\ \tau(i) = \tau_m \end{bmatrix} = \frac{\text{Trace}\{\Theta^{(m')} \Lambda^{i-i'} \Theta^{(m)} \Lambda^{N_1+i'-i}\}}{\text{Trace}\{\Lambda^{N_1}\}}. \quad (11.1.17)$$

Formulae can also be obtained for the joint probabilities for the states of other subsets of the slices.⁴

The expectation value of any function $f_i := f(\tau(i))$ of the state of a single slice $\mathcal{N}_i$ is

$$\langle f_i \rangle = \sum_{m=1}^M \text{Prob}[\tau(i) = \tau_m] f(\tau_m) = \frac{\text{Trace}\{\Psi \Lambda^{N_1}\}}{\text{Trace}\{\Lambda^{N_1}\}} =: \mathcal{F}_1(\mathcal{N}), \quad (11.1.18)$$

given in terms of the Hermitian operator

$$\Psi := YFX, \quad \text{with} \quad F := \sum_{m=1}^M |m\rangle f(\tau_m) \langle m|. \quad (11.1.19)$$

Similarly

$$\begin{aligned} \langle f_{i'} f_i \rangle &= \sum_{m=1}^M \sum_{m'=1}^M \text{Prob} \begin{bmatrix} \tau(i') = \tau_{m'} \\ \tau(i) = \tau_m \end{bmatrix} f(\tau_{m'}) f(\tau_m) \\ &= \frac{\text{Trace}\{\Psi \Lambda^{i-i'} \Psi \Lambda^{N_1+i'-i}\}}{\text{Trace}\{\Lambda^{N_1}\}} =: \mathcal{F}_2(\mathcal{N}; i - i'). \end{aligned} \quad (11.1.20)$$

The expectation value $\langle f_{i'} f_i \rangle$ is dependent on i and i' only through the difference $i - i'$. The pair correlation function between the values of $f(\tau)$ at slices $\mathcal{N}_{i'}$ and $\mathcal{N}_i$ is given (Eq.(2.5.9)) by

$$\Gamma_2(\mathcal{N}; f_{i'}, f_i) = \langle f_{i'} f_i \rangle - \langle f_{i'} \rangle \langle f_i \rangle = \mathcal{F}_2(\mathcal{N}; i - i') - [\mathcal{F}_1(\mathcal{N})]^2. \quad (11.1.21)$$

We now obtain expressions for $\Gamma_2(\mathcal{N}; f_{i'}, f_i)$ in the limit $N_1 \rightarrow \infty$, for the cases A–C given above. Since the results for case A can be obtained, in the limit $p \rightarrow 1$, from B and, when $\omega \rightarrow 1$, from C it need not be treated separately. It is straightforward to show that

⁴ Given that the energy of the system arises from the interaction between a pair of neighbouring slices, the calculation of the energy-energy correlation function involves probabilities for the states of two pairs of slices. The developments needed to calculate the energy-energy correlation function and a corresponding correlation length are similar to those given in this and the next section.

$$\mathcal{F}_1(\mathcal{L}_1) \simeq \begin{cases} p^{-1} \sum_{m=1}^p \langle m | \Psi | m \rangle, & \text{case B,} \\ \omega^{-1} \sum_{m=1}^{\omega} \langle m | \Psi | m \rangle, & \text{case C,} \end{cases} \quad (11.1.22)$$

$$\mathcal{F}_2(\mathcal{L}_1; s) \simeq \mathcal{F}_2^{(1)}(\mathcal{L}_1; s) + \mathcal{F}_2^{(2)}(\mathcal{L}_1; s), \quad (11.1.23)$$

where

$$\mathcal{F}_2^{(1)}(\mathcal{L}_1; s) = \begin{cases} p^{-1} \sum_{m=1}^p \sum_{m'=1}^p |\langle m | \Psi | m' \rangle|^2 \exp(2\pi i s [m - m'] / p), & \text{case B,} \\ \omega^{-1} \sum_{m=1}^{\omega} \sum_{m'=1}^{\omega} |\langle m | \Psi | m' \rangle|^2, & \text{case C,} \end{cases} \quad (11.1.24)$$

$$\mathcal{F}_2^{(2)}(\mathcal{L}_1; s) = \begin{cases} p^{-1} \sum_{m=p+1}^M \sum_{m'=1}^p \Omega_m^s |\langle m | \Psi | m' \rangle|^2 \exp(-2\pi i s [m' - 1] / p), & \text{case B,} \\ \omega^{-1} \sum_{m=\omega+1}^M \sum_{m'=1}^{\omega} \Omega_m^s |\langle m | \Psi | m' \rangle|^2. & \text{case C.} \end{cases} \quad (11.1.25)$$

Then

$$\Gamma_2(\mathcal{L}_1; f_i, f_{i'}) = \Gamma_2^{(1)}(\mathcal{L}_1; f_i, f_{i'}) + \mathcal{F}_2^{(2)}(\mathcal{L}_1; i - i'), \quad (11.1.26)$$

where

$$\Gamma_2^{(1)}(\mathcal{L}_1; f_i, f_{i'}) = \begin{cases} p^{-2} \sum_{m=1}^p \sum_{m'=1}^p [p |\langle m | \Psi | m' \rangle|^2 \exp(2\pi i [m - m'] [i - i'] / p) \\ \quad - \langle m | \Psi | m \rangle \langle m' | \Psi | m' \rangle], & \text{case B,} \\ \omega^{-2} \sum_{m=1}^{\omega} \sum_{m'=1}^{\omega} \{\omega |\langle m | \Psi_f | m' \rangle|^2 - \langle m | \Psi_f | m \rangle \langle m' | \Psi_f | m' \rangle\}. & \text{case C.} \end{cases} \quad (11.1.27)$$

In both cases

$$\lim_{|i-i'|\rightarrow\infty} \mathcal{F}_2^{(2)}(\mathcal{L}_1; i - i') = 0, \quad (11.1.28)$$

but $\Gamma_2^{(1)}(\mathcal{L}_1; f_i, f_{i'})$ does not decay except in the Case A limit, $p = 1$ or $\omega = 1$, or when it becomes zero because of special symmetry properties.

From (4.2.4), the correlation length with respect to f is

$$\tau^2(\mathcal{L}_1; f) = \frac{a^2}{2} \frac{\sum_{i=0}^{\infty} i^2 \Gamma_2(\mathcal{L}_1; f_i, f_0)}{\sum_{i=0}^{\infty} \Gamma_2(\mathcal{L}_1; f_i, f_0)}, \quad (11.1.29)$$

where a is a length parameter measuring the distance between $\mathcal{N}_i$ and $\mathcal{N}_{i+1}$ for any i , and we have used the result $c(0) = c(1) = \frac{1}{2}$ obtained in Sect. 5.6.3. Since the summation is over *all* pairs of slices the implicit assumption, underlying interest in this correlation length, is that the ground state for every slice is the same. That is the ground state is homogeneous, like that of the spin- $\frac{1}{2}$ ferromagnetic Ising model or the ν -state Potts model with $R > 0$.

It is clear from the above discussion that the summations in (11.1.29) converge only when $\Gamma_2^{(1)}(\mathcal{L}_1; f_i, f_{i'}) = 0$ and, in general this is the case when the transfer matrix is non-reducible and primitive. Then, for $N_1 \rightarrow \infty$,

$$\Gamma_2(\mathcal{L}_1; f_i, f_0) = \sum_{m=2}^M \Omega_m^i |\langle m | \Psi | 1 \rangle|^2 \quad (11.1.30)$$

and

$$\tau^2(\mathcal{L}_1; f) = \frac{a^2}{2} \frac{\sum_{m=2}^M \frac{\langle m | \Psi | 1 \rangle^2 \Omega_m [1 + \Omega_m]}{[1 - \Omega_m]^3}}{\sum_{m=2}^M \frac{|\langle m | \Psi | 1 \rangle|^2}{1 - \Omega_m}}. \quad (11.1.31)$$

If the interest is in a phase transition to a periodic ground state of period η , we need a correlation length based on the correlation between sites at a lattice distance which is an integer multiple of η . The expressions (11.1.29) and (11.1.31) are replaced by

$$\tau_\eta^2(\mathcal{L}_1; f) = \frac{a^2}{2} \frac{\sum_{i=0}^{\infty} (i\eta)^2 \Gamma_2(\mathcal{L}_1; f_{i\eta}, f_0)}{\sum_{i=0}^{\infty} \Gamma_2(\mathcal{L}_1; f_{i\eta}, f_0)}, \quad (11.1.32)$$

$$\tau_\eta^2(\mathcal{L}_1; f) = \frac{(\eta a)^2}{2} \frac{\sum_{m=2}^{\infty} \frac{|\langle m|\Psi|1\rangle|^2 (\Omega_m)^\eta \{1 + (\Omega_m)^\eta\}}{\{1 - (\Omega_m)^\eta\}^3}}{\sum_{m=2}^{\infty} \frac{|\langle m|\Psi|1\rangle|^2}{1 - (\Omega_m)^\eta}}. \quad (11.1.33)$$

In the complex plane $\mathbb{C}_\Omega$, the scaled eigenvalues Ω_m are complex numbers lying within the unit-radius spectral circle $\tilde{SC}(V)$. Their number $M = \nu^{N_2}$ increases with increasing N_2 and they will move within the disc as the temperature, or any other thermodynamic variable, is varied. It is clear that the correlation length given by (11.1.31) becomes large when a scaled eigenvalue approaches the point $1 \in \mathbb{C}_\Omega$. This, according to the discussion in Sect. 17.2.4 must mean that V becomes reducible, with two diagonal blocks containing the degenerate eigenvalues Λ_1 . In fact, unless this Ω_m is real and the approach is along the real axis the situation will involve a conjugated pair of eigenvalues resulting in a degeneracy of three. The corresponding situation for (11.1.33) occurs when, given that Ω_1 is already equal to one, a scaled eigenvalue approaches $\exp(2\pi i k/\eta)$, for some $k = 1, \dots, \eta - 1$. But as we seen in Sect. 17.2.4, this can occur only if eigenvalues tend to all the values $\Lambda_1 \exp(2\pi i k/\eta)$, $k = 1, \dots, \eta - 1$. The transfer matrix becomes periodic with period η .

Suppose that a conjugate pair of scaled eigenvalues $\Omega_2 := |\Omega_2| \exp(i\theta)$ and $\Omega_3 := |\Omega_2| \exp(-i\theta)$ are such that, for some (small) $\varepsilon > 0$ and $\tilde{N}$,

$$|\Omega_m - 1| < \varepsilon, \quad \forall \quad N_2 > \tilde{N}, \quad m = 2, 3. \quad (11.1.34)$$

Then

$$\Omega_m = \exp\{\ln[\Omega_m]\} \simeq 1 + \ln[\Omega_m]. \quad (11.1.35)$$

The summations in (11.1.31) are dominated by these two terms. Since $\langle 2|\Psi|1\rangle$ and $\langle 3|\Psi|1\rangle$ are constructed from the conjugate eigenvectors of V , $\langle 3|\Psi|1\rangle = \overline{\langle 2|\Psi|1\rangle}$ giving

$$\tau(f, \mathcal{L}_1) \simeq -a\{\ln |\Omega_2|\}^{-1}, \quad (11.1.36)$$

for any $f(\tau)$ which is a function of the state τ of a single slice.

In the case of the ground-state of periodicity η

$$|\Omega_m - \exp(2\pi i[m-1]/\eta)| < \varepsilon, \quad \forall \quad N_2 > \tilde{N}, \quad m = 2, 3, \dots, \eta, \quad (11.1.37)$$

giving

$$(\Omega_m)^\eta = \exp\{\ln[(\Omega_m)^\eta]\} \simeq 1 + \ln[(\Omega_m)^\eta]. \quad (11.1.38)$$

The summations in (11.1.33) are dominated by these terms giving

$$\mathbf{r}_\eta^2(\mathcal{L}_1; f) \simeq (\eta a)^2 \frac{\sum_{m=2}^{\eta} \langle m | \Psi | 1 \rangle \{\ln[(\Omega_m)^\eta]\}^{-3}}{\sum_{m=2}^{\eta} \langle m | \Psi | 1 \rangle \{\ln[(\Omega_m)^\eta]\}^{-1}}. \quad (11.1.39)$$

The approach to the spectral circle $\widetilde{SC}(V)$ will, of course, occur for eigenvalues which are (apart from one approaching the point -1 if η is even) conjugate and approaching conjugate points on the circle. Assuming that they are approaching with equal magnitude $|\Omega_2|$ with

$$(\Omega_m)^\eta = \exp \left[i \text{Arg}\{(\Omega_m)^\eta\} + \eta \ln |\Omega_2| \right], \quad (11.1.40)$$

and

$$|\text{Arg}\{(\Omega_m)^\eta\}| \ll \eta |\ln |\Omega_2||, \quad (11.1.41)$$

(11.1.39) yields again the formula (11.1.36), which is that used for the correlation length in many investigations which employ transfer matrices (e.g. the phenomenological renormalization method of Sect. 15.10).

11.2 Incipient Phase Transitions

Given that the elements of the transfer matrix can be expressed in terms of some suitable Boltzmann factor $\mathfrak{X}$, it follows that the partition function $Z(\mathcal{N}, \mathfrak{X})$ for finite N_1 and N_2 is a polynomial in $\mathfrak{X}$ with positive coefficients. By the argument used on p. 253 for the simple lattice fluid it follows that the system can have no phase transitions. A modified version of Theorem 8.4.1 can then be used to show that no phase transitions occur in classical one-dimensionally infinite systems with finite-range interaction.⁵ Here one must distinguish between the *onset* of long-range order consequent upon the occurrence of a phase transition and the *existence* of a long-range ordered state. As we shall see in Sect. 11.3.2, a state with degenerate eigenvalues, for which the transfer matrix must be reducible, or with other eigenvalues on the

⁵ As has been noted in Sect. 3.2.1 a phase transition does occur in the one-dimensional *quantum* Ising model with a transverse field and long-range order is possible in systems with an *infinite-range* interaction (van Enter and Zegarliński 1997).

spectral circle (particularly an eigenvalue equal to but of opposite sign to the leading eigenvalue) for which the transfer matrix is periodic, is indicative of a state of long-range order with long-range correlations and divergent correlation length. This can occur at zero temperature without it being the result of a finite-temperature phase transition.

Here we are interested in the occurrence of phase transitions at non-zero temperature. Since such do occur, in systems which are more than one-dimensionally infinite, this must involve the double limit when both N_1 and N_2 become large. Ree and Chesnut (1966) have argued that a necessary condition for a phase transition at $\mathfrak{X} = \mathfrak{X}_c$ is that, in the double limit,

$$|\Omega_m(\mathfrak{X})|^{N_2} \rightarrow L(\mathfrak{X}), \quad \text{for some } m \neq 1 \text{ and } L(\mathfrak{X}_c) \neq 0. \quad (11.2.1)$$

To achieve a condition like (11.2.1) it is, of course, necessary to specify the precise way in which the double limit is taken. In any event the condition

$$\lim_{N_2 \rightarrow \infty} |\Omega_m(\mathfrak{X}_c)| = 1, \quad \text{for some } m \neq 1, \quad (11.2.2)$$

is implied by (11.2.1) and is in turn a necessary condition for a phase transition at $\mathfrak{X} = \mathfrak{X}_c$.

We now understand that a phase transition to an ordered phase in a system with interactions only between neighbouring slices $\mathcal{N}_i$ can occur only in the limit of an infinite system, where both N_1 and N_2 become large. It is also clear from (11.1.36) that, for a semi-infinite system ($N_1 \rightarrow \infty$), the correlation length becomes large at some $\mathfrak{X} = \mathfrak{X}'$ when one or more of the scaled eigenvalues approaches a point on $\widetilde{\mathcal{SC}}(\mathcal{V})$. For N_2 finite and the transfer matrix irreducible and primitive all the scaled eigenvalues lie within the circle and the correlation length will be a continuous function of $\mathfrak{X}$ with a maximum at $\mathfrak{X} = \mathfrak{X}'$. A point of closest approach to $\widetilde{\mathcal{SC}}(\mathcal{V})$ by one or more scaled eigenvalues can be regarded as an *incipient phase transition*. If this eigenvalue $\Omega_m(\mathfrak{X})$ approaches the point 1, then an incipient phase transition is to a homogeneous state. If the approach is to the point $\exp(2\pi i k/\eta)$, on $\widetilde{\mathcal{SC}}(\mathcal{V})$, for some k , there will be simultaneous approaches to all the points $\exp(2\pi i k/\eta)$, $k = 1, 2, \dots, \eta - 1$, indicative of a transition to a ground state of period η .

11.3 Using Symmetry Properties

Let $\mathcal{Q} := \{I, Q^{(1)}, \dots, Q^{(n-1)}\}$ be the largest group such that each $Q^{(s)}$ maps a state τ_m of a slice of the system into another state $Q^{(s)}\tau_m$, with $Q^{(s)}\tau_m \neq Q^{(s)}\tau_{m'}$ if $\tau_m \neq \tau_{m'}$, and for which $\hat{h}(\tau_m, \tau_{m'})$, and consequently $V(\tau_m, \tau_{m'})$, is left invariant. When periodic boundary conditions are applied in the j direction, so that the lattice forms a cylinder, the symmetry group $\mathcal{Q}$ will normally include the cyclic group corresponding to rotation about the axis of the cylinder. Other subgroups of $\mathcal{Q}$ will

correspond to any symmetry of the Hamiltonian arising from permutations of the states of the microsystems. For the standard ν -state Potts model $\mathcal{Q}$ will contain the symmetric group $\mathcal{S}_\nu$.

11.3.1 Block-Diagonalization of the Transfer Matrix

The effect of $\mathcal{Q}$ on the states $\{\tau_m\}$ is given, using the representation $\mathcal{R}(\mathcal{Q}) := \{\mathbf{I}, \mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(n-1)}\}$ of M -dimensional permutation matrices,⁶ by

$$\mathbf{Q}^{(s)} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_M \end{pmatrix} = \begin{pmatrix} \mathbf{Q}^{(s)} \tau_1 \\ \mathbf{Q}^{(s)} \tau_2 \\ \vdots \\ \mathbf{Q}^{(s)} \tau_M \end{pmatrix}, \quad s = 0, 1, \dots, n-1 \quad (11.3.1)$$

and the symmetry property of the transfer matrix can be represented in the form

$$\mathbf{Q}^{(s)} \mathbf{V} [\mathbf{Q}^{(s)}]^\dagger = \mathbf{V}, \quad \text{or equivalently} \quad [\mathbf{Q}^{(s)}, \mathbf{V}] = 0, \quad s = 0, 1, \dots, n-1. \quad (11.3.2)$$

The irreducible representations of $\mathcal{Q}$ are denoted by $\Theta(k)$, for $k = 1, 2, \dots, \kappa$, with $e(k)$ being the dimension of $\Theta(k)$.

Equation (11.3.2) encapsulates the symmetry properties of the transfer matrix with respect to the representation $\mathcal{R}(\mathcal{Q})$ of the group $\mathcal{Q}$, and (17.3.10) asserts that a matrix $\mathbf{U}$ exists which reduces all the matrices of $\mathcal{R}(\mathcal{Q})$ to block-diagonal form. It can be shown that, if the matrix $\mathbf{U}$ is known, it can be used, together with row and column permutations, to reduce $\mathbf{V}$ to a block-diagonal form, each block being one of a set of $m(k)$ identical matrices identified with the representation $\Theta(k)$. This reduction is to the unique *canonical form* of $\mathbf{V}$. However, in general achieving this can be quite difficult and a simpler and often quite sufficient reduction is that given in (17.3.15) by the matrix $\mathbf{W}$ since, as can be easily shown:

Theorem 11.3.1 *If $\mathbf{W}$ reduces each of the matrices of the group $\mathcal{Q}$ to block-diagonal form, with one block for each irreducible representation $\Theta(k)$, as shown in (17.3.15), then the transfer matrix satisfying (11.3.2) is reduced in a similar way. That is*

$$\mathbf{W}^\dagger \mathbf{V} \mathbf{W} = \tilde{\mathbf{V}} := \tilde{\mathbf{V}}(1) \oplus \tilde{\mathbf{V}}(2) \oplus \dots \oplus \tilde{\mathbf{V}}(\kappa). \quad (11.3.3)$$

Of course, if $m(k) = 1$ for each representation then $\mathbf{U} = \mathbf{W}$ and reduction to the canonical form is complete. Otherwise a further reduction can be made of the block for $\Theta(k)$ to $m(k)$ identical sub-blocks.

Any matrix which, like $\mathbf{V}$, satisfies (11.3.2) is said to be $\mathcal{Q}$ -invariant. The problem of reducing such a matrix to block-diagonal form is simplified if the states of the

⁶ For a brief account of group representation theory and of the notation used here see Sect. 17.3.1.

system are ordered into *equivalence classes*, closed under the operations of the group $\mathcal{Q}$ and not further divisible into closed subsets. Using the notation described in Sect. 17.3.2 with the equivalence classes denoted by $\mathfrak{E}(\alpha)$, $\alpha = 1, 2, \dots, \gamma$, it can be shown (Theorem 17.3.7) that the equivalence class $\mathfrak{E}(\alpha)$ can be generated by applying all the elements of $\mathcal{Q}$ to any one state of $\mathfrak{E}(\alpha)$ and the number of states $p(\alpha)$ in $\mathfrak{E}(\alpha)$ is a divisor of n . It follows, from (11.3.1), that $\mathbf{Q}^{(s)}$ has the block-diagonal form (17.3.23) with $\mathcal{R}(\alpha; \mathcal{Q}) := \{\mathbf{I}(\alpha), \mathbf{Q}^{(1)}(\alpha), \dots, \mathbf{Q}^{(n-1)}(\alpha)\}$, a $p(\alpha)$ -dimensional representation of $\mathcal{Q}$ operating on the equivalence class $\mathfrak{E}(\alpha)$, for $\alpha = 1, 2, \dots, \gamma$. The formula (17.3.24) of the same form as (17.3.15) now applies to the matrices $\mathbf{Q}^{(s)}(\alpha)$. The matrix $\mathbf{W}(\alpha)$ have the effect of reducing $\mathbf{Q}^{(s)}(\alpha)$ to block-diagonal form with the blocks corresponding to the irreducible representations in order and in Sect. 17.3.3 we describe a method for deriving it.

Suppose that $\mathbf{V}$ is expressed in the form

$$\mathbf{V} := \begin{pmatrix} \mathbf{V}(1, 1) & \mathbf{V}(1, 2) & \cdots & \mathbf{V}(1, \gamma) \\ \mathbf{V}(2, 1) & \mathbf{V}(2, 2) & \cdots & \mathbf{V}(2, \gamma) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}(\gamma, 1) & \mathbf{V}(\gamma, 2) & \cdots & \mathbf{V}(\gamma, \gamma) \end{pmatrix}, \quad (11.3.4)$$

where $\mathbf{V}(\alpha, \alpha')$ is the submatrix with row and column numbers in the equivalence classes α and α' respectively. The matrix

$$\check{\mathbf{V}} := [\mathbf{W}(1) \oplus \mathbf{W}(2) \oplus \cdots \oplus \mathbf{W}(\gamma)]^\dagger \mathbf{V} [\mathbf{W}(1) \oplus \mathbf{W}(2) \oplus \cdots \oplus \mathbf{W}(\gamma)] \quad (11.3.5)$$

has submatrices $\check{\mathbf{V}}(\alpha, \alpha') = \mathbf{W}^\dagger(\alpha) \mathbf{V}(\alpha, \alpha') \mathbf{W}(\alpha')$ which, from (11.3.2), (17.3.23) and (17.3.24), satisfy the equations

$$[\tilde{\mathbf{Q}}^{(s)}(\alpha)]^\dagger \check{\mathbf{V}}(\alpha, \alpha') \tilde{\mathbf{Q}}^{(s)}(\alpha') = \check{\mathbf{V}}(\alpha, \alpha'), \quad s = 1, 2, \dots, n-1. \quad (11.3.6)$$

Using analysis similar to that in the proof of Theorem 11.3.1, it can be shown that the rows and columns of $\check{\mathbf{V}}(\alpha, \alpha')$ can be indexed by the pair (k, j) with $j = 1, 2, \dots, w(\alpha; k)$ and

$$\langle k, j | \check{\mathbf{V}}(\alpha, \alpha') | k', j' \rangle = \delta^{Kr}(k - k') \langle k, j | \check{\mathbf{V}}(\alpha, \alpha') | k, j' \rangle. \quad (11.3.7)$$

For every irreducible representation $\Theta(k)$ which occurs in both $\mathfrak{E}(\alpha)$ and $\mathfrak{E}(\alpha')$ there will be $w(\alpha; k) \times w(\alpha'; k)$ elements of $\check{\mathbf{V}}(\alpha, \alpha')$ which can be non-zero. The (possibly) non-zero elements of $\check{\mathbf{V}}(\alpha, \alpha')$ are given by

$$\langle k, j | \check{\mathbf{V}}(\alpha, \alpha') | k, j' \rangle = \sum_{m=1}^{p(\alpha)} \sum_{m'=1}^{p(\alpha')} \overline{W}_{mj}^{(k)}(\alpha) \langle m | \mathbf{V}(\alpha, \alpha') | m' \rangle W_{m'j'}^{(k)}(\alpha'), \quad (11.3.8)$$

where the quantities $W_{mj}^{(k)}(\alpha)$ are defined in Sect. 17.3.3. In particular, from (17.3.43), for the one-dimension symmetric representation $w(\alpha; 1) = 1$; each submatrix contributes exactly one term given by

$$\langle 1, 1 | \check{V}(\alpha, \alpha') | 1, 1 \rangle = \frac{1}{\sqrt{p(\alpha)p(\alpha')}} \sum_{m=1}^{p(\alpha)} \sum_{m'=1}^{p(\alpha')} \langle m | V(\alpha, \alpha') | m' \rangle. \quad (11.3.9)$$

The permutation matrix Π of equation (17.3.25) collects them together so that

$$\Pi \check{V} \Pi = \tilde{V}, \quad (11.3.10)$$

has the block form given in (11.3.3).

11.3.1.1 Eigenvalues and Eigenvectors

The eigenvalues of $\tilde{V}$ given by (11.3.3), are the same as those of V . Each eigenvalue can, therefore, be identified with some irreducible representation $\Theta(k)$ within which an ordering in decreasing magnitude is applied. Thus, for V , we replace the notation (Λ_k, y_k, x_k) , $k = 1, 2, \dots, M$ for eigenvalues and vectors with $(\Lambda_\ell(k), y_\ell(k), x_\ell(k))$, $k = 1, 2, \dots, \kappa$, $\ell = 1, 2, \dots, w(k)$. From the commutativity condition (11.3.2) and Theorem 17.2.4 it follows that:

Theorem 11.3.2 *If $(y_\ell(k), x_\ell(k))$ is a non-degenerate eigenvector pair of V it is also an eigenvector pair of $Q^{(s)}$ with eigenvalue $\chi^{(k)}(Q^{(s)})/e(k)$, for $s = 0, 1, \dots, n-1$.*

The M -dimensional vector v is a *vector-invariant* of Q if

$$Q^{(s)} v = v, \quad \forall s = 0, 1, \dots, n-1. \quad (11.3.11)$$

Suppose the elements of v are ordered in equivalence classes with $\{Q^{(s)}\}$ in the block-diagonal form (17.3.23), so that

$$v = \begin{pmatrix} v(1)j(1) \\ v(2)j(2) \\ \vdots \\ v(\gamma)j(\gamma) \end{pmatrix}, \quad (11.3.12)$$

where $j(\alpha)$ is the vector of dimension $p(\alpha)$ with every entry unity. Then:

Theorem 11.3.3 *Non-degenerate eigenvectors $y_\ell(k)$ and $x_\ell(k)$ with $k \neq 1$ are orthogonal to all vector-invariants of the group.*

Proof From Theorem 11.3.2 and (11.3.11), for some vector-invariant v ,

$$[\mathbf{y}_\ell(\mathbf{k})]^T \mathbf{v} = [\mathbf{y}_\ell(\mathbf{k})]^T \mathbf{Q}^{(s)} \mathbf{v} = \chi^{(\mathbf{k})}(\mathbf{Q}^{(s)}) [\mathbf{y}_\ell(\mathbf{k})]^T \mathbf{v} / e(\mathbf{k}), \quad (11.3.13)$$

for $s = 0, 1, \dots, n-1$. Since the only representation for which $\chi^{(\mathbf{k})}(\mathbf{Q}^{(s)}) = e(\mathbf{k})$, for all $s = 0, 1, \dots, n-1$, is $\Theta(1)$, it follows that $[\mathbf{y}_\ell(\mathbf{k})]^T \mathbf{v} = 0$, unless $\mathbf{k} = 1$, with the same analysis applying to the right eigenvector $\mathbf{x}_\ell(\mathbf{k})$. $\square$

As we have seen, given that $\mathbf{V}$ is irreducible and primitive (Case A, Sect. 11.1.2), the principal eigenvalue is positive and of larger magnitude than all other eigenvalues with positive left and right eigenvectors. But such eigenvectors cannot be orthogonal to vector invariants of the form (11.3.12) with $f(\alpha) > 0$ for $\alpha = 1, 2, \dots, \gamma$. It follows that:

Theorem 11.3.4 *For such a $\mathbf{V}$ the principal eigenvalue is in the one-dimensional symmetric representation (with designation $\Lambda_1(1)$).*

This is a particular useful result, since, if we need only the largest eigenvalue, it can be obtained by solving the eigenvalue problem for $\tilde{\mathbf{V}}(1)$ and this matrix can be obtained very simply from $\mathbf{V}$ using (11.3.9), thus yielding the partition function and the free-energy density in the limit $N_1 \rightarrow \infty$.

However, even an approximate calculation of the correlation function, given, from (11.1.30), by

$$\Gamma_2(\mathcal{L}_1; f_i, f_{i'}) = \sum_{\{\mathbf{k}, \ell\}}^{[1,1]} |\langle \mathbf{k}, \ell | \Psi | 1 \rangle|^2 [\Omega_\ell(\mathbf{k})]^{i-i'}, \quad (11.3.14)$$

where ‘[1, 1]’ denotes the fact that the largest eigenvalue is excluded from the summation, and

$$|\langle \mathbf{k}, \ell | \Psi | 1 \rangle|^2 = \langle 1, 1 | \mathbf{Y} \mathbf{F} \mathbf{X} | \mathbf{k}, \ell \rangle \langle \mathbf{k}, \ell | \mathbf{Y} \mathbf{F} \mathbf{X} | 1, 1 \rangle, \quad (11.3.15)$$

still requires the computation of the largest sub-maximum eigenvalues. There is, in general, no reason to suppose that these lie in the one-dimensional symmetric matrix block. Even though block-diagonalization still has the advantage of allowing the handling of a set of smaller matrices, it is in principle necessary to compute the eigenvalues for each block. However, there is an exceptional case. If $\mathbf{F}$ is $\mathbf{Q}$ -invariant, meaning that function $f(\tau_m)$ has the same value for all states in an equivalence class, $\langle 1, 1 | \mathbf{Y} \mathbf{F} \mathbf{X} | \mathbf{k}, \ell \rangle = 0$ for $\mathbf{k} \neq 1$. The correlation function and the correlation length are given by just the eigenvalues and vectors in $\Theta(1)$. So the problem is reduced to looking for the largest eigenvalues in $\tilde{\mathbf{V}}(1)$, without the need to consider the other matrix blocks.

11.3.2 Applications

As illustrations of the methods described here we consider two examples with a positive transfer matrix, the spin- $\frac{1}{2}$ Ising model and the 3-state Potts model on a square lattice wrapped around a cylinder with N_1 rings of N_2 sites. In each case the symmetry group for lattice rotations and reflections is $\mathcal{C}_{N_2}$ and the symmetry groups arising from the nature of the models are $\mathcal{S}_2$ and $\mathcal{S}_3$ respectively. Given the advances in algebraic computing software in recent years, block-diagonalization of the transfer matrix is not particularly advantageous for small values of N_2 . However, as N_2 increases, a factor which determines the virtue of block-diagonal is the difference between the matrix size, v_2^N , with $v = 2$ and 3 for the Ising and Potts models respectively, and the size $m(1; N_2)$ of the one-dimension block.

As an indication of the numbers involved, we consider the Ising model, ignoring, for simplicity, reflections, so that the symmetry group is $\mathcal{Q} = \mathcal{C}_{N_2} \times \mathcal{S}_2$.⁷ By counting the states left invariant by this group, it is not difficult to see that, for $N_2 \geq 2$,

$$m(1; N_2) = \begin{cases} 2^{\frac{1}{2}(N_2-3)} + \frac{1}{4N_2} \sum_{s=0}^{N_2-1} 2^{[s, N_2]}, & N_2 \text{ odd,} \\ 2^{\frac{1}{2}N_2-1} + \frac{1}{4N_2} \sum_{s=0}^{N_2-1} 2^{[s, N_2]} \Delta(N_2, s), & N_2 \text{ even,} \end{cases} \quad (11.3.16)$$

where

$$\Delta(N_2, s) = 2 - \text{Parity} \left(\frac{N_2}{[s, N_2]} \right) \quad (11.3.17)$$

and $[s, N_2]$ is the greatest common divisor of s and N_2 . As examples: $2^4 = 16$, $m(1; 4) = 4$; $2^{10} = 1024$, $m(1; 10) = 44$.

Example 11.3.1 For the Ising model the states of pairs, ordered in the equivalence classes, are

$$\left\{ \begin{bmatrix} +1 \\ +1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}, \left\{ \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix} \right\} \quad (11.3.18)$$

and, with $K := 2J/T$, $\mathfrak{X} := \exp(K)$, the transfer matrix takes the form

$$V = \begin{pmatrix} \mathfrak{X}^2 & 1 & 1 & 1 \\ 1 & \mathfrak{X}^2 & 1 & 1 \\ 1 & 1 & 1 & \mathfrak{X}^{-2} \\ 1 & 1 & \mathfrak{X}^{-2} & 1 \end{pmatrix}. \quad (11.3.19)$$

⁷ Including reflections will, of course, reduce $m(1; N_2)$.

For the symmetry group $\mathcal{Q} := \mathcal{S}_2^{(m)} \times \mathcal{C}_2$, the conjugacy classes, irreducible representations and characters are given in Example 17.3.3, with

$$\begin{aligned} R(I) &= \begin{pmatrix} I_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & I_2 \end{pmatrix}, & R(Q^{(m)}) &= \begin{pmatrix} J_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & J_2 \end{pmatrix}, \\ R(C_2) &= \begin{pmatrix} I_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & J_2 \end{pmatrix}, & R(Q^{(m)}C_2) &= \begin{pmatrix} J_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & I_2 \end{pmatrix}, \end{aligned} \quad (11.3.20)$$

where I_2 and $\mathbf{0}_2$ are the two-dimensional unit and zero matrices, respectively, and

$$J_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (11.3.21)$$

From the character table the multiplicities of the irreducible representations in the four-dimensional representation of states are $m(A) = 2$, $m(B_1) = 1$, $m(B_2) = 0$ and $m(B_3) = 1$. The matrix which reduces the transfer matrix to diagonal form is $W \oplus W$ where

$$W = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad (11.3.22)$$

yielding

$$\tilde{V}(A) = \begin{pmatrix} \mathfrak{x}^2 + 1 & 2 \\ 2 & 1 + \mathfrak{x}^{-2} \end{pmatrix}, \quad (11.3.23)$$

$$\tilde{V}(B_1) = \mathfrak{x}^2 - 1, \quad \tilde{V}(B_3) = 1 - \mathfrak{x}^{-2}. \quad (11.3.24)$$

The characteristic equation of the A is

$$\Lambda^2 - (\mathfrak{x} + \mathfrak{x}^{-1})^2 \Lambda + (\mathfrak{x} - \mathfrak{x}^{-1})^2 = 0, \quad (11.3.25)$$

giving

$$\left. \begin{aligned} \Lambda_1(A; 2, \mathfrak{x}) \\ \Lambda_2(A; 2, \mathfrak{x}) \end{aligned} \right\} = \frac{1}{2} \left[(\mathfrak{x} + \mathfrak{x}^{-1})^2 \pm \sqrt{(\mathfrak{x} + \mathfrak{x}^{-1})^4 - 4(\mathfrak{x} - \mathfrak{x}^{-1})^2} \right], \quad (11.3.26)$$

$$\Lambda_1(B_1; 2, \mathfrak{x}) = \mathfrak{x}^2 - 1, \quad (11.3.27)$$

$$\Lambda_1(B_3; 2, \mathfrak{x}) = 1 - \mathfrak{x}^{-2}. \quad (11.3.28)$$

Example 11.3.2 For the 3-state Potts model the Hamiltonian is given by (3.6.49), where $\sigma(\mathbf{r})$ takes three consecutive values. In the convention of Sect. 3.6.4 these would be 0, 1, 2. However it is more convenient here to choose, as in Sect. 5.4.2, the spin-1 Ising values $-1, 0, 1$.

With $N_2 = 1$ the symmetry group is $\mathfrak{S}_3$:

We choose the same three group generators as in Sect. 5.4.2: I , $Q^{(m)}$, which simply interchanges $+1$ and -1 and $Q^{(*)}$, given by (5.4.3), which cyclically permutes the states. The three-dimensional representation of $\mathfrak{S}_3$ operating on $\sigma := (-1, 0, +1)^T$ has

$$R(I) = I_3, \quad R(Q^{(*)}) = S_3, \quad R(Q^{(m)}) = J_3, \quad (11.3.29)$$

where

$$S_3 := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad J_3 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \quad (11.3.30)$$

The characters for the conjugacy classes defined as in Table 5.1 are $\chi(Q_1) = 3$, $\chi(Q_2) = 0$ and $\chi(Q_3) = 1$, and it follows from (17.3.13) that the multiplicities of the irreducible representations are $m(A) = 1$, $m(B) = 0$, $m(E) = 1$. The reduction matrix

$$W_3 := \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega_3 & \omega_3^2 \\ 1 & \omega_3^2 & \omega_3 \end{pmatrix}, \quad \omega_3 := \exp(2\pi i/3), \quad (11.3.31)$$

is given by Ree and Chesnut (1966), with

$$W_3^\dagger S_3 W_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega_3 & 0 \\ 0 & 0 & \omega_3^2 \end{pmatrix}, \quad W_3^\dagger J_3 W_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \omega_3^2 \\ 1 & \omega_3^2 & 0 \end{pmatrix}. \quad (11.3.32)$$

Using the Hamiltonian (3.6.49) the transfer matrix is

$$V_3 = \begin{pmatrix} \mathfrak{X} & 1 & 1 \\ 1 & \mathfrak{X} & 1 \\ 1 & 1 & \mathfrak{X} \end{pmatrix} \quad (11.3.33)$$

where $\mathfrak{X} := \exp(K)$, $K := R/T$ and

$$W_3^\dagger V_3 W_3 = \tilde{V}_3 = \begin{pmatrix} \mathfrak{X} + 2 & 0 & 0 \\ 0 & \mathfrak{X} - 1 & 0 \\ 0 & 0 & \mathfrak{X} - 1 \end{pmatrix}. \quad (11.3.34)$$

Since there is only one equivalence class the eigenvalue $\Lambda_1(A; \mathfrak{X}) = \mathfrak{X} + 2$ for the representation A can, according to (11.3.9), also be obtained as a weighted average of the elements of V_3 . It is also fortunately the case that the two-dimensional block associated with E is diagonal; $\Lambda_1(E; \mathfrak{X}) = \Lambda_2(E; \mathfrak{X}) = \mathfrak{X} - 1$ and the right and left eigenvectors of V_3 are respectively the columns of W_3 and the rows of $W_3^\dagger$. From (11.1.14) the free energy per lattice site, in the limit $N \rightarrow \infty$, $\mathcal{N} \rightarrow \mathcal{L}_1$, is

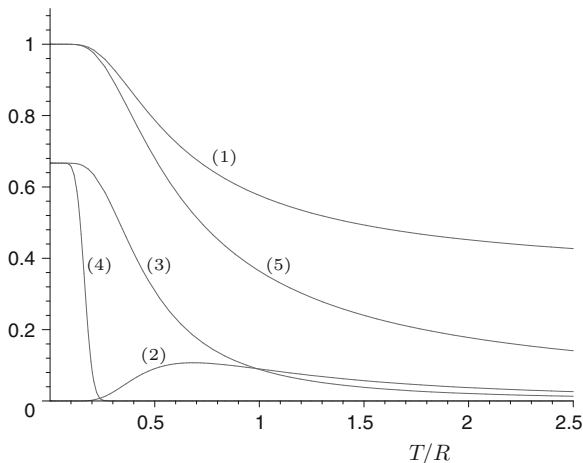


Fig. 11.1 Plots of (1) $-u(\mathcal{L}_1, K)$, (2) $c(\mathcal{L}_1, K)/k_B$, (3) $\Gamma_2(\mathcal{L}_1; \sigma(i), \sigma(i+2))$ (4) $\Gamma_2(\mathcal{L}_1; \sigma(i), \sigma(i+100))$ and (5) $\Omega_1(E; K)$, against $T/R = 1/K$

$$\phi_0(\mathcal{L}_1, K) = -\ln[\Lambda_1(A; \mathfrak{X})] = -\ln[\exp(K) + 2] \quad (11.3.35)$$

and, from (2.3.27) and (2.3.29), the internal energy per lattice site (in units of R) and the heat capacity are respectively

$$u(\mathcal{L}_1, K) := \frac{\partial \phi_0}{\partial K} = -\frac{\exp(K)}{\exp(K) + 2}, \quad (11.3.36)$$

$$c(\mathcal{L}_1, K) := -k_B K^2 \left(\frac{\partial u}{\partial K} \right) = 2k_B \left[\frac{K}{\exp(K) + 2} \right]^2. \quad (11.3.37)$$

We see that $c(\mathcal{L}_1, K) \rightarrow 0$ as $T \rightarrow 0$ and as $T \rightarrow \infty$. A plot against temperature is shown as plot (2) in Fig. 11.1.

It is straightforward using (11.1.18)–(11.1.20), (11.1.26) and $X := W_3$, $Y := W_3^\dagger$, to show that, in the limit $N \rightarrow \infty$,

$$\langle \sigma(i) \rangle = 0, \quad \langle \sigma(i) \sigma(i+s) \rangle = \frac{2}{3} [\Omega_1(E; K)]^s, \quad (11.3.38)$$

$$\langle \sigma^2(i) \rangle = \frac{2}{3}, \quad \langle \sigma^2(i) \sigma^2(i+s) \rangle = \frac{4}{9} + \frac{2}{9} [\Omega_1(E; K)]^s, \quad (11.3.39)$$

$$\Gamma_2(\mathcal{L}_1; \sigma(i), \sigma(i+s)) = \frac{2}{3} [\Omega_1(E; K)]^s, \quad (11.3.40)$$

$$\Gamma_2(\mathcal{L}_1; \sigma^2(i), \sigma^2(i+s)) = \frac{2}{9} [\Omega_1(E; K)]^s, \quad (11.3.41)$$

where

$$\Omega_1(E; K) := \frac{\Lambda_1(E; \mathfrak{K})}{\Lambda_1(A; \mathfrak{K})} = \frac{\exp(K) - 1}{\exp(K) + 2}. \quad (11.3.42)$$

From (11.1.29),

$$\mathfrak{r}^2(\mathcal{L}_1; \sigma) = \mathfrak{r}^2(\mathcal{L}_1; \sigma^2) = \frac{1}{18}a^2[\exp(K) - 1][2\exp(K) + 1]. \quad (11.3.43)$$

The first quantities in (11.3.38) and (11.3.39) correspond respectively to the magnetization and density when $\mathcal{H} = \mu = 0$. Thus we see, as expected that this model has no spontaneous magnetization and the expected density of microsystems in each of the three states is $\frac{1}{3}$. Using (3.6.59), (11.3.36), (11.3.38) and (11.3.39) confirms the expected result that $u = -\langle \delta^{\text{Kr}}(\sigma(i) - \sigma(i+1)) \rangle$.

Both correlation functions (11.3.40) and (11.3.41) are, for an s th neighbour pair, independent of position on the lattice and decrease as s increases (cf. plots (3) and (4) in Fig. 11.1). For any fixed value of s , they tend to one as $T \rightarrow 0$ and to zero as $T \rightarrow \infty$. The picture which emerges is that, although the expectation values of the total numbers of microsystems in each state remain the same, blocks of microsystems in the same state form and steadily increase in length as temperature decreases. In an ordered ground state $u(\mathcal{L}_1, \infty) = -\frac{1}{2}z = -1$ and the approach of $u(\mathcal{L}_1, K)$ to this limiting value corresponds to the number of unlike first-neighbour pairs, which represent block boundaries, becoming very small relative to the number of like pairs at low temperatures. Although there is no *long-range order* in the sense of a majority of microsystems being in the same state, the properties of the system at low temperatures are dominated by the short-range order.

With $N_2 = 2$ the symmetry group is $\mathfrak{S}_3 \times \mathfrak{C}_2$:

The states of a ring, arranged in equivalence classes, are

$$\left(\left\{ \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} +1 \\ +1 \end{bmatrix} \right\}, \right. \\ \left. \left\{ \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\} \right) \quad (11.3.44)$$

and the irreducible representations and character table for the symmetry group are given in Example 17.3.4. Using a representative matrix for an operator in each of the conjugacy classes, $\mathbf{R}(\mathbf{I}) = \mathbf{I}_9$ and

$$\begin{aligned}
R(Q^{(*)}) &= \begin{pmatrix} S_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & S_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & S_3 \end{pmatrix}, & R(Q^{(m)}) &= \begin{pmatrix} J_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & S_3 \\ \mathbf{0}_3 & S_3^2 & \mathbf{0}_3 \end{pmatrix}, \\
R(C_2) &= \begin{pmatrix} I_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & J_3 \\ \mathbf{0}_3 & J_3 & \mathbf{0}_3 \end{pmatrix}, & R(Q^{(*)}C_2) &= \begin{pmatrix} S_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & S_3 J_3 \\ \mathbf{0}_3 & S_3 J_3 & \mathbf{0}_3 \end{pmatrix}, \\
R(Q^{(m)}C_2) &= \begin{pmatrix} J_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & S_3 J_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & J_3 S_3 \end{pmatrix},
\end{aligned} \tag{11.3.45}$$

where S_3 and J_3 are given by (11.3.30). Thus the characters for this representation are $\chi(Q_1) = 9$, $\chi(Q_2) = 0$, $\chi(Q_3) = 1$, $\chi(Q_4) = 3$, $\chi(Q_5) = 0$ and $\chi(Q_6) = 3$, and it follows from (17.3.13) that the multiplicities of the irreducible representations are $m(A) = 1$, $m(B_1) = m(B_2) = 0$, $m(B_3) = 1$, $m(E_1) = 3$, $m(E_2) = 0$. The transfer matrix is

$$V_9 = \left(\begin{array}{ccc|ccc|ccc} \mathfrak{x}^4 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x}^2 \\ \mathfrak{x}^2 & \mathfrak{x}^4 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x}^2 \\ \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x}^4 & \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} \\ \hline \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x}^2 & 1 & 1 & \mathfrak{x} & \mathfrak{x} & 1 \\ \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x}^2 & 1 & \mathfrak{x}^2 & 1 & \mathfrak{x} & 1 & \mathfrak{x} \\ \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x}^2 & 1 & 1 & \mathfrak{x}^2 & 1 & \mathfrak{x} & \mathfrak{x} \\ \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x} & \mathfrak{x} & 1 & \mathfrak{x}^2 & 1 & 1 \\ \mathfrak{x} & \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & 1 & \mathfrak{x} & 1 & \mathfrak{x}^2 & 1 \\ \mathfrak{x}^2 & \mathfrak{x}^2 & \mathfrak{x} & 1 & \mathfrak{x} & \mathfrak{x} & 1 & 1 & \mathfrak{x}^2 \end{array} \right). \tag{11.3.46}$$

Since the first three and latter six ring-states form equivalence classes under the operations of the group, it follows from (11.3.9) the two-dimensional matrix block for the representation A can be obtained by adding over the elements in the four parts of V_9 (indicated by the partitioning lines) and dividing by the appropriated weighting factors. To obtain the one-dimensional block for B_3 and the six-dimensional block for E_1 it is necessary to determine the block-reduction matrix for the group matrices (11.3.45). Here we simply give the eigenvalues which are easily determined, either directly or after block-diagonalization, using an algebraic computing program.

$$\begin{aligned}
\Lambda_{1,2}(A; \mathfrak{X}) &= \frac{1}{2}[g_1(\mathfrak{X}) \pm \sqrt{g_2(\mathfrak{X})}], \\
\Lambda_1(B_3; \mathfrak{X}) &= (1 - \mathfrak{X})^2, \\
\Lambda_{1,2}(E_1; \mathfrak{X}) &= \Lambda_{3,4}(E_1; \mathfrak{X}) = \frac{1}{2}\mathfrak{X}(\mathfrak{X} - 1)[g_3(\mathfrak{X}) \pm \sqrt{g_4(\mathfrak{X})}], \\
\Lambda_5(E_1; \mathfrak{X}) &= \Lambda_6(E_1; \mathfrak{X}) = (\mathfrak{X} + 2)(\mathfrak{X} - 1),
\end{aligned} \tag{11.3.47}$$

where

$$\begin{aligned}
g_1(\mathfrak{X}) &:= \mathfrak{X}^4 + 3\mathfrak{X}^2 + 2\mathfrak{X} + 3, \\
g_2(\mathfrak{X}) &:= (\mathfrak{X}^4 + 2\mathfrak{X}^3 + 5\mathfrak{X}^2 - 2\mathfrak{X} + 3)(\mathfrak{X}^4 - 2\mathfrak{X}^3 + \mathfrak{X}^2 + 6\mathfrak{X} + 3), \\
g_3(\mathfrak{X}) &:= \mathfrak{X}^2 + \mathfrak{X} + 1, \\
g_4(\mathfrak{X}) &:= \mathfrak{X}^4 + 2\mathfrak{X}^3 - \mathfrak{X}^2 - 2\mathfrak{X} + 9.
\end{aligned} \tag{11.3.48}$$

From (11.1.14) the free energy per lattice site in the limit $\mathcal{N} \rightarrow \mathcal{L}_1$ is

$$\phi_0(\mathcal{L}_1, K) = -\frac{1}{2} \ln[\Lambda_1(A; \mathfrak{X})]. \tag{11.3.49}$$

The formulae for the internal energy and heat capacity are as given in the leading equalities in (11.3.36) and (11.3.37) and these quantities are plotted in Fig. 11.2.

With $N_2 = 3$ the symmetry group is $\mathfrak{S}_3 \times \mathfrak{C}_{3v}$:

The lattice slices consist of rings of three sites. Appropriately ordered the lattice states now fall into five equivalence classes meaning that there are five eigenvalues

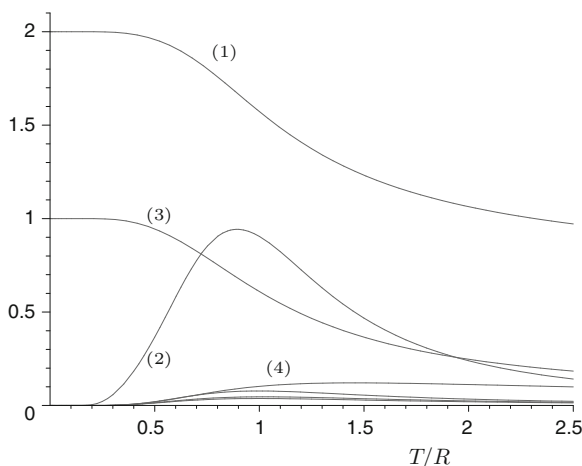


Fig. 11.2 Plots of (1) $-u(\mathcal{L}_1, K)$, (2) $c(\mathcal{L}_1, K)/k_B$, (3) the largest scaled eigenvalue, (4) a group of seven curves representing the remaining scaled eigenvalues, against $T/R = 1/K$

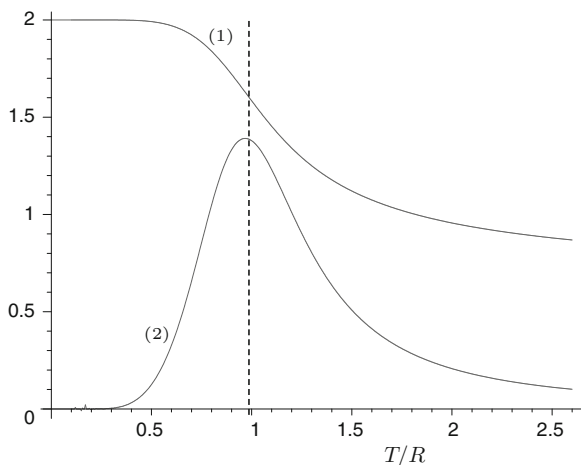
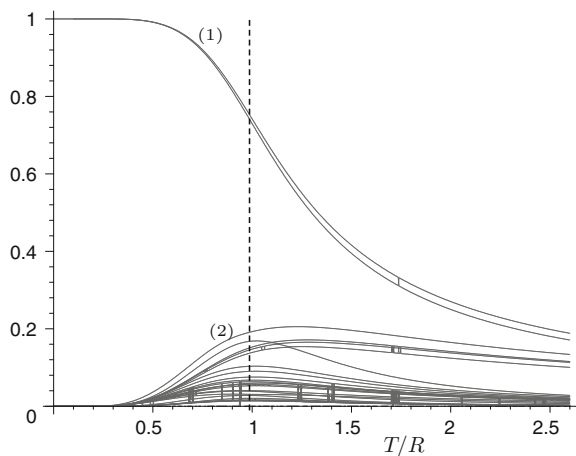


Fig. 11.3 Plots of (1) $-u(\mathcal{L}_1, K)$, (2) $c(\mathcal{L}_1, K)/k_B$, against $T/R = 1/K$. The broken line indicates the location of the critical temperature $T_c/R = 0.9950$, given by (9.2.28), of the two-dimensional square-lattice model

Fig. 11.4 Plots of (1) the two largest scaled eigenvalues, (2) the remaining twenty-four scaled eigenvalues, against $T/R = 1/K$



in the one-dimensional symmetric representation A. The details of the analysis of this case, which follows the same lines as that $N_2 = 1, 2$, are omitted. The internal energy density and heat capacity are shown in Fig. 11.3. The location of the incipient phase transition given by the maximum in the heat capacity approximates quite well, even for this small value of N_2 , the exact value of the critical temperature. The scaled eigenvalues are shown in Fig. 11.4. Here we begin to see the behaviour which is evident in the discussion of the exact eigenvalues for the spin- $\frac{1}{2}$ Ising model in Sect. 11.4.3. One or more scaled eigenvalues (two in this case) approach the limiting

value of unity at a certain temperature, which provides another approximation for the critical temperature, while the remaining scaled eigenvalues peak in the same vicinity.

11.4 Analysis in the Complex Plane: The Wood Method

Here we consider $\mathbb{C}_{\mathfrak{X}}$, the complex plane of a Boltzmann factor $\mathfrak{X}$, associated with an interaction energy (rather than a field) and in terms of which the characteristic equation of the transfer matrix is

$$\text{Det}\{\mathbf{I}_M \Lambda - \mathbf{V}\} = \mathcal{F}(N_2, \mathfrak{X}; \Lambda) = 0, \quad (11.4.1)$$

where

$$\mathcal{F}(N_2, \mathfrak{X}; \Lambda) := \Lambda^M + f_{M-1}(N_2, \mathfrak{X})\Lambda^{M-1} + \cdots + f_0(N_2, \mathfrak{X}) \quad (11.4.2)$$

and $\mathfrak{X} := \exp(nK)$, n being chosen for convenience; ($n = 2$ for the case of the spin- $\frac{1}{2}$ Ising model in Sect. 11.4.3).⁸ In general the coefficients $f_k(N_2, \mathfrak{X})$, $k = 0, 1, \dots, M$, can be sums of monomials in $\mathfrak{X}$ with negative and positive exponents. We can, however, without loss of generality, suppose that a constant has been added to all the interaction energies so that the coefficients are analytic polynomial functions of $\mathfrak{X}$.⁹

As in Sects. 11.1 and 11.2 the eigenvalues of $\mathbf{V}$ are, at this point, denoted by $\Lambda_m(N_2, \mathfrak{X})$, $m = 1, 2, \dots, M$, without distinguishing their association with the irreducible representations of the symmetry group of the system. The partition function is given by (11.1.6) which is exact for finite N_1 and N_2 and can be expanded to give a polynomial in $\mathfrak{X}$ with positive real coefficients. The eigenvalues will be finite in $\mathbb{C}_{\mathfrak{X}}$, except possibly at the point at infinity. They may however, have a complicated branch-point structure. This cancels from the summation in the partition function to produce an analytic function with a finite set of zeros. We denote this set by $\mathbb{O}(N_1, N_2)$.

11.4.1 Evolution of Partition Function Zeros

Any two eigenvalues $\Lambda_m(N_2, \mathfrak{X})$ and $\Lambda_{m'}(N_2, \mathfrak{X})$ will have the same magnitude on a line (if it exists) for which $\mathfrak{X}$ satisfies

$$\Lambda_m(N_2, \mathfrak{X}) = \varpi \Lambda_{m'}(N_2, \mathfrak{X}), \quad \varpi := \exp(i\vartheta), \quad 0 \leq \vartheta < 2\pi. \quad (11.4.3)$$

⁸ The system may have a number of couplings, but these can, as we have already seen, be included by using parameter ratios of K .

⁹ For an example of this see the discussion of the square-lattice Ising model in Sect. 11.4.3.

In general the order of the relative magnitudes of the eigenvalues will interchange across this line. The set of all curves defined by (11.4.3) for all m and all $m' \neq m$ is denoted by $\mathcal{D}_{N_2}$. This forms a complicated network in the plane $\mathbb{C}_{\mathfrak{X}}$. Let $\mathcal{M}_{N_2}$ be the subset of $\mathcal{D}_{N_2}$ for which the eigenvalues satisfying (11.4.3) are of maximum magnitude. It follows that, for any $\mathfrak{X} \notin \mathcal{M}_{N_2}$, there is a unique eigenvalue of maximum magnitude. This we denote by $\Lambda_{\max}(N_2, \mathfrak{X})$.¹⁰ We shall now show that, in the limit $N_1 \rightarrow \infty$, the zeros $\mathbb{O}(N_1, N_2)$ converge onto the curves $\mathcal{M}_{N_2}$. To make this idea precise the following theorem is needed:

Theorem 11.4.1 *For every $\mathfrak{X} \notin \mathcal{M}_{N_2}$ and any ϵ satisfying the condition $|\Lambda_{\max}(N_2, \mathfrak{X})| > \epsilon > 0$, there exists an $N_0 > 0$ such that, for all $N_1 > N_0$, $|Z(N_1, N_2, \mathfrak{X})|^{1/N_1} > \epsilon$.*

Proof From (11.1.6),

$$|Z(N_1, N_2, \mathfrak{X})|^{1/N_1} = |\Lambda_{\max}(N_2, \mathfrak{X})| \left| \sum_{m=1}^M [\Omega_m(N_2, \mathfrak{X})]^{N_1} \right|^{1/N_1}, \quad (11.4.4)$$

where $\{\Omega_m(N_2, \mathfrak{X})\}$ are the scaled eigenvalues defined in (11.1.10). Since the transfer matrix $\mathbf{V}$ is non-singular for any $\mathfrak{X}$, not all of its eigenvalues are zero. In particular it is possible, for the given ϵ , to find a $B > 0$ such that $|\Lambda_{\max}(N_2, \mathfrak{X})| > B > \epsilon$. Since $\mathfrak{X} \notin \mathcal{M}_{N_2}$ only one scaled eigenvalue has modulus of unity. Choosing

$$\eta := \frac{B - \epsilon}{B(M - 1)} > 0, \quad (11.4.5)$$

it is, therefore, possible to find an N_0 such that

$$|\Omega_k(N_2, \mathfrak{X})|^{N_1} < \eta, \quad \forall N_1 > N_0 \quad (11.4.6)$$

and every scaled eigenvalue with modulus less than unity. It follows, from (11.4.4)–(11.4.6), that

$$|Z(N_1, N_2, \mathfrak{X})|^{1/N_1} > B[1 - (M - 1)\eta]^{1/N_1} > B[1 - (M - 1)\eta] = \epsilon. \quad (11.4.7)$$

□

This theorem establishes the result that no point $\mathfrak{X} \notin \mathcal{M}_{N_2}$ can be a member of $\mathbb{O}(N_1, N_2)$ for arbitrarily large N_1 . Thus that the set

$$\lim_{N_1 \rightarrow \infty} \mathbb{O}(N_1, N_2) = \mathbb{O}(\mathcal{L}_1, N_2) \in \mathcal{M}_{N_2}. \quad (11.4.8)$$

¹⁰ Bearing in mind that the definition is ambiguous on the curves of $\mathcal{M}_{N_2}$ across which there is a change of analytic form.

Further insight into this situation can be gained if we consider a curve defined by (11.4.3), for particular values $m = m_0$ and $m' = m'_0$. If this curve belongs to $\mathcal{M}_{N_2}$, then, in the limit $N_1 \rightarrow \infty$,

$$Z(N_1, N_2, \mathfrak{X}) \simeq [\Lambda_{m_0}(N_2, \mathfrak{X})]^{N_1} [1 + \exp(i\vartheta N_1)]. \quad (11.4.9)$$

This asymptotic expression is zero at the values $\vartheta = (2r + 1)\pi/N_1$, $r = 0, 1, \dots, N_1 - 1$, which correspond to a set of points on the curve, with density which increases with N_1 . For $M = 2$, as would be the case for the one-dimensional Ising model, this argument establishes the fact that the zeros of the partition function lie on the single curve given by (11.4.3) for all N_1 . For $M > 2$, it adds understanding rather than rigour to the result established by Theorem 11.4.1.

On $\mathbb{R}^{(+)}$, the positive real axis of $\mathbb{C}_{\mathfrak{X}}$, Perron's theorem applies (Sect. 17.2) and $\Lambda_{\max}(N_2, \mathfrak{X})$ denotes (unambiguously) a real positive non-degenerate eigenvalue. Let the eigenvalues be ordered so that $\Lambda_1(N_2, \mathfrak{X})$ is identified by the condition that $\Lambda_1(N_2, \mathfrak{X}) = \Lambda_{\max}(N_2, \mathfrak{X})$, for $\mathfrak{X} \in \mathbb{R}^{(+)}$. Although curves belonging to $\mathcal{D}_{N_2}$ can cross or touch $\mathbb{R}^{(+)}$, none that do can belong to $\mathcal{M}_{N_2}$. We now define $\mathcal{E}_{N_2}$ as the subset of $\mathcal{M}_{N_2}$ consisting of points which can be reached by a continuous path from some point of $\mathbb{R}^{(+)}$ without first crossing some other member of $\mathcal{M}_{N_2}$. It must be the case that, along the curves of $\mathcal{E}_{N_2}$, $\Lambda_1(N_2, \mathfrak{X})$ is one of the eigenvalues which satisfies (11.4.3).

Suppose that $\mathbb{W}$ is a region in $\mathbb{C}_{\mathfrak{X}}$ containing a segment of $\mathbb{R}^{(+)}$, but none of the members of $\mathcal{O}(N_1, N_2)$ for all N_2 and all N_1 greater than some N_0 . Then, by a slight modification of Theorem 8.4.1, it follows that $\phi(\mathcal{L}, \mathfrak{X})$, the free energy per lattice site for the infinite system $\mathcal{L}$ (with both $N_1 \rightarrow \infty$ and $N_2 \rightarrow \infty$), is a regular function of $\mathfrak{X}$ in $\mathbb{W}$. No singularity of $\phi(\mathcal{L}, \mathfrak{X})$ can occur at a real value of $\mathfrak{X} \in \mathbb{W} \cap \mathbb{R}^{(+)}$. For any fixed value of N_2 it follows from Theorem 11.4.1 that N_0 can be chosen so that, if $\mathbb{W} \cap \mathcal{E}_{N_2} = \emptyset$, then $\mathcal{O}(N_1, N_2) \cap \mathbb{W} = \emptyset$ for $N_1 > N_0$. The final conclusion of this argument is that singularities of the infinite system $\mathcal{L}$ are possible only at real values of $\mathfrak{X}$ to which some points of $\mathcal{E}_{N_2}$ converge as $N_2 \rightarrow \infty$.

11.4.2 Connection Curves and Cross-Block Curves

Suppose, as in Sect. 11.3, that the irreducible representations of the symmetry group $\mathcal{Q}$ are $\Theta(k)$, $k = 1, 2, \dots, \kappa$. The transfer matrix can be block-diagonalized into blocks corresponding to these representations with, in the case of a representation $\Theta(k)$ of dimension $e(k) > 1$, a further reduction into $e(k)$ identical blocks. The characteristic equation take the form

$$\prod_{k=1}^{\kappa} [\text{Det}\{\mathbf{I}_{m(k)}\Lambda - \mathbf{V}'(k)\}]^{e(k)} =: \mathcal{F}(N_2, \mathfrak{X}; \Lambda) = 0, \quad (11.4.10)$$

where $V'(k)$ is one of the $e(k)$ identical blocks obtained by block-diagonalizing the matrix $\tilde{V}(k)$ which appears in (11.3.3). The factors in the product in (11.4.10) are polynomials of degree $m(k)$ with

$$\text{Det}\{\mathbf{I}_{m(k)}\Lambda - V'(k)\} = \mathcal{G}(k; N_2, \mathfrak{X}; \Lambda), \quad (11.4.11)$$

where

$$\mathcal{G}(k; N_2, \mathfrak{X}; \Lambda) := \Lambda^{m(k)} + g_{m(k)-1}(k; N_2, \mathfrak{X})\Lambda^{m(k)-1} + \cdots + g_0(k; N_2, \mathfrak{X}). \quad (11.4.12)$$

Thus we have

$$\mathcal{F}(N_2, \mathfrak{X}; \Lambda) = \prod_{k=1}^{\kappa} [\mathcal{G}(k; N_2, \mathfrak{X}; \Lambda)]^{e(k)}. \quad (11.4.13)$$

The coefficients in (11.4.12) are polynomials in $\mathfrak{X}$ and, if the factorization (11.4.13) were complete in the sense that no further reduction into factors with polynomial coefficients were possible, then the eigenvalue solutions $\Lambda_\ell(k; N_2, \mathfrak{X})$, $\ell = 1, 2, \dots, m(k)$, of

$$\mathcal{G}(k; N_2, \mathfrak{X}; \Lambda) = 0 \quad (11.4.14)$$

would be the branches of an (*irreducible*) algebraic function (Sect. 17.6). There are two reasons why this might not be the case. Firstly, it is sometimes quite difficult to determine the complete symmetry group of the system and any further ‘hidden’ symmetry would lead to a factorization of $\mathcal{G}(k; N_2, \mathfrak{X}; \Lambda)$. Secondly, we have no guarantee that the full symmetry group will lead to a complete factorization of the characteristic equation. An example of this is given, in Sect. 11.4.3, by the square-lattice Ising model. For the moment we shall suppose that $\mathcal{G}(k; N_2, \mathfrak{X}; \Lambda)$ is irreducible. The only modification needed if this were not the case, would be an additional index to distinguish the separate factors. The partition function is now expressible in the form

$$Z(N_1, N_2, \mathfrak{X}) = \sum_{k=1}^{\kappa} \sum_{\ell=1}^{m(k)} [\Lambda_\ell(k; N_2, \mathfrak{X})]^{N_1} \quad (11.4.15)$$

and the curves in $\mathcal{D}_{N_2}$ are distinguished by their representations. The set for which $\mathfrak{X}$ satisfies the equation

$$\Lambda_\ell(k; N_2, \mathfrak{X}) = \varpi \Lambda_{\ell'}(k'; N_2, \mathfrak{X}), \quad \varpi = \exp(i\vartheta), \quad 0 \leq \vartheta < 2\pi, \quad (11.4.16)$$

for all $\ell = 1, 2, \dots, m(k)$, $\ell' = 1, 2, \dots, m(k')$ is denoted by $\mathcal{C}_{N_2}(k, k')$. For reasons made clear in Sect. 17.6, $\mathcal{C}_{N_2}(k, k)$ are called *connection curves* and $\mathcal{C}_{N_2}(k, k')$, with $k \neq k'$ are called *cross-block curves*. The total set

$$\bigcup_{\{k, k'\}} \mathcal{C}_{N_2}(k, k') = \mathcal{D}_{N_2}. \quad (11.4.17)$$

On $\mathbb{R}^{(+)}$, $\Lambda_{\max}(N_2, \mathfrak{X}) = \Lambda_1(1; N_2, \mathfrak{X})$. So $\mathcal{E}_{N_2}$ consists of curves in the set $\mathcal{C}_{N_2}(k, 1)$ for some k .

In principle an investigation of the evolution of the set $\mathcal{E}_{N_2}$ for increasing N_2 should provide some indication of the critical point of the system in the limit $N_2 \rightarrow \infty$. However, the dimension M of the transfer matrix will grow very rapidly with N_2 , with all the resulting computational difficulties. These can be lessened by using block-diagonalization of the transfer matrix, particularly if $\mathcal{E}_{N_2}$ consists of connection curves in $\mathcal{C}_{N_2}(1, 1)$ rather than cross-block curves in $\mathcal{C}_{N_2}(k, 1)$ with $k \neq 1$. Then it will be necessary to calculate only the one-dimensional symmetric block of the transfer matrix, which is straightforward using formula (11.3.9). Unfortunately this is not necessarily the case, as we shall see in Sect. 11.4.3 for the square-lattice first-neighbour spin- $\frac{1}{2}$ Ising model and as was also shown for the square-lattice Ising model with equal first- and second-neighbour interactions by Wood (1987). The interesting and important feature of this method is the argument that, as $N_2 \rightarrow \infty$, $\mathcal{E}_{N_2}$ converges on to that part $\mathcal{C}_{N_2}^*(1, 1)$ of $\mathcal{C}_{N_2}(1, 1)$ which involves $\Lambda_1(1; N_2, \mathfrak{X})$. This means that, for finite N_2 , the curves of $\mathcal{C}_{N_2}^*(1, 1)$ *rather than* $\mathcal{E}_{N_2}$ can be used to obtain an approximation to the critical properties of the infinite system. Two arguments can be advanced for the truth of this contention. The first of a general nature and the second based on calculations for particular systems. Our discussion in Sect. 11.4.3 for the square-lattice spin- $\frac{1}{2}$ Ising model falls in the latter category.

In a general context the argument¹¹ is based on a comparison between two different finite systems, one with the toroidal boundary conditions used here and the other with boundaries and a contribution to the Hamiltonian $\hat{b}(\tau(N_1 - 1), \tau(0))$ as is given in (11.1.1). This results in a modification to the partition function formula (11.4.15) in which N_1 is replaced by $N_1 - 1$ and the terms are weighted by coefficients $\beta(k; N_2, \mathfrak{X})$ dependent on the boundary conditions. Finally it can be shown that a particular symmetry choice of boundary conditions will yield $\beta(k; N_2, \mathfrak{X}) = \delta^{kr}(k - 1)$ with

$$Z(N_1, N_2, \mathfrak{X}) = \sum_{\ell=1}^{m(1)} [\Lambda_{\ell}(1; N_2, \mathfrak{X})]^{N_1-1}. \quad (11.4.18)$$

Both this partition function and (11.4.15) give the same limit for the free energy per lattice site. This means that the critical properties of the two systems in the limit $N_2 \rightarrow \infty$ are identical and are also the same as that of a *reduced* system with toroidal boundary conditions, with the transfer matrix consisting of just the one-dimensional symmetric block of that for the full system, and $N_1 - 1$ rather than N_1 slices. A sequence of finite cases of each of these three systems can be understood to give a succession of approximations to the same infinite system. This is illustrated in Fig. 12.7 for the square-lattice Ising model, with $N_2 = 1000$. It can be seen that

¹¹ Described in detail in Lavis and Bell (1999b, Sects. 4.9 and 4.10)

both the ratio of the two largest eigenvalues and the two largest eigenvalues *in the one-dimensional symmetric representation* approach unity at the critical point.

11.4.3 The Spin- $\frac{1}{2}$ Square-Lattice Ising Model

In this case the symmetry group is $\mathcal{G} := \mathcal{C}_{N_2V} \times \mathcal{S}_2^{(m)}$, where (Sect. 17.3.4) $\mathcal{C}_{N_2V}$ is the group of rotations of the lattice and reflections through planes containing the lattice axis and $\mathcal{S}_2^{(m)}$ corresponds to spin inversion. To illustrate the methods described in this section we look ahead to the exact formula for the transfer matrix eigenvalues given in Sect. 12.2.7, restricting attention to the isotropic case $K^{(h)} = K^{(ch)} = K$, $\mathfrak{X}_h = \mathfrak{X}_{ch} = \mathfrak{X} = \exp(2K)$. To ensure that the coefficients in the characteristic equation (11.4.1) are polynomials in positive powers of $\mathfrak{X}$ we add the constant $-J$ to each first-neighbour pair interaction, giving $-2J$ and zero for a pair of parallel and anti-parallel spins, respectively. This means that each term of the transfer matrix and hence each eigenvalue is multiplied by a factor $\mathfrak{X}^{N_2}$ giving, from (12.2.149),¹²

$$\Lambda^{(s)}(N_2, \mathfrak{X}; \boldsymbol{\gamma}) = [\mathfrak{X}(\mathfrak{X}^2 - 1)]^{\frac{1}{2}N_2} \prod_{\{\psi\}^{(s)}} \exp\left(\frac{1}{2} \gamma_\psi \alpha \omega(\mathfrak{X}; \psi)\right), \quad (11.4.19)$$

where, from (18.3.111), (12.2.136) and (12.2.137), $\{\psi\}^{(s)}$ is the set of values

$$\psi_j^{(s)} := \frac{\pi(4j - s - 1)}{4N_2}, \quad j = 1, 2, \dots, N_2, \quad (11.4.20)$$

$$\omega(\mathfrak{X}; \psi) := \ln \left[W(\mathfrak{X}; \psi) + \sqrt{W^2(\mathfrak{X}; \psi) - 1} \right], \quad (11.4.21)$$

$$W(\mathfrak{X}; \psi) := \frac{(\mathfrak{X}^2 + 1)^2 - 2\mathfrak{X}(\mathfrak{X}^2 - 1) \cos(2\psi)}{2\mathfrak{X}(\mathfrak{X}^2 - 1)} \quad (11.4.22)$$

and $\alpha = \pm 1$ is chosen so that $\alpha \omega(\mathfrak{X}; \psi) > 0$ at low temperatures. As is shown in Sect. 12.2.7 this means that $\alpha = 1$ if $\psi \neq 0$ and, for $\psi = 0$, $\alpha = 1$ for $T < T_c$ and $\alpha = -1$ for $T > T_c$ where $\mathfrak{X}_c = \exp(2J/T_c)$ is given by $\mathfrak{X}_c(\mathfrak{X}_c - 2) = 1$.

From (11.4.21) and (11.4.22) $\exp[\omega(\mathfrak{X}; \psi)]$ will have branch-points corresponding to zeros of the functions $W(\mathfrak{X}; \psi) \mp 1$ unless they are both perfect squares, which, from (12.2.151), is the case when $\psi = 0$, which is a member of $\{\psi\}^{(-)}$, or $\psi = \frac{1}{2}\pi$, which is a member of $\{\psi\}^{(-)}$ if N_2 is even and $\{\psi\}^{(+)}$ if N_2 is odd.¹³ In all other cases the zeros of these quantities give eight branch-points for $\exp[\omega(\mathfrak{X}; \psi)]$. These branch-points will appear in the eigenvalues given by (11.4.19) unless γ_ψ and $\gamma_{\pi-\psi}$

¹² This formula, which is that derived by Onsager (1944) is obtained by neglecting the factors $\ell^{(s)}(\boldsymbol{\gamma})$ and $(k/2)^{n_2}$ in (12.2.149), which are a consequence of the method of derivation.

¹³ For deriving the eigenvalues in Sect. 12.2 we have assumed that N_2 is even but this is not needed for the standard Onsager derivation.

have opposite signs. In that case the branch-points are eliminated. With φ defined in terms of $\mathfrak{X}$ by

$$\mathfrak{X} := \exp(i\varphi) \pm \sqrt{\exp(2i\varphi) + 1}, \quad (11.4.23)$$

the zeros of $W(\mathfrak{X}; \psi) \mp 1$ are given, respectively by¹⁴

$$\cos \varphi = \begin{cases} \cos^2(\psi) \\ -\sin^2(\psi). \end{cases} \quad \psi \in \psi_j^{(s)}. \quad (11.4.24)$$

If the characteristic polynomial of any model is expressed, in the form (11.4.13), as a product of irreducible polynomials, then we know that every branch of an algebraic function must be an eigenvalue in the same irreducible representation. However, as we shall now see the converse is not necessarily the case. Not all eigenvalues in the same irreducible representation are branches of the same algebraic function. For the present model the eigenvalues which are the branches of a single algebraic function can be identified in the sets $s = 1$ and $s = -1$ of (11.4.19) from their branch-points. Given any eigenvalue, another branch of the same algebraic function can be obtained by changing the signs of γ_ψ and $\gamma_{\pi-\psi}$. On $\mathbb{R}^{(+)}$ the largest eigenvalue is $\Lambda_{\max}(N_2, \mathfrak{X}) = \Lambda^{(+)}(N_2, \mathfrak{X}; \mathbf{j})$, where $\mathbf{j} = (1, 1, \dots, 1)$. This is, therefore, a branch of an algebraic function of degree

$$d(N_2) = \begin{cases} 2^{\frac{1}{2}N_2}, & N_2 \text{ even,} \\ 2^{\frac{1}{2}(N_2-1)}, & N_2 \text{ odd.} \end{cases} \quad (11.4.25)$$

Using (11.3.16), we see that $m(1; N_2) = d(N_2)$ for $N_2 = 2, 3, \dots, 6$, but $m(1; 7) = 9 > d(7) = 8$ and $m(1; 8) = 18 > d(8) = 16$. For $N_2 > 6$ the one-dimensional symmetric representation contains eigenvalues, with a different branch-point structure from $\Lambda_{\max}(N_2, \mathfrak{X})$. However, it is known that the eigenvalues of any representation are either entirely in the set $s = 1$ or $s = -1$, according respectively to whether the representation in question is formed from the one-dimensional symmetric or anti-symmetric representation of $S_2^{(m)}$. It follows that all the eigenvalues of the one-dimensional symmetric representation are members of the set $s = 1$.

We now return to (11.4.23) and (11.4.24) which define the branch-points of the eigenvalues. If φ_0 is a solution of the first of the equations (11.4.24) in the interval $[0, \frac{1}{2}\pi]$, then $\varphi_1 \in [\frac{1}{2}\pi, \pi]$ satisfies the second equation when $\cos \varphi_1 = \cos \varphi_0 - 1$. The connection curve in $\mathbb{C}_{\mathfrak{X}}$ described by (11.4.23) with the positive sign and $\varphi \in [\varphi_0, \varphi_1]$ has a cusp at $\varphi = \frac{1}{2}\pi$, $\mathfrak{X} = i$. Another connection curve is given by taking $\varphi \in [-\varphi_1, -\varphi_0]$. The two connection curves corresponding to the negative sign in (11.4.23) can be obtained by applying the mapping $\mathfrak{X} \rightarrow -1/\mathfrak{X}$. It is not difficult to

¹⁴ Meaning that the four roots are given by taking each sign in (11.4.23) and the two roots obtained by solving the quadratic for $\exp(i\varphi)$ given by the appropriate case of (11.4.24).

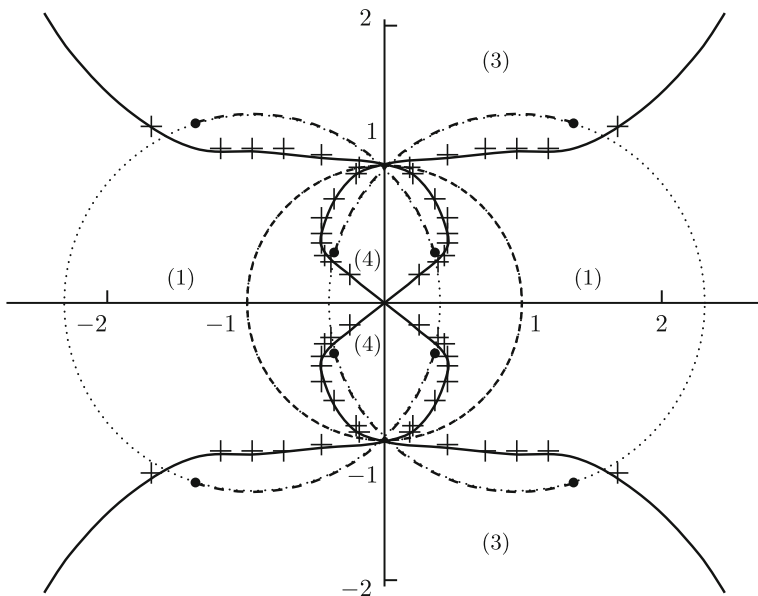


Fig. 11.5 The connection and cross-block curves, in the plane $\mathbb{C}_{\mathfrak{X}}$, for the Ising model with $N_2 = 2$. The zeros of $Z(16, 2, \mathfrak{X})$ are marked by + and the region in which $\Lambda_1(k; 2, \mathfrak{X})$ is the eigenvalue of largest magnitude is labelled (k)

show that these four curves form pairs of crossing curves lying on the *Fisher circles*

$$\mathfrak{X} = \sqrt{2} \exp(i\theta) \pm 1. \quad (11.4.26)$$

These connection curves for $\psi = \pi/4$, $N_2 = 2$, when $\varphi_0 = \frac{1}{3}\pi$, $\varphi_1 = \frac{2}{3}\pi$ are shown in Fig. 11.5 as chain lines terminating at the eight branch-points which are denoted by $\bullet$. The Fisher circles (11.4.26) are completed with dotted lines.

The second largest eigenvalue is $\Lambda_{\max}^{(-)}(N_2, \mathfrak{X})$, given by (12.2.167). $\mathcal{E}_{N_2}$ therefore consists of cross-block curves. Although it is possible that the zeros of $Z(N_1, N_2, \mathfrak{X})$ all lie on the part of $\mathcal{M}_{N_2}$ which is not $\mathcal{E}_{N_2}$ it is more plausible to expect, (as in the case $N_2 = 2$ investigated below), that some of the zeros $\mathbb{O}(N_1, N_2)$ for any N_1 will lie on $\mathcal{E}_{N_2}$. It can, however, be shown (Fisher 1965; Wood 1985) that the limiting locus of partition function zeros for this model is the pair of Fisher circles (11.4.26). It must, therefore, be concluded that, in the double limit $N_1 \rightarrow \infty$, $N_2 \rightarrow \infty$, the curve $\mathcal{E}_{N_2}$ closes in on the circles (11.4.26), which according to Theorem 11.4.1 contain the zeros of the partition function $Z^{(1)}(N_1, N_2, \mathfrak{X})$, defined by (11.4.18). These zeros lie on the connection curve $\mathcal{C}_{N_2}^*(1, 1)$

$$\Lambda_{\max}^{(-)}(N_2, \mathfrak{X}) = \varpi \Lambda_{\text{submax}}^{(-)}(N_2, \mathfrak{X}), \quad (11.4.27)$$

where $\Lambda_{\text{submax}}^{(-)}(N_2, \mathfrak{X})$, given by (12.2.165), is the second largest member of the set (11.4.19) with $s = -1$. This curve has branch-points on the circle $\mathfrak{X} = \sqrt{2} \exp(i\theta) + 1$ at the pair of points

$$\begin{aligned} \mathfrak{X} = \exp \left\{ \pm i \arccos \left[\cos^2 \left(\frac{\pi}{2N_2} \right) \right] \right\} \\ + \sqrt{\exp \left\{ \pm 2i \arccos \left[\cos^2 \left(\frac{\pi}{2N_2} \right) \right] \right\} + 1}. \end{aligned} \quad (11.4.28)$$

As $N_2 \rightarrow \infty$, $\mathfrak{X} \rightarrow 1 + \sqrt{2}$, which is equivalent to (9.2.36), the formula for the critical temperature of the square-lattice Ising ferromagnet. All the connection curves form a set of overlapping arcs on the circles (11.4.26) and in this limit branch-points of all the connection curves approach the critical point which becomes for them a point of accumulation. The critical point can also be seen as a result of analytically continuing $\mathcal{C}_{N_2}^*(1, 1)$ through the branch-points to $\mathbb{R}^{(+)}$.

The remarkable feature of this result is that, although the critical point is for the system in the thermodynamic limit of infinite N_1 and N_2 , it applies to the eigenvalues for any finite N_2 . This can be seen by a more detailed examination of a special case.

Example 11.4.1 The eigenvalues for $N_2 = 2$ are given in Example 11.3.1. Those results can be utilized, but it must be noted that here, in order to ensure that all the elements of the transfer matrix are polynomials in $\mathfrak{X}$, we must multiply the eigenvalue (11.3.28) by $\mathfrak{X}^2$, and the quadratic (11.3.25) must be modified by multiplying the linear term by $\mathfrak{X}^2$ and the constant term by $\mathfrak{X}^4$. This having been done the connection curves $\mathcal{C}_2(1, 1)$ are given, either using (11.4.3) and (11.3.26) or the resultant formula (17.6.19) with the quadratic (11.3.25). They have the structure described above for the general N_2 case and are shown in Fig. 11.5 by chain lines. From (11.3.28) the cross-block curve $\mathcal{C}_2(3, 4)$ is the circle $|\mathfrak{X}| = 1$ (represented by a broken line in Fig. 11.5). From (11.3.26) and (11.3.28) the cross-block curves $\mathcal{C}_2(1, 3)$ and $\mathcal{C}_2(1, 4)$ are both given by the solutions of

$$\mathfrak{X}^2(\mathfrak{X}^2 - 1) \exp(2i\vartheta) - (\mathfrak{X}^2 + 1)^2 \exp(i\vartheta) + \mathfrak{X}^2 - 1 = 0. \quad (11.4.29)$$

These curves are represented by continuous lines in Fig. 11.5. On that part of the curves for which $|\mathfrak{X}| > 1$, $|\Lambda_1(A; 2, \mathfrak{X})| = |\Lambda_1(B_1; 2, \mathfrak{X})|$ and when $|\mathfrak{X}| < 1$, $|\Lambda_1(A; 2, \mathfrak{X})| = |\Lambda_1(B_3; 2, \mathfrak{X})|$. The exterior and interior of the circle $|\mathfrak{X}| = 1$ correspond respectively to the ferromagnetic ($J > 0$) and antiferromagnetic ($J < 0$) models. The inversion symmetry of Fig. 11.5 with respect to this circle corresponds to the invariance of the free energy of the square-lattice Ising model with respect to interchange of the sign of J . It follows from the topological structure of Fig. 11.5 that, for $J > 0$, the curve $|\Lambda_1(A; 2, \mathfrak{X})| = |\Lambda_1(B_3; 2, \mathfrak{X})|$ forms the subset $\mathcal{M}_2$ along which there are two eigenvalues of maximum magnitude. In a similar way $|\Lambda_1(A; 2, \mathfrak{X})| = |\Lambda_1(B_3; 2, \mathfrak{X})|$ gives the subset $\mathcal{M}_2$ when $J < 0$. The eigenvalues of maximum magnitude in each of the regions formed by the curves $\mathcal{M}_2$ are shown

in Fig. 11.5. The curves of $\mathcal{E}_2$ consist of that part of $\mathcal{M}_2$ to the right of the imaginary axis. The conclusion for general N_2 that $\mathcal{E}_{N_2}$ consists of cross-block curves is, therefore, explicitly confirmed in this case. According to Theorem 11.4.1 the zeros of $Z(N_1, 2, \mathfrak{X})$ converge on to $\mathcal{M}_2$ as $N_1 \rightarrow \infty$. The zeros of $Z(16, 2, \mathfrak{X})$ are indicated by crosses in Fig. 11.5. Even with this comparatively small value of N_1 the conclusion of the theorem is supported.

We need at this point to distinguish those features of the behaviour of the square-lattice Ising model which can be expected to appear in a wide class of lattice models and those which arise because of its special properties. The characteristic of general validity is the convergence with increasing N_2 of $\mathcal{E}_{N_2}$ on to $\mathcal{C}_{N_2}^*(1, 1)$, with the analytic continuation of $\mathcal{C}_{N_2}^*(1, 1)$ to $\mathbb{R}^{(+)}$ yielding an approximation to the critical point. A special feature of the Ising model is that $\mathcal{C}_{N_2}^*(1, 1)$ lies on the *same* curves (11.4.26) for all values on N_2 . This means that the *exact* critical temperature can be obtained using any finite value of N_2 . The reason for this can be found in the self-duality of this model. With

$$\mathfrak{S} := \sinh(2K) = \frac{1}{2}(\mathfrak{X} - \mathfrak{X}^{-1}), \quad (11.4.30)$$

the duality transformation (9.2.32) is given by

$$\mathfrak{S} \longleftrightarrow \mathfrak{S}^* = 1/\mathfrak{S} \quad (11.4.31)$$

and the curves given by

$$\mathfrak{S} = \exp(i\varphi), \quad -\pi \leq \varphi < \pi, \quad (11.4.32)$$

are invariant under the duality transformation, which takes $\varphi \rightarrow -\varphi$. Equations (11.4.23) give the solutions of (11.4.32) and it follows that all the branch-points and all the connection curves lie on the self-dual curves (11.4.32) for all N_2 .

Duality also leads to the appearance of a curve which contains $\mathcal{C}_{N_2}^*(1, 1)$ for all N_2 in the case of the triangular Ising model and the square-lattice $\mu = 0$ ν -state Potts model (Wood et al. 1987). The interesting question concerning the use of the method to uncover approximate, or hitherto unsuspected symmetries, leading to critical points given in terms of algebraic numbers, has been discussed by Wood and Ball (1990).

11.4.4 Critical Points and Exponents

As has been indicated above, the procedure, in the absence of any special symmetry properties, is to analytically continue the connection curve $\mathcal{C}_{N_2}^*(1, 1)$ to $\mathbb{R}^{(+)}$ to give an approximation to the critical point. Two different approaches can be taken to this task.

In the first it is supposed that the appropriate branch of the connection curve has been identified by

$$\Lambda_1(1; N_2, \mathfrak{X}) = \varpi \Lambda_2(1; N_2, \mathfrak{X}), \quad (11.4.33)$$

with known expressions in terms of $\mathfrak{X}$ for the eigenvalues. With

$$f(\mathfrak{X}) = \frac{\Lambda_1(1; N_2, \mathfrak{X})}{\Lambda_2(1; N_2, \mathfrak{X})}, \quad (11.4.34)$$

the connection curve is given by $|f(\mathfrak{X})| = 1$, with $f(\mathfrak{X}) = 1$ at the branch-points. The analytic continuation of the connection curve through the branch-points, called the *extension curve*, is defined by $\text{Arg}\{f(\mathfrak{X})\} = 0$. Now let

$$\begin{aligned} \mathfrak{X} &:= r \exp(i\theta), \\ f(\mathfrak{X}) &:= \mathfrak{F}(r, \theta) \exp[i\Theta(r, \theta)]. \end{aligned} \quad (11.4.35)$$

The extension curve is given by

$$\Theta(r, \theta) = 0, \quad (11.4.36)$$

with normal in the direction

$$\hat{n}(r, \theta) = \left(\frac{\partial \Theta}{\partial r}, \frac{1}{r} \frac{\partial \Theta}{\partial \theta} \right). \quad (11.4.37)$$

Except at the branch-points the Cauchy–Riemann conditions

$$\frac{\partial \mathfrak{F}}{\partial r} = \frac{\mathfrak{F}}{r} \frac{\partial \Theta}{\partial \theta}, \quad \frac{1}{r} \frac{\partial \mathfrak{F}}{\partial \theta} = -\mathfrak{F} \frac{\partial \Theta}{\partial r}, \quad (11.4.38)$$

are satisfied. The approximation to the critical point is given by the intersection of the extension curve and $\mathbb{R}^{(+)}$. If this intersection is orthogonal then, from (11.4.37) and (11.4.38)

$$\frac{\partial \Theta}{\partial \theta} = 0, \quad \frac{\partial \mathfrak{F}}{\partial r} = 0. \quad (11.4.39)$$

The second of these conditions, identifies $\mathfrak{X}_c$ as a turning point of $|f(\mathfrak{X})|$ along $\mathbb{R}^{(+)}$.

The second approach, which maps out the whole of $\mathcal{C}_{N_2}(1, 1)$ and its extension curves, is to begin with the defining polynomial $\mathcal{G}(1; N_2, \mathfrak{X}; \Lambda)$, given by (11.4.12), for the eigenvalues in the one-dimensional symmetric representation. According to the method described in Sect. 17.6 we now form the polynomial

$$\mathcal{H}(1; N_2, \mathfrak{X}; \Lambda) := \frac{\mathcal{G}(1; N_2, \varpi \mathfrak{X}; \Lambda) - \mathcal{G}(1; N_2, \mathfrak{X}; \Lambda)}{\Lambda(\varpi - 1)}. \quad (11.4.40)$$

The connection curves and their extensions are given by solving the equation

$$\text{Res}\{\mathfrak{Z}, \mathcal{G}, \mathcal{H}\} = 0. \quad (11.4.41)$$

Example 11.4.2 Using the second procedure, we consider the hard-square lattice fluid of particles of fugacity $\mathfrak{Z} := \exp(\mu/T)$ on the square lattice (Sect. 8.4.3). With $N_2 = 4$ the symmetry group is $\mathcal{C}_{4v}$ and it is not difficult to show that the eigenvalues of the one-dimensional symmetric representation satisfy the cubic equation

$$\Lambda^3 - \Lambda^2(\mathfrak{Z}^2 + 3\mathfrak{Z} + 1) + \Lambda\mathfrak{Z}(\mathfrak{Z}^2 - \mathfrak{Z} - 1) + \mathfrak{Z}^3 = 0. \quad (11.4.42)$$

In this case the resultant equation (11.4.41) gives

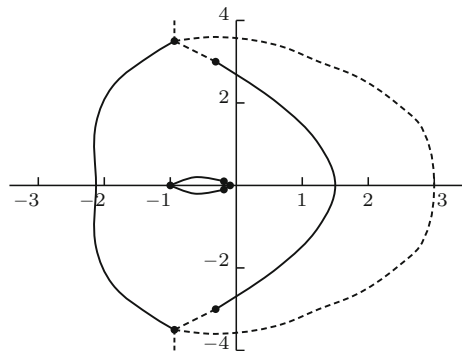
$$\begin{aligned} 0 = & \mathfrak{Z}^8 + 4\mathfrak{Z}^7 + 2(9 - 2\varsigma)(1 + \varsigma)\mathfrak{Z}^6 + 8(5 + 5\varsigma - \varsigma^2)\mathfrak{Z}^5 \\ & + 2(41 + 52\varsigma - 4\varsigma^3)\mathfrak{Z}^4 + 4(24 + 25\varsigma + 4\varsigma^2)\mathfrak{Z}^3 \\ & + 4(13 + 8\varsigma + \varsigma^2)\mathfrak{Z}^2 + 4(3 + \varsigma)\mathfrak{Z} + 1, \end{aligned} \quad (11.4.43)$$

where

$$\varsigma = \frac{1}{2}(\varpi + \varpi^{-1}) = \cos \vartheta. \quad (11.4.44)$$

The connection curves $\mathcal{C}_4(1, 1)$ are the roots of (11.4.43) with $|\varsigma| \leq 1$. The extension curves are given by continuing the curves into the region $|\varsigma| > 1$, where ϑ becomes imaginary. These curves are shown in Fig. 11.6. The extensions of the connection curves through the branch-points at $-0.881 \pm i3.473$ meet $\mathbb{R}^{(+)}$ at $\mathfrak{Z}_c = 3.016$, with $\varsigma_c = 5.186$. These values can be computed from the condition that $\mathfrak{Z}_c$ is the meeting point of two extension curves approaching from above and below. It, therefore, corresponds to a value of ς for which (11.4.43) has a repeated root. This can be found by the using the resultant with respect to $\mathfrak{Z}$ of (11.4.43) and its derivative. The estimate $\mathfrak{Z}_c = 3.016$ for the critical fugacity can be compared (rather unfavourably) with the high-density series result $\mathfrak{Z}_c = 3.7962 \pm 0.0001$ quoted in Sect. 8.4.3.

Fig. 11.6 The connection curves $\mathcal{C}_4(1, 1)$ and their extensions for the hard-square lattice fluid. The connection curves are shown by *continuous lines* and their extensions by *broken lines*



However, it has been shown by Wood et al. (1987) that for $N_2 = 6$ the comparison is much improved with $\mathfrak{Z}_c = 3.730$.

In the picture of the development of critical behaviour originating in the work of Yang and Lee the critical point of a one-parameter lattice system corresponds, in the thermodynamic limit, to a point of accumulation of the zeros of the partition function. We have now shown that the critical point can also be identified as the limit of some of the branch-points of $C_{N_2}^*(1, 1)$ as $N_2 \rightarrow \infty$. Suppose that there are s of these branch-points each with cycle number two. They will all be the termination of connection curves involving $\Lambda_1(1; N_2, \mathfrak{X})$ which coalesce at $\mathfrak{X}_c$ as $N_2 \rightarrow \infty$. Wood (1987) has argued that the result is to form a branch-point of $\Lambda_1(1; \infty, \mathfrak{X})$ with cycle number $p \leq s$. It follows, from the discussion of algebraic functions in Sect. 17.6 and of scaling at a critical point in Sect. 4.5, that the scaling exponent y_2 in a two-dimensional system is given by

$$\frac{2}{y_2} = \frac{q}{p}. \quad (11.4.45)$$

Wood (1987) has suggested that p is likely to take its maximum value s and that q should be some multiple t of $2(s - 1)$, giving the result

$$y_2 = \frac{s}{t(s - 1)}. \quad (11.4.46)$$

This then gives, from (4.5.39) and (4.5.53),

$$\alpha = \frac{2[s - t(s - 1)]}{s}, \quad \nu = \frac{t(s - 1)}{s}. \quad (11.4.47)$$

The first of these formulae can give rise to a singularity in heat capacity only when $t = 1$. For this case the hard-hexagon and 3-state Potts models correspond to $s = 6$, $\alpha = \frac{1}{3}$ and the K-type ($\alpha = \frac{1}{2}$) transition exhibited by the dimer model described in Sect. 13.2.2 corresponds to $s = 4$. In the case of the three-spin Ising model on the triangular lattice the polynomial of which $\Lambda_1(1; \mathfrak{X}, \infty)$ is a branch is known (Baxter and Wu 1973, 1974; Baxter 1974; Joyce 1975a, b). This leads directly to the result (Wood 1987) $2/y_2 = \frac{4}{3}$, $\alpha = \frac{2}{3}$ corresponding to $t = 1$, $s = 3$ in formulae (11.4.46) and (11.4.47). As we have seen in Sect. 11.4.3, for the spin- $\frac{1}{2}$ Ising model on a square lattice, the critical point corresponds, in the limit $N_2 \rightarrow \infty$, to a point of accumulation of an infinite number of branch-points. Thus we have $s = \infty$, which together with $t = 1$ yields $\alpha = 0$, which is consistent with the logarithmic singularity in the heat capacity.

A theory of finite-size scaling of partition function zeros was developed by Itzykson et al. (1983). For a two-dimensional $N_2 \times N_2$ lattice system they proposed the scaling form $|\mathfrak{X}_0 - \mathfrak{X}_c| \sim N_2^{-z}$, for the zero $\mathfrak{X}_0$ closest to the critical value $\mathfrak{X}_c$ in the complex plane $\mathbb{C}_{\mathfrak{X}}$ of some Boltzmann variable $\mathfrak{X}$ for which the

scaling exponent is z . Given that boundary conditions ensure that $\mathcal{E}_{N_2}$ involves only eigenvalues in the one-dimensional symmetric representation, the points of this set closest to $\mathbb{R}^{(+)}$ will be branch-points of $\Lambda_1(1; N_2, \mathfrak{X})$, occurring in conjugate pairs. They correspond to bounds on the asymptotic distribution of partition function zeros. Using this insight Williams and Lavis (1996) proposed a modified version of the formula of Itzykson et al. (1983) for a $N_1 \times N_2$ lattice in the limit $N_1 \rightarrow \infty$. With the conjugate pair of branch-points closest to $\mathfrak{X}_c$ in $\mathbb{C}_{\mathfrak{X}}$ denoted by $\mathfrak{X}_*$ and $\overline{\mathfrak{X}}_*$, the proposed form was $|\mathfrak{X}_* - \mathfrak{X}_c| \sim N_2^{-z}$, for the calculation of z . Support for this method, when $\mathfrak{X}$ is the thermal Boltzmann factor, is given by the behaviour of the rounding exponent w defined in (4.10.18). The quantity N_2^{-w} is of the order of the size of the neighbourhood in which the correlation length exceeds the lattice width (Fisher and Barber 1972). This should be of the same order as the radius of convergence of a power-series expansion of the free energy about the critical point, which will be given by the location of the nearest branch-point of $\Lambda_1(1; N_2, \mathfrak{X})$. The equality (4.10.19), which in the present notation is equivalent to $w = z$, is thought to be satisfied in a wide class of systems, particularly in the case of a cylinder with free boundary conditions at the ends (Fisher and Ferdinand 1967).

Chapter 12

Transfer Matrices: Exactly Solved Models

12.1 A General Eight-Vertex Model

As described in Sect. 3.8.1 the arrows of this model are on a rapidity lattice $\mathcal{N}^{\otimes}$ consisting of a finite number of straight lines, which here we suppose is bounded by a simple closed curve $\mathcal{C}$, as shown in Fig. 16.15. Each lattice edge is the boundary between a shaded and an unshaded face these being decorated with one site of the dual lattices $\mathcal{N}$ and $\mathcal{N}^*$, respectively. Boltzmann weights are given in terms of the spins on the sites of $\mathcal{N}$ and $\mathcal{N}^*$ by (3.8.9) and in terms of edge variables on the edges of $\mathcal{N}^{\otimes}$ by (3.8.10).¹

It is clear that the partition function is invariant under any smooth deformation which, while keeping $\mathcal{C}$ fixed, distorts the lattice without:

- (i) changing the order of the points of intersection of the lattice lines with $\mathcal{C}$,
- (ii) a vertex passing through a lattice line.

Condition (i) is equivalent to preventing vertices passing through $\mathcal{C}$. The interest here is in systems for which the partition function is invariant under the more ‘intrusive’ transformation where condition (ii) is no longer demanded.

Definition 12.1.1 The eight-vertex model on $\mathcal{N}^{\otimes}$ is *Z-invariant* if the partition function is unchanged under any orthogonal translations of lattice lines which conserve the order of the intersections on $\mathcal{C}$.

As we noted in Sect. 3.8.1.3, this eight-vertex model factors into two Ising models when the four-spin interaction is zero and the definition of Z-invariance carries over to the Ising models on $\mathcal{N}$ and $\mathcal{N}^*$. An example of the transformation under which the partition function of a Z-invariant model must remain unchanged is shown in

¹ As indicated in Sect. 3.8.1, the edge variables can be used to assign, using the direction and shading rules, two different arrangements of arrows on the edges of $\mathcal{N}^{\otimes}$. Although these arrows give the popular picture of vertex models they play no direct role in the mathematical analysis. This is achieved in terms of edge variables or spins. We shall, therefore, make little reference to the arrows, and to the alternative direction and shading rules, in this chapter.

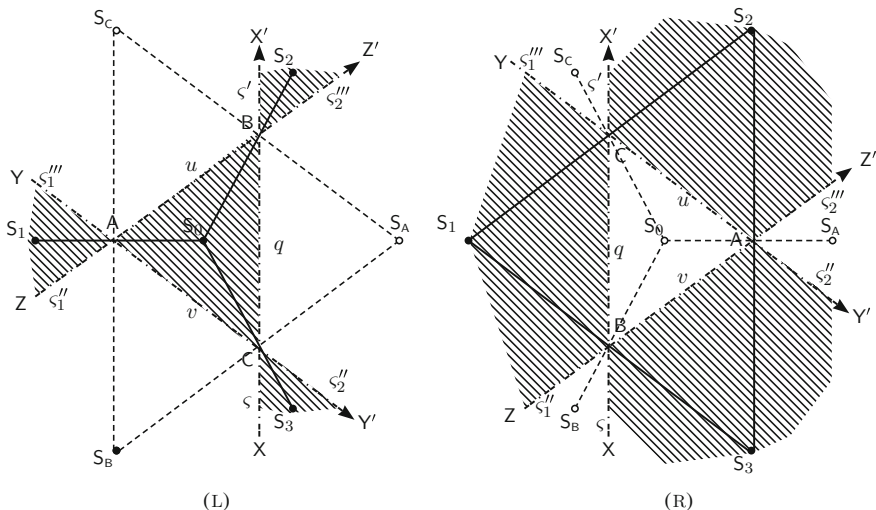


Fig. 12.1 A translation from (L) to (R) of the line XX' . The lines XX' , YY' and ZZ' are given directions and rapidities $\bar{\varrho}$, ϱ and ϱ' , respectively

Fig. 12.1 where A, B and C lie within $\mathcal{C}$ and, in going from (L) to (R), the line XX' passes through A.

12.1.1 A Generalized Star-Triangle Transformation

In Fig. 12.1, XX' , YY' and ZZ' are lines and A, B and C are sites of $\mathcal{N}^{(4)}$; S_1 , S_2 and S_3 are sites of $\mathcal{N}$ and S_A , S_B and S_C are sites of $\mathcal{N}^*$, with σ_0 the spin on site S_0 , for $Q = A, B, C$ and σ_j the spin on site S_j , for $j = 1, 2, 3$. The spin on site S_0 is σ_0 . The couplings between spins are labelled using the angle bisected by the line between the spins. Thus, for example, K_{ZBX} is the coupling between spins σ_0 and σ_2 in Fig. 12.1 (L) and σ_0 and σ_B in Fig. 12.1 (R). The four-spin couplings are $K_A^{(4)}$, $K_B^{(4)}$ and $K_C^{(4)}$. Neglecting the constant term which does not play a role in the analysis, the Boltzmann weights for the three sites in (L) are

$$\begin{aligned} \mathfrak{V}_A(\sigma_1, \sigma_C, \sigma_0, \sigma_B) &:= \exp(K_{YAZ}\sigma_1\sigma_0 + K_{YAZ'}\sigma_B\sigma_C + K_A^{(4)}\sigma_1\sigma_C\sigma_0\sigma_B), \\ \mathfrak{V}_B(\sigma_0, \sigma_C, \sigma_2, \sigma_A) &:= \exp(K_{ZBX}\sigma_0\sigma_2 + K_{ZBX'}\sigma_A\sigma_C + K_B^{(4)}\sigma_0\sigma_C\sigma_2\sigma_A), \\ \mathfrak{V}_C(\sigma_B, \sigma_0, \sigma_A, \sigma_3) &:= \exp(K_{XCY}\sigma_B\sigma_A + K_{XCY'}\sigma_3\sigma_0 + K_C^{(4)}\sigma_B\sigma_0\sigma_A\sigma_3) \end{aligned} \quad (12.1.1)$$

and, distinguishing them with a prime, for (R)

$$\begin{aligned}
\mathfrak{Y}'_A(\sigma_0, \sigma_2, \sigma_A, \sigma_3) &:= \exp(K_{YAZ}\sigma_0\sigma_A + K_{YAZ'}\sigma_3\sigma_2 + K_A^{(4)}\sigma_0\sigma_2\sigma_A\sigma_3), \\
\mathfrak{Y}'_B(\sigma_B, \sigma_1, \sigma_0, \sigma_3) &:= \exp(K_{ZBX}\sigma_B\sigma_0 + K_{ZBX'}\sigma_3\sigma_1 + K_B^{(4)}\sigma_B\sigma_1\sigma_0\sigma_3), \\
\mathfrak{Y}'_C(\sigma_1, \sigma_C, \sigma_2, \sigma_0) &:= \exp(K_{XC_Y}\sigma_1\sigma_2 + K_{XC_Y'}\sigma_0\sigma_C + K_C^{(4)}\sigma_1\sigma_C\sigma_2\sigma_0).
\end{aligned} \tag{12.1.2}$$

Apart from representing a translation of the lattice line XX' through the site A, the transformation in Fig. 12.1 from the left-hand figure (L) to right-hand figure (R) can be regarded as a star-triangle transformation on sites of $\mathcal{N}$ in which summation over σ_0 removes the site S_0 together with a triangle-star transformation on sites of $\mathcal{N}^*$ which introduces the site S_0 . Given that the couplings are unchanged by this procedure the system is said to locally satisfy *generalized star-triangle (GST) conditions* if the partition function is unaffected. More specifically, since the rest of the lattice is left unchanged, the contributions to the partition function for the configurations in Fig. 12.1 (L) and (R) with the spins apart from σ_0 fixed must be the same. For (L) this is

$$\begin{aligned}
\mathcal{Z}(\sigma_A, \sigma_B, \sigma_C, \sigma_1, \sigma_2, \sigma_3) \\
&:= \sum_{\sigma_0=\pm 1} \tilde{\mathfrak{Y}}_A(\sigma_1, \sigma_C, \sigma_0, \sigma_B) \tilde{\mathfrak{Y}}_B(\sigma_0, \sigma_C, \sigma_2, \sigma_A) \tilde{\mathfrak{Y}}_C(\sigma_B, \sigma_0, \sigma_A, \sigma_3) \\
&= 2 \exp(K_{YAZ'}\sigma_B\sigma_C + K_{ZBX'}\sigma_C\sigma_A + K_{XC_Y}\sigma_A\sigma_B) \\
&\quad \times \cosh(K_{YAZ}\sigma_1 + K_{ZBX}\sigma_2 + K_{XC_Y'}\sigma_3 \\
&\quad + K_A^{(4)}\sigma_1\sigma_B\sigma_C + K_B^{(4)}\sigma_2\sigma_C\sigma_A + K_C^{(4)}\sigma_3\sigma_A\sigma_B).
\end{aligned} \tag{12.1.3}$$

It is clear from Fig. 12.1 that the transformation between (L) and (R) is equivalent to the interchange $(\sigma_A, \sigma_B, \sigma_C) \longleftrightarrow (\sigma_1, \sigma_2, \sigma_3)$. So the contribution to the partition function for (R) is

$$\mathcal{Z}'(\sigma_A, \sigma_B, \sigma_C, \sigma_1, \sigma_2, \sigma_3) = \mathcal{Z}(\sigma_1, \sigma_2, \sigma_3, \sigma_A, \sigma_B, \sigma_C) \tag{12.1.4}$$

and the GST condition is

$$\mathcal{Z}(\sigma_A, \sigma_B, \sigma_C, \sigma_1, \sigma_2, \sigma_3) = \mathcal{Z}(\sigma_1, \sigma_2, \sigma_3, \sigma_A, \sigma_B, \sigma_C). \tag{12.1.5}$$

12.1.2 The Solution to the GST Transformation and the Elliptic Variable Formulation

The formula (12.1.5) gives sixty-four equations, with the number reduced to thirty-two because the equations are unchanged by negating all the spin values. However, there are further symmetries because the sets $\{\sigma_A, \sigma_B, \sigma_C\}$ and $\{\sigma_1, \sigma_2, \sigma_3\}$ can be negated independently and the members of the pairs $\{\sigma_A, \sigma_1\}$, $\{\sigma_B, \sigma_2\}$ and $\{\sigma_C, \sigma_3\}$ can be interchanged simultaneously. This leaves eight equations of which

Table 12.1 Spin values for the six independent equations derived from (12.1.5)

$\sigma_{\mathbf{A}}$	$\sigma_{\mathbf{B}}$	$\sigma_{\mathbf{C}}$	σ_1	σ_2	σ_3
+1	+1	+1	+1	+1	-1
+1	+1	+1	+1	-1	+1
+1	+1	+1	+1	-1	-1
+1	+1	-1	+1	-1	+1
+1	+1	-1	+1	-1	-1
+1	-1	+1	+1	-1	-1

two are identities. The spin values for the remaining six equations are given in Table 12.1. With the notation $(K_1, K_2, K_3) := (K_{\text{YAZ}}, K_{\text{ZBX}}, K_{\text{XC}'})$, $(K'_1, K'_2, K'_3) := (K_{\text{YAZ}'}, K_{\text{ZBX}'}, K_{\text{XC}'})$ and $(K^{(4)}_1, K^{(4)}_2, K^{(4)}_3) := (K^{(4)}_{\mathbf{A}}, K^{(4)}_{\mathbf{B}}, K^{(4)}_{\mathbf{C}})$, with the index j being modulo 3, and

$$\begin{aligned} C(x_1, x_2, x_3, x_4, x_5, x_6) := & \cosh(x_1 + x_2 + x_3 + x_4 + x_5 + x_6) \\ & \times \cosh(x_1 - x_2 - x_3 + x_4 + x_5 + x_6), \end{aligned} \quad (12.1.6)$$

these can be simplified to the form

$$\begin{aligned} \exp(4K'_j) &= \frac{C(K_j, K_{j+1}, K_{j+2}, -K^{(4)}_j, -K^{(4)}_{j+1}, K^{(4)}_{j+2})}{C(K_j, K_{j+1}, -K_{j+2}, K^{(4)}_j, K^{(4)}_{j+1}, -K^{(4)}_{j+2})} \\ &= \frac{C(K_j, K_{j+1}, K_{j+2}, -K^{(4)}_j, K^{(4)}_{j+1}, -K^{(4)}_{j+2})}{C(K_j, K_{j+1}, -K_{j+2}, K^{(4)}_j, -K^{(4)}_{j+1}, K^{(4)}_{j+2})}, \quad j = 1, 2, 3. \end{aligned} \quad (12.1.7)$$

The three equalities between the second and third expressions in (12.1.7) can be taken as defining $K^{(4)}_1, K^{(4)}_2$ and $K^{(4)}_3$ as functions of K_1, K_2 and K_3 . However, these equations are not independent since they are satisfied by the infinity of solutions given by

$$K^{(4)}_1 = K^{(4)}_2 = K^{(4)}_3 = K^{(4)}, \quad (12.1.8)$$

for all $K^{(4)}$. When $K^{(4)} = 0$, the transformation decomposes into two star-triangle transformations of the form (9.4.10).

No solutions to (12.1.7) are known (and possibly do not exist) except when (12.1.8) is satisfied. We, therefore, follow Baxter (1978) in assuming this condition when (12.1.7) gives, for $j = 1, 2, 3$,

$$\tanh(2K'_j) = \frac{\sinh(2K_{j+1}) \sinh(2K_{j+2}) - \sinh(2K_j) \sinh(2K^{(4)})}{\cosh(2K_j) \cosh(2K^{(4)}) + \cosh(2K_{j+1}) \cosh(2K_{j+2})}, \quad (12.1.9)$$

and²

$$\Delta_1 = \Delta_2 = \Delta_3, \quad (12.1.10)$$

where

$$\Delta_j := \sinh(2K_j) \sinh(2K'_j) + \tanh(2K^{(4)}) \cosh(2K_j) \cosh(2K'_j). \quad (12.1.11)$$

Referring to Fig. 12.1 and noting the directions of the lattice lines, it can be seen that, for both (L) and (R), K_1 and K'_1 are the forward and backward couplings at site **A** and K_2 and K'_2 are the forward and backward couplings at site **B**. But at site **C** the situation is reversed with K'_3 and K_3 the forward and backward couplings. This means, by reference to (3.8.11), that the Boltzmann weights c_3 and d_3 are interchanged. However, this does not affect the conclusion, using (3.8.12), that

$$\Delta_j = \frac{a_j^2 + b_j^2 - c_j^2 - d_j^2}{2(a_j b_j + c_j d_j)}. \quad (12.1.12)$$

In a similar way, if

$$\Gamma_j := \tanh(2K_j^{(4)}), \quad (12.1.13)$$

then

$$\Gamma_1 = \Gamma_2 = \Gamma_3, \quad (12.1.14)$$

with, from (3.8.13),

$$\Gamma_j = \frac{a_j b_j - c_j d_j}{a_j b_j + c_j d_j}. \quad (12.1.15)$$

Conditions (12.1.10) and (12.1.14) are sufficient to ensure that (12.1.5) is satisfied, when Δ_j and Γ_j are expressed in terms of Ising couplings by (12.1.11) and (12.1.13) or in terms of Boltzmann weights by (12.1.12) and (12.1.15). Although the common values Δ and Γ for Δ_j and Γ_j at every site fix the values of $K_j^{(4)} = K^{(4)}$ the values of K_j and K'_j vary from site to site.

It is well-known from the work of Onsager for the Ising model and Baxter for the eight-vertex model that considerable illumination and simplification can be achieved if the analysis is formulated in terms of Jacobi elliptic functions. In following the same route it is necessary to place some restrictions on the ranges of the parameters used

² The analysis in this chapter is essentially that of Baxter (1978). However, as explained in Sect. 3.8.1, notational changes have been made to bring alignment with Baxter (1982b). These are listed in Table 3.2.

in this discussion. To do so we rely on the results of Sects. 3.8.1.4 and 9.3 for the regular square-lattice model. There it was shown that the condition

$$a > b + c + d, \quad (12.1.16)$$

defines the low-temperature zone³ $\mathcal{R}_L(I)$, of frozen ferromagnetic ordering, in the phase space of (a, b, c, d) , with $a = b + c + d$ being the transition surface $\mathcal{C}(I)$ to the disordered high-temperature zone $\mathcal{R}_H(I)$. It was further shown that symmetry transformations, including the low-to-high-temperature weak-graph transformation, allow the partition function throughout phase space to be derived from that within $\mathcal{R}_L(I)$. With this motivation we impose the condition (12.1.16), which, from (3.8.11), becomes in terms of the couplings

$$\exp[2K^{(4)}] \sinh[K^{(f)} + K^{(b)}] \geq \cosh[K^{(f)} - K^{(b)}]. \quad (12.1.17)$$

As was shown in Sect. 9.3.3, $-1 < \Gamma < 1$ throughout the phase space and $\Delta > 1$ in $\mathcal{R}_L(I)$ with $\Delta = 1$ on $\mathcal{C}(I)$. It follows that the (positive) elliptic modulus k defined by⁴

$$k^2 := \frac{1 - \Gamma^2}{\Delta^2 - \Gamma^2} = \frac{1}{\Delta^2 \cosh^2(2K^{(4)}) - \sinh^2(2K^{(4)})}, \quad (12.1.18)$$

satisfies the condition $0 < k < 1$ in $\mathcal{R}_L(I)$, with $k = 1$ on $\mathcal{C}(I)$. Following Baxter (1978), elliptic parameters λ , α_j and α'_j are defined, in terms of the Jacobi amplitude function (Eq. (18.3.36)), by

$$\begin{aligned} 2K^{(4)} &=: i \operatorname{am}(i[\mathcal{K}' - \lambda]|k), \\ 2K_j &=: -i \operatorname{am}(i[\mathcal{K}' - \alpha_j]|k), & 2K'_j &=: -i \operatorname{am}(i[\mathcal{K}' - \alpha'_j]|k). \end{aligned} \quad (12.1.19)$$

Whatever the value of $k \in [0, 1]$, and whatever are the real values of the couplings, these equations will have, as is seen in Sect. 18.3.5, a solution in the range $[0, 2\mathcal{K}']$, for each of the elliptic parameters. From (18.3.38) and (18.3.45),

$$\begin{aligned} \sinh(2K^{(4)}) &= -i \operatorname{ns}(i\lambda|k)/k, & \cosh(2K^{(4)}) &= i \operatorname{ds}(i\lambda|k)/k, \\ \sinh(2K_j) &= i \operatorname{ns}(i\alpha_j|k)/k, & \cosh(2K_j) &= i \operatorname{ds}(i\alpha_j|k)/k, \\ \sinh(2K'_j) &= i \operatorname{ns}(i\alpha'_j|k)/k, & \cosh(2K'_j) &= i \operatorname{ds}(i\alpha'_j|k)/k. \end{aligned} \quad (12.1.20)$$

³ As indicated in the footnote on p. 297 the term ‘region’ is used to refer to the parts I, II, III, IV of the phase space defined in Sect. 3.8.1.4. Subsets of the regions are referred to as ‘zones’.

⁴ As in Sect. 18.3 we use the notation $\mathcal{K} := \mathcal{K}(k)$ and $\mathcal{K}' := \mathcal{K}(k')$ to denote the complete elliptic integral and its complement for modulus k .

Then, from (12.1.13) and (12.1.18),

$$\Gamma = -\text{nd}(\text{i } \lambda | \mathbf{k}), \quad (12.1.21)$$

$$\Delta = \text{cd}(\text{i } \lambda | \mathbf{k}). \quad (12.1.22)$$

Inserting from (12.1.20) and (12.1.22) into (12.1.11) and using (18.3.61),

$$\begin{aligned} k^2 \text{sn}(\text{i } \alpha_j | \mathbf{k}) \text{sn}(\text{i } \alpha'_j | \mathbf{k}) \{ \text{cn}(\text{i } [\alpha_j + \alpha'_j] | \mathbf{k}) - \text{cn}(\text{i } \lambda | \mathbf{k}) \} \\ = \text{dn}(\text{i } \lambda | \mathbf{k}) - \text{dn}(\text{i } [\alpha_j + \alpha'_j] | \mathbf{k}), \end{aligned} \quad (12.1.23)$$

which is satisfied by

$$\alpha_j + \alpha'_j = \lambda. \quad (12.1.24)$$

Substituting from (12.1.20) into (12.1.9) gives

$$\begin{aligned} k^2 \text{cn}(\text{i } \alpha'_j | \mathbf{k}) \text{sn}(\text{i } \alpha_{j+1} | \mathbf{k}) \text{sn}(\text{i } \alpha_{j+2} | \mathbf{k}) \\ = \text{dn}(\text{i } \alpha_{j+1} | \mathbf{k}) \text{dn}(\text{i } \alpha_{j+2} | \mathbf{k}) - \text{dn}(\text{i } \alpha'_j | \mathbf{k}). \end{aligned} \quad (12.1.25)$$

Comparing with (18.3.61) it can be seen that (12.1.25) is satisfied by

$$\alpha'_j = \alpha_{j+1} + \alpha_{j+2} \quad (12.1.26)$$

and, from (12.1.24),

$$\alpha_1 + \alpha_2 + \alpha_3 = \lambda, \quad (12.1.27)$$

$$\alpha'_1 + \alpha'_2 + \alpha'_3 = 2\lambda. \quad (12.1.28)$$

The GST transformation (12.1.5) is satisfied if the couplings are related to elliptic variables $\mathbf{k}$, λ , α_j , α'_j , for $j = 1, 2, 3$, by (12.1.20) and these satisfy (12.1.24) and (12.1.27), or equivalently (12.1.28). In general there are no other restrictions on the elliptic variables, which can be real or complex.

When $K^{(4)} = 0$ the transformation decomposes into (local) star-triangle transformations on $\mathcal{N}$ and $\mathcal{N}^*$. We note that, for this case, from (18.3.55), (18.3.56), (12.1.11), (12.1.21) and (12.1.22),

$$\lambda = \mathcal{K}', \quad \Delta = \sinh(2K_j) \sinh(2K'_j) = 1/\mathbf{k}, \quad \Gamma = 0. \quad (12.1.29)$$

12.1.3 Z-Invariance

Let the elliptic variables be re-labelled in terms of the angle notation initially used for the couplings. Then (12.1.24), becomes

$$\alpha_{\text{YAZ}} = \lambda - \alpha_{\text{YAZ}'}, \quad \alpha_{\text{ZBX}} = \lambda - \alpha_{\text{ZBX}'}, \quad \alpha_{\text{XCY}'} = \lambda - \alpha_{\text{XCY}}. \quad (12.1.30)$$

α_{YAZ} and $\alpha_{\text{YAZ}'}$ are supplementary angle variables at the site A, as defined in Sect. 16.3, with α_{ZBX} and $\alpha_{\text{ZBX}'}$, and $\alpha_{\text{XCY}'}$ and α_{XCy} being the corresponding pairs for B and C respectively. Then (12.1.27) and (12.1.28), re-expressed in the form

$$\alpha_{\text{YAZ}} + \alpha_{\text{ZBX}} + \alpha_{\text{XCy}'} = \lambda, \quad (12.1.31)$$

$$\alpha_{\text{YAZ}'} + \alpha_{\text{ZBX}'} + \alpha_{\text{XCy}} = 2\lambda, \quad (12.1.32)$$

are the conditions on the interior and exterior sums, respectively, for the triangle ABC. Of course, (12.1.32) follows from (12.1.30) and (12.1.31), and the latter is the condition that the partition function is left unchanged as A passes through the line.

For the system to be Z-invariant GST conditions must hold for every orthogonal translation of a lattice line through a vertex. Let the lines of $\mathcal{N}^{\otimes}$ be L_i , $i = 1, 2, \dots$. Not all the vertices V_{ij} of intersections of pairs of lines L_i and L_j belong to $\mathcal{N}^{\otimes}$ (are inside $\mathfrak{C}$), but for those that are there are couplings K_{ij} , $\bar{K}_{ij}$ and $K^{(4)}$ which determine the elliptic parameters λ and k and the supplementary angle variables α_{ij} and $\bar{\alpha}_{ij}$.⁵ The eight-vertex model on $\mathcal{N}^{\otimes}$ is Z-invariant if the supplementary angle variables at each vertex and the interior angles of each triangle of $\mathcal{N}^{\otimes}$ sum to λ . It is convenient to extend $\mathcal{N}^{\otimes}$ to $\mathcal{E}(\mathcal{N}^{\otimes})$. This extended lattice is created by lengthening the lines of $\mathcal{N}^{\otimes}$ to include all vertices V_{ij} and beyond to a new boundary $\mathfrak{C}'$. At the new vertices angle variables can be *defined* to satisfy the Z-invariance conditions and then couplings defined so that the eight-vertex model is Z-invariant. The eight-vertex model on $\mathcal{E}(\mathcal{N}^{\otimes})$ is Z-invariant if and only if it is invariant on $\mathcal{N}^{\otimes}$.⁶

In Sect. 16.3, IS(c) was used to denote the sum of the interior angle variables of a polygon c of m edges of $\hat{\mathfrak{g}}$ (now replaced by $\mathcal{N}^{\otimes}$ or $\mathcal{E}(\mathcal{N}^{\otimes})$) and it follows from Theorem 16.3.1 that the eight-vertex model on the graph or lattice is Z-invariant if and only the supplementary angle variables at a vertex sum to λ and $\text{IS}(c) = (m - 2)\lambda$, for all polygons formed from edges and vertices of $\mathcal{N}^{\otimes}$. In the proof of this theorem a *base line*, denoted by L_0 , was used to assign to each line L_i of the extended graph a number q_i , called the *rapidity*, and a direction (orientation). This relied on the base line being such that it intersected every line of the extended graph. In the present situation there may be a choice of a number of lines of $\mathcal{E}(\mathcal{N}^{\otimes})$ for the role of base line. Any line not parallel to any other line of $\mathcal{E}(\mathcal{N}^{\otimes})$ will suffice, or if such does not exist an extra line will need to be added. In the latter case angle variables α_{i0} and $\bar{\alpha}_{i0}$ at its points of intersection with the lines of $\mathcal{E}(\mathcal{N}^{\otimes})$ are defined as they were when $\mathcal{N}^{\otimes}$ was augmented to $\mathcal{E}(\mathcal{N}^{\otimes})$. With an orientation given to the base line and consequently to all the lines of $\mathcal{E}(\mathcal{N}^{\otimes})$, as described Sect. 16.3, the notation α_{ij} and $\bar{\alpha}_{ij}$ is refined so that these angle variables correspond respectively to the angles interior and exterior to the two orientation directions. Then the rapidities are simply $q_i := \alpha_{i0}$ giving

⁵ The bar notation is used rather than primes since the supplementary angles are no longer associated with being interior or exterior angles of a particular triangle.

⁶ Although, of course, certain displacements of lattice lines for $\mathcal{E}(\mathcal{N}^{\otimes})$ will not be allowed on $\mathcal{N}^{\otimes}$, since they lead to vertices passing through $\mathfrak{C}$.

$$\alpha_{ij} := \varrho_j - \varrho_i, \quad \bar{\alpha}_{ij} := \lambda - \varrho_j + \varrho_i. \quad (12.1.33)$$

Of course, the rapidities will depend on the choice of base line. But they will differ by an additive quantity determined by setting the rapidity of the base line to zero.

As an example, if ϱ_x , ϱ_y and ϱ_z denote the rapidities of XX' , YY' and ZZ' respectively,

$$\begin{aligned} \alpha_{yAz} &= \varrho_z - \varrho_y, & \alpha_{yAz'} &= \lambda - \varrho_z + \varrho_y, \\ \alpha_{zBx} &= \varrho_x - \varrho_z, & \alpha_{zBx'} &= \lambda - \varrho_x + \varrho_z, \\ \alpha_{xCy} &= \varrho_x - \varrho_y, & \alpha_{xCy'} &= \lambda - \varrho_x + \varrho_y, \end{aligned} \quad (12.1.34)$$

which satisfy (12.1.30)–(12.1.32).

The conclusion of this analysis, derived through Sects. 12.1.1, 12.1.2, and 16.3 is that:

Theorem 12.1.1 *The eight-vertex model on $\mathcal{N}^{(4)}$ is Z -invariant if and only if orientations, and rapidities ϱ_i can be chosen for the lines L_i , $i = 1, 2, \dots$ of $\mathcal{N}^{(4)}$, so that the couplings $K^{(4)}$, K_{ij} and $\bar{K}_{ij}$ are given at every vertex V_{ij} by*

$$\begin{aligned} K^{(4)} &= \frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda]|k), \\ K_{ij} &= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \varrho_j + \varrho_i]|k), \\ \bar{K}_{ij} &= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda + \varrho_j - \varrho_i]|k), \end{aligned} \quad (12.1.35)$$

for some parameter λ and elliptic modulus k .

Suppose that the spins around the vertex V_{ij} are $\sigma_{ij}^{(k)}$, $k = a, b, c, d$, where $k = a, c$ correspond to the spins on a bond like that shown in Fig. 16.16a and $k = b, d$ for a bond like that shown in Fig. 16.16b. Neglecting any constant term, the Hamiltonian is

$$\begin{aligned} \hat{H}(K_{ij}, \bar{K}_{ij}, K_{ij}^{(4)}; \sigma) &= - \sum_{\{ij\}} \{ K_{ij} \sigma_{ij}^{(a)} \sigma_{ij}^{(c)} + \bar{K}_{ij} \sigma_{ij}^{(b)} \sigma_{ij}^{(d)} \\ &\quad + K_{ij}^{(4)} \sigma_{ij}^{(a)} \sigma_{ij}^{(b)} \sigma_{ij}^{(c)} \sigma_{ij}^{(d)} \}, \end{aligned} \quad (12.1.36)$$

where we have (temporarily) generalized⁷ to the case of different four-spin couplings on the different sites. Then, for V_{ij} , the local correlations which contribute to the enthalpy are

⁷ See Sect. 2.5 for a similar case of generalization.

$$\begin{aligned}
\langle \sigma_{ij}^{(a)} \sigma_{ij}^{(c)} \rangle &= -N^{\otimes} \frac{\partial \phi^{(G)}}{\partial K_{ij}}, & \langle \sigma_{ij}^{(b)} \sigma_{ij}^{(d)} \rangle &= -N^{\otimes} \frac{\partial \phi^{(G)}}{\partial \bar{K}_{ij}}, \\
\langle \sigma_{ij}^{(a)} \sigma_{ij}^{(b)} \sigma_{ij}^{(c)} \sigma_{ij}^{(d)} \rangle &= -N^{\otimes} \frac{\partial \phi^{(G)}}{\partial K_{ij}^{(4)}},
\end{aligned}
\tag{12.1.37}$$

where the generalized free energy per vertex is

$$\phi^{(G)}(\{K_{ij}\}, \{\bar{K}_{ij}\}, \{K_{ij}^{(4)}\}); \tag{12.1.38}$$

that is a function of the whole sets of couplings.⁸ As $K_{ij}^{(4)} \rightarrow K^{(4)}$,

$$\phi^{(G)}(\{K_{ij}\}, \{\bar{K}_{ij}\}, \{K_{ij}^{(4)}\}) \rightarrow \phi(\{K_{ij}\}, \{\bar{K}_{ij}\}, K^{(4)}). \tag{12.1.39}$$

Differentiation with respect to $K^{(4)}$ now ‘picks up’ the equal four-spin correlations at every lattice site and the third of Eqs. (12.1.37) is replaced by

$$\langle \sigma_{ij}^{(a)} \sigma_{ij}^{(b)} \sigma_{ij}^{(c)} \sigma_{ij}^{(d)} \rangle = -\frac{\partial \phi}{\partial K^{(4)}}. \tag{12.1.40}$$

Given that V_{ij} is deep in the lattice, it is reasonable to suppose that these correlations will be independent of the couplings on and position of a lattice line L' far away near $\mathcal{C}'$.⁹ However, for a Z -invariant system, L' can¹⁰ be orthogonally translated to approach V_{ij} without altering the partition function and hence the correlations. Indeed any orthogonal translations can be applied to lattice lines apart from L_i and L_j without changing the correlations. It follows that they must be dependent at most on $\varrho_i, \varrho_j, \lambda$ and k and be independent of the arrangement of the other lines of the lattice. They must, therefore, be the same as that for the regular eight-vertex model with lattice lines forming parallelograms with rapidities ϱ_i on the lines parallel to L_i and rapidities ϱ_j on the lines parallel to L_j . That is with couplings $K_{ij}, \bar{K}_{ij}$ and $K^{(4)}$ at every vertex. This regular eight-vertex model corresponds to the limits $K_{gh} \rightarrow K_{ij}, \bar{K}_{gh} \rightarrow \bar{K}_{ij}$ for all vertices V_{gh} with

$$\phi(\{K_{ij}\}, \{\bar{K}_{ij}\}, K^{(4)}) \rightarrow \phi^{(R)}(K_{ij}, \bar{K}_{ij}, K^{(4)}) =: \phi_{ij}^{(R)} \tag{12.1.41}$$

and the regular lattice correlation functions are given by

⁸ Which, exceptionally in this case, we have indicated by the enclosing brackets $\{\cdot\}$.

⁹ Meaning that in order to get from V_{ij} to any part of L' many other lattice lines need to be crossed.

¹⁰ As long as $\mathcal{C}'$ is drawn so that no vertices of $\mathcal{E}(\mathcal{N}^{\otimes})$ pass through it in the process.

$$\begin{aligned}
\langle \sigma_{ij}^{(a)} \sigma_{ij}^{(c)} \rangle^{(R)} &= -\frac{\partial \phi_{ij}^{(R)}}{\partial K_{ij}}, & \langle \sigma_{ij}^{(b)} \sigma_{ij}^{(d)} \rangle^{(R)} &= -\frac{\partial \phi_{ij}^{(R)}}{\partial \bar{K}_{ij}}, \\
\langle \sigma_{ij}^{(a)} \sigma_{ij}^{(b)} \sigma_{ij}^{(c)} \sigma_{ij}^{(d)} \rangle^{(R)} &= -\frac{\partial \phi_{ij}^{(R)}}{\partial K^{(4)}}.
\end{aligned}
\tag{12.1.42}$$

Returning now to the generalized model, with small variations δK_{ij} , $\delta \bar{K}_{ij}$ and $\delta K_{ij}^{(4)}$ in all the couplings,

$$\begin{aligned}
\delta \phi^{(G)} &= \sum_{\{ij\}} \left\{ \frac{\partial \phi^{(G)}}{\partial K_{ij}} \delta K_{ij} + \frac{\partial \phi^{(G)}}{\partial \bar{K}_{ij}} \delta \bar{K}_{ij} + \frac{\partial \phi^{(G)}}{\partial K_{ij}^{(4)}} \delta K_{ij}^{(4)} \right\} \\
&= \frac{1}{N^{(*)}} \sum_{\{ij\}} \left\{ \frac{\partial \phi_{ij}^{(R)}}{\partial K_{ij}} \delta K_{ij} + \frac{\partial \phi_{ij}^{(R)}}{\partial \bar{K}_{ij}} \delta \bar{K}_{ij} + \frac{\partial \phi_{ij}^{(R)}}{\partial K_{ij}^{(4)}} \delta K_{ij}^{(4)} \right\} \\
&= \frac{1}{N^{(*)}} \sum_{\{ij\}} \delta \phi_{ij}^{(R)}.
\end{aligned}
\tag{12.1.43}$$

So $N^{(*)} \phi^{(G)} - \sum_{\{ij\}} \delta \phi_{ij}^{(R)}$ is stationary with respect to these small variations if the system is Z-invariant. However, while keeping the system Z-invariant, the couplings can be increased. In the limit the only term in the partition function both for the general and regular models has all the spins aligned. So for all Z-invariant coupling values it must be the case that

$$\phi^{(G)}(\{K_{ij}\}, \{\bar{K}_{ij}\}, \{K_{ij}^{(4)}\}) = \frac{1}{N^{(*)}} \sum_{\{ij\}} \phi^{(R)}(K_{ij}, \bar{K}_{ij}, K_{ij}^{(4)}).
\tag{12.1.44}$$

This decomposition of the free-energy density of the general Z-invariant eight-vertex model is the “key result” of Baxter (1978) since it means that the solution of the regular model by Baxter (1972b) implies the solution of the general Z-invariant model.

12.1.4 Edge Variables and Matrix Formulation

The expression for the Boltzmann weight at a site in terms of spin variables can be transformed into one in terms of edge variables using (3.8.2). With edge variables, spins and rapidities as shown in Fig. 12.2a, the pair and four-spin couplings are given, from (12.1.35), by

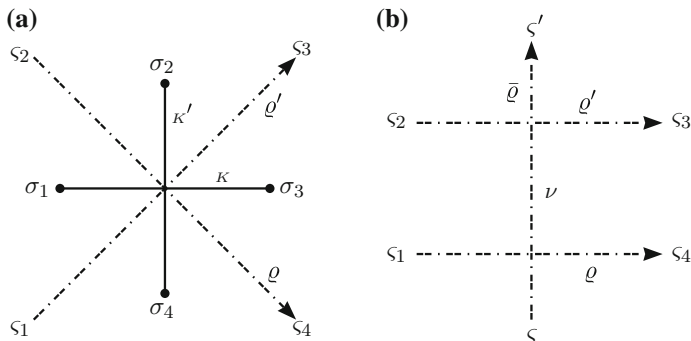


Fig. 12.2 Edges, couplings and spins at a lattice site

$$\begin{aligned}
 K &= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \varrho' + \varrho]|\mathbf{k}), & K' &= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda + \varrho' - \varrho]|\mathbf{k}), \\
 K^{(4)} &= \frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda]|\mathbf{k}),
 \end{aligned}
 \tag{12.1.45}$$

and the four-dimensional matrix $U_0(K, K')$ with elements parameterized by pairs of edge variables is defined by¹¹

$$\left\langle \begin{smallmatrix} s_2 \\ s_1 \end{smallmatrix} \middle| U_0(K, K') \middle| \begin{smallmatrix} s_3 \\ s_4 \end{smallmatrix} \right\rangle := \begin{cases} \exp(K^{(0)} + K s_1 s_4 + K' s_1 s_2 + K^{(4)} s_1 s_3), & \text{if } s_1 s_2 s_3 s_4 = 1, \\ 0, & \text{if } s_1 s_2 s_3 s_4 \neq 1. \end{cases}
 \tag{12.1.46}$$

It is clear that these elements are invariant under a cyclic permutation of the edge variables by two steps and by one step accompanied by an interchange of K and K' . At this point it is convenient to re-represent the matrix $U_0(K, K')$, using (12.1.45) and Fig. 12.2a, in terms of rapidities as $U_0(\varrho' - \varrho)$, when $U_0(K', K)$ becomes $U_0(\lambda - \varrho' + \varrho)$. Then

$$\left\langle \begin{smallmatrix} s_2 \\ s_1 \end{smallmatrix} \middle| U_0(\varrho' - \varrho) \middle| \begin{smallmatrix} s_3 \\ s_4 \end{smallmatrix} \right\rangle = \left\langle \begin{smallmatrix} s_4 \\ s_3 \end{smallmatrix} \middle| U_0(\varrho' - \varrho) \middle| \begin{smallmatrix} s_1 \\ s_2 \end{smallmatrix} \right\rangle = \left\langle \begin{smallmatrix} s_3 \\ s_2 \end{smallmatrix} \middle| U_0(\lambda - \varrho' + \varrho) \middle| \begin{smallmatrix} s_4 \\ s_1 \end{smallmatrix} \right\rangle.
 \tag{12.1.47}$$

We shall also need the four-dimensional matrix $U_2\left(\frac{s'}{s}; \frac{\bar{\varrho} - \varrho'}{\bar{\varrho} - \varrho}\right)$ with elements

$$\left\langle \begin{smallmatrix} s_2 \\ s_1 \end{smallmatrix} \middle| U_2\left(\frac{s'}{s}; \frac{\bar{\varrho} - \varrho'}{\bar{\varrho} - \varrho}\right) \middle| \begin{smallmatrix} s_3 \\ s_4 \end{smallmatrix} \right\rangle := \sum_{\{v\}} \left\langle \begin{smallmatrix} s_1 \\ s \end{smallmatrix} \middle| U_0(\bar{\varrho} - \varrho) \middle| \begin{smallmatrix} v \\ s_4 \end{smallmatrix} \right\rangle \left\langle \begin{smallmatrix} s_2 \\ v \end{smallmatrix} \middle| U_0(\bar{\varrho} - \varrho') \middle| \begin{smallmatrix} s' \\ s_3 \end{smallmatrix} \right\rangle,
 \tag{12.1.48}$$

¹¹ For the sake of brevity constant and four-spin couplings are not included as arguments.

as represented by Fig. 12.2b. Then the contribution to the partition functions from the triangle in Fig. 12.1(L) is

$$\begin{aligned} \mathcal{Z}(\varsigma_1'', \varsigma_1''', \varsigma, \varsigma', \varsigma_2'', \varsigma_2''') &= \sum_{\{v, \mu\}} \left\langle \frac{\varsigma_1'''}{\varsigma_1''} \left| U_0(\varrho' - \varrho) \right| \frac{\mu}{v} \right\rangle \left\langle \frac{\mu}{v} \left| U_2 \left(\frac{\varsigma'}{\varsigma}; \frac{\bar{\varrho} - \varrho'}{\bar{\varrho} - \varrho} \right) \right| \frac{\varsigma_2'''}{\varsigma_2''} \right\rangle \\ &= \left\langle \frac{\varsigma_1'''}{\varsigma_1''} \left| U_0(\varrho' - \varrho) U_2 \left(\frac{\varsigma'}{\varsigma}; \frac{\bar{\varrho} - \varrho'}{\bar{\varrho} - \varrho} \right) \right| \frac{\varsigma_2'''}{\varsigma_2''} \right\rangle. \end{aligned} \quad (12.1.49)$$

In a similar way the contribution to the partition functions from the triangle in Fig. 12.1(R) is

$$\mathcal{Z}'(\varsigma_1'', \varsigma_1''', \varsigma, \varsigma', \varsigma_2'', \varsigma_2') = \left\langle \frac{\varsigma_2'''}{\varsigma_2'} \left| U_2 \left(\frac{\varsigma'}{\varsigma}; \frac{\bar{\varrho} - \varrho}{\bar{\varrho} - \varrho'} \right) U_0(\varrho' - \varrho) \right| \frac{\varsigma_2'''}{\varsigma_2''} \right\rangle \quad (12.1.50)$$

and the GST transformation is equivalent to

$$U_0(\varrho' - \varrho) U_2 \left(\frac{\varsigma'}{\varsigma}; \frac{\bar{\varrho} - \varrho'}{\bar{\varrho} - \varrho} \right) = U_2 \left(\frac{\varsigma'}{\varsigma}; \frac{\bar{\varrho} - \varrho}{\bar{\varrho} - \varrho'} \right) U_0(\varrho' - \varrho) \quad \forall \varsigma, \varsigma'. \quad (12.1.51)$$

12.1.5 Square-Lattice Models

For simplicity the rest of the chapter is restricted to models on the square lattice. In particular the square-lattice version of the model of Baxter (1978) as shown in Fig. 12.3 and its special cases which are versions of the Ising model. Adopting the notation of Chap. 11, but with $\mathcal{N}$ replaced by $\mathcal{N}^{\otimes}$, the sites of $\mathcal{N}^{\otimes}$ are labelled by the index pairs (i, j) , where $i = 1, 2, \dots, N_1$ and $j = 1, 2, \dots, N_2$ are respectively the vertical and horizontal indices. Periodic boundary conditions are applied in the horizontal direction, so that the lattice is in the form of a vertical cylinder, and, straining the notions of vertical and horizontal a little,¹² also in the vertical direction, so the indices i and j are, respectively integers modulo N_1 and N_2 . Sites of sublattices $\mathcal{N}_{\text{E}}^{\otimes}$ and $\mathcal{N}_{\text{O}}^{\otimes}$, as defined in Sect. 3.8.1, correspond respectively to $(i - j)$ being even or odd and it is clear that in order for the sublattice structure to coexist with the periodic boundary conditions both N_1 and N_2 must be even.¹³ The index i labels a slice of the lattice which is either a row (ring) of $\mathcal{N}^{\otimes}$ or equivalently the overlaid zigzag rows of $\mathcal{N}$ and $\mathcal{N}^*$. Given that the system is Z -invariant with the same four-spin coupling

¹² See the footnote on p. 60.

¹³ For the Ising model in Sect. 12.2 it is necessary to impose the more stringent conditions that N_1 and N_2 are divisibly by four, but for the treatment of the square-lattice eight-vertex model in Sect. 12.3 no restrictions on N_1 and N_2 need be applied.

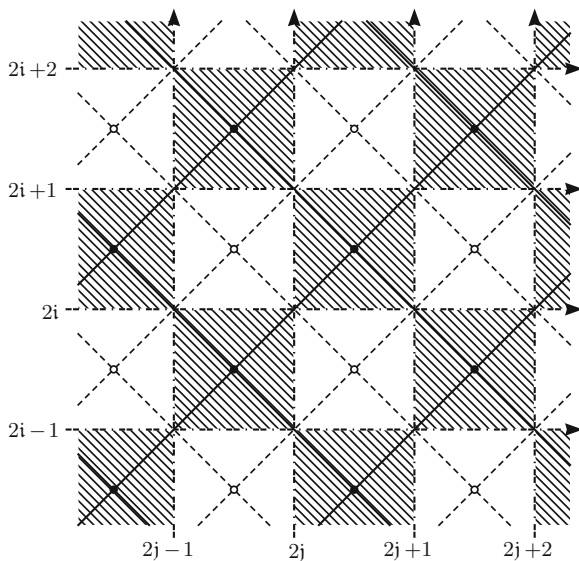


Fig. 12.3 The square lattice version of the model of Baxter (1978). The sites of $\mathcal{N}$ and $\mathcal{N}^*$ are respectively on the *shaded* and *unshaded* faces of $\mathcal{N}^{\otimes}$. The sublattices $\mathcal{N}_E^{\otimes}$ and $\mathcal{N}_O^{\otimes}$ of $\mathcal{N}^{\otimes}$ have (as in Fig. 3.3) *shaded* faces on the *top-right* (and *bottom-left*) and *top-left* (and *bottom-right*) respectively

$$K^{(4)} := \frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda]|\mathbf{k}), \quad (12.1.52)$$

at every site, directions and rapidities q_i and $\bar{q}_j$ can be assigned to the horizontal and vertical lattice lines and, from (12.1.35), the couplings are related to the rapidities by

$$\begin{aligned} K_{ij}^{(f)} &:= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \bar{q}_j + q_i]|\mathbf{k}), \\ K_{ij}^{(b)} &:= -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - \lambda + \bar{q}_j - q_i]|\mathbf{k}). \end{aligned} \quad (12.1.53)$$

12.2 Square-Lattice Ising Models

The work on the square-lattice Ising model which evolved from the work of Onsager pictured the lattice with horizontal and vertical edges; this we shall refer to as the *standard square-lattice model*.¹⁴ Of course, the orientation of the lattice does not affect the partition function but it does affect the transfer matrix and its eigenvalues

¹⁴ Baxter (1978, p. 14) calls this the ‘regular’ model but we shall use this term to refer to the special case which is presented by Baxter (1982b, Chap. 7).

as can be seen from the presentation of Baxter (1982b, Chap. 7) where the lattice edges are at angles of 45° to the horizontal. This can be called the *checkerboard Ising model* because the lattice sites can be viewed as located at the centres of the black or white squares of a checkerboard (draughts board).¹⁵

The application of transfer matrices to the special case $K^{(4)} = 0$ of the Baxter (1978) model, described in this chapter, presents two possibilities. With the eight-vertex model on a kagomé lattice the spin representation divides into Ising models on triangular and honeycomb lattices and with the eight-vertex model on a square lattice the spin representation divides into two equivalent Ising models on square lattices oriented so that the diagonals of the squares are horizontal and vertical. For simplicity we concentrate solely on the latter case with the method of solution following that of Baxter (1982b, Chap. 7) and Baxter (1986). The lattice of each Ising model consists of the $\frac{1}{2}N_1N_2$ sites on alternate faces of the eight-vertex model, with periodic boundary conditions in both lattice directions and N_1 and N_2 even. So each is a checkerboard Ising model. For reasons which become apparent it will be convenient to assume that both N_1 and N_2 are divisible by four¹⁶ and the integers

$$n_1 := \frac{1}{4}N_1, \quad n_2 := \frac{1}{4}N_2 \quad (12.2.1)$$

will be used. As the model is Z -invariant, with the couplings being functions of rapidity differences, it is often more convenient to use the variables¹⁷

$$u := \bar{\varrho} - \varrho, \quad u' := \bar{\varrho} - \varrho'. \quad (12.2.2)$$

Now $\lambda = \mathcal{K}'$ and the system is defined in terms of u and the elliptic modulus k ,¹⁸ with, from (12.1.53), the forward and backward couplings (at the same site of $\mathcal{N}^{(*)}$) given by

$$\begin{aligned} \exp[2K^{(f)}(u)] &= \exp\{-i \operatorname{am}(i[\mathcal{K}' - u]|k)\}, \\ \exp[2K^{(b)}(u)] &= \exp\{-i \operatorname{am}(i u|k)\}. \end{aligned} \quad (12.2.3)$$

From (18.3.89),

$$\begin{aligned} \sinh[2K^{(f)}(u)] &= -i \operatorname{sn}(i[\mathcal{K}' - u]|k), \quad \sinh[2K^{(b)}(u)] = -i \operatorname{sn}(i u|k), \\ \cosh[2K^{(f)}(u)] &= \operatorname{cn}(i[\mathcal{K}' - u]|k), \quad \cosh[2K^{(b)}(u)] = \operatorname{cn}(i u|k), \end{aligned} \quad (12.2.4)$$

¹⁵ Although, as we saw in Sect. 3.8.1.5 and as will be evident below, the term is usually used in rather more restrictive circumstances.

¹⁶ See the footnote on p. 393. Since a motivation of this analysis is to develop a method which allow the limit $N_1, N_2 \rightarrow \infty$, this is not a serious restriction.

¹⁷ Although, since both the horizontal rapidity ϱ and the vertical rapidity $\bar{\varrho}$ can differ between sites it is sometimes necessary to display them both explicitly.

¹⁸ In the interests of brevity this latter dependence is not always displayed.

where

$$K^{(f)}(u) = K^{(b)}(\mathcal{K}' - u) \quad (12.2.5)$$

and, from (18.3.50),

$$\sinh[2K^{(f)}(u)] \sinh[2K^{(b)}(u)] = 1/k, \quad (12.2.6)$$

for all u , as has already been shown in the context of the GST transformation in (12.1.29).

Couplings which satisfy (12.2.4)–(12.2.6), with $u = \bar{\varrho} - \varrho$, for interactions between sites on two lattice pairs which pass through the $\mathcal{N}^{(*)}$ site of intersection of lines with rapidities ϱ and $\bar{\varrho}$, are said to be *conjugate*. In the present context it can be seen that the pairs are one each between sites of $\mathcal{N}$ and $\mathcal{N}^*$,¹⁹ with $K^{(f)}(u)$ and $K^{(b)}(u)$ between site-pairs where, with respect to the directions of the rapidity lines the first divides the interior and the second the exterior angle. From (18.3.52),²⁰

$$\begin{aligned} K^{(f)}(-u) &= K^{(f)}(u) \pm \frac{1}{2}i\pi, & K^{(b)}(-u) &= -K^{(b)}(u), \\ \sinh[2K^{(f)}(-u)] &= -\sinh[2K^{(f)}(u)], & \sinh[2K^{(b)}(-u)] &= -\sinh[2K^{(b)}(u)], \\ \cosh[2K^{(f)}(-u)] &= -\cosh[2K^{(f)}(u)], & \cosh[2K^{(b)}(-u)] &= \cosh[2K^{(b)}(u)], \end{aligned} \quad (12.2.7)$$

$$\begin{aligned} K^{(x)}(u + 2\mathcal{K}') &= -K^{(x)}(u) \pm \frac{1}{2}i\pi, \\ \sinh[2K^{(x)}(u + 2\mathcal{K}')] &= \sinh[2K^{(x)}(u)], \end{aligned} \quad (12.2.8)$$

$$\begin{aligned} \cosh[2K^{(x)}(u + 2\mathcal{K}')] &= -\cosh[2K^{(x)}(u)], \\ K^{(x)}(u + 2i\mathcal{K}) &= K^{(x)}(u) \pm \frac{1}{2}i\pi, \\ \sinh[2K^{(x)}(u + 2i\mathcal{K})] &= -\sinh[2K^{(x)}(u)], \\ \cosh[2K^{(x)}(u + 2i\mathcal{K})] &= -\cosh[2K^{(x)}(u)], \end{aligned} \quad (12.2.9)$$

where $x = f, b$. From (18.3.38) and (18.3.44)–(18.3.47),

$$\exp[2K^{(f)}(u)] = W(u|k), \quad \exp[2K^{(b)}(u)] = W(\mathcal{K}' - u|k), \quad (12.2.10)$$

¹⁹ However, when, as below, we take all the horizontal rapidities on the lattice to be equal these relationships imply ones between backward and forward couplings across vertical neighbouring $\mathcal{N}^{(*)}$ sites for pairs on the same lattice ($\mathcal{N}$ of $\mathcal{N}^*$).

²⁰ The choice of signs in these formulae arise because couplings occur as the arguments of exponentials.

where

$$W(u|k) := \frac{i [\operatorname{dn}(iu|k) + 1]}{\operatorname{ksn}(iu|k)} = \frac{i[\sqrt{k'} \Theta(iu + \mathcal{K}|k) + \Theta(iu|k)]}{\sqrt{k} H(iu|k)}, \quad (12.2.11)$$

which is meromorphic with poles at $u = 2\ell\mathcal{K}' + 2\ell'i\mathcal{K}$ for all integer values of ℓ and ℓ' . From (18.3.2), (18.3.6) and (18.3.36),

$$\int_{2K^{(b)}}^{\infty} \frac{d\omega}{\sqrt{1+k^2 \sinh^2(\omega)}} = \int_0^{2K^{(b)}} \frac{d\omega}{\sqrt{1+k^2 \sinh^2(\omega)}} = u, \quad (12.2.12)$$

implying that, if both couplings are real and positive, u is real with $0 < u < \mathcal{K}'$.

We now suppose that the vertical rapidity has one of two values $\bar{\varrho}_E$ and $\bar{\varrho}_O$ according as the index j is even or odd. We shall later make a similar distinction between values of the horizontal rapidities but for the moment it is sufficient simply to use the one variable ϱ .

Referring to Fig. 12.3, the lattice line of spins on sites of $\mathcal{N}$ and $\mathcal{N}^*$ immediately above the line of $\mathcal{N}^{(*)}$ labelled i is also labelled with the index i . The spins on the sites of this line are denoted by $\tilde{\sigma}^{(i)} := (\tilde{\sigma}_1^{(i)}, \tilde{\sigma}_2^{(i)}, \dots, \tilde{\sigma}_{4n_2}^{(i)})$, where $\tilde{\sigma}_j^{(i)}$ is on the site immediately to the right of the vertical lattice line of $\mathcal{N}^{(*)}$ labelled j . It is now useful to give different variables for the spins on $\mathcal{N}$ and $\mathcal{N}^*$. This is done by setting

$$\begin{aligned} \tau_j^{(i)} &:= \tilde{\sigma}_{2j-1}^{(i)}, & \sigma_j^{(i)} &:= \tilde{\sigma}_{2j}^{(i)}, & i \text{ even,} \\ \sigma_j^{(i)} &:= \tilde{\sigma}_{2j-1}^{(i)}, & \tau_j^{(i)} &:= \tilde{\sigma}_{2j}^{(i)}, & i \text{ odd.} \end{aligned} \quad (12.2.13)$$

The effect of this redesignation (Fig. 12.4) is to have spins labelled with σ 's on $\mathcal{N}$ and spins labelled with τ 's on $\mathcal{N}^*$. Around the vertical lattice line $2j$ of $\mathcal{N}^{(*)}$ there are zigzag lines of spins on $\mathcal{N}$ and $\mathcal{N}^*$ all with the index j .

In spin-representation two forms of the eight-vertex transfer matrix $V(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ must be distinguished according to whether the row index i is even or odd. These will be denoted respectively by $V_E(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ and $V_O(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$, with²¹

$$\begin{aligned} \langle \sigma, \tau | V_E(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma', \tau' \rangle &:= \langle \sigma | \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle \langle \tau | \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \tau' \rangle, \\ \langle \sigma, \tau | V_O(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma', \tau' \rangle &:= \langle \sigma | \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle \langle \tau | \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \tau' \rangle \end{aligned} \quad (12.2.14)$$

²¹ The index i is omitted and spins below and above the lattice line are denoted by σ and τ and σ' and τ' respectively. The use of the letters 'B' and 'A' for matrices signify their locations below and above the lattice line.

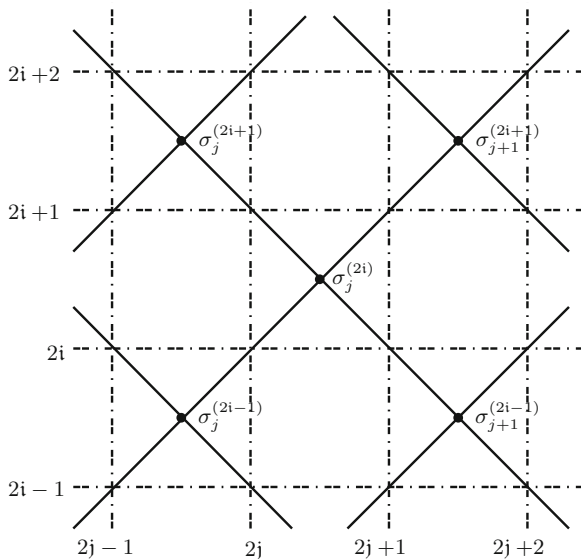


Fig. 12.4 The relabelling of the spins on the sites of $\mathcal{N}$ according to (12.2.13)

and

$$\langle \sigma | \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle := \prod_{j=1}^{2n_2} \exp \left[K^{(f)}(\bar{\varrho}_E - \varrho) \sigma_j \sigma'_j + K^{(b)}(\bar{\varrho}_O - \varrho) \sigma_{j+1} \sigma'_{j+1} \right],$$

$$\langle \sigma | \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle := \prod_{j=1}^{2n_2} \exp \left[K^{(b)}(\bar{\varrho}_E - \varrho) \sigma_j \sigma'_j + K^{(f)}(\bar{\varrho}_O - \varrho) \sigma_j \sigma'_{j+1} \right],$$
(12.2.15)

where the index j is an integer modulo $2n_2$. Equation (12.2.14) give the Kronecker-product forms

$$V_E(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) = \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) \otimes \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho),$$

$$V_O(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) = \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) \otimes \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho),$$
(12.2.16)

where the first term in each product is that of $\mathcal{N}$. The partition function of the eight-vertex model is related to that of the disconnected Ising models on $\mathcal{N}$ and $\mathcal{N}^*$. Given that the value of the horizontal rapidity alternates between rows with ϱ_E and ϱ_O on even and odd number rows, respectively,

$$\begin{aligned}
Z^{(8\text{M})}(\mathcal{N}^{(\otimes)}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) &= \text{Trace}\{[\mathbf{V}_E(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathbf{V}_O(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1}\} \\
&= Z^{(\text{IM})}(\mathcal{N}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) Z^{(\text{IM})}(\mathcal{N}^{\star}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O),
\end{aligned} \tag{12.2.17}$$

where

$$\begin{aligned}
Z^{(\text{IM})}(\mathcal{N}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) &:= \text{Trace}\{[\mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1}\}, \\
Z^{(\text{IM})}(\mathcal{N}^{\star}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) &:= \text{Trace}\{[\mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1}\}.
\end{aligned} \tag{12.2.18}$$

The first of these equations is the partition function of Baxter's (1986) checkerboard Ising model on the lattice $\mathcal{N}$ which we introduced in Sect. 3.8.1.5.

12.2.1 The Modified Checkerboard Ising Model

The star of sites shown in Fig. 12.4, with summation over $\sigma_j^{(2i)}$, is a contribution to $\langle \sigma | \mathbf{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathbf{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O) | \sigma' \rangle$. Using

$$\sigma_0 := \sigma_j^{(2i)}, \quad \sigma_1 := \sigma_j^{(2i-1)}, \quad \sigma_2 := \sigma_j^{(2i+1)}, \quad \sigma_3 := \sigma_{j+1}^{(2i+1)}, \quad \sigma_4 := \sigma_{j+1}^{(2i-1)},$$

this star is reproduced with the respective couplings in Fig. 12.5. Its vertex weight is

$$\begin{aligned}
\mathfrak{V}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) &= 2 \cosh[K^{(\text{f})}(\bar{\varrho}_E - \varrho_E)\sigma_1 + K^{(\text{b})}(\bar{\varrho}_E - \varrho_O)\sigma_2 \\
&\quad + K^{(\text{f})}(\bar{\varrho}_O - \varrho_O)\sigma_3 + K^{(\text{b})}(\bar{\varrho}_O - \varrho_E)\sigma_4].
\end{aligned} \tag{12.2.19}$$

With the notation

$$\begin{aligned}
K_{\text{B}}^{(\text{f})} &:= K^{(\text{f})}(\bar{\varrho}_E - \varrho_E), & K_{\text{B}}^{(\text{b})} &:= K^{(\text{b})}(\bar{\varrho}_O - \varrho_E), \\
K_{\text{A}}^{(\text{f})} &:= K^{(\text{f})}(\bar{\varrho}_O - \varrho_O), & K_{\text{A}}^{(\text{b})} &:= K^{(\text{b})}(\bar{\varrho}_E - \varrho_O),
\end{aligned} \tag{12.2.20}$$

and horizontal and vertical couplings and diagonal couplings across the square set to zero, this is the vertex weight (3.8.20) which was shown to satisfy the free-fermion condition (3.8.21). From (12.2.4),

$$\begin{aligned}
K_{\text{B}}^{(\text{f})} &= -\frac{1}{2}i \text{am}(i[\mathcal{K}' - \bar{\varrho}_E + \varrho_E]|\mathbf{k}), & K_{\text{B}}^{(\text{b})} &= -\frac{1}{2}i \text{am}(i[\bar{\varrho}_O - \varrho_E]|\mathbf{k}), \\
K_{\text{A}}^{(\text{f})} &= -\frac{1}{2}i \text{am}(i[\mathcal{K}' - \bar{\varrho}_O + \varrho_O]|\mathbf{k}), & K_{\text{A}}^{(\text{b})} &= -\frac{1}{2}i \text{am}(i[\bar{\varrho}_E - \varrho_O]|\mathbf{k}).
\end{aligned} \tag{12.2.21}$$

The four couplings can be chosen independently. Since the rapidities occur as differences they form three variables

$$u_1 := \frac{1}{2}[(\bar{\varrho}_E + \bar{\varrho}_O) - (\varrho_E + \varrho_O)], \quad u_2 := \frac{1}{2}(\bar{\varrho}_E - \bar{\varrho}_O), \quad u_3 := \frac{1}{2}(\varrho_E - \varrho_O), \tag{12.2.22}$$

(say), which together with the elliptic modulus $\mathbf{k}$ are an independent equivalent set.

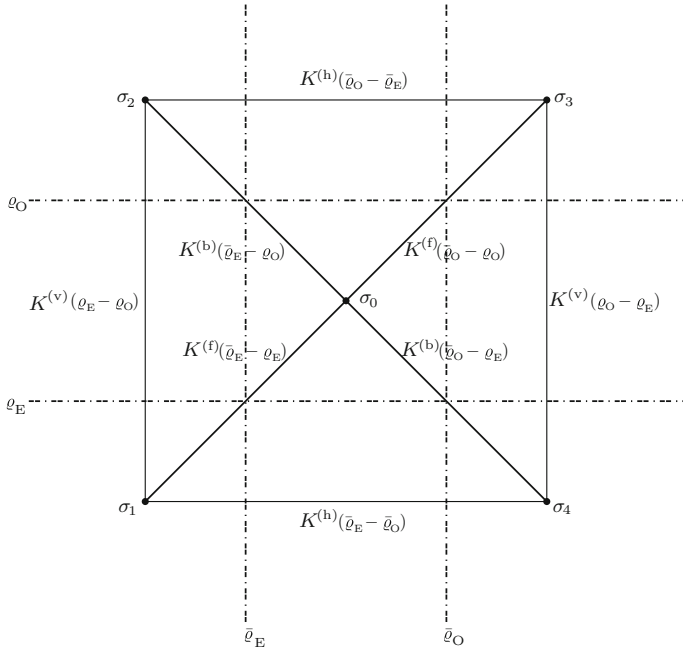


Fig. 12.5 The star of sites with vertex weight $\mathfrak{V}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and supplementary vertex weight $\mathfrak{V}'(\bar{\varrho}_O, \bar{\varrho}_E; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4)$

The free-fermion condition continues to hold if couplings between the horizontal and vertical pairs are included and we showed in Sect. 3.8.1.5 that the eight-vertex conditions (3.8.27) can be satisfied if the vertex weight (12.2.19) is modified by including a supplementary multiplicative factor

$$\mathfrak{V}'(\sigma_1, \sigma_2, \sigma_3, \sigma_4) := \exp\left\{\frac{1}{2}[K^{(h)}(\sigma_1\sigma_4 - \sigma_2\sigma_3) + K^{(v)}(\sigma_1\sigma_2 - \sigma_3\sigma_4)]\right\}, \quad (12.2.23)$$

with $K^{(h)}$ and $K^{(v)}$ satisfying (3.8.25) and (3.8.26). Substituting into the right-hand side of these equations using (12.2.4) it can be shown that

$$\tanh[2K^{(h)}(\bar{\varrho}_E - \bar{\varrho}_O)] = \frac{\begin{bmatrix} \sinh[2K^{(f)}(\bar{\varrho}_O - \varrho_O)] \sinh[2K^{(b)}(\bar{\varrho}_E - \varrho_O)] \\ - \sinh[2K^{(f)}(\bar{\varrho}_E - \varrho_E)] \sinh[2K^{(b)}(\bar{\varrho}_O - \varrho_E)] \end{bmatrix}}{\begin{bmatrix} \cosh[2K^{(f)}(\bar{\varrho}_O - \varrho_O)] \cosh[2K^{(b)}(\bar{\varrho}_E - \varrho_O)] \\ + \cosh[2K^{(f)}(\bar{\varrho}_E - \varrho_E)] \cosh[2K^{(b)}(\bar{\varrho}_O - \varrho_E)] \end{bmatrix}}, \quad (12.2.24)$$

$$\tanh[2K^{(v)}(\varrho_o - \varrho_E)] = \frac{\begin{bmatrix} \sinh[2K^{(f)}(\bar{\varrho}_o - \varrho_o)] \sinh[2K^{(b)}(\bar{\varrho}_o - \varrho_E)] \\ - \sinh[2K^{(f)}(\bar{\varrho}_E - \varrho_E)] \sinh[2K^{(b)}(\bar{\varrho}_E - \varrho_o)] \end{bmatrix}}{\begin{bmatrix} \cosh[2K^{(f)}(\bar{\varrho}_o - \varrho_o)] \cosh[2K^{(b)}(\bar{\varrho}_o - \varrho_E)] \\ + \cosh[2K^{(f)}(\bar{\varrho}_E - \varrho_E)] \cosh[2K^{(b)}(\bar{\varrho}_E - \varrho_o)] \end{bmatrix}}, \quad (12.2.25)$$

where

$$K^{(h)}(u) := -\frac{1}{2}i \operatorname{am}(i u | k), \quad K^{(v)}(u) := -\frac{1}{2}i \operatorname{am}(i u | k). \quad (12.2.26)$$

Both $K^{(h)}$ and $K^{(v)}$ are of the same functional form as $K^{(b)}$, but it is useful, in the interests of clarity, to retain for them a distinguishing notation. It is also worth repeating (12.2.23) to display the rapidity dependence. Thus

$$\begin{aligned} \mathfrak{Y}'(\bar{\varrho}_E, \bar{\varrho}_o; \varrho_E, \varrho_o; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \\ = \exp\left\{\frac{1}{2}[K^{(h)}(\bar{\varrho}_E - \bar{\varrho}_o)(\sigma_1\sigma_4 - \sigma_2\sigma_3) + K^{(v)}(\varrho_o - \varrho_E)(\sigma_1\sigma_2 - \sigma_3\sigma_4)]\right\}. \end{aligned} \quad (12.2.27)$$

Because $K^{(h)}(u)$ and $K^{(v)}(u)$ are odd functions it can be seen that the changes of sign in the horizontal and vertical pairs of couplings simply correspond to interchanges between the members of their respective pairs of rapidities. It is important to note that the inclusion of the additional Boltzmann factor (12.2.27), whether or not the couplings are given by (12.2.26), does not affect the Hamiltonian on the whole lattice. So the partition function of this modified checkerboard Ising model is exactly the same as that of the unmodified model.

At this point it is worth pausing and considering a curious feature of the couplings $K^{(h)}(u)$ and $K^{(v)}(u)$ as they arise in (12.2.24) and (12.2.25) respectively. This is that, unlike $K^{(f)}(u)$ and $K^{(b)}(u)$, their arguments are the differences of rapidities on *non-intersecting lines of $\mathcal{N}^{\otimes}$* . A conceptual way to resolve this puzzle is to think of extending the vertical and horizontal rapidity lines in Fig. 12.5, respectively, upwards and to the right so that they intersect at infinity with the coupling pairs passing horizontally and vertically through the intersections. Then conjugate couplings

$$K^{(ch)}(u) := -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - u] | k), \quad K^{(cv)}(u) := -\frac{1}{2}i \operatorname{am}(i[\mathcal{K}' - u] | k), \quad (12.2.28)$$

can be defined. The coupling $K^{(ch)}(u)$ is between a *vertical pair* and $K^{(cv)}(u)$ between a *horizontal pair*.²²

²² Although, of course, they differ from $K^{(v)}(u)$ and $K^{(h)}(u)$ respectively.

We now investigate the existence of a different set of couplings which also satisfy the free-fermion and the eight-vertex conditions. The simplest way to do this is to use the transformations

$$\begin{aligned}\bar{S}_1 : (\bar{\varrho}_E, \bar{\varrho}_O) &\longrightarrow (\bar{\varrho}_O, \bar{\varrho}_E), & \bar{S}_2 : (\bar{\varrho}_E, \bar{\varrho}_O) &\longrightarrow (\mathcal{K}' - \bar{\varrho}_E, \mathcal{K}' - \bar{\varrho}_O), \\ S_1 : (\varrho_E, \varrho_O) &\longrightarrow (\varrho_O, \varrho_E), & S_2 : (\varrho_E, \varrho_O) &\longrightarrow (-\varrho_E, -\varrho_O).\end{aligned}\tag{12.2.29}$$

$\bar{S}_1$ and $\bar{S}_2$ change the sign of $K^{(h)}$ and S_1 and S_2 change the sign of $K^{(v)}$. The coupling transformations (a) and (b) of (3.8.28) are, respectively, equivalent to $\bar{S}_1 S_1$ and $\bar{S}_1 \bar{S}_2 S_2$. Consider now three transformations applied to (12.2.24) and (12.2.25):

- (i) S_1 which interchanges ϱ_E and ϱ_O .
- (ii) $\bar{S}_1$ which interchanges $\bar{\varrho}_E$ and $\bar{\varrho}_O$.
- (iii) $\bar{S}_2 S_2$ which has the effect of interchanging $K^{(f)}(u)$ and $K^{(b)}(u)$ for all the arguments u on the right-hand sides of (12.2.24) and (12.2.25) and changing the signs of $K^{(h)}$ and $K^{(v)}$.

These transformations each in a different way (respectively, translating the star of sites a lattice space vertically, translating the star of sites a lattice space horizontally, interchanging the interacting pair at each site of $\mathcal{N}^{(\otimes)}$) map between $\mathcal{N}$ and $\mathcal{N}^*$. Equations produced by transforming (12.2.24) and (12.2.25) are the same for (i), (ii) and (iii) (after some interchanges in the order of terms). These transformations yield a second solution for a set of couplings which satisfy the eight-vertex condition (3.8.5) and the free-fermion condition (3.8.21), it being precisely the dual solution on $\mathcal{N}^*$.

It is now important to investigate the relationship between the vertex weights for these two sets of couplings. Specifically the problem is to find a function $\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O)$ so that

$$\frac{\mathfrak{Y}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \mathfrak{Y}'(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4)}{\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O)}\tag{12.2.30}$$

is invariant, for all $\sigma_1, \sigma_2, \sigma_3$ and σ_4 , under the transformations (i), (ii) and (iii). After some manipulation (Baxter 1986) it can be shown that, to within an arbitrary multiplicative constant,²³

$$\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) := g^{(f)}(\bar{\varrho}_E - \varrho_E) g^{(b)}(\bar{\varrho}_O - \varrho_E) g^{(b)}(\bar{\varrho}_E - \varrho_O) g^{(f)}(\bar{\varrho}_O - \varrho_O),\tag{12.2.31}$$

where

$$g^{(x)}(u) := \{2 \sinh[2K^{(x)}(u)]\}^{\frac{1}{4}}, \quad x = f, b.\tag{12.2.32}$$

²³ For comparison with the account of the uniform case in Baxter (1982b) it should be noted, from (12.2.37), that $\mathfrak{C}(\bar{\varrho}, \bar{\varrho}; \varrho, \varrho) = 2/\sqrt{k}$.

12.2.2 Properties of the Transfer Matrices

It is convenient at this stage to absorb the quantity $\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O)$ into $\mathfrak{Y}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4)$ by dividing the Boltzmann weight with coupling $K^{(x)}(u)$ by $g^{(x)}(u)$, for all u and $x = f, b$. Denoting the vertex weight with this modification by $\mathfrak{Y}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4)$ the invariance conditions for (i), (ii) and (iii) take the respective forms

$$(i) \quad \mathfrak{Y}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \mathfrak{Y}'(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \\ = \mathfrak{Y}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \mathfrak{Y}'(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E; \sigma_1, \sigma_2, \sigma_3, \sigma_4), \quad (12.2.33)$$

$$(ii) \quad = \mathfrak{Y}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \mathfrak{Y}'(\bar{\varrho}_O, \bar{\varrho}_E; \varrho_E, \varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4), \quad (12.2.34)$$

$$(iii) \quad = \mathfrak{Y}(\mathcal{K}' - \bar{\varrho}_O, \mathcal{K}' - \bar{\varrho}_E, -\varrho_E, -\varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4) \\ \times \mathfrak{Y}'(\mathcal{K}' - \bar{\varrho}_O, \mathcal{K}' - \bar{\varrho}_E, -\varrho_E, -\varrho_O; \sigma_1, \sigma_2, \sigma_3, \sigma_4). \quad (12.2.35)$$

These conditions lead to relationships between transfer matrices modified in a similar way, so that

$$\langle \sigma | \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle := \prod_{j=1}^{2n_2} \frac{\exp [K^{(f)}(\bar{\varrho}_E - \varrho) \sigma_j \sigma'_j + K^{(b)}(\bar{\varrho}_O - \varrho) \sigma_{j+1} \sigma'_j]}{\{4 \sinh[2K^{(f)}(\bar{\varrho}_E - \varrho)] \sinh[2K^{(b)}(\bar{\varrho}_O - \varrho)]\}^{\frac{1}{2}n_2}}, \\ \langle \sigma | \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle := \prod_{j=1}^{2n_2} \frac{\exp [K^{(b)}(\bar{\varrho}_E - \varrho) \sigma_j \sigma'_j + K^{(f)}(\bar{\varrho}_O - \varrho) \sigma_j \sigma'_{j+1}]}{\{4 \sinh[2K^{(f)}(\bar{\varrho}_O - \varrho)] \sinh[2K^{(b)}(\bar{\varrho}_E - \varrho)]\}^{\frac{1}{2}n_2}}. \quad (12.2.36)$$

Using coupling parameters, which are allowed to have any complex values, it clear from (12.2.36) that

$$\mathring{\mathbf{B}}(K^{(b)}, K^{(f)}) = \mathring{\mathbf{A}}^T(K^{(f)}, K^{(b)}), \quad \mathring{\mathbf{A}}(K^{(b)}, K^{(f)}) = \mathring{\mathbf{B}}^T(K^{(f)}, K^{(b)}), \quad (12.2.37)$$

which, using (12.2.5), can be expressed in terms of rapidities as

$$\mathring{\mathbf{B}}(-\bar{\varrho}_E, -\bar{\varrho}_O; -\mathcal{K}' - \varrho) = \mathring{\mathbf{A}}^T(\bar{\varrho}_E, \bar{\varrho}_O; \varrho), \\ \mathring{\mathbf{A}}(-\bar{\varrho}_E, -\bar{\varrho}_O; -\mathcal{K}' - \varrho) = \mathring{\mathbf{B}}^T(\bar{\varrho}_E, \bar{\varrho}_O; \varrho). \quad (12.2.38)$$

The vertical interactions cancel when the matrices $\mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ and $\mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ are formed but the horizontal interactions still play a role for (ii) and (iii), so we need the matrix $\mathbf{D}(\bar{\varrho}_E - \bar{\varrho}_O)$ given by

$$\langle \sigma | \mathbf{D}(\bar{\varrho}_E - \bar{\varrho}_O) | \sigma' \rangle := \delta^{\text{Kr}}(\sigma - \sigma') \prod_{j=1}^{2n_2} \exp \left[K^{(\text{h})}(\bar{\varrho}_E - \bar{\varrho}_O) \sigma_j \sigma_{j+1} \right], \quad (12.2.39)$$

for which

$$\mathbf{D}^{-1}(\bar{\varrho}_E - \bar{\varrho}_O) = \mathbf{D}(\bar{\varrho}_O - \bar{\varrho}_E). \quad (12.2.40)$$

Using (12.2.37) and (12.2.40) the invariance properties (12.2.33), (12.2.34) and (12.2.35) give

$$\begin{aligned} & \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho') \\ \text{(i)} \quad &= \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho') \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho), \end{aligned} \quad (12.2.41)$$

$$\text{(ii)} \quad = \mathbf{D}(\bar{\varrho}_O - \bar{\varrho}_E) \mathring{\mathbf{B}}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho) \mathring{\mathbf{A}}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho') \mathbf{D}(\bar{\varrho}_E - \bar{\varrho}_O), \quad (12.2.42)$$

$$\text{(iii)} \quad = \mathbf{D}(\bar{\varrho}_O - \bar{\varrho}_E) \mathring{\mathbf{B}}(-\bar{\varrho}_E, -\bar{\varrho}_O; -\mathcal{K}' - \varrho) \mathring{\mathbf{A}}(-\bar{\varrho}_E, -\bar{\varrho}_O; -\mathcal{K}' - \varrho') \mathbf{D}(\bar{\varrho}_E - \bar{\varrho}_O). \quad (12.2.43)$$

We now explore some other transformation properties of $\mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ and $\mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$:

- Given the periodic boundary conditions in the horizontal direction the system is invariant under the rotation group $\mathcal{C}_{2n_2} := \{I, C, C^2, \dots, C^{2n_2-1}\}$, where C is a rotation about the vertical axis of the lattice cylinder through an angle π/n_2 (equivalent to a translation to the right). For $\mathring{\mathbf{B}}$ the lower (unprimed) indices in (12.2.36) remain unchanged and the upper (primed) indices increase by one with the reverse being the case for $\mathring{\mathbf{A}}$.²⁴ The orthogonal matrix $\mathbf{C} := \mathbf{R}(C)$ given by

$$\langle \sigma | \mathbf{C} | \sigma' \rangle := \prod_{j=1}^{2n_2} \delta^{\text{Kr}}(\sigma_{j+1} - \sigma'_j), \quad (12.2.44)$$

is a representational matrix for the 2^{2n_2} -dimensional representation of row states and

$$\mathbf{C} \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) = \mathring{\mathbf{B}}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho), \quad \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) \mathbf{C}^T = \mathring{\mathbf{A}}(\bar{\varrho}_O, \bar{\varrho}_E; \varrho), \quad (12.2.45)$$

giving

$$[\mathbf{C}, \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)] = \mathbf{0}, \quad [\mathbf{C}, \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)] = \mathbf{0}, \quad (12.2.46)$$

²⁴ Remember that the matrix ordering in a product from left to right corresponds to bottom to top in terms of the lattice.

and

$$\mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho) \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho') = \mathring{\mathbf{A}}(\bar{\varrho}_0, \bar{\varrho}_E, \varrho) \mathring{\mathbf{B}}(\bar{\varrho}_0, \bar{\varrho}_E; \varrho'). \quad (12.2.47)$$

From (12.2.41) and (12.2.47)

$$\mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho') \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho) = \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho) \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho'). \quad (12.2.48)$$

- The system is also invariant under the group $\mathcal{S}_2 := \{I, S\}$, where S inverts all the spins on a row and the orthogonal and symmetric matrix $\mathbf{S} := \mathbf{R}(S)$ is given by

$$\langle \sigma | S | \sigma' \rangle := \delta^{Kr}(\sigma + \sigma'). \quad (12.2.49)$$

Its effect is most easily expressed if the rapidities are replaced as arguments by the pair $(K^{(f)}, K^{(b)})$. Then

$$\begin{aligned} \mathbf{S} \mathring{\mathbf{B}}(K^{(f)}, K^{(b)}) &= \mathring{\mathbf{B}}(K^{(f)}, K^{(b)}) \mathbf{S} = \mathring{\mathbf{B}}(-K^{(f)}, -K^{(b)}), \\ \mathbf{S} \mathring{\mathbf{A}}(K^{(f)}, K^{(b)}) &= \mathring{\mathbf{A}}(K^{(f)}, K^{(b)}) \mathbf{S} = \mathring{\mathbf{A}}(-K^{(f)}, -K^{(b)}). \end{aligned} \quad (12.2.50)$$

Equation (12.2.8) can now be used to express this result in terms of rapidities, but to do so we need to know for $\mathring{\mathbf{B}}$ that $\sum_{j=1}^{2n_2} (\sigma_j \sigma'_j + \sigma_{j+1} \sigma'_j)$ is divisible by four to eliminate the effects of the $\frac{1}{2}\pi i$ factors. Consider the two sets of spins σ and σ' ordered in the sequence $\sigma_1, \sigma'_1, \sigma_2, \sigma'_2, \dots$ of length $4n_2$. If η and η' are respectively the number of unlike (σ_j, σ'_j) and $(\sigma'_j, \sigma_{j+1})$ pairs

$$\sum_{j=1}^{2n_2} (\sigma_j \sigma'_j + \sigma_{j+1} \sigma'_j) = 4n_2 - 2(\eta + \eta') \quad (12.2.51)$$

and the required result follows from the observation that $\eta + \eta'$ must be even. A similar argument is applicable to $\mathring{\mathbf{A}}$ but now the ordered sequence of spins is $\sigma'_1, \sigma_1, \sigma'_2, \sigma_2, \dots$ and ζ and ζ' denote respectively the number of unlike (σ'_j, σ_j) and $(\sigma_j, \sigma'_{j+1})$ pairs. Then

$$\mathbf{S} \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho) = \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho - 2\mathcal{K}'), \quad (12.2.52)$$

$$\mathbf{S} \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho) = \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho - 2\mathcal{K}'),$$

$$[\mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho), \mathbf{S}] = [\mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_0; \varrho), \mathbf{S}] = \mathbf{0}. \quad (12.2.53)$$

- Using η , η' , ζ and ζ' as defined above, (12.2.36) yields

$$\langle \sigma | \mathring{\mathbf{B}}(K^{(\text{f})}, K^{(\text{b})}) | \sigma' \rangle = \frac{\exp[2(n_2 - \eta)K^{(\text{f})} + 2(n_2 - \eta')K^{(\text{b})}]}{\{4 \sinh[2K^{(\text{f})}] \sinh[2K^{(\text{b})}]\}^{\frac{1}{2}n_2}}, \quad (12.2.54)$$

$$\langle \sigma | \mathring{\mathbf{A}}(K^{(\text{f})}, K^{(\text{b})}) | \sigma' \rangle = \frac{\exp[2(n_2 - \zeta)K^{(\text{b})} + 2(n_2 - \zeta')K^{(\text{f})}]}{\{4 \sinh[2K^{(\text{f})}] \sinh[2K^{(\text{b})}]\}^{\frac{1}{2}n_2}},$$

and thus

$$\begin{aligned} \mathring{\mathbf{B}}(K^{(\text{f})} \pm \tfrac{1}{2}i\pi, K^{(\text{b})} \pm \tfrac{1}{2}i\pi) &= \mathring{\mathbf{B}}(K^{(\text{f})}, K^{(\text{b})}), \\ \mathring{\mathbf{A}}(K^{(\text{f})} \pm \tfrac{1}{2}i\pi, K^{(\text{b})} \pm \tfrac{1}{2}i\pi) &= \mathring{\mathbf{A}}(K^{(\text{f})}, K^{(\text{b})}), \end{aligned} \quad (12.2.55)$$

where the signs in the two arguments can be chosen independently. In terms of rapidities, from (12.2.6), (12.2.10), (12.2.11) and (12.2.20), (12.2.54) becomes

$$\begin{aligned} \langle \sigma | \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle &= \left[\tfrac{1}{2} Y(\bar{\varrho}_E - \varrho, \bar{\varrho}_O - \varrho | k) \right]^{n_2} \\ &\times [W(\bar{\varrho}_E - \varrho | k)]^{n_2 - \eta} [W(\mathcal{K}' - \bar{\varrho}_O + \varrho | k)]^{n_2 - \eta'}, \end{aligned} \quad (12.2.56)$$

$$\begin{aligned} \langle \sigma | \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle &= \left[\tfrac{1}{2} Y(\bar{\varrho}_O - \varrho, \bar{\varrho}_E - \varrho | k) \right]^{n_2} \\ &\times [W(\bar{\varrho}_O - \varrho | k)]^{n_2 - \zeta} [W(\mathcal{K}' - \bar{\varrho}_E + \varrho | k)]^{n_2 - \zeta'}, \end{aligned}$$

where

$$Y(u', u | k) := \{-\text{sn}(i[\mathcal{K}' - u'] | k) \text{sn}(i u | k)\}^{-\frac{1}{2}}. \quad (12.2.57)$$

From (12.2.9), (12.2.55) is equivalent to

$$\mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho + 2i\mathcal{K}) = \mathring{\mathbf{B}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho), \quad \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho + 2i\mathcal{K}) = \mathring{\mathbf{A}}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho). \quad (12.2.58)$$

- From (12.2.36), and using the notation of (12.2.20),

$$\begin{aligned} &\langle \sigma | \mathring{\mathbf{B}}(K_A^{(\text{f})}, K_A^{(\text{b})}) \mathring{\mathbf{A}}(K_B^{(\text{f})}, K_B^{(\text{b})}) | \sigma' \rangle \\ &= \prod_{j=1}^{2n_2} \frac{\cosh[K_A^{(\text{f})}\sigma_j + K_B^{(\text{f})}\sigma'_{j+1} + K_A^{(\text{b})}\sigma_{j+1} + K_B^{(\text{b})}\sigma'_j]}{\{\sinh[2K_A^{(\text{f})}] \sinh[2K_A^{(\text{b})}] \sinh[2K_B^{(\text{f})}] \sinh[2K_B^{(\text{b})}]\}^{\frac{1}{4}}}. \end{aligned} \quad (12.2.59)$$

The case where the couplings are related by (3.8.22) is of particular interest since, as can be seen from (3.8.23) and also from (12.2.59), the matrix $\mathring{\mathbf{B}}(K_A^{(\text{f})}, K_A^{(\text{b})}) \mathring{\mathbf{A}}(K_B^{(\text{f})}, K_B^{(\text{b})})$ is rather sparse. In fact it is not difficult to show that

$$\begin{aligned} \mathring{B}(K^{(f)}, K^{(b)}) \mathring{A}(-K^{(b)}, K^{(f)} \pm \tfrac{1}{2}i\pi) &= \left[-\frac{\sinh(2K^{(f)})}{\sinh(2K^{(b)})} \right]^{n_2} \mathbf{I} \\ &+ \left[-\frac{\sinh(2K^{(b)})}{\sinh(2K^{(f)})} \right]^{n_2} \mathbf{S}. \end{aligned} \quad (12.2.60)$$

Using (12.2.5), (12.2.7) and (12.2.20), (3.8.22) can be translated into relationships between rapidities and then, using (12.2.4) and (18.3.44), (12.2.60) can be re-expressed in the form

$$\begin{aligned} \mathring{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) \mathring{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho - \mathcal{K}') &= \left[-\frac{\text{sn}(i[\mathcal{K}' - \bar{\varrho}_E + \varrho]|\mathbf{k})}{\text{sn}(i[\bar{\varrho}_O - \varrho]|\mathbf{k})} \right]^{n_2} \mathbf{I} \\ &+ \left[-\frac{\text{sn}(i[\bar{\varrho}_O - \varrho]|\mathbf{k})}{\text{sn}(i[\mathcal{K}' - \bar{\varrho}_E + \varrho]|\mathbf{k})} \right]^{n_2} \mathbf{S}. \end{aligned} \quad (12.2.61)$$

12.2.3 The Reduction to Regular Ising Models

The regular case of the present model (Baxter 1982b, Chap. 7) corresponds to $K_A^{(f)} = K_B^{(f)}$, $K_A^{(b)} = K_B^{(b)}$, which is achieved by setting $\varrho_O = \varrho_E$, $\bar{\varrho}_O = \bar{\varrho}_E$. In this section we show that interchanges of the rows and columns of $\mathcal{N}^{(*)}$ can be used to decompose the partition function of the general case to a product of partition functions of four regular models. From (12.2.18),²⁵

$$\begin{aligned} Z^{(\text{im})}(\mathcal{N}; \bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) &= [\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O)]^{4n_1 n_2} \mathcal{Z}_{2n_2}^{(2n_1)}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O), \\ Z^{(\text{im})}(\mathcal{N}^*; \bar{\varrho}_O, \bar{\varrho}_E; \varrho_E, \varrho_O) &= [\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E)]^{4n_1 n_2} \mathcal{Z}_{2n_2}^{(2n_1)}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E), \end{aligned} \quad (12.2.62)$$

where

$$\begin{aligned} \mathcal{Z}_{2n_2}^{(2n_1)}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O) &= \mathcal{Z}_{2n_2}^{(2n_1)}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E) \\ &:= \text{Trace}\{[\mathring{B}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{A}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1}\}. \end{aligned} \quad (12.2.63)$$

The commutation relations (12.2.41) and (12.2.48) interchange rapidities on rows of the lattice and their repeated application is used to bring together rows with the same rapidity. Thus

²⁵ As a temporary measure the width $2n_2$ and height $2n_1$ have been explicitly displayed. (The number of lattice sites is, of course, $8n_1 n_2$, which can be seen from the fact that the terms in the trace are each equivalent to two rows of sites.)

$$\begin{aligned}
& [\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1} \\
&= [\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E)]^{n_1} [\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{n_1} \\
&= [\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{n_1} [\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E)]^{n_1}.
\end{aligned} \tag{12.2.64}$$

The second and third lines of (12.2.64) are a commutation relation indicating that the factors are simultaneously diagonalizable with eigenvectors independent of ϱ . Thus for large n_1 , when the traces are dominated by the largest eigenvalue

$$\begin{aligned}
& \text{Trace}\{[\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{2n_1}\} \\
& \simeq \text{Trace}\{[\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E)]^{n_1}\} \\
& \quad \times \text{Trace}\{[\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O) \mathring{\mathbf{A}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O)]^{n_1}\}.
\end{aligned} \tag{12.2.65}$$

The assumption is simply that the contributions to the free energy from the interfaces between the parts of the system with horizontal rapidities ϱ_E and ϱ_O is small for large n_1 .

The next task is to carry out a similar procedure for columns. Clearly this is possible because of the invariance property of the star contribution (12.2.30) to the partition function under interchange of either row or column rapidities. This means that the two equivalent ways of expressing the partition function as a matrix product, by vertical multiplication of rows (as we have represented it here), or by horizontal multiplication of columns, both allow reordering of the rapidities. A simple way to see the argument is as follows. Think of the system as consisting of connected vertical units. These can either be taken as vertical columns of stars based on a pair of rapidity lines with rapidities $\bar{\varrho}_E$ and $\bar{\varrho}_O$ on the left and right or as vertical columns of diamonds based on a pair of rapidity lines with rapidities $\bar{\varrho}_O$ and $\bar{\varrho}_E$ on the left and right. We know that the rapidities on the stars can be interchanged. That they can also be interchanged on the diamonds follows from the observation that diamonds translated one lattice space upwards or downwards (remembering that now all the horizontal rapidities are the same) become stars. This means that we can freely shuffle the vertical rapidities to bring like rapidities together. The matrix elements of $\mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O, \varrho)$ can be reformulated as

$$\begin{aligned}
\langle \sigma | \mathring{\mathbf{B}}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) | \sigma' \rangle &= \prod_{j=1}^{n_2} \mathfrak{X}^{(\text{f})}(\bar{\varrho}_E - \varrho; \sigma_j, \sigma'_j) \mathfrak{X}^{(\text{b})}(\bar{\varrho}_E - \varrho; \sigma'_j, \sigma_{j+1}) \\
&\quad \times \prod_{j=n_2+1}^{2n_2} \mathfrak{X}^{(\text{f})}(\bar{\varrho}_O - \varrho; \sigma_j, \sigma'_j) \mathfrak{X}^{(\text{b})}(\bar{\varrho}_O - \varrho; \sigma'_j, \sigma_{j+1}).
\end{aligned} \tag{12.2.66}$$

where

$$\mathfrak{X}^{(x)}(u; \sigma, \sigma') := \frac{\exp[K^{(x)}(u)\sigma\sigma']}{g^{(x)}(u)}, \quad x = f, b. \quad (12.2.67)$$

There is one linking term at the end of the first summation tying it to the second and (with j an integer modulo $2n_2$) a term at the end of the second summation linking it to the beginning of the first summation. For large n_2 these linking terms make a negligible contribution to the free energy and we approximate this expression by cutting the links. This simply involved resetting the integer j to modulo n_2 and relabelling the spins for $j = n_2 + 1, n_2 + 2, \dots, 2n_2$. This gives

$$\mathring{B}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) = \mathring{B}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho) \otimes \mathring{B}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho). \quad (12.2.68)$$

In a similar way

$$\mathring{A}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho) = \mathring{A}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho) \otimes \mathring{A}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho) \quad (12.2.69)$$

and

$$\begin{aligned} & \text{Trace}\{[\mathring{B}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)\mathring{A}_{2n_2}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)]^{n_1}\} \\ &= \text{Trace}\{[\mathring{B}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho)\mathring{A}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho)]^{n_1}\} \\ & \quad \times \text{Trace}\{[\mathring{B}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho)\mathring{A}_{n_2}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho)]^{n_1}\}. \end{aligned} \quad (12.2.70)$$

Then, from (12.2.63), (12.2.65) and (12.2.70),

$$\begin{aligned} \mathcal{Z}_{2n_2}^{(2n_1)}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_O, \varrho_E) &= \mathcal{Z}_{n_2}^{(n_1)}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho_E, \varrho_E) \mathcal{Z}_{n_2}^{(n_1)}(\bar{\varrho}_O, \bar{\varrho}_O; \varrho_E, \varrho_E) \\ & \quad \times \mathcal{Z}_{n_2}^{(n_1)}(\bar{\varrho}_E, \bar{\varrho}_E; \varrho_O, \varrho_O) \mathcal{Z}_{n_2}^{(n_1)}(\bar{\varrho}_O, \bar{\varrho}_O; \varrho_O, \varrho_O). \end{aligned} \quad (12.2.71)$$

12.2.4 Transfer Matrix Eigenvectors

Assume that $\mathring{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$, $\mathring{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)$ and the products $\mathring{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)\mathring{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho')$ and $\mathring{A}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)\mathring{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho')$ are simple (Sect. 17.2). Let $\mathbf{X}_{BA}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho, \varrho')$ and $\mathbf{X}_{AB}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho, \varrho')$ be the respective matrices of the right eigenvectors (as columns). It can easily be shown (Example 17.2.1) that the two product matrices have the same eigenvalues, denoted by $\mathbf{\Lambda}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho, \varrho')$, and that

$$\mathbf{X}_{BA}^{-1}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho, \varrho')\mathring{B}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho)\mathbf{X}_{AB}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho', \varrho) \text{ is diagonal} \quad (12.2.72)$$

and

$$\mathbf{X}_{\text{AB}}^{-1}(\bar{\varrho}_{\text{E}}, \bar{\varrho}_{\text{O}}; \varrho', \varrho) \mathring{\mathbf{A}}(\bar{\varrho}_{\text{E}}, \bar{\varrho}_{\text{O}}; \varrho') \mathbf{X}_{\text{BA}}(\bar{\varrho}_{\text{E}}, \bar{\varrho}_{\text{O}}; \varrho, \varrho') \text{ is diagonal.} \quad (12.2.73)$$

It follows that the eigenvector matrices are respectively independent of ϱ' and ϱ . Taking the argument a step further, by choosing $\varrho = \varrho' = \bar{\varrho}_{\text{O}}$ in $\mathring{\mathbf{B}}(\bar{\varrho}_{\text{E}}, \bar{\varrho}_{\text{O}}; \varrho)$ and $\mathring{\mathbf{A}}(\bar{\varrho}_{\text{E}}, \bar{\varrho}_{\text{O}}; \varrho')$, it follows from (12.2.36) that the eigenvector matrices are functions of the single rapidity variable $\bar{\varrho}_{\text{E}} - \bar{\varrho}_{\text{O}}$.

12.2.5 Notational Changes

From this point we shall specialize to the case with only one horizontal rapidity ϱ so it is convenient to adopt a generalization of the notation of (12.2.2) with

$$u := \frac{1}{2}(\bar{\varrho}_{\text{E}} + \bar{\varrho}_{\text{O}}) - \varrho, \quad \bar{u} := \frac{1}{2}(\bar{\varrho}_{\text{E}} - \bar{\varrho}_{\text{O}}), \quad (12.2.74)$$

modifying the arguments of the matrices and the eigenvalues and eigenvectors accordingly. In particular we need to express the function $Y(u, u')$, defined by (12.2.57), in terms of $\bar{u}$ and u (Eq. (18.3.100)).

It is also useful to bring together (12.2.4) and (12.2.20) to give expressions for the couplings in Fig. 12.5 now that $\varrho_{\text{E}} = \varrho_{\text{O}} = \varrho$. From (12.2.20), for the lower row, related to the matrix $\mathbf{B}(\bar{u}; u)$,

$$K_{\text{B}}^{(\text{f})} = K^{(\text{f})}(u + \bar{u}), \quad K_{\text{B}}^{(\text{b})} = K^{(\text{b})}(u - \bar{u}), \quad (12.2.75)$$

with

$$\sinh(2K_{\text{B}}^{(\text{f})}) = \frac{i}{k \operatorname{sn}(i[u + \bar{u}]|k)}, \quad \sinh(2K_{\text{B}}^{(\text{b})}) = -i \operatorname{sn}(i[u - \bar{u}]|k), \quad (12.2.76)$$

$$\cosh(2K_{\text{B}}^{(\text{f})}) = \frac{i \operatorname{dn}(i[u + \bar{u}]|k)}{k \operatorname{sn}(i[u + \bar{u}]|k)}, \quad \cosh(2K_{\text{B}}^{(\text{b})}) = \operatorname{cn}(i[u - \bar{u}]|k)$$

and for the upper row, related to the matrix $\mathbf{A}(\bar{u}; u)$,

$$K_{\text{A}}^{(\text{f})} = K^{(\text{f})}(u - \bar{u}), \quad K_{\text{A}}^{(\text{b})} = K^{(\text{b})}(u + \bar{u}), \quad (12.2.77)$$

with

$$\sinh(2K_{\text{A}}^{(\text{f})}) = \frac{i}{k \operatorname{sn}(i[u - \bar{u}]|k)}, \quad \sinh(2K_{\text{A}}^{(\text{b})}) = -i \operatorname{sn}(i[u + \bar{u}]|k), \quad (12.2.78)$$

$$\cosh(2K_{\text{A}}^{(\text{f})}) = \frac{i \operatorname{dn}(i[u - \bar{u}]|k)}{k \operatorname{sn}(i[u - \bar{u}]|k)}, \quad \cosh(2K_{\text{A}}^{(\text{b})}) = \operatorname{cn}(i[u + \bar{u}]|k).$$

As we have observed above, in the general case $\varrho_E \neq \varrho_O$ the four couplings $K_B^{(f)}$, $K_A^{(f)}$, $K_B^{(b)}$, $K_A^{(b)}$ are independent with the corresponding set of combinations of rapidities u_1 , u_2 , u_3 , given by (12.2.22), forming together with the elliptic modulus k an equivalent independent set. Now that the condition $\varrho_E = \varrho_O$, giving $u_1 = u$, $u_2 = \bar{u}$, $u_3 = 0$, has been imposed, the couplings are, from (12.2.76) and (12.2.78), related by the conditions

$$\sinh(2K_B^{(f)}) \sinh(2K_A^{(b)}) = \sinh(2K_A^{(f)}) \sinh(2K_B^{(b)}) = 1/k, \quad (12.2.79)$$

$$\frac{\sinh(2K_B^{(b)})}{\sinh(2K_B^{(f)})} = \frac{\sinh(2K_A^{(b)})}{\sinh(2K_A^{(f)})} = -P(\bar{u}, u|k), \quad (12.2.80)$$

where

$$P(\bar{u}, u|k) := k \operatorname{sn}(i[u + \bar{u}]|k) \operatorname{sn}(i[u - \bar{u}]|k) = \frac{P_H(\bar{u}, u|k)}{P_\Theta(\bar{u}, u|k)}, \quad (12.2.81)$$

with $P_H(\bar{u}, u|k)$ and $P_\Theta(\bar{u}, u|k)$ defined in Sect. 18.3.7. The relationship (12.2.79) can be compared to (12.2.6), which, applying in the general case, is between coupling on $\mathcal{N}$ and its dual lattice $\mathcal{N}^*$. Thus the condition $\varrho_E = \varrho_O$ imposes a general form of the self-duality of the square-lattice Ising model (Sect. 9.2.2.1).

With $\varrho_E = \varrho_O$, the couplings on the vertical sides of the square in Fig. 12.5 are zero. The horizontal couplings $K^{(h)}$ and $-K^{(h)}$ on the lower and upper sides of the square, respectively, with

$$K^{(h)} = K^{(h)}(2\bar{u}), \quad K^{(ch)} = K^{(h)}(\mathcal{K}' - 2\bar{u}), \quad (12.2.82)$$

and

$$\sinh(2K^{(h)}) = -i \operatorname{sn}(2i\bar{u}|k), \quad \sinh(2K^{(ch)}) = \frac{i}{k \operatorname{sn}(2i\bar{u}|k)}, \quad (12.2.83)$$

$$\cosh(2K^{(h)}) = \operatorname{cn}(2i\bar{u}|k), \quad \cosh(2K^{(ch)}) = \frac{i \operatorname{dn}(2i\bar{u}|k)}{k \operatorname{sn}(2i\bar{u}|k)},$$

are present, although, as we have indicated above, their presence does not alter the partition function. Their role is to facilitate the calculation and their presence in the final form for the eigenvalues is as dependent quantities, with the conjugate pair $K^{(h)}$ and $K^{(ch)}$ related by (12.2.82) and, from (18.3.45)–(18.3.50), (18.3.62) and (12.2.83),

$$\cosh(2K^{(ch)}) = \coth(2K^{(h)}) \sinh(2K_B^{(f)}) \sinh(2K_B^{(b)}) + \cosh(2K_B^{(f)}) \cosh(2K_B^{(b)}). \quad (12.2.84)$$

Changing the sign of $\bar{u}$ gives

$$\cosh(2K^{(\text{ch})}) = \coth(2K^{(\text{h})}) \sinh(2K_{\text{A}}^{(\text{f})}) \sinh(2K_{\text{A}}^{(\text{b})}) - \cosh(2K_{\text{A}}^{(\text{f})}) \cosh(2K_{\text{A}}^{(\text{b})}) \quad (12.2.85)$$

and equating the right-hand sides of these two equations gives, in this modified notation, (12.2.24).

12.2.6 Transfer Matrix Eigenvalues

12.2.6.1 The Separate Matrices

Using (12.2.41) and (12.2.47),

$$X(\bar{u}) := X_{\text{BA}}(\bar{u}), \quad X_{\text{AB}}(\bar{u}) = X(-\bar{u}) = D(\bar{u})X(\bar{u}) \quad (12.2.86)$$

and (12.2.72) and (12.2.73) take the form

$$\begin{aligned} X^{-1}(\bar{u})\mathring{B}(\bar{u}, u)X(-\bar{u}) &= \Lambda_{\text{B}}(\bar{u}, u), \\ X^{-1}(-\bar{u})\mathring{A}(\bar{u}, u)X(\bar{u}) &= \Lambda_{\text{A}}(\bar{u}, u), \end{aligned} \quad (12.2.87)$$

where $\Lambda_{\text{B}}(\bar{u}, u)$ and $\Lambda_{\text{A}}(\bar{u}, u)$ are diagonal matrices. From (12.2.45) and (12.2.52),

$$\begin{aligned} X^{-1}(\bar{u})CX(\bar{u})\Lambda_{\text{A}}(\bar{u}, u) &= \Lambda_{\text{B}}(-\bar{u}, u), \\ X^{-1}(\bar{u})SX(\bar{u})\Lambda_{\text{Z}}(\bar{u}, u) &= \Lambda_{\text{Z}}(\bar{u}, u + 2\mathcal{K}'), \end{aligned} \quad (12.2.88)$$

where here and in later equations $\text{Z} = \text{B}, \text{A}$. So $X(\bar{u})$ is the matrix of right eigenvectors of C and S with

$$X^{-1}(\bar{u})CX(\bar{u}) = \Xi, \quad (12.2.89)$$

$$X^{-1}(\bar{u})SX(\bar{u}) = \Sigma, \quad (12.2.90)$$

where Ξ is the diagonal matrix with elements which are complex $2n_2$ -th roots of unity and Σ is the diagonal matrix with elements equal to $+1$ or -1 . Thus we have, from (12.2.58) and (12.2.88)–(12.2.90),

$$\Xi \Lambda_{\text{A}}(\bar{u}, u) = \Lambda_{\text{B}}(-\bar{u}, u), \quad (12.2.91)$$

$$\Sigma \Lambda_{\text{Z}}(\bar{u}, u) = \Lambda_{\text{Z}}(\bar{u}, u + 2\mathcal{K}'), \quad (12.2.92)$$

$$\Lambda_{\text{Z}}(\bar{u}, u) = \Lambda_{\text{Z}}(\bar{u}, u + 2i\mathcal{K}). \quad (12.2.93)$$

As was noted in Sect. 18.3.2 the four theta functions are entire. So, from (12.2.11) and (12.2.56),

$$\langle \sigma | \mathbf{B}(\bar{u}, u) | \sigma' \rangle = \frac{f_B(\bar{u}, u; \sigma, \sigma' | k)}{[Q(\bar{u}, u | k)]^{n_2}}, \quad \langle \sigma | \mathbf{A}(\bar{u}, u) | \sigma' \rangle = \frac{f_A(\bar{u}, u; \sigma, \sigma' | k)}{[Q(-\bar{u}, u | k)]^{n_2}}, \quad (12.2.94)$$

where $f_B(\bar{u}, u; \sigma, \sigma' | k)$ and $f_A(\bar{u}, u; \sigma, \sigma' | k)$ are entire functions of u and $Q(v, u | k)$ is defined in (18.3.94). The 2^{2n_2} eigenvalues of $\mathbf{B}(\bar{u}, u)$ appear in a certain order in the diagonal of $\Lambda_B(\bar{u}, u)$, thus associating them each with an element $s = +1$ or -1 of Σ ; this is the *parity* of the eigenvalue. The eigenvalues will be indexed by the pair $\{s, m\}$, with $m = 1, 2, \dots, 2^{2n_2-1}$, and written as $\Lambda_B^{(s)}(\bar{u}, u; m | k)$ with a corresponding notation for the eigenvalues of $\mathbf{A}(\bar{u}, u)$. It follows from (12.2.87) that $\Lambda_B^{(s)}(\bar{u}, u; m | k)$ and $\Lambda_A^{(s)}(\bar{u}, u; m | k)$ are linear combinations of the elements of $\mathbf{B}(\bar{u}, u)$ and $\mathbf{A}(\bar{u}, u)$, respectively, with coefficients dependent on $\bar{u}$ and k but not on u . The eigenvalues $\Lambda_B^{(s)}(\bar{u}, u; m | k)$ and $\Lambda_A^{(s)}(\bar{u}, u; m | k)$ are associated with a complex $2n_2$ -th root of unity denoted by $\xi_m^{(s)}$, which is the corresponding element of $\mathbf{C}$, and, to satisfy (12.2.91), we let

$$\begin{aligned} \Lambda_B^{(s)}(\bar{u}, u; m | k) &:= [\xi_m^{(s)}]^{\frac{1}{2}} [Y(\bar{u}, u | k)]^{n_2} \Upsilon^{(s)}(\bar{u}, u; m | k), \\ \Lambda_A^{(s)}(\bar{u}, u; m | k) &:= [\xi_m^{(s)}]^{-\frac{1}{2}} [Y(-\bar{u}, u | k)]^{n_2} \Upsilon^{(s)}(-\bar{u}, u; m | k). \end{aligned} \quad (12.2.95)$$

Then from (12.2.92), (12.2.93), (18.3.102) and (18.3.103)

$$\Upsilon^{(s)}(\bar{u}, u; m | k) = \Upsilon^{(s)}(\bar{u}, u + 2i\mathcal{K}; m | k), \quad (12.2.96)$$

$$\Upsilon^{(s)}(\bar{u}, u; m | k) = s \Upsilon^{(s)}(\bar{u}, u + 2\mathcal{K}'; m | k). \quad (12.2.97)$$

So $\Upsilon^{(s)}(\bar{u}, u; m) \in \text{DP}(2i\mathcal{K}, 2\mathcal{K}'; 1, s)$.²⁶ A small shift ϵ in the real direction, so that the period rectangle becomes

$$\mathbb{W}'(2i\mathcal{K}, 2\mathcal{K}') := \{u : \epsilon < \Re(u) < 2\mathcal{K}' + \epsilon, 0 < \Im(u) < 2\mathcal{K}\}, \quad (12.2.98)$$

is necessary only when it is needed to satisfy (12.2.99) if $\bar{u}$ is an integer. It follows from (12.2.94) that it is meromorphic with poles of order n_2 at $u = (2p_1 - 1)\mathcal{K}' + 2p_2\mathcal{K}i + \bar{u}$ and $u = 2q_1\mathcal{K}' + 2q_2\mathcal{K}i - \bar{u}$, for all integer values of p_1, p_2, q_1 and q_2 . $\mathbb{W}'(2i\mathcal{K}, 2\mathcal{K}')$ will contain two poles, from these sets with $p_2 = q_2 = 0$ and $p_1 = p_1^*(\bar{u})$ and $q_1 = q_1^*(\bar{u})$, the unique integers satisfying

$$\begin{aligned} \epsilon - \bar{u} < 2\mathcal{K}'p_1^*(\bar{u}) < 2\mathcal{K}' - \bar{u} + \epsilon, \\ \epsilon + \bar{u} - \mathcal{K}' < 2\mathcal{K}'q_1^*(\bar{u}) < \mathcal{K}' + \bar{u} + \epsilon. \end{aligned} \quad (12.2.99)$$

²⁶ The class of doubly-periodic functions is defined in Sect. 18.2.

Then, from Theorem 18.2.3, $\Upsilon^{(s)}(\bar{u}, u; m|k)$ has exactly $2n_2$ zeros (including each according to its multiplicity) in the interior of $\mathbb{W}'(2i\mathcal{K}, 2\mathcal{K}')$ with none on the boundary. Let these zeros be $u_1^{(s)}(\bar{u}; m), u_2^{(s)}(\bar{u}; m), \dots, u_{2n_2}^{(s)}(\bar{u}; m)$. Then, from Theorem 18.3.1 and (18.3.96),

$$\Upsilon^{(s)}(\bar{u}, u; m|k) = A_1^{(s)}(\bar{u}; m) \exp\left(\frac{\pi \aleph^{(s)}(\bar{u}; m)u}{2\mathcal{K}}\right) \prod_{j=1}^{2n_2} \frac{H(i[u - u_j^{(s)}(\bar{u}; m)]|k)}{\sqrt{Q(\bar{u}, u|k)}}. \quad (12.2.100)$$

To satisfy (12.2.96) and (12.2.97) $\aleph^{(s)}(\bar{u}; m) - n_2$ must be an even integer with²⁷

$$\frac{1}{\mathcal{K}'} \sum_{j=1}^{2n_2} \Re\{u_j^{(s)}(\bar{u}; m)\} = \aleph^{(s)}(\bar{u}; m), \quad (12.2.101)$$

$$\frac{1}{\mathcal{K}} \sum_{j=1}^{2n_2} \Im\{u_j^{(s)}(\bar{u}; m)\} \equiv -\frac{1}{2}(1-s) \pmod{2}. \quad (12.2.102)$$

From (18.3.23),

$$\begin{aligned} \Upsilon^{(s)}(-\bar{u}, u + \mathcal{K}'; m|k) &= (-1)^{n_2} A_1^{(s)}(-\bar{u}; m) \exp(-\pi n_2 \mathcal{K}'/2\mathcal{K}) \\ &\times \exp\left(\frac{\pi[\aleph^{(s)}(-\bar{u}; m) - 2n_2]u}{2\mathcal{K}}\right) \prod_{j=1}^{2n_2} \frac{H(i[u + \mathcal{K}' - u_j^{(s)}(-\bar{u}; m)]|k)}{\sqrt{Q(\bar{u}, u|k)}}. \end{aligned} \quad (12.2.103)$$

12.2.6.2 The Product Formula

From (12.2.61), (12.2.90) and (12.2.91), and using the notation of Sect. 18.3.7 for combinations of theta functions,

$$\Lambda_B(\bar{u}, u) \Lambda_A(\bar{u}, u + \mathcal{K}') = \frac{[P_\Theta(\bar{u}, u|k)]^{2n_2} \mathbf{I} + [P_H(\bar{u}, u|k)]^{2n_2} \Sigma}{[Q(\bar{u}, u|k)Q(-\bar{u}, u|k)]^{n_2}}. \quad (12.2.104)$$

From (18.3.100)–(18.3.108),

$$\Lambda_B^{(s)}(\bar{u}, u; m|k) \Lambda_A^{(s)}(\bar{u}, u + \mathcal{K}'; m|k) = \frac{s[P_\Theta(\bar{u}, u|k)]^{2n_2}}{[Q(\bar{u}, u|k)Q(-\bar{u}, u|k)]^{n_2}} G^{(s)}(\bar{u}, u|k), \quad (12.2.105)$$

²⁷ Using (18.3.96) these results also follow directly from Theorem 18.3.1 with $\tau = 1$, $\tau' = s$.

where $G^{(s)}(\bar{u}, u|k)$ is given by (18.3.109) and, from (12.2.95), (18.3.101)–(18.3.103),

$$\Upsilon^{(s)}(\bar{u}, u; m|k) \Upsilon^{(s)}(-\bar{u}, u + \mathcal{K}'; m|k) = \frac{s}{k^{n_2}} \left[\frac{P_{\Theta}(\bar{u}, u|k)}{Q(\bar{u}, u|k)} \right]^{2n_2} G^{(s)}(\bar{u}, u|k). \quad (12.2.106)$$

A factorization of $G^{(s)}(\bar{u}, u|k)$ is derived in Sect. 18.3.7, which, together with (12.2.106), yields

$$\begin{aligned} \Upsilon^{(s)}(\bar{u}, u; m|k) \Upsilon^{(s)}(-\bar{u}, u + \mathcal{K}'; m|k) &= \frac{1}{k^{n_2}} C^{(s)}(\bar{u}) \\ &\times \prod_{j=1}^{2n_2} \frac{H(i u - \zeta(\bar{u}; \psi_j^{(s)}) + \frac{1}{2} i \mathcal{K}'|k) H(i u + \zeta(\bar{u}; \psi_j^{(s)}) - \frac{1}{2} i \mathcal{K}'|k)}{Q(\bar{u}, u|k)}, \end{aligned} \quad (12.2.107)$$

where

$$\psi_j^{(s)} := \frac{\pi(4j - s - 1)}{8n_2}, \quad j = 1, 2, \dots, 2n_2, \quad (12.2.108)$$

and, from (18.3.119),

$$C^{(s)}(\bar{u}) = s \prod_{j=1}^{2n_2} \frac{\Theta(2i \bar{u}|k) \Theta(0|k)}{H(i \bar{u} - \zeta(\bar{u}; \psi_j^{(s)}) + \frac{1}{2} i \mathcal{K}'|k) H(i \bar{u} + \zeta(\bar{u}; \psi_j^{(s)}) - \frac{1}{2} i \mathcal{K}'|k)}. \quad (12.2.109)$$

From (18.3.104) and (18.3.115), $\zeta(\bar{u}; \psi)$ is given by

$$k \operatorname{sn}(i \bar{u} - \zeta(\bar{u}; \psi) + \frac{1}{2} i \mathcal{K}'|k) \operatorname{sn}(i \bar{u} + \zeta(\bar{u}; \psi) - \frac{1}{2} i \mathcal{K}'|k) = -\exp(2i \psi). \quad (12.2.110)$$

The solution of this equation is a crucial part of the determination of the eigenvalues.

12.2.6.3 Determining Zeros

The product of (12.2.100) and (12.2.103) must be equivalent to (12.2.107). The task is to match terms in the numerator of the product together with a contribution from the exponential factor with terms in the numerator of (12.2.107). From (12.2.108) and (18.3.116),

$$\psi_{2n_2-j+1}^{(+1)} = \pi - \psi_j^{(+1)}, \quad \zeta(\bar{u}; \psi_{2n_2-j+1}^{(+1)}) = -\zeta(\bar{u}; \psi_j^{(+1)}),$$

$$j = 1, 2, \dots, n_2, \quad (12.2.111)$$

$$\psi_{2n_2-j}^{(-1)} = \pi - \psi_j^{(-1)}, \quad \zeta(\bar{u}; \psi_{2n_2-j}^{(-1)}) = -\zeta(\bar{u}; \psi_j^{(-1)}),$$

$$j = 1, 2, \dots, n_2 - 1, \quad (12.2.112)$$

$$\psi_{n_2}^{(-1)} = \frac{1}{2}\pi, \quad \zeta(\bar{u}; \psi_{n_2}^{(-1)}) = 0, \quad \psi_{2n_2}^{(-1)} = \pi, \quad \zeta(\bar{u}; \psi_{2n_2}^{(-1)}) = \mathcal{K}.$$

$$(12.2.113)$$

Apart from the terms $\psi = \frac{\pi}{2}$ and $\psi = \pi$ which appear in the product for $s = -1$, the terms are in pairs $(\psi, \pi - \psi)$. In the numerator in (12.2.107) for a particular value of ψ we select four roots

$$u = -i \zeta(\bar{u}; \psi) - \frac{1}{2}\mathcal{K}', \quad u = +i \zeta(\bar{u}; \psi) + \frac{1}{2}\mathcal{K}',$$

$$u = +i \zeta(\bar{u}; \psi) - \frac{1}{2}\mathcal{K}', \quad u = -i \zeta(\bar{u}; \psi) + \frac{1}{2}\mathcal{K}'. \quad (12.2.114)$$

If we take $u_j^{(s)}(\bar{u}; m) = -i \zeta(\bar{u}; \psi_j^{(s)}) \pm \frac{1}{2}\mathcal{K}'$ in the numerator in (12.2.100), then, since $\zeta(-\bar{u}; \psi_j^{(s)}) = \zeta(\bar{u}; \psi_j^{(s)})$, the root in $\mathbb{W}'(2i\mathcal{K}, 2\mathcal{K}')$ in the numerator in (12.2.103) is $u = -i \zeta(\bar{u}; \psi_j^{(s)}) \mp \frac{1}{2}\mathcal{K}'$. Thus we have an independent choice of signs for each $\psi_j^{(s)}$. Let

$$u_j^{(s)}(\bar{u}; m) = -i \zeta(\bar{u}; \psi_j^{(s)}) - \frac{1}{2}\gamma_j \mathcal{K}', \quad j = 1, 2, \dots, 2n_2, \quad (12.2.115)$$

for some particular choice of γ_j , equal to -1 or $+1$, for $j = 1, 2, \dots, 2n_2$. Noting that the choice of γ_j for the term with $\pi - \psi_j^{(s)}$ is independent of that for $\psi_j^{(s)}$, an eigenvalue is identified by a set of γ 's. The number of choices is subject to the constraint (12.2.101), which now takes the form

$$\frac{1}{2} \sum_{j=1}^{2n_2} (1 - \gamma_j) = n_2 + \aleph^{(s)}(\bar{u}; m). \quad (12.2.116)$$

So $n_2 + \aleph^{(s)}(\bar{u}; m)$ and $n_2 - \aleph^{(s)}(\bar{u}; m)$ are respectively the even number of $\gamma_j = -1$ and $+1$. To satisfy this equation, $2n_2 - 1$ γ_j 's can be chosen independently. And (together with s) the eigenvalues can be indexed by a appropriately constrained set $\{\gamma_j\}$, which, for convenience we represent by the $2n_2$ -dimensional vector $\boldsymbol{\gamma}$ that *henceforth replaces the parameter m* . Now consider the sum of the imaginary parts of the roots $u_j^{(s)}(\bar{u}; \boldsymbol{\gamma})$. For $s = 1$ they cancel in pairs and for $s = -1$ they cancel in pairs except for the terms $\zeta(\bar{u}; \psi_{n_2}^{(-)}) = 0$ and $\zeta(\bar{u}; \psi_{2n_2}^{(-)}) = \mathcal{K}$. Thus

$$\sum_{j=1}^{2n_2} \zeta(\bar{u}; \psi_j^{(s)}) = \frac{1}{2}(1-s)\mathcal{K}. \quad (12.2.117)$$

This is just a stronger version of (12.2.102). The condition

$$\prod_{\{\psi\}^{(s)}} k \operatorname{sn}(i\bar{u} \pm \zeta(\bar{u}; \psi) + \frac{1}{2}i\mathcal{K}'|k) \operatorname{sn}(i\bar{u} \pm \zeta(\bar{u}; \psi) - \frac{1}{2}i\mathcal{K}'|k) = 1, \quad (12.2.118)$$

where the product is over the values of ψ given by (12.2.108), follows directly from (18.3.50) when the choice of sign in each term is the same. However, the condition $\zeta(\bar{u}; \pi - \psi) = -\zeta(\bar{u}; \psi)$ allows independent choices of signs in the two terms. From (12.2.100), (12.2.115) and (12.2.116),

$$\Upsilon^{(s)}(\bar{u}, u; \boldsymbol{\gamma}|k) = A_1^{(s)}(\bar{u}; \boldsymbol{\gamma}) \prod_{\{\psi\}^{(s)}} \exp\left(-\frac{\pi\gamma_\psi u}{4\mathcal{K}}\right) \frac{H(iu - \zeta(\bar{u}; \psi) + \frac{1}{2}\gamma_\psi i\mathcal{K}'|k)}{\sqrt{Q(\bar{u}, u|k)}} \quad (12.2.119)$$

and, from (18.3.25),

$$\Upsilon^{(s)}(\bar{u}, u; \boldsymbol{\gamma}|k) = A_2^{(s)}(\bar{u}; \boldsymbol{\gamma}) \prod_{\{\psi\}^{(s)}} \exp\left(\frac{\pi\gamma_\psi u}{4\mathcal{K}}\right) \frac{\Theta(iu - \zeta(\bar{u}; \psi) - \frac{1}{2}\gamma_\psi i\mathcal{K}'|k)}{\sqrt{Q(\bar{u}, u|k)}}, \quad (12.2.120)$$

where

$$A_2^{(s)}(\bar{u}; \boldsymbol{\gamma}) = (-1)^{n_2} A_1^{(s)}(\bar{u}; \boldsymbol{\gamma}) \prod_{\{\psi\}^{(s)}} \exp\left(\frac{\pi i \zeta(\bar{u}; \psi) \gamma_\psi}{2\mathcal{K}}\right). \quad (12.2.121)$$

From (12.2.95), (18.3.44) and (18.3.100),

$$\begin{aligned} \Upsilon^{(s)}(\bar{u}, u; \boldsymbol{\gamma}|k) \Upsilon^{(s)}(-\bar{u}, u; \boldsymbol{\gamma}|k) &= X_1^{(s)}(\bar{u}, u; \boldsymbol{\gamma}|k) \\ &\times \prod_{\{\psi\}^{(s)}} \sqrt{k} \operatorname{sn}(iu - \zeta(\bar{u}; \psi) + \frac{1}{2}\gamma_\psi i\mathcal{K}'|k), \end{aligned} \quad (12.2.122)$$

where

$$\begin{aligned} X_1^{(s)}(\bar{u}, u; \boldsymbol{\gamma}|k) &:= A_1^{(s)}(\bar{u}; \boldsymbol{\gamma}) A_2^{(s)}(-\bar{u}; \boldsymbol{\gamma}) \\ &\times \prod_{\{\psi\}^{(s)}} \frac{\Theta(iu - \zeta(\bar{u}; \psi) + \frac{1}{2}i\mathcal{K}'|k) \Theta(iu - \zeta(\bar{u}; \psi) - \frac{1}{2}i\mathcal{K}'|k)}{\sqrt{Q(\bar{u}, u|k)Q(-\bar{u}, u|k)}}. \end{aligned} \quad (12.2.123)$$

$X_1^{(s)}(\bar{u}, u; \boldsymbol{\gamma} | k) \in \text{DP}(2i\mathcal{K}, 2\mathcal{K}'; 1, 1)$ and it can be seen from (18.3.108), (18.3.110) and (18.3.118) that it has the same poles and zeros as $F^{(s)}(\bar{u}, u | k)$, defined by (18.3.106). Thus, from Theorem 18.2.4 and (18.3.107), $X_1^{(s)}(\bar{u}, u; \boldsymbol{\gamma} | k)$ can be replaced by

$$X_2^{(s)}(\bar{u}, u; \boldsymbol{\gamma} | k) := A_3^{(s)}(\bar{u}; \boldsymbol{\gamma}) \left\{ \left[\frac{\text{sn}(i[\mathcal{K}' - u - \bar{u}] | k)}{\text{sn}(i[u - \bar{u}] | k)} \right]^{n_2} + s \left[\frac{\text{sn}(i[u - \bar{u}] | k)}{\text{sn}(i[\mathcal{K}' - u - \bar{u}] | k)} \right]^{n_2} \right\}. \quad (12.2.124)$$

From (12.2.107)–(12.2.109), (12.2.122) and (12.2.124)

$$\begin{aligned} & \Upsilon^{(s)}(\bar{u}; \bar{u}; \boldsymbol{\gamma} | k) \Upsilon^{(s)}(-\bar{u}; \bar{u}; \boldsymbol{\gamma} | k) \Upsilon^{(s)}(\bar{u}; \bar{u} + \mathcal{K}'; \boldsymbol{\gamma} | k) \Upsilon^{(s)}(-\bar{u}; \bar{u} + \mathcal{K}'; \boldsymbol{\gamma} | k) \\ & \simeq \begin{cases} \frac{1}{k^{2n_2}} \left[\frac{\Theta(0 | k) \Theta(2i\bar{u} | k)}{H(0 | k) H(i\bar{u} | k)} \right]^{2n_2} \\ \text{and} \\ \frac{s}{k^{2n_2}} [A_3^{(s)}(\bar{u}; \boldsymbol{\gamma})]^2 \left[\frac{\Theta(0 | k) \Theta(2i\bar{u} | k)}{H(0 | k) H(i\bar{u} | k)} \right]^{2n_2}, \end{cases} \quad (12.2.125) \end{aligned}$$

from which it follows that $A_3^{(s)}(\bar{u}; \boldsymbol{\gamma}) = 1/\sqrt{s}$. So, from (12.2.95), (12.2.122), (18.3.100) and (12.2.124),

$$\Lambda_A^{(s)}(\bar{u}, u; \boldsymbol{\gamma} | k) \Lambda_B^{(s)}(\bar{u}, u; \boldsymbol{\gamma} | k) = \frac{(-1)^{n_2}}{\sqrt{s}} F^{(s)}(\bar{u}, u | k) \prod_{\{\psi\}^{(s)}} S(\bar{u}, u; \psi; \alpha, \gamma_\psi | k), \quad (12.2.126)$$

where

$$S(\bar{u}; v; \psi; \alpha, \gamma) := \sqrt{k} \text{sn}(\alpha i v - \zeta(\bar{u}; \psi) + \frac{1}{2} \gamma i \mathcal{K}' | k), \quad \alpha, \gamma = \pm 1. \quad (12.2.127)$$

The parameter $\alpha = \pm 1$ can be chosen with either sign in (12.2.126) (see (12.2.129)).

12.2.6.4 The General Case

The remaining task is to obtain an expression for $S(\bar{u}; v; \psi; \alpha, \gamma | k)$ using (12.2.110). Let

$$C(\bar{u}; v; \psi; \alpha, \gamma | k) := \text{cds}(\alpha i v - \zeta(\bar{u}; \psi) + \frac{1}{2} \gamma i \mathcal{K}' | k), \quad \alpha, \gamma = \pm 1. \quad (12.2.128)$$

From (18.3.50), (18.3.51) and (18.3.116),²⁸

$$\begin{aligned} S(\bar{u}; v; \psi; \alpha, -\gamma|k) &= [S(\bar{u}; v; \psi; \alpha, \gamma|k)]^{-1}, \\ S(\bar{u}; v; \pi - \psi; \alpha, -\gamma|k) &= -\mathcal{C}\{S(\bar{u}; v; \psi; \alpha, -\gamma|k)\} \\ &= -S(\bar{u}; v; \psi; -\alpha, \gamma|k), \end{aligned} \quad (12.2.129)$$

$$\begin{aligned} C(\bar{u}; v; \psi; \alpha, -\gamma|k) &= -C(\bar{u}; v; \psi; \alpha, \gamma|k), \\ C(\bar{u}; v; \pi - \psi; \alpha, -\gamma|k) &= -\mathcal{C}\{C(\bar{u}; v; \psi; \alpha, -\gamma|k)\} \\ &= -C(\bar{u}; v; \psi; -\alpha, \gamma|k). \end{aligned} \quad (12.2.130)$$

Equation (12.2.110), which can now be expressed in the form²⁹

$$\bar{S}(\bar{u}; \psi; \alpha, \gamma|k) \bar{S}(\bar{u}; \psi; -\alpha, \gamma|k) = \exp(2\gamma i \psi), \quad (12.2.131)$$

is expanded using (18.3.57) to give

$$\begin{aligned} \bar{C}(\bar{u}; \psi; \alpha, \gamma|k) &= \alpha \operatorname{cds}(2i \bar{u}|k) \\ &+ \alpha \frac{\exp(2\gamma i \psi)[k \operatorname{sn}^2(2i \bar{u}|k) - \bar{S}^2(\bar{u}; \psi; \alpha, -\gamma|k)]}{\operatorname{sn}(2i \bar{u}|k)}. \end{aligned} \quad (12.2.132)$$

Using the first of Eqs. (12.2.129) and (12.2.130) yields a quadratic for $\bar{S}^2(\bar{u}; \psi; \alpha, \gamma|k)$ and hence a formula for $\bar{C}(\bar{u}; \psi; \alpha, \gamma|k)$. The solution of the former must be chosen to satisfy (12.2.129) and hence

$$\bar{S}(\bar{u}; \psi; \alpha, \gamma|k) = \exp(\gamma i \psi) \left\{ W(\bar{u}; \psi|k) + \alpha \gamma \sqrt{W^2(\bar{u}; \psi|k) - 1} \right\}^{\frac{1}{2}}, \quad (12.2.133)$$

$$\bar{C}(\bar{u}; \psi; \alpha, \gamma|k) = \frac{\gamma \{ \alpha i k \operatorname{sn}^2(2i \bar{u}|k) \sin(2\psi) + \sqrt{W^2(\bar{u}; \psi|k) - 1} \}}{\operatorname{sn}(2i \bar{u}|k)}, \quad (12.2.134)$$

where

$$W(\bar{u}; \psi|k) := \operatorname{cn}(2i \bar{u}|k) \operatorname{dn}(2i \bar{u}|k) + k \operatorname{sn}^2(2i \bar{u}|k) \cos(2\psi). \quad (12.2.135)$$

²⁸ To avoid confusion with the overline notation the complex conjugate is here denoted by $\mathcal{C}\{\cdot\}$.

²⁹ The notation is simplified for S and C by denoting them by $\bar{S}$ and $\bar{C}$, respectively, when the first two arguments (compressed to one) are $\bar{u}$.

Translating into coupling form using (12.2.75)–(12.2.78),³⁰

$$W(K^{(h)}, K^{(ch)}; \psi) = \frac{\cosh(2K^{(h)}) \cosh(2K^{(ch)}) - \sinh(2K^{(h)}) \cos(2\psi)}{\sinh(2K^{(ch)})}. \quad (12.2.136)$$

With

$$\begin{aligned} \omega(K^{(h)}, K^{(ch)}; \psi) &:= \ln \left[W(K^{(h)}, K^{(ch)}; \psi) + \sqrt{W^2(K^{(h)}, K^{(ch)}; \psi) - 1} \right], \\ &= \operatorname{arccosh}[W(K^{(h)}, K^{(ch)}; \psi)], \end{aligned} \quad (12.2.137)$$

(12.2.133) and (12.2.134) become

$$\bar{S}(K^{(h)}, K^{(ch)}; \psi; \alpha, \gamma) = \exp \left(\gamma \left[i\psi + \frac{1}{2}\alpha \omega(K^{(h)}, K^{(ch)}; \psi) \right] \right), \quad (12.2.138)$$

$$\bar{C}(K^{(h)}, K^{(ch)}; \psi; \alpha, \gamma) = \gamma \left\{ \frac{\sinh(\omega(K^{(h)}, K^{(ch)}; \psi))}{\sinh(2K^{(h)})} - \frac{\alpha \sin(2\psi)}{\sinh(2K^{(ch)})} \right\}. \quad (12.2.139)$$

From (18.3.57) and (12.2.127),

$$S(\bar{u}, u; \psi; 1, \gamma | k) = \bar{S}(\bar{u}; \psi; \alpha, \gamma | k) D(\bar{u}, u; \psi; \alpha, \gamma | k), \quad (12.2.140)$$

for both $\alpha = \pm 1$, where

$$D(\bar{u}, u; \psi; \alpha, \gamma | k) := \frac{\left\{ \operatorname{cn}(i[u - \alpha \bar{u}] | k) \operatorname{dn}(i[u - \alpha \bar{u}] | k) \right\}}{1 - k \bar{S}^2(\bar{u}; \psi; \alpha, \gamma | k) \operatorname{sn}^2(i[u - \alpha \bar{u}] | k)}. \quad (12.2.141)$$

Then, from (12.2.129) and (12.2.140),

$$D(\bar{u}, u; \psi; \alpha, \gamma | k) D(\bar{u}, u; \psi; \alpha, -\gamma | k) = 1,$$

$$D(\bar{u}, u; \psi; \alpha, \gamma | k) D(\bar{u}, u; \psi; -\alpha, -\gamma | k) = -\frac{1}{\bar{S}(\bar{u}; \psi; \alpha, \gamma | k) \bar{S}(\bar{u}; \pi - \psi; \alpha, \gamma | k)}. \quad (12.2.142)$$

Translating into coupling form using (12.2.75)–(12.2.78),

³⁰ Henceforth we shall interchange freely between elliptic and coupling arguments.

$$\begin{aligned}
& D(K_A^{(f)}, K_B^{(b)}, K^{(h)}, K^{(ch)}; \psi; 1, 1) \\
&= \frac{\left\{ \begin{array}{c} \cosh(2K_A^{(f)}) \cosh(2K_B^{(b)}) \\ -i \bar{C}(K^{(h)}, K^{(ch)}; \psi; 1, 1) \sinh(2K_A^{(f)}) \sinh(2K_B^{(b)}) \end{array} \right\}}{\sinh(2K_A^{(f)}) + \bar{S}^2(K^{(h)}, K^{(ch)}; \psi; 1, 1) \sinh(2K_B^{(b)})},
\end{aligned} \tag{12.2.143}$$

$$\begin{aligned}
& D(K_B^{(f)}, K_A^{(b)}, K^{(h)}, K^{(ch)}; \psi; -1, 1) \\
&= \frac{\left\{ \begin{array}{c} \cosh(2K_B^{(f)}) \cosh(2K_A^{(b)}) \\ -i \bar{C}(K^{(h)}, K^{(ch)}; \psi; -1, 1) \sinh(2K_B^{(f)}) \sinh(2K_A^{(b)}) \end{array} \right\}}{\sinh(2K_B^{(f)}) + \bar{S}^2(K^{(h)}, K^{(ch)}; \psi; -1, 1) \sinh(2K_A^{(b)})}.
\end{aligned} \tag{12.2.144}$$

Finally, from (12.2.133) and (12.2.140),

$$S(\bar{u}, u; \psi; 1, \gamma | k) := i \exp(\gamma i \psi) \{ D(\bar{u}, u; \psi; 1, 1 | k) D(\bar{u}, u; \psi; -1, 1 | k) \}^{\frac{1}{2} \gamma}, \tag{12.2.145}$$

which yields an explicit form for $\Lambda_B^{(s)}(\bar{u}, u; \gamma | k)$ and $\Lambda_A^{(s)}(\bar{u}, u; \gamma | k)$ either in terms of elliptic variables or couplings.

Baxter (1986, p. 22) gives separate formulae for $\Lambda_B^{(s)}$ and $\Lambda_A^{(s)}$ in the more general case $\varrho_E \neq \varrho_O$. To compare with the results given here it is necessary to combine the terms for ψ and $\pi - \psi$ in the product in (12.2.126) and expand $F^{(s)}(\bar{u}, u | k)$ using (18.3.108)–(18.3.110).

The results given by Baxter (1982b, Sect. 7.9) for the regular model $\bar{u} = 0$ can be easily extracted. It can also be shown, as a confirmation of the results of Sect. 12.2.3, that the general model is given, for large n_2 , as a combination of regular models with parameter u replaced by $u + \bar{u}$ and $u - \bar{u}$ respectively.

12.2.7 The Standard Model

The eigenvalues for the standard Ising model first solved by Onsager (1944) can be extracted from this analysis if, in one of the two pairs of diagonal couplings $(K_B^{(f)}, K_A^{(b)})$ and $(K_A^{(f)}, K_B^{(b)})$ in Fig. 12.5, one member is taken to the limit of zero and the other to infinity. The latter has the effect of binding the spin on the centre site to that of the corner site, being an effective merging of the two sites. This produces a square-lattice model with horizontal coupling $K^{(h)}$ and vertical coupling $K^{(ch)}$ except that the ‘vertical’ lattice lines are translated so that they wind around the lattice-column. The lattice now has $2n_1$ rows of $2n_2$ sites and we return to the notation of Chap. 11 by using the variables $N_1 := 2n_1$ and $N_2 := 2n_2$. In order for there to be

a periodic boundary condition in the vertical direction it must be the case that N_1 is some multiple of N_2 .

The reduction to the standard model can be carried out in four ways. Referring to (12.2.76), (12.2.78) and (12.2.83):

- (i) $u \rightarrow \bar{u} : K_A^{(f)} \rightarrow \infty, K_B^{(b)} \rightarrow 0, K_B^{(f)} \rightarrow K^{(ch)}, K_A^{(b)} \rightarrow K^{(h)};$
- (ii) $u \rightarrow -\bar{u} : K_B^{(f)} \rightarrow \infty, K_A^{(b)} \rightarrow 0, K_A^{(f)} \rightarrow K^{(ch)}, K_B^{(b)} \rightarrow K^{(h)};$
- (iii) $u \rightarrow \mathcal{K}' - \bar{u} : K_A^{(b)} \rightarrow \infty, K_B^{(f)} \rightarrow 0, K_B^{(b)} \rightarrow K^{(ch)}, K_A^{(f)} \rightarrow K^{(h)};$
- (iv) $u \rightarrow \mathcal{K}' + \bar{u} : K_B^{(b)} \rightarrow \infty, K_A^{(f)} \rightarrow 0, K_A^{(b)} \rightarrow K^{(ch)}, K_B^{(f)} \rightarrow K^{(h)}.$

Of course, this limit will lead to an infinite contribution to the eigenvalues of the transfer matrix.

In case (i), from (12.2.10), (12.2.11) and (18.3.16), the infinite contribution is

$$\exp(N_2 K_A^{(f)}) \simeq \left[\frac{-8\mathcal{K}k'}{\pi k H^2(0|k)} \right]^{\frac{1}{4}N_2}. \quad (12.2.146)$$

From (12.2.79), the elliptic modulus is

$$k(K^{(h)}, K^{(ch)}) = [\sinh(2K^{(h)}) \sinh(2K^{(ch)})]^{-1} \quad (12.2.147)$$

and, from (18.3.106),

$$F^{(s)}(\bar{u}, u|k) \simeq \exp(N_2 K_A^{(f)}) \left[-\frac{k \sinh(2K^{(ch)})}{2} \right]^{\frac{1}{2}N_2}. \quad (12.2.148)$$

The eigenvalues are the products of the A and B eigenvalues of the general model. From (12.2.126), (12.2.138) and (12.2.146),

$$\begin{aligned} \Lambda^{(s)}(N_2, K^{(h)}, K^{(ch)}; \boldsymbol{\gamma}) &:= \frac{\Lambda_B^{(s)}(\bar{u}; \bar{u}; \boldsymbol{\gamma}|k) \Lambda_A^{(s)}(\bar{u}; \bar{u}; \boldsymbol{\gamma}|k)}{\exp(N_2 K_A^{(f)})} \\ &= \ell^{(s)}(\boldsymbol{\gamma}) \left(\frac{k}{2} \right)^{\frac{1}{2}N_2} [2 \sinh(2K^{(ch)})]^{\frac{1}{2}N_2} \prod_{\{\psi\}^{(s)}} \exp\left(\frac{1}{2} \gamma_\psi \alpha \omega(K^{(h)}, K^{(ch)}; \psi)\right), \end{aligned} \quad (12.2.149)$$

where

$$\ell^{(s)}(\boldsymbol{\gamma}) := \prod_{\{\psi\}^{(s)}} \exp([\gamma_\psi - 1]i \psi). \quad (12.2.150)$$

This is the result derived by Onsager (1944) apart from:

- (1) The constant factors $\ell^{(s)}(\boldsymbol{\gamma})$, for $s = \pm 1$ and all $\boldsymbol{\gamma}$, which are the eigenvalues of the symmetry matrix $\mathbf{C}$ defined in Sect. 12.2.2. Their presence is a consequence of the winding of the lattice lines around the vertical cylinder.³¹ Since N_1 is a multiple of N_2 , $[\ell^{(s)}(\boldsymbol{\gamma})]^{N_1} = 1$, so this factor does not affect the free energy of the system.
- (2) The factor $(k/4)^{\frac{1}{2}N_2}$ is a consequence of the division of the matrix elements by $\mathfrak{C}(\bar{\varrho}_E, \bar{\varrho}_O; \varrho_E, \varrho_O)$ (see the footnote on p. 402).

12.2.7.1 Analysis of the Functions $\omega(K^{(h)}, K^{(ch)}; \psi)$

The sign of α in (12.2.149) is chosen so that $\alpha \omega(K^{(h)}, K^{(ch)}; \psi)$ is positive at low temperatures and has a continuous derivative for all temperatures.³² It is now convenient to use the variables $\mathfrak{X}_h := \exp(2K^{(h)})$ and $\mathfrak{X}_{ch} := \exp(2K^{(ch)})$.³³ From (12.2.136),

$$W(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi) \pm 1 = \begin{cases} \frac{(\mathfrak{X}_h \mathfrak{X}_{ch} + \mathfrak{X}_h \pm \mathfrak{X}_{ch} \mp 1)^2}{2\mathfrak{X}_h(\mathfrak{X}_{ch}^2 - 1)} - \frac{\mathfrak{X}_{ch}(\mathfrak{X}_h^2 - 1)}{\mathfrak{X}_h(\mathfrak{X}_{ch}^2 - 1)} [\cos(2\psi) + 1], \\ \text{or} \\ \frac{(\mathfrak{X}_h \mathfrak{X}_{ch} - \mathfrak{X}_h \pm \mathfrak{X}_{ch} \pm 1)^2}{2\mathfrak{X}_h(\mathfrak{X}_{ch}^2 - 1)} - \frac{\mathfrak{X}_{ch}(\mathfrak{X}_h^2 - 1)}{\mathfrak{X}_h(\mathfrak{X}_{ch}^2 - 1)} [\cos(2\psi) - 1]. \end{cases} \quad (12.2.151)$$

It follows that the quantity under the square sign defining $\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi)$ in (12.2.137) is a perfect square if $\psi = 0$ or $\psi = \frac{1}{2}\pi$. We observe that

$$\mathfrak{X}_h \mathfrak{X}_{ch} - \mathfrak{X}_h \pm \mathfrak{X}_{ch} \mp 1 > 0, \quad \mathfrak{X}_h \mathfrak{X}_{ch} - \mathfrak{X}_h + \mathfrak{X}_{ch} + 1 > 0 \quad (12.2.152)$$

and

$$\mathfrak{X}_h \mathfrak{X}_{ch} - \mathfrak{X}_h - \mathfrak{X}_{ch} - 1 \geq 0, \quad T \leq T_c(\tau), \quad (12.2.153)$$

where, with $\tau = K^{(ch)}/K^{(h)}$, $T_c(\tau)$ is the critical temperature given by

$$\sinh(2K^{(h)}) \sinh(2K^{(ch)}) = 1, \quad (12.2.154)$$

³¹ If $\mathbf{V}$ is the transfer matrix of the standard spin- $\frac{1}{2}$ Ising model (Sect. 11.4.3) then the transfer matrix of the present model is (apart from the factors $(k/2)^{\frac{1}{2}N_2}$) $\mathbf{V}\mathbf{C}^\dagger$.

³² And (to emphasis an obvious, but important, point) the surd in the first version of (12.2.137) is the *positive* square root.

³³ Henceforth, for the functions displayed, we freely interchange between arguments $(K^{(h)}, K^{(ch)})$ and $(\mathfrak{X}_h, \mathfrak{X}_{ch})$, whichever seems more appropriate.

for the square-lattice model with unequal couplings.³⁴ It is also evident from (12.2.147) that $k = 1$ corresponds to the critical point in the formulation in terms of elliptic variables, with $0 < k \leq 1$ corresponding to $T \leq T_c$. From (12.2.137), (12.2.151) and (12.2.153),

$$\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; 0) = \pm \ln \left[\frac{\mathfrak{X}_h(\mathfrak{X}_{ch} - 1)}{\mathfrak{X}_{ch} + 1} \right], \quad (12.2.155)$$

where the plus and minus signs apply respectively to the regions $T < T_c(\tau)$ and $T > T_c(\tau)$. This means that $\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; 0) \geq 0$ for all temperatures with minimum values of zero with a cusp at the critical temperature. However, we need $\alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; 0)$ to be a smooth function of the couplings (or temperature) which is positive at low temperatures. Thus we need to choose $\alpha = 1$ for $T < T_c(\tau)$ and $\alpha = -1$ for $T > T_c(\tau)$ giving

$$\alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; 0) = \ln \left[\frac{\mathfrak{X}_h(\mathfrak{X}_{ch} - 1)}{\mathfrak{X}_{ch} + 1} \right], \quad (12.2.156)$$

In the case of $\psi = \frac{1}{2}\pi$, the situation is simpler since $\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \frac{1}{2}\pi) > 0$ and, with $\alpha = 1$,

$$\alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \frac{1}{2}\pi) = \ln \left[\frac{\mathfrak{X}_h(\mathfrak{X}_{ch} + 1)}{\mathfrak{X}_{ch} - 1} \right]. \quad (12.2.157)$$

It is not difficult to show that $\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi) > 0$, with $\psi > 0$, is positive with a single minimum with respect to variations of T , tending to infinity, for $T \rightarrow 0$ and $T \rightarrow \infty$ and, as a function of ψ , is increasing for $0 \leq \alpha \leq \frac{1}{2}\pi$ with $\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi) = \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \pi - \psi)$ (Fig. 12.6).

In terms of the variable $\mathfrak{X}_h$ and $\mathfrak{X}_{ch}$ the duality transformation (9.2.32)³⁵ takes the form

$$\mathfrak{X}_h \leftrightarrow \mathfrak{X}_h^* = \frac{\mathfrak{X}_{ch} + 1}{\mathfrak{X}_{ch} - 1}, \quad \mathfrak{X}_{ch} \leftrightarrow \mathfrak{X}_{ch}^* = \frac{\mathfrak{X}_h + 1}{\mathfrak{X}_h - 1} \quad (12.2.158)$$

and, from (12.2.136) and (12.2.137),

$$\omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi) = \omega(\mathfrak{X}_h^*, \mathfrak{X}_{ch}^*; \psi). \quad (12.2.159)$$

Thus, from this and from (12.2.156),

³⁴ In the case $K^{(ch)} = K^{(ev)} = K_s$, (12.2.154) reduces to the critical coupling equation (9.2.36) with solution $T_c(1) = 2.269185J$.

³⁵ Generalized to the case $K^{(h)} \neq K^{(ch)}$.

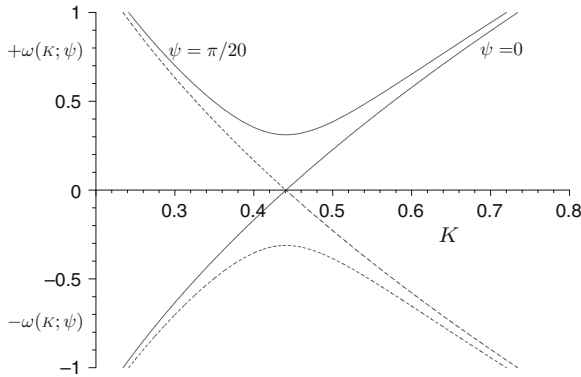


Fig. 12.6 Plots of $+\omega(K^{(h)}, K^{(ch)}; \psi)$ and $-\omega(K^{(h)}, K^{(ch)}; \psi)$ for $K^{(h)} = K^{(ch)} = K$. The branches which are $\alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi)$ are given by continuous lines

$$\begin{aligned} \alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; \psi) &= \alpha \omega(\mathfrak{X}_h^*, \mathfrak{X}_{ch}^*; \psi), & \psi \neq 0, \\ \alpha \omega(\mathfrak{X}_h, \mathfrak{X}_{ch}; 0) &= -\alpha \omega(\mathfrak{X}_h^*, \mathfrak{X}_{ch}^*; 0). \end{aligned} \quad (12.2.160)$$

It follows that the eigenvalues, given by (12.2.149) for $s = +1$ where the set $\{\psi\}^{(s)}$ does not include $\psi = 0$, satisfy the condition

$$\frac{\Lambda^{(+)}(N_2, K^{(h)}, K^{(ch)}; \gamma)}{[k \sinh(2K^{(ch)})]^{\frac{1}{2}N_2}} = \frac{\Lambda^{(+)}(N_2, K^{(h)*}, K^{(ch)*}; \gamma)}{[k^* \sinh(2K^{(ch)*})]^{\frac{1}{2}N_2}}, \quad (12.2.161)$$

where the elliptic modulus k is the function of $K^{(h)}$ and $K^{(ch)}$ given by (12.2.147) and k^* is the corresponding function of $K^{(h)*}$ and $K^{(ch)*}$.³⁶ The presence of the term $\psi = 0$ in the product in (12.2.149) for $s = -1$ prevents a duality relationship like (12.2.161) applying in that case.

The symmetry group used in the analysis of the checkerboard Ising model in Sect. 12.2.2 is $\mathcal{C}_{N_2} \times \mathcal{S}_2^{(m)}$ and this is inherited by the present special case.³⁷ Block diagonalization of transfer matrices as described in Sect. 11.3 allows the division of eigenvalues into sets corresponding to the irreducible representations of the symmetry group. In this case, although this has not been explicitly evident in our method of analysis, the sets corresponding to the one-dimensional symmetric and antisymmetric representations of $\mathcal{S}_2^{(m)}$ are given respectively by $s = 1$ and $s = -1$. It then follows from Theorem 11.3.4, that the largest eigenvalue is in the set $s = 1$. Since, for this case, $\alpha \omega(K^{(h)}, K^{(ch)}; \psi) > 0$, for all $\{\psi\}^{(+)}$,

³⁶ The usual Onsager result is obtained by setting $k = k^* = 2$.

³⁷ However, the larger symmetry group $\mathcal{C}_{N_{2Y}} \times \mathcal{S}_2^{(m)}$ of the square-lattice spin- $\frac{1}{2}$ Ising model in Sect. 11.4.3 with rings of N_2 sites and symmetry reflections through planes containing the lattice axis no longer applies because of the winding of the lattice around the cylinder.

$$\Lambda_{\max}(N_2, K^{(\text{h})}, K^{(\text{ch})}) = \Lambda^{(+)}(N_2, K^{(\text{h})}, K^{(\text{ch})}; \mathbf{j}), \quad (12.2.162)$$

where $\mathbf{j} := (1, 1, \dots, 1)$.

Example 12.2.1 For $N_2 = 2$ and $K^{(\text{h})} = K^{(\text{ch})} = K$, $\mathfrak{X} = \exp(2K)$ the block-diagonalization of the transfer matrix is given in Example 11.3.1, where, in the notation used in Chap. 11, the eigenvalues in (11.3.26) and (11.3.28) are labelled by representation. From (12.2.149),

$$\begin{aligned} \Lambda^{(+)}(2, \mathfrak{X}; (+, +)) &= \frac{1}{2}k \Lambda_1(A; 2, \mathfrak{X}), \quad \Lambda^{(+)}(2, \mathfrak{X}; (-, -)) = \frac{1}{2}k \Lambda_2(A; 2, \mathfrak{X}), \\ \Lambda^{(-)}(2, \mathfrak{X}; (+, +)) &= \frac{1}{2}k \Lambda_1(B_1; 2, \mathfrak{X}), \quad \Lambda^{(-)}(2, \mathfrak{X}; (-, -)) = -\frac{1}{2}k \Lambda_1(B_3; 2, \mathfrak{X}). \end{aligned} \quad (12.2.163)$$

The change in sign in the last relationship arises from $\ell^{(+)}(\pm, \pm) = \ell^{(-)}(+, +) = 1$, $\ell^{(-)}(-, -) = -1$. It is easy to see that the eigenvalues for $s = +1$ satisfy the duality relationship (12.2.161), but that those for $s = -1$ do not satisfy a corresponding relationship.

12.2.7.2 Eigenvalue Degeneracy

To determine an asymptotic expression for $\Lambda_{\max}(K^{(\text{h})}, K^{(\text{ch})})$ when $N_2 \rightarrow \infty$ we apply the Euler-Maclaurin asymptotic formula³⁸ to give

$$\begin{aligned} &\sum_{j=1}^{N_2} \omega(K^{(\text{h})}, K^{(\text{ch})}; \pi[2j-1]/[2N_2]) \\ &\simeq \frac{N_2}{\pi} \int_0^\pi \omega(K^{(\text{h})}, K^{(\text{ch})}; \psi) d\psi - 2 \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \omega^{(2k-1)}(K^{(\text{h})}, K^{(\text{ch})}; 0). \end{aligned} \quad (12.2.164)$$

As we have observed above a consequence of Perron's theorem is that the largest eigenvalue $\Lambda_{\max}(N_2, K^{(\text{h})}, K^{(\text{ch})})$, for finite N_2 , is non-degenerate; meaning (Sect. 11.2) that the system, even in the limit $N_1 \rightarrow \infty$, cannot exhibit a phase transition. Thus the only possibility for the occurrence of a phase transition is that, at some value

³⁸ If $w(z)$ is a smooth function for $0 \leq z \leq n$ then

$$\sum_{m=0}^n w(m) \simeq \int_0^n w(z) dz + \frac{1}{2}[w(0) + w(n)] + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} [w^{(2k-1)}(n) - w^{(2k-1)}(0)],$$

where B_{2k} are the even Bernoulli numbers, (Forder 1949; Goodwin 1949).

of the couplings $(K^{(h)}, K^{(ch)})$, $\Lambda_{\max}(N_2, K^{(h)}, K^{(ch)})$ becomes degenerate in the limit $N_2 \rightarrow \infty$. To investigate this we need the second largest eigenvalue and we have no guarantee that this belongs to the set corresponding to the one-dimensional symmetric representation.³⁹ There are two possibilities:

The second largest eigenvalue is in the one-dimensional symmetric representation:

For this we use the subscript ‘submax’ and the eigenvalue is given by reversing the signs of the first and last γ_ψ ’s in (12.2.149). Thus

$$\Lambda_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)}) = \Lambda^{(+)}(N_2, K^{(h)}, K^{(ch)}; (-, +, +, \dots, +, -)), \quad (12.2.165)$$

giving

$$\begin{aligned} \Omega_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)}) &:= \frac{\Lambda_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)})}{\Lambda_{\max}^{(+)}(N_2, K^{(h)}, K^{(ch)})} \\ &\simeq \exp\{-2\omega(K^{(h)}, K^{(ch)}; 0)\}, \quad \text{as } N_2 \rightarrow \infty. \end{aligned} \quad (12.2.166)$$

It follows from the form of $\omega(K^{(h)}, K^{(ch)}; 0)$, described above and shown in Fig. 12.6 that $\Upsilon_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)})$ has a maximum at the coupling values which satisfy (12.2.154) and which give the critical temperature $T = T_c(\tau)$, and that this is a single point of degeneracy between $\Lambda_{\max}^{(+)}(N_2, K^{(h)}, K^{(ch)})$ and $\Lambda_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)})$ in the limit $N_2 \rightarrow \infty$.

The largest eigenvalue is in the one-dimensional anti-symmetric representation:

This is

$$\Lambda_{\max}^{(-)}(N_2, K^{(h)}, K^{(ch)}) = \Lambda^{(-)}(N_2, K^{(h)}, K^{(ch)}; (+, \dots, +)), \quad (12.2.167)$$

for which we need

³⁹ This is one of the complicating features of investigations, even for one-dimensional models with large transfer matrices, where the ratio of the two largest eigenvalues is used to locate incipient phase transitions (Lavis 1976).

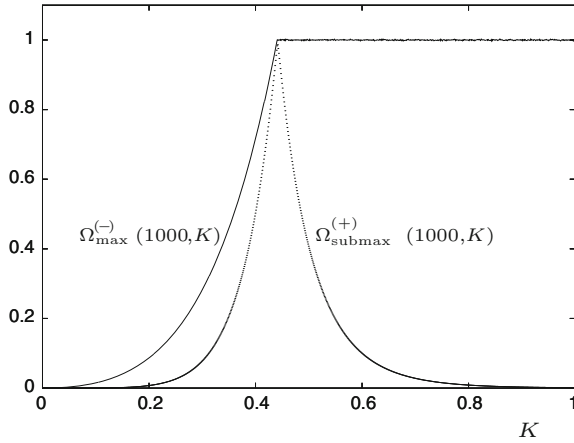


Fig. 12.7 Plots of $\Omega_{\max}^{(-)}(N_2, K^{(h)}, K^{(ch)})$ and $\Omega_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)})$ with $K^{(h)} = K^{(ch)} = K$ and $N_2 = 1000$

$$\begin{aligned}
 \sum_{j=1}^{N_2-1} \omega(K^{(h)}, K^{(ch)}; \pi j/N_2) \pm \omega(K^{(h)}, K^{(ch)}; 0) &= \frac{N_2}{\pi} \int_0^\pi \omega(K^{(h)}, K^{(ch)}; \psi) d\psi \\
 &\quad - \omega(K^{(h)}, K^{(ch)}; 0) \pm \omega(K^{(h)}, K^{(ch)}; 0) \\
 &\quad - 2 \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \omega^{(2k-1)}(K^{(h)}, K^{(ch)}; 0), \tag{12.2.168}
 \end{aligned}$$

where the plus and minus signs correspond respectively to $T > T_c(\tau)$ and $T < T_c(\tau)$, giving

$$\begin{aligned}
 \Omega_{\max}^{(-)}(N_2, K^{(h)}, K^{(ch)}) &:= \frac{\Lambda_{\max}^{(-)}(N_2, K^{(h)}, K^{(ch)})}{\Lambda_{\max}^{(+)}(N_2, K^{(h)}, K^{(ch)})} \\
 &\simeq \begin{cases} \exp\{-\omega(K^{(h)}, K^{(ch)}; 0)\}, & T \geq T_c(\tau), \\ 1, & T \leq T_c(\tau), \end{cases} \quad \text{as } N_2 \rightarrow \infty. \tag{12.2.169}
 \end{aligned}$$

Thus, it has been shown that, for large N_2 ,

$$\Lambda_{\max}^{(+)}(N_2, K^{(h)}, K^{(ch)}) \geq \Lambda_{\max}^{(-)}(N_2, K^{(h)}, K^{(ch)}) > \Lambda_{\text{submax}}^{(+)}(N_2, K^{(h)}, K^{(ch)}), \tag{12.2.170}$$

with the equality applying in the limit $N_2 \rightarrow \infty$ when $T \leq T_c(\tau)$. An example of this for $N_2 = 1000$ and $K^{(h)} = K^{(ch)} = K$ is shown in Fig. 12.7, and it can be shown that the inequalities (but not the equality) in (12.2.170) are valid for all $N_2 > 0$.

12.2.7.3 The Free-Energy Density

We are now in the position to obtain the zero-field free-energy density for the two-dimensionally infinite system. However, the lattice differs from the ‘normal’ square lattice $\mathcal{L}_s$ in that the vertical lattice lines are rotated through an angle of 45° . To signify this we denote the lattice as $\mathcal{L}_{cs}$ retaining the notation $K^{(ch)}$ (rather than $K^{(v)}$) for the coupling in this direction.

Equations (11.1.11) and (11.1.14) give the zero-field free-energy density in the limit $N_1 \rightarrow \infty$ when, in the limit $N_2 \rightarrow \infty$, the largest eigenvalue of the transfer matrix is non-degenerate. However, it can also be applied to the present case since the degeneracy of the largest eigenvalue below the critical temperature, contributes a vanishingly small contribution $-\ln(2)/N$. Thus we have, from (11.1.14), (12.2.136), (12.2.137), (12.2.147) and (12.2.164),

$$\begin{aligned} \phi_1^{(IM)}(\mathcal{L}_{cs}, K^{(h)}, K^{(ch)}, 0) &= - \lim_{N_2 \rightarrow \infty} \frac{1}{N_2} \ln[\Lambda^{(+)}(N_2, K^{(h)}, K^{(ch)}; \mathbf{j})] \\ &= -\frac{1}{2} \ln[k(K^{(h)}, K^{(ch)}) \sinh(2K^{(ch)})] \\ &\quad - \frac{1}{2\pi} \int_0^\pi \operatorname{arccosh} \left\{ \frac{\cosh(2K^{(h)}) \cosh(K^{(ch)}) - \sinh(2K^{(h)}) \cos(2\varphi)}{\sinh(2K^{(ch)})} \right\} d\varphi. \end{aligned} \quad (12.2.171)$$

This formula can be re-expressed in a more symmetric form by using the identity

$$\operatorname{arccosh}(x) = \frac{1}{\pi} \int_0^\pi \ln[2x - 2\cos(\varphi)] d\varphi \quad (12.2.172)$$

to give

$$\begin{aligned} \phi_1^{(IM)}(\mathcal{L}_{cs}, K^{(h)}, K^{(ch)}, 0) &= -\frac{1}{2} \ln[2k(K^{(h)}, K^{(ch)})] \\ &\quad - \frac{1}{2\pi^2} \int_0^\pi d\varphi_1 \int_0^\pi d\varphi_2 \ln\{\cosh(2K^{(h)}) \cosh(2K^{(ch)}) \\ &\quad - \sinh(2K^{(h)}) \cos(\varphi_1) + \sinh(2K^{(ch)}) \cos(\varphi_2)\}. \end{aligned} \quad (12.2.173)$$

For reasons explained above, the formulae derived from the standard presentation of Onsager’s result are given by replacing $k(K^{(h)}, K^{(ch)})$ by 2.⁴⁰ With this identification, in the isotropic case $K^{(h)} = K^{(ch)} = K$, a change of variables $\omega_1 = \frac{1}{2}(\varphi_1 + \varphi_2)$, $\omega_2 = \frac{1}{2}(\varphi_1 - \varphi_2)$ can be used to express (12.2.173) in the form

⁴⁰ See, for example, Domb (1960, Sect. 3.5.1) or Lavis and Bell (1999a, Sect. 8.10).

$$\begin{aligned}\phi_1^{(\text{Im})}(\mathcal{L}_{\text{CS}}, K, 0) &= -\ln[2 \cosh(2K)] \\ &\quad - \frac{1}{\pi^2} \int_0^{\frac{1}{2}\pi} d\omega_2 \int_0^\pi d\omega_1 \ln[1 - f(K) \cos(\omega_1) \cos(\omega_2)],\end{aligned}\quad (12.2.174)$$

where

$$f(K) := \frac{2 \sinh(2K)}{\cosh^2(2K)}. \quad (12.2.175)$$

The inner integral in (12.2.174) can be performed by using

$$\int_0^\pi \ln[y - x \cos(\omega)] d\omega = \pi \ln \left[\frac{y + (y^2 - x^2)^{1/2}}{2} \right], \quad x < y, \quad (12.2.176)$$

and a trivial transformation from ω_2 to $\psi = \frac{1}{2}\pi - \omega_2$ gives

$$\begin{aligned}\phi_1^{(\text{Im})}(\mathcal{L}_{\text{CS}}, K, 0) &= -\frac{1}{2} \ln[2 \cosh^2(2K)] \\ &\quad - \frac{1}{\pi} \int_0^{\frac{1}{2}\pi} \ln \left[1 + \{1 - f^2(K) \sin^2(\psi)\}^{1/2} \right] d\psi.\end{aligned}\quad (12.2.177)$$

From (2.3.27), (18.3.3) and (18.3.4), the internal energy per lattice site is

$$u = -\coth(2K) \left\{ 1 + \frac{2}{\pi} \mathcal{K}(k_2) [2 \tanh^2(2K) - 1] \right\}, \quad (12.2.178)$$

where the elliptic modulus $k_2(K) = f(K)$,⁴¹ for which, with $K \geq 0$, $0 \leq f(K) \leq 1$, with $f(K_c) = 1$. It is straightforward to show using (18.3.10) that

$$\frac{2}{\pi} \mathcal{K}(k_2) [2 \tanh^2(2K) - 1] \simeq -\frac{1}{\sqrt{2}} \Delta K \ln(\Delta K) \rightarrow 0, \quad \text{as } K \rightarrow K_c, \quad (12.2.179)$$

where $\Delta K := K - K_c$, and hence that $u_c = -\sqrt{2}$. From (2.3.28), (2.3.32) and (18.3.8), the zero-field heat capacity is

$$\begin{aligned}c_{\mathcal{H}} &= \frac{k_B}{\pi} [2K \coth(2K)]^2 \{ \mathcal{K}(k_2) - \mathcal{E}(k_2) \\ &\quad - \text{sech}^2(2K) [\frac{\pi}{2} + \mathcal{K}(k_2) [2 \tanh^2(2K) - 1]] \}.\end{aligned}\quad (12.2.180)$$

⁴¹ The use of this notation is to give conformity with (12.3.292).

The singular contribution at the critical point is given by the leading term which, using (18.3.10), yields

$$c_{\mathcal{H}} \simeq \frac{32k_B K_c^2}{\pi} \ln |\Delta K|, \quad \text{as } K \rightarrow K_c. \quad (12.2.181)$$

12.3 The Square-Lattice Eight-Vertex Model

In this discussion it will be necessary to make use of a number of different elliptic moduli, k , which is now relabelled k_0 , and k_1 , k_2 and k_3 , related to k_0 by Landen transformations, as described in Sect. 18.3.4.⁴²

The lattice is a vertical cylinder with the sites on N_1 rings of N_2 sites and the restriction in Sect. 12.1.5, that these integers must be even, can now be relaxed. In Sect. 12.1.3 it was shown, following Baxter (1978), that the free-energy density of the general Z -invariant eight-vertex model decomposes into a sum of those of regular square-lattice models. For such models all the vertical rapidities are $\bar{q}$. We shall, henceforth, suppose that this is the case, using the variables u and u' given by (12.2.2). Given the couplings ($K^{(t)}$, $K^{(b)}$, $K^{(d)}$), it follows from Theorem 12.1.1, that the Z -invariance of the model implies that the equations

$$\begin{aligned} \exp[2K^{(t)}] &= \exp\{-i \operatorname{am}(i[\mathcal{K}'_0 - u]|k_0)\}, \\ \exp[2K^{(b)}] &= \exp\{-i \operatorname{am}(i[\mathcal{K}'_0 - \lambda + u]|k_0)\}, \\ \exp[2K^{(d)}] &= \exp\{i \operatorname{am}(i[\mathcal{K}'_0 - \lambda]|k_0)\}, \end{aligned} \quad (12.3.1)$$

can be solved for u , λ and k_0 . Thus the couplings $K^{(t)}$ and $K^{(b)}$ are functions of u , λ and k_0 ⁴³ and $K^{(d)}$ of λ and k_0 . The Boltzmann weights (a , b , c , d) given by (3.8.11) have a similar dependence, with, as we have already seen in (12.1.22), Δ and Γ dependent simply on λ and k_0 .

12.3.1 The Low-Temperature Zone $\mathcal{R}_t(I)$

It should again be emphasized that the elliptic formulation (12.3.1) is predicated on the condition (12.1.16), which confines its applicability to $\mathcal{R}_t(I)$. In the following we shall show how it can be extended to the rest of phase space. The first step is to use

⁴² A similar notation is adopted for complete elliptic integrals and nomes.

⁴³ Or their transformed equivalents. Sometimes, in the interests of brevity, we shall specifically expose dependence on only one variable whose changing values are to be noted.

the variables (w_1, w_2, w_3, w_4) defined in Sect. 9.3.3 and satisfying the inequality (9.3.24). In $\mathcal{R}_L(1)$ these variables are related to the Boltzmann weights by

$$\begin{aligned} w_1 &= \frac{1}{2}(a + b), & w_2 &= \frac{1}{2}(a - b), \\ w_3 &= \frac{1}{2}(c + d), & w_4 &= \frac{1}{2}(c - d), \end{aligned} \quad (12.3.2)$$

(see the first line of Table 9.1) with

$$\Delta^* = \frac{a^2 + b^2 - c^2 - d^2}{2[ab + cd]} = \Delta, \quad \Gamma^* = \frac{ab - cd}{ab + cd} = \Gamma. \quad (12.3.3)$$

Since the Boltzmann weights are functions of the arbitrary constant coupling $K^{(0)}$ their ratios are the quantities of importance and this similarly applies to the w_j 's.

In the many accounts of the eight-vertex model, it has proved to be the case that elliptic forms with different moduli are convenient in different circumstances for expressing the ratios $w_1 : w_2 : w_3 : w_4$ and the variables Δ^* and Γ^* . Here we continue to use the elliptic modulus k_0 and its complement k'_0 , but from Sect. 12.3.5 onwards a more transparent formulation is achieved by using the Landen-transformed moduli k_1 and k'_1 . For the sake of clarity the forms for the w_j 's, Δ^* and Γ^* in both formulations are given in Table 12.2.

For k_0 and k'_0 more succinct expressions are achieved by using

$$\xi := \frac{1}{2}\lambda, \quad z := \frac{1}{2}\lambda - u, \quad (12.3.4)$$

rather than λ and u . The forms for the w_j 's, Δ^* and Γ^* for these moduli are given in the first two columns of Table 12.2. The inequalities (9.3.24) and (9.3.31) are predicated on the quantities involved being real and, as shown in Sect. 18.3.3, this will be the case only if both members of the pair (ξ, z) are real or purely imaginary. Since the analysis of Sect. 12.1 assumes that λ and u are real, then so are ξ and z and the inequalities hold if⁴⁴

$$|z| < \xi < K'_0. \quad (12.3.5)$$

From (12.3.2), the Boltzmann weights are

⁴⁴ Note, however, that the formulation given below in terms of the variable pair $(\hat{v}, \hat{\eta})$ allows for complex values. This implies the extension of (z, ξ) to complex values.

Table 12.2 Low-temperature elliptic expressions for the ratios $w_1 : w_2 : w_3 : w_4$ and the variables Δ^* and Γ^* for different moduli

Elliptic Moduli	k_0	k'_0	k_1	k'_1
w_1	1	$\frac{\text{cn}(z k'_0)}{\text{cn}(\xi k'_0)}$	$1 - k_1 \text{sn}(\eta + v k_1) \text{sn}(\eta - v k_1)$	$1 + k_1 \text{sc}(i[\eta + v] k'_1) \text{sc}(i[\eta - v] k'_1)$
w_2	$\frac{\text{dn}(iz k_0)}{\text{dn}(i\xi k_0)}$	$\frac{\text{dn}(z k'_0)}{\text{dn}(\xi k'_0)}$	$1 + k_1 \text{sn}(\eta + v k_1) \text{sn}(\eta - v k_1)$	$1 - k_1 \text{sc}(i[\eta + v] k'_1) \text{sc}(i[\eta - v] k'_1)$
w_3	$\frac{\text{cn}(iz k_0)}{\text{cn}(i\xi k_0)}$	1	$\frac{\text{sn}(\eta + v k_1) + \text{sn}(\eta - v k_1)}{\text{sn}(2\eta k_1)}$	$\frac{\text{sc}(i[\eta + v] k'_1) + \text{sc}(i[\eta - v] k'_1)}{\text{sc}(2i\eta k'_1)}$
w_4	$\frac{\text{sn}(iz k_0)}{\text{sn}(i\xi k_0)}$	$\frac{\text{sn}(z k'_0)}{\text{sn}(\xi k'_0)}$	$\frac{\text{sn}(\eta + v k_1) - \text{sn}(\eta - v k_1)}{\text{sn}(2\eta k_1)}$	$\frac{\text{sc}(i[\eta + v] k'_1) - \text{sc}(i[\eta - v] k'_1)}{\text{sc}(2i\eta k'_1)}$
Δ^*	$\text{cd}(2i\xi k_0)$	$\text{nd}(2\xi k'_0)$	$\frac{\text{cn}(2\eta k_1) \text{dn}(2\eta k_1)}{1 - k_1 \text{sn}^2(2\eta k_1)}$	$\frac{\text{nc}(2i\eta k'_1) \text{dc}(2i\eta k'_1)}{1 + k_1 \text{sc}^2(2i\eta k'_1)}$
Γ^*	$-\text{nd}(2i\xi k_0)$	$-\text{cd}(2\xi k'_0)$	$-\frac{1 + k_1 \text{sn}^2(2\eta k_1)}{1 - k_1 \text{sn}^2(2\eta k_1)}$	$-\frac{1 - k_1 \text{sc}^2(2i\eta k'_1)}{1 + k_1 \text{sc}^2(2i\eta k'_1)}$

$$\begin{aligned}
a(z, \xi|k_0) &= \frac{\text{dn}(i\xi|k_0) + \text{dn}(iz|k_0)}{\text{dn}(i\xi|k_0)}, \\
b(z, \xi|k_0) &= \frac{\text{dn}(i\xi|k_0) - \text{dn}(iz|k_0)}{\text{dn}(i\xi|k_0)}, \\
c(z, \xi|k_0) &= \frac{\text{sn}(i\xi|k_0) \text{cn}(iz|k_0) + \text{sn}(iz|k_0) \text{cn}(i\xi|k_0)}{\text{sn}(i\xi|k_0) \text{cn}(i\xi|k_0)}, \\
d(z, \xi|k_0) &= \frac{\text{sn}(i\xi|k_0) \text{cn}(iz|k_0) - \text{sn}(iz|k_0) \text{cn}(i\xi|k_0)}{\text{sn}(i\xi|k_0) \text{cn}(i\xi|k_0)}.
\end{aligned} \tag{12.3.6}$$

12.3.2 The Low-Temperature Zone $\mathcal{R}_L(\text{III})$

This formulation can now be extended to the low-temperature zones $\mathcal{R}_L(\text{II})$, $\mathcal{R}_L(\text{III})$ and $\mathcal{R}_L(\text{IV})$. The forms of the w_j 's in these zones will, apart from a multiplicative factor, be the same but there will be a transformation of variables corresponding to $\mathcal{T}_1$, $\mathcal{T}_2$ or $\mathcal{T}_3$, respectively, which maps from the relationships (12.3.2) and (12.3.3) into the corresponding ones, given in Table 9.1, for these low-temperature zones.

In the original work of Baxter and the more recent work of Fabricius and McCoy the ‘principal regime’ (Baxter 1982b, p. 224) is chosen to be $\mathcal{R}_L(\text{III})$ achieved by the mapping $\mathcal{T}_2$ from $\mathcal{R}_L(\text{I})$.⁴⁵ Here

$$w_1 = \frac{1}{2}(c + d), \quad w_2 = \frac{1}{2}(c - d), \quad (12.3.7)$$

$$w_3 = \frac{1}{2}(a + b), \quad w_4 = \frac{1}{2}(a - b),$$

$$\Delta^* = \frac{c^2 + d^2 - a^2 - b^2}{2[cd + ab]} = -\Delta, \quad \Gamma^* = \frac{cd - ab}{cd + ab} = -\Gamma. \quad (12.3.8)$$

Let

$$\hat{z} := z + \mathcal{K}'_0, \quad \hat{\xi} := \xi + \mathcal{K}'_0. \quad (12.3.9)$$

Then, from (12.3.53),

$$\begin{aligned} a(\hat{z}, \hat{\xi} | k_0) &= E_0(z, \xi | k_0) c(z, \xi | k_0), & b(\hat{z}, \hat{\xi} | k_0) &= E_0(z, \xi | k_0) d(z, \xi | k_0), \\ c(\hat{z}, \hat{\xi} | k_0) &= E_0(z, \xi | k_0) a(z, \xi | k_0), & d(\hat{z}, \hat{\xi} | k_0) &= E_0(z, \xi | k_0) b(z, \xi | k_0), \end{aligned} \quad (12.3.10)$$

where

$$E_0(z, \xi | k_0) = \frac{\text{sn}(i\xi | k_0)}{\text{sn}(iz | k_0)} = \frac{1}{E_0(\hat{z}, \hat{\xi} | k_0)}. \quad (12.3.11)$$

The transformation $\mathcal{T}_2$ corresponds to the change of variables $(\xi, \eta) \rightarrow (\hat{\xi}, \hat{\eta})$. Similar changes of variables correspond to the transformations $\mathcal{T}_1$ and $\mathcal{T}_3$ into the zones $\mathcal{R}_L(\text{II})$ and $\mathcal{R}_L(\text{IV})$. In the cases of $\mathcal{R}_L(\text{III})$ and $\mathcal{R}_L(\text{IV})$ there is a change of sign of Δ and Γ , but not of Δ^* and Γ^* , as defined by (9.3.30).

12.3.3 The Transfer Matrix

In this section and in 12.3.4 it is convenient to use the variables ξ and z , given by (12.3.4).⁴⁶ To carry out the programme described in Chap. 11 for calculating the partition function in the limit $N_1 \rightarrow \infty$ the transfer matrix must be obtained. Suppose

⁴⁵ It may be useful to note that, for our variables $(\hat{z}, \hat{\xi}, k'_0, \hat{v}, \hat{\eta}, k_1)$ in $\mathcal{R}_L(\text{III})$, Baxter (1972b, 1973a, b, c) uses the symbols $(V, \zeta, \ell, v, \eta, k)$, this also being the notation of Fabricius and McCoy (2003, 2005) and McCoy (2010), and Lavis and Bell (1999b, Chap. 5) use $(\mathcal{U}, \zeta, \ell, \mathcal{u}, \xi, \ell'_1)$. Baxter (1982b) uses just the variables equivalent to our $(\hat{v}, \hat{\eta}, k_1)$ for which his notation is $(-\frac{1}{2}iv, \frac{1}{2}i\lambda, k)$.

⁴⁶ Beginning with Sect. 12.3.5 the variables associated with k_1 , and, on occasion, k_2 and k_3 are more useful choices.

for the moment that the horizontal rapidities are different, so that the transfer matrix on the i th row is a function of $z_i = \frac{1}{2}\lambda - u_i$, where $u_i := \bar{q} - q_i$. From (11.1.6),

$$Z(\mathcal{N}^{\otimes}) = \text{Trace} \left\{ \prod_{i=1}^{N_1} \mathbf{V}(z_i) \right\}. \quad (12.3.12)$$

The next equality in (11.1.6), which reduces the partition function to a sum of powers of eigenvalues, depends on the reduction of $\mathbf{V}(z)$ to diagonal form as shown in (11.1.4). In this case we have a product of matrices parameterized by z_i and the reduction can be effected only if, assuming that $\mathbf{V}(z)$ is simple for all z , there exists a non-singular matrix $\mathbf{X}$, independent of z_i , such that

$$\mathbf{X}^{-1} \mathbf{V}(z_i) \mathbf{X} = \mathbf{\Lambda}(z_i), \quad i = 1, 2, \dots, N_1. \quad (12.3.13)$$

A necessary and sufficient condition for this (Sect. 17.2) is

$$[\mathbf{V}(z), \mathbf{V}(z')] = \mathbf{0}, \quad \text{for all } z, z'. \quad (12.3.14)$$

A matrix $\mathbf{X}$ satisfying (12.3.13) is called an eigenmatrix of $\mathbf{V}(z_i)$ and condition (12.3.14) implies that the set $\mathcal{S}_E(\mathbf{V})$ of eigenmatrices is the same for all z . It can be shown (Sect. 12.3.5) that $\mathbf{V}(-z) = \mathbf{V}^T(z)$. So (12.3.14) also implies (Sect. 17.2) that $\mathbf{V}(z)$ is normal and the eigenmatrix $\mathbf{X}$ is unitary. That condition (12.3.14) holds depends, as we shall now show, on the system being Z -invariant.⁴⁷

An element of $\mathbf{V}(z)$ is denoted by $\langle \varsigma | \mathbf{V}(z) | \varsigma' \rangle$, where $\varsigma := (\varsigma_1, \dots, \varsigma_{N_2})$, with ς_j being the edge variable below the j th site of the row; ς' denotes in a similar way the set of edge variables above the line of sites. Let the horizontal edge variable to the right of the j th site be ς_j'' . The matrix element is clearly a sum of products of the Boltzmann weights in a row. With the vertical edge variables fixed the sum will consist of only two non-zero terms since once ς_j'' is fixed the remaining horizontal edge variables are fixed by the eight-vertex constraint condition (3.8.1). Alternatively the eight-vertex condition can be used to extend the summation over all 2^{N_2} terms $\varsigma_j'' = \pm 1$, $j = 1, 2, \dots, N_2$. From (12.1.46),

$$\left\langle \varsigma_{j-1}'' \left| U_0(z) \right| \varsigma_j'' \right\rangle = \begin{cases} \mathfrak{X}(z; \varsigma_j, \varsigma_{j-1}'', \varsigma_j', \varsigma_j''), & \text{if } \varsigma_j \varsigma_{j-1}'' \varsigma_j' \varsigma_j'' = 1, \\ 0, & \text{if } \varsigma_j \varsigma_{j-1}'' \varsigma_j' \varsigma_j'' \neq 1, \end{cases} \quad (12.3.15)$$

where, from (3.8.10),

$$\mathfrak{X}(z; \varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4) = \exp[K^{(0)} + K^{(f)}(z)\varsigma_2\varsigma_3 + K^{(b)}(z)\varsigma_3\varsigma_4 + K^{(4)}\varsigma_1\varsigma_3]. \quad (12.3.16)$$

⁴⁷ This argument can equally well be represented in terms of edge variables or spins; we shall use the former. The alternative spin formulation is similar to that for the Ising model given by Baxter (1982b, Sect. 7.3).

We shall also need

$$\langle v | U_1 \left(\begin{smallmatrix} \zeta' \\ \zeta \end{smallmatrix}; z \right) | v' \rangle := \left\langle \begin{smallmatrix} v \\ \zeta \end{smallmatrix} \middle| U_0(z) \middle| \begin{smallmatrix} \zeta' \\ v' \end{smallmatrix} \right\rangle. \quad (12.3.17)$$

The row and column indices of U_0 are ordered in the pairs

$$\begin{pmatrix} +1 \\ +1 \end{pmatrix}, \quad \begin{pmatrix} +1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ +1 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad (12.3.18)$$

and for U_1 by $+1$ and -1 . From (3.8.11) and (12.3.15),

$$U_0(z) = \begin{pmatrix} a(z) & 0 & 0 & d(z) \\ 0 & c(z) & b(z) & 0 \\ 0 & b(z) & c(z) & 0 \\ d(z) & 0 & 0 & a(z) \end{pmatrix}. \quad (12.3.19)$$

When the four-dimensional matrix U_0 was introduced in Sect. 12.1.4, it was not with respect to a particular lattice structure and the edge variables were simply labelled in clockwise order. This is still the case in (12.3.17), but now there is a distinction between the vertical variables ζ and ζ' used as indices for the transfer matrix elements and the horizontal variables used below to construct those elements in the trace in (12.3.23). It will be convenient if our notation reflects this distinction, particular as, in Sect. 12.3.6, the horizontal variables will be allowed to represent a range of values different from ± 1 . To achieve this result the ordering of the variables in the bra on the left of U_0 in (12.3.17) are interchanged so that both in this bra and (as is already the case) in the ket on the right the upper and lower variables are for the vertical and horizontal variables respectively. Thus

$$\left\langle \begin{smallmatrix} \zeta \\ v \end{smallmatrix} \middle| J U_0(z) \middle| \begin{smallmatrix} \zeta' \\ v' \end{smallmatrix} \right\rangle = \left\langle \begin{smallmatrix} v \\ \zeta \end{smallmatrix} \middle| U_0(z) \middle| \begin{smallmatrix} \zeta' \\ v' \end{smallmatrix} \right\rangle. \quad (12.3.20)$$

where

$$J := I_1 \oplus \Sigma_x \oplus I_1, \quad (12.3.21)$$

with Σ_x the Pauli matrix given by (3.2.6), so that

$$J U_0(z) = \begin{pmatrix} U_1 \left(\begin{smallmatrix} +1 \\ +1 \end{smallmatrix}; z \right) & U_1 \left(\begin{smallmatrix} -1 \\ +1 \end{smallmatrix}; z \right) \\ U_1 \left(\begin{smallmatrix} +1 \\ -1 \end{smallmatrix}; z \right) & U_1 \left(\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}; z \right) \end{pmatrix} = \begin{pmatrix} a(z) & 0 & 0 & d(z) \\ 0 & b(z) & c(z) & 0 \\ 0 & c(z) & b(z) & 0 \\ d(z) & 0 & 0 & a(z) \end{pmatrix}. \quad (12.3.22)$$

The elements of the transfer matrix are

$$\langle \zeta | V(z) | \zeta' \rangle = \text{Trace} \left\{ U_1 \left(\begin{smallmatrix} \zeta'_1 \\ \zeta_1 \end{smallmatrix}; z \right) \cdots U_1 \left(\begin{smallmatrix} \zeta'_{N_2} \\ \zeta_{N_2} \end{smallmatrix}; z \right) \right\} \quad (12.3.23)$$

and

$$\langle \zeta | V(z) V(z') | \zeta' \rangle = \text{Trace} \left\{ U_2 \left(\begin{smallmatrix} \zeta'_1 & z' \\ \zeta_1 & z \end{smallmatrix} \right) \cdots U_2 \left(\begin{smallmatrix} \zeta'_{N_2} & z' \\ \zeta_{N_2} & z \end{smallmatrix} \right) \right\}, \quad (12.3.24)$$

where, from (12.1.48) and (17.2.14),

$$U_2 \left(\begin{smallmatrix} \zeta' & z' \\ \zeta & z \end{smallmatrix} \right) = \sum_{\{v\}} U_1 \left(\begin{smallmatrix} v \\ \zeta \end{smallmatrix}; z \right) \otimes U_1 \left(\begin{smallmatrix} \zeta' \\ v \end{smallmatrix}; z \right). \quad (12.3.25)$$

To establish the commutation relation (12.3.14) we need a non-singular matrix Y , independent of ζ and ζ' , but, in general, dependent on z and z' , such that

$$U_2 \left(\begin{smallmatrix} \zeta' & z' \\ \zeta & z \end{smallmatrix} \right) = Y^{-1}(z, z') U_2 \left(\begin{smallmatrix} \zeta' & z \\ \zeta & z' \end{smallmatrix} \right) Y(z, z'). \quad (12.3.26)$$

12.3.4 Analysis in Terms of Pauli Matrices

For easy of presentation we relabel the Pauli matrices given by (3.2.6) as

$$\Sigma_1 := I_2, \quad \Sigma_2 := \Sigma_z, \quad \Sigma_3 := \Sigma_x, \quad \Sigma_4 := \Sigma_y. \quad (12.3.27)$$

12.3.4.1 The Commutativity Relationship

In $\mathcal{R}_L(I)$ the Boltzmann weights are related to the w_j 's by (12.3.2) and it is straightforward to verify that

$$\left\langle q \left| U_0(z) \right| \begin{smallmatrix} \zeta' \\ q' \end{smallmatrix} \right\rangle = \sum_{\ell=1}^4 w_\ell(z) \langle \zeta | \Sigma_\ell | \zeta' \rangle \langle q | \Sigma_\ell | q' \rangle. \quad (12.3.28)$$

Then, from (12.3.25),

$$\begin{aligned} \left\langle \begin{smallmatrix} \zeta_2 \\ \zeta_1 \end{smallmatrix} \left| U_2 \left(\begin{smallmatrix} \zeta' & z' \\ \zeta & z \end{smallmatrix} \right) \right| \begin{smallmatrix} \zeta_3 \\ \zeta_4 \end{smallmatrix} \right\rangle &= \sum_{j=1}^4 \sum_{k=1}^4 w_j(z) w_k(z') \langle \zeta | \Sigma_j \Sigma_k | \zeta' \rangle \\ &\quad \times \langle \zeta_1 | \Sigma_j | \zeta_4 \rangle \langle \zeta_2 | \Sigma_k | \zeta_3 \rangle. \end{aligned} \quad (12.3.29)$$

Suppose now that

$$\left\langle \begin{smallmatrix} v \\ \varsigma \end{smallmatrix} \middle| Y(z, z') \middle| \begin{smallmatrix} \varsigma' \\ v' \end{smallmatrix} \right\rangle = \sum_{\ell=1}^4 y_{\ell}(z, z') \langle \varsigma | \Sigma_{\ell} | \varsigma' \rangle \langle v | \Sigma_{\ell} | v' \rangle. \quad (12.3.30)$$

Then

$$\begin{aligned} \left\langle \begin{smallmatrix} \varsigma_2 \\ \varsigma_1 \end{smallmatrix} \middle| Y(z, z') U_2 \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \begin{smallmatrix} z' \\ z \end{smallmatrix} \right) \middle| \begin{smallmatrix} \varsigma_3 \\ \varsigma_4 \end{smallmatrix} \right\rangle &= \sum_{i=1}^4 \sum_{j=1}^4 \sum_{k=1}^4 y_i(z, z') w_j(z) w_k(z') \\ &\times \langle \varsigma | \Sigma_j \Sigma_k | \varsigma' \rangle \langle \varsigma_1 | \Sigma_i \Sigma_k | \varsigma_3 \rangle \langle \varsigma_2 | \Sigma_i \Sigma_j | \varsigma_4 \rangle, \end{aligned} \quad (12.3.31)$$

$$\begin{aligned} \left\langle \begin{smallmatrix} \varsigma_2 \\ \varsigma_1 \end{smallmatrix} \middle| U_2 \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \begin{smallmatrix} z' \\ z \end{smallmatrix} \right) Y(z, z') \middle| \begin{smallmatrix} \varsigma_3 \\ \varsigma_4 \end{smallmatrix} \right\rangle &= \sum_{i=1}^4 \sum_{j=1}^4 \sum_{k=1}^4 y_i(z, z') w_j(z) w_k(z') \\ &\times \langle \varsigma | \Sigma_k \Sigma_j | \varsigma' \rangle \langle \varsigma_1 | \Sigma_k \Sigma_i | \varsigma_3 \rangle \langle \varsigma_2 | \Sigma_j \Sigma_i | \varsigma_4 \rangle. \end{aligned} \quad (12.3.32)$$

Evaluating the elements of the Pauli matrices in (12.3.31) and (12.3.32), it is straightforward to show that (12.3.26) is satisfied if and only if

$$\begin{aligned} &y_i(z, z') w_j(z) w_k(z') - y_j(z, z') w_i(z) w_{\ell}(z') \\ &+ y_k(z, z') w_{\ell}(z) w_i(z') - y_m(z, z') w_k(z) w_j(z') = 0 \end{aligned} \quad (12.3.33)$$

holds for all permutations (i, j, k, ℓ) of $(1, 2, 3, 4)$. There are six such distinct relationships which can be regarded as linear homogeneous equations in $y_i(z, z')$, $i = 1, \dots, 4$. A solution exists if the determinant of the coefficients of any four of them is zero. It can be shown that this is the case if and only if the ratio function, defined by (9.3.28), and for which we now introduce dependence on z , satisfies

$$W(i, j|k, \ell; z) = W(i, j|k, \ell; z'), \quad (12.3.34)$$

for all permutations (i, j, k, ℓ) of $(1, 2, 3, 4)$. Since, as was indicated in Sect. 9.3.3, only two of these ratio functions are independent and can be expressed as functions of the elliptic parameters Δ and Γ , (12.3.34) is satisfied and the matrix $Y(z, z')$, given by (12.3.30) exists. This argument, a form of which is given by Baxter (1972b, Appendix B) is included for its illuminating interest. However, here it is strictly unnecessary since, comparing (12.1.51) and (12.3.26) it can be seen that

$$Y(z, z') = U_0(z' - z). \quad (12.3.35)$$

12.3.4.2 The Relationship to the XYZ Chain

In a derivation mirroring that of (12.3.28)

$$\left\langle \begin{smallmatrix} v \\ \varsigma \end{smallmatrix} \middle| U_0(z) \middle| \begin{smallmatrix} \varsigma' \\ v' \end{smallmatrix} \right\rangle = \sum_{\ell=1}^4 g_{\ell}(z) \langle v | \Sigma_{\ell} | \varsigma' \rangle \langle \varsigma | \Sigma_{\ell} | v' \rangle, \quad (12.3.36)$$

where

$$\begin{aligned} g_1(z) &:= \frac{1}{2} [w_1(z) + w_2(z) + w_3(z) + w_4(z)], \\ g_2(z) &:= \frac{1}{2} [w_1(z) + w_2(z) - w_3(z) - w_4(z)], \\ g_3(z) &:= \frac{1}{2} [w_1(z) - w_2(z) + w_3(z) - w_4(z)], \\ g_4(z) &:= \frac{1}{2} [w_1(z) - w_2(z) - w_3(z) + w_4(z)]. \end{aligned} \quad (12.3.37)$$

From Table 12.2,

$$\begin{aligned} g_1(z) &= g(+1, +1, +1; z), & g_2(z) &= g(+1, -1, -1; z), \\ g_3(z) &= g(-1, +1, -1; z), & g_4(z) &= g(-1, -1, +1; z), \end{aligned} \quad (12.3.38)$$

where

$$g(p, q, r; z) := \frac{1}{2} \left\{ \frac{\text{cn}(z|k'_0)}{\text{cn}(\xi|k'_0)} + p \frac{\text{dn}(z|k'_0)}{\text{dn}(\xi|k'_0)} + q + r \frac{\text{sn}(z|k'_0)}{\text{sn}(\xi|k'_0)} \right\}. \quad (12.3.39)$$

From (18.3.40) and Table 12.2⁴⁸

$$\begin{aligned} g_1(\xi) &= 2, & \dot{g}_1(\xi) &= \text{ds}(2\xi|k'_0)(1 + \Delta^* + \Gamma^*), \\ g_2(\xi) &= 0, & \dot{g}_2(\xi) &= \text{ds}(2\xi|k'_0) \Gamma^*, \\ g_3(\xi) &= 0, & \dot{g}_3(\xi) &= -\text{ds}(2\xi|k'_0), \\ g_4(\xi) &= 0, & \dot{g}_4(\xi) &= \text{ds}(2\xi|k'_0) \Delta^*. \end{aligned} \quad (12.3.40)$$

⁴⁸ The dot here denotes differentiation with respect to z . The differences between these quantities and those given by Baxter (1972a, Eq. (3.9)) are due to the different ordering of the Boltzmann factors (compare (12.3.2) with Baxter (1972b, Eq. (2.5)).

Then

$$\langle \nu | U_1 \left(\frac{\zeta'}{\zeta}; \xi \right) | \nu' \rangle = \left\langle \frac{\nu}{\zeta} \left| U_0(\xi) \right| \frac{\zeta'}{\nu'} \right\rangle = \delta^{\text{Kr}}(\nu - \zeta') \delta^{\text{Kr}}(\zeta - \nu'), \quad (12.3.41)$$

giving, from (12.3.23),

$$\langle \zeta | V(\xi) | \zeta' \rangle = \delta^{\text{Kr}}(\zeta_1 - \zeta'_2) \cdots \delta^{\text{Kr}}(\zeta_{N_2-1} - \zeta'_{N_2}) \delta^{\text{Kr}}(\zeta_{N_2} - \zeta'_1), \quad (12.3.42)$$

$$\langle \zeta | V^{-1}(\xi) | \zeta' \rangle = \delta^{\text{Kr}}(\zeta_2 - \zeta'_1) \cdots \delta^{\text{Kr}}(\zeta_{N_2} - \zeta'_{N_2-1}) \delta^{\text{Kr}}(\zeta_1 - \zeta'_{N_2}). \quad (12.3.43)$$

In a similar way

$$\begin{aligned} \langle \zeta | \dot{V}(\xi) | \zeta' \rangle &= \sum_{j=1}^{N_2} \left\{ \delta^{\text{Kr}}(\zeta_1 - \zeta'_2) \cdots \delta^{\text{Kr}}(\zeta_{j-1} - \zeta'_j) \right. \\ &\quad \times \left. \left\langle \frac{\zeta_{j-1}}{\zeta_j} \left| \dot{U}_0(\xi) \right| \frac{\zeta'_j}{\zeta'_{j+1}} \right\rangle \delta^{\text{Kr}}(\zeta_{j+1} - \zeta'_{j+2}) \cdots \delta^{\text{Kr}}(\zeta_{N_2} - \zeta'_1) \right\}. \end{aligned} \quad (12.3.44)$$

Now consider the matrix

$$L(z) := V^{-1}(z) \dot{V}(z), \quad (12.3.45)$$

for which it can easily be shown that

$$[V(z), L(z')] = \mathbf{0}. \quad (12.3.46)$$

From (12.3.43)–(12.3.45),

$$\begin{aligned} \langle \zeta | L(\xi) | \zeta' \rangle &= \sum_{j=1}^{N_2} \left\{ \delta^{\text{Kr}}(\zeta_1 - \zeta'_1) \cdots \delta^{\text{Kr}}(\zeta_{j-1} - \zeta'_{j-1}) \right. \\ &\quad \times \left. \left\langle \frac{\zeta_j}{\zeta_{j+1}} \left| \dot{U}_0(\xi) \right| \frac{\zeta'_j}{\zeta'_{j+1}} \right\rangle \delta^{\text{Kr}}(\zeta_{j+2} - \zeta'_{j+2}) \cdots \delta^{\text{Kr}}(\zeta_{N_2} - \zeta'_1) \right\} \\ &= \sum_{\ell=1}^4 \dot{g}_\ell(\xi) \sum_{j=1}^{N_2} \left\{ \delta^{\text{Kr}}(\zeta_1 - \zeta'_1) \cdots \delta^{\text{Kr}}(\zeta_{j-1} - \zeta'_{j-1}) \right. \\ &\quad \times \left. \langle \zeta_j | \Sigma_\ell | \zeta'_j \rangle \langle \zeta_{j+1} | \Sigma_\ell | \zeta'_{j+1} \rangle \delta^{\text{Kr}}(\zeta_{j+2} - \zeta'_{j+2}) \cdots \delta^{\text{Kr}}(\zeta_{N_2} - \zeta'_1) \right\}. \end{aligned} \quad (12.3.47)$$

Thus

$$\begin{aligned} L(\xi) = N_2 \dot{g}_1(\xi) \mathbf{I} + \sum_{j=1}^{N_2} \{ \dot{g}_3(\xi) \Sigma_x(j) \Sigma_x(j+1) + \dot{g}_4(\xi) \Sigma_y(j) \Sigma_y(j+1) \\ + \dot{g}_2(\xi) \Sigma_z(j) \Sigma_z(j+1) \}. \end{aligned} \quad (12.3.48)$$

Apart the leading diagonal term, $L(\xi)$ is the Hamiltonian of the zero-field spin- $\frac{1}{2}$ XYZ model given by (3.2.4), with

$$(J_x, J_y, J_z) = ds(2\xi|k'_0) (1, -\Delta^*, -\Gamma^*). \quad (12.3.49)$$

The commutation relationship (12.3.46) then implies that the eight-vertex transfer matrix and the XYZ Hamiltonian have a complete set of common eigenvectors; the result due to Sutherland (1970). By keeping Δ^* and Γ^* fixed and varying the Boltzmann weights it is easy to see that the minimum (ground-state) eigenvalue of the XYZ Hamiltonian corresponds to the maximum eigenvalue of V , this, of course, being that which is needed to obtain the free energy.

Special cases arise when two (or more) of J_x, J_y and J_z are equal or one (or more) is zero. The conditions for this can be seen from the expressions for Δ^* and Γ^* in Table 12.2. In order for Δ^* and Γ^* to be real ξ can be either real or imaginary and we include both cases:

$$\begin{aligned} \text{(i)} \quad \xi = i\mathcal{K}_0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (1, 1, 1), \\ \text{(ii)} \quad \xi = \mathcal{K}'_0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (1, -1, -1), \\ \text{(iii)} \quad \xi = 0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (1, -1, 1), \\ \text{(iv)} \quad \xi = \frac{1}{2}i\mathcal{K}_0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (k'_0, 0, 1), \\ \text{(v)} \quad \xi = \frac{1}{2}\mathcal{K}'_0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (k_0, -1, 0), \\ \text{(vi)} \quad k_0 = 0 & \quad \text{gives} \quad (J_x, J_y, J_z) = (1, -\cosh(2\xi), 1), \\ \text{(vii)} \quad k_0 = 1 & \quad \text{gives} \quad (J_x, J_y, J_z) = (1, -1, \cos(2\xi)). \end{aligned} \quad (12.3.50)$$

These cases of the Heisenberg spin chain are discussed in Sect. 3.2.1. (i) is the isotropic case and (ii) and (iii) with parameters differing in sign, but not magnitude, from this have the same eigenvalues but different ground states. (iv) and (v) are cases of the XY model except when k_0 takes the value 1 or zero, respectively, when the system reduces to the Ising model. (vi) and (vii) are cases of the XYZ model except when $\xi = 0$ or (for (vii)) when $\xi = \frac{1}{2}\pi$, when they reduce to the XXZ model.

12.3.5 Analysis of the Transfer Matrix

12.3.5.1 A Change of Variables and the High-Temperature Zone

The formulation of the rest of the analysis of the eight-vertex model, both in the original papers of Baxter (1972b, 1973a, b, c) and in the recent work of Fabricius and McCoy (2003, 2005, 2007), Fabricius (2007), Bazhanov and Mangazeev (2007), Roan (2007), Fabricius and McCoy (2009), McCoy (2010) uses the elliptic parametrization with the Landen-transformed modulus k_1 , related to k_0 by (18.3.75). We do the same, with (v, η) in $\mathcal{R}_L(\text{I})$, transformed to $(\hat{v}, \hat{\eta})$ in $\mathcal{R}_L(\text{III})$.

For k_1 and k'_1 it is convenient to use variables initially set to be purely imaginary. Let

$$v := \frac{iz}{1+k_1} = \frac{iz\mathcal{K}_1}{\mathcal{K}_0} = \frac{iz\mathcal{K}'_1}{2\mathcal{K}'_0}, \quad \eta := \frac{i\xi}{1+k_1} = \frac{i\xi\mathcal{K}_1}{\mathcal{K}_0} = \frac{i\xi\mathcal{K}'_1}{2\mathcal{K}'_0}, \quad (12.3.51)$$

with, from (12.3.5),

$$|v| < -i\eta < \frac{1}{2}\mathcal{K}'_1. \quad (12.3.52)$$

The third and fourth columns of Table 12.2 are then derived from the first two columns using (18.3.82) and (18.3.83). From (12.3.2), in $\mathcal{R}_L(\text{I})$,

$$\begin{aligned} a(v, \eta|k_1) &= 1, & b(v, \eta|k_1) &= k_1 \operatorname{sn}(v + \eta|k_1) \operatorname{sn}(v - \eta|k_1), \\ c(v, \eta|k_1) &= \frac{\operatorname{sn}(v + \eta|k_1)}{\operatorname{sn}(2\eta|k_1)}, & d(v, \eta|k_1) &= -\frac{\operatorname{sn}(v - \eta|k_1)}{\operatorname{sn}(2\eta|k_1)}. \end{aligned} \quad (12.3.53)$$

These expressions for the Boltzmann weights are arbitrary to within a multiplicative factor and differ from those given by (12.3.6) not just in the change of variables but also with respect to such a factor. It is useful, at the outset of the rather technical manipulations needed to derive the eigenvalues of the transfer matrix, to choose a ‘canonical form’ for the Boltzmann weights. This will be done using *Baxter’s modified theta functions*⁴⁹ defined in Sect. 18.3.8. As we saw in Sect. 9.3.1 the partition function of the eight-vertex model is invariant under the unphysical transformations $\mathcal{T}_a, \mathcal{T}_b, \mathcal{T}_c$ and $\mathcal{T}_d$, which change the sign of the corresponding Boltzmann factors. The use of such a transformation turns out to be crucial for the solution of the model. In zone $\mathcal{R}_L(\text{I})$ the need is to change the sign of d .⁵⁰ With this change implemented, and using (12.3.53) and (18.3.44),⁵¹ the Boltzmann weights are given in terms of the

⁴⁹ These functions, which include two new integer parameters m and m' , were introduced by Baxter (1973a). The earlier work of Baxter (1972b), using the standard theta functions, corresponds to the case $m = 1, m' = 0$.

⁵⁰ In zone $\mathcal{R}_L(\text{III})$ it is b which Baxter (1972b) negates.

⁵¹ Which applies equally well to Baxter’s modified theta functions.

functions $\Psi_{\mathfrak{B}}(\alpha, \beta; v, \eta|k_1)$, defined in (18.3.150) and (18.3.151), by

$$\begin{aligned} a(v, \eta|k_1) &:= \Psi_{\mathfrak{B}}(+1, +1; v, \eta|k_1), & b(v, \eta|k_1) &:= \Psi_{\mathfrak{B}}(+1, -1; v, \eta|k_1), \\ c(v, \eta|k_1) &:= \Psi_{\mathfrak{B}}(-1, +1; v, \eta|k_1), & d(v, \eta|k_1) &:= \Psi_{\mathfrak{B}}(-1, -1; v, \eta|k_1). \end{aligned} \quad (12.3.54)$$

We now consider the effect of the *Bazhanov–Mangazeev transformation* described in Sect. 18.3.4. It should first be noted, from (18.3.163), that the Boltzmann weights (12.3.54) satisfy (12.3.2) with the identification

$$\begin{aligned} w_1 &:= W_{\mathfrak{B}}(+1, +1; v, \eta|k_1) = a(v, \eta|k_1) + b(v, \eta|k_1), \\ w_2 &:= W_{\mathfrak{B}}(+1, -1; v, \eta|k_1) = a(v, \eta|k_1) - b(v, \eta|k_1), \\ w_3 &:= W_{\mathfrak{B}}(-1, +1; v, \eta|k_1) = c(v, \eta|k_1) + d(v, \eta|k_1), \\ w_4 &:= W_{\mathfrak{B}}(-1, -1; v, \eta|k_1) = c(v, \eta|k_1) - d(v, \eta|k_1). \end{aligned} \quad (12.3.55)$$

Then, from (18.3.165),

$$\begin{aligned} W_{\mathfrak{B}}(+1, +1; v_3, \eta_3|k_3) &= \Pi_{\mathfrak{B}}(v, \eta|k_1)[c(v, \eta|k_1) + d(v, \eta|k_1)], \\ W_{\mathfrak{B}}(+1, -1; v_3, \eta_3|k_3) &= \Pi_{\mathfrak{B}}(v, \eta|k_1)[a(v, \eta|k_1) + b(v, \eta|k_1)], \\ W_{\mathfrak{B}}(-1, +1; v_3, \eta_3|k_3) &= \Pi_{\mathfrak{B}}(v, \eta|k_1)[c(v, \eta|k_1) - d(v, \eta|k_1)], \\ W_{\mathfrak{B}}(-1, -1; v_3, \eta_3|k_3) &= \Pi_{\mathfrak{B}}(v, \eta|k_1)[a(v, \eta|k_1) - b(v, \eta|k_1)], \end{aligned} \quad (12.3.56)$$

where k_3 is related to k_1 by (18.3.80) and

$$v_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - v]}{\mathcal{K}'_1}, \quad \eta_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - \eta]}{\mathcal{K}'_1}. \quad (12.3.57)$$

From Table 9.1 it can be seen that this transformation is $\mathcal{T}_2\mathcal{T}_{\text{wg}}$, which maps first from $\mathcal{R}_{\text{L}}(\text{I})$ into the high-temperature zone $\mathcal{R}_{\text{O}}(\text{I})$ and then into $\mathcal{R}_{\text{O}}(\text{III})$.

12.3.5.2 Symmetry and Commutativity

The rows and columns of the transfer matrix are indexed by the vectors $\underline{\varsigma}$, which store the edge variables of the row of N_2 vertical edges and each element $\langle \underline{\varsigma} | V(v, \eta) | \underline{\varsigma}' \rangle$ of the transfer matrix is a polynomial of two terms of degree N_2 in the Boltzmann weights. In each term the total number of c's and d's is even, meaning that the total number of a's and b's is of the parity of N_2 . The two terms differ by a change of sign of the horizontal edge variables (a reversal of horizontal arrows). This has the effect of the transformation $\mathcal{T}_1$ of (9.3.3) and, from (18.3.153),

$$\mathbf{V}(v \pm i\mathcal{K}'_1, \eta) = \{-\mathfrak{E}(-m[i\mathcal{K}'_1 \pm 2v]|k_1)\}^{N_2} \mathbf{V}(v, \eta), \quad (12.3.58)$$

where

$$\mathfrak{E}(u|k_1) := \exp\left(\frac{i\pi u}{\mathfrak{U}_{\mathfrak{B}1}}\right), \quad \mathfrak{U}_{\mathfrak{B}1} := \mathfrak{U}(2m, m'|k_1). \quad (12.3.59)$$

The element $\langle \boldsymbol{\varsigma} | \mathbf{V}^T(v, \eta) | \boldsymbol{\varsigma}' \rangle = \langle \boldsymbol{\varsigma}' | \mathbf{V}(v, \eta) | \boldsymbol{\varsigma} \rangle$ of the transposed matrix obtained by interchanging the vertical edge variables at each site of the row has the same effect as the transformation $\mathcal{T}_0$ of (9.3.2), namely an interchange of the c's and d's. From (18.3.152),

$$\mathbf{V}^T(v, \eta) = (-1)^{N_2} \mathbf{V}(v, -\eta) = \mathbf{V}(-v, \eta). \quad (12.3.60)$$

Finally we note, also from (18.3.152), that

$$\mathbf{V}(v + 2\mathcal{K}_1, \eta) = \{\mathfrak{E}(2m'[\mathcal{K}_1 + v]|k_1)\}^{N_2} \mathbf{V}(v, \eta). \quad (12.3.61)$$

It has so far been unnecessary to impose any particular ordering on the index vectors $\boldsymbol{\varsigma}$. However, now it becomes useful to do so and the method we use to produce the sequence $\boldsymbol{\varsigma}^{(1)}, \boldsymbol{\varsigma}^{(2)}, \dots, \boldsymbol{\varsigma}^{(2^{N_2})}$ is recursive⁵²:

- (i) For $N_2 = 1$, $\boldsymbol{\varsigma}^{(1)} := (+1)$, $\boldsymbol{\varsigma}^{(2)} := (-1)$.
- (ii) To generate the index variables for $N_2 + 1$ from those for N_2 add +1 on the right to each vector and then repeat the process adding -1 on the right.

Example 12.3.1 For $N_2 = 2$,

$$\begin{aligned} \boldsymbol{\varsigma}^{(1)} &= (+1, +1), & \boldsymbol{\varsigma}^{(2)} &= (-1, +1), \\ \boldsymbol{\varsigma}^{(3)} &= (+1, -1), & \boldsymbol{\varsigma}^{(4)} &= (-1, -1), \end{aligned} \quad (12.3.62)$$

and for $N_2 = 3$,

$$\begin{aligned} \boldsymbol{\varsigma}^{(1)} &= (+1, +1, +1), & \boldsymbol{\varsigma}^{(2)} &= (-1, +1, +1), \\ \boldsymbol{\varsigma}^{(3)} &= (+1, -1, +1), & \boldsymbol{\varsigma}^{(4)} &= (-1, -1, +1), \\ \boldsymbol{\varsigma}^{(5)} &= (+1, +1, -1), & \boldsymbol{\varsigma}^{(6)} &= (-1, +1, -1), \\ \boldsymbol{\varsigma}^{(7)} &= (+1, -1, -1), & \boldsymbol{\varsigma}^{(8)} &= (-1, -1, -1). \end{aligned} \quad (12.3.63)$$

The symmetry group of the model is $\mathcal{S}_2^{(e)} \times \mathcal{S}_{N_2}$, where $\mathcal{S}_2^{(e)}$ corresponds to the change of sign of edge variables in a row (reversal of arrow directions) and $\mathcal{S}_{N_2}$ is the group

⁵² Contrary to our convention we here denote the index vectors as *rows* containing the edge variables listed from left to right.

of all permutations of a row of edge variables (which leaves invariant the number of up and down arrow directions). We treat these two groups separately and consider their 2^{N_2} -dimensional representations of matrices operating on the column vector $(\varsigma^{(1)} \varsigma^{(2)} \dots \varsigma^{(2^{N_2})})^\top$.

• Since

$$\varsigma^{(k)} = -\varsigma^{(2^{N_2}-k+1)}, \quad k = 1, 2, \dots, 2^{N_2}, \quad (12.3.64)$$

the representation matrix for the arrow-reversal operation is $R(S_2) := S_x$, where

$$\langle \varsigma^{(k)} | S_x | \varsigma^{(k')} \rangle := \delta^{kr} (k + k' - 2^{N_2} - 1) \quad (12.3.65)$$

and

$$[S_x, V(v, \eta)] = \mathbf{0}. \quad (12.3.66)$$

In terms of the Pauli matrices defined in (3.2.6),

$$S_x = \Sigma_x \otimes \Sigma_x \otimes \dots \otimes \Sigma_x \otimes \Sigma_x. \quad (12.3.67)$$

Since

$$W_2^\dagger \Sigma_x W_2 = \Sigma_z, \quad (12.3.68)$$

where

$$W_2 := \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad (12.3.69)$$

it follows that

$$W^\dagger S_x W = S_z, \quad (12.3.70)$$

where

$$W := W_2 \otimes W_2 \otimes \dots \otimes W_2 \otimes W_2 \quad (12.3.71)$$

and

$$S_z := \Sigma_z \otimes \Sigma_z \otimes \dots \otimes \Sigma_z \otimes \Sigma_z \quad (12.3.72)$$

is diagonal with an equal number of entries $+1$ and -1 . This means that this reducible representation of $S^{(e)}$ contains an equal number of copies of the

symmetric and antisymmetric one-dimensional representations of the group. Let P be the permutation matrix with the effect that

$$S'_z := P^\dagger S_z P = P^\dagger W^\dagger S_x W P \quad (12.3.73)$$

has all the symmetric representation $+1$'s appearing before the antisymmetric -1 's. It follows from Theorem 11.3.1 that

$$V_x(v, \eta) := P^\dagger W^\dagger V(v, \eta) W P \quad (12.3.74)$$

is block-diagonal in two blocks to which are attached the variable $\tau_x = +1$ and $\tau_x = -1$ corresponding to the symmetric and antisymmetric representations.

- For $\zeta^{(k)}$, let

$$\pi(\zeta^{(k)}) := \prod_{j=1}^{N_2} \zeta_j^{(k)}. \quad (12.3.75)$$

This variable assigns parities to the vectors $\zeta^{(k)}$ and it follows from the method of ordering of these vectors given above that

$$\langle \zeta^{(k)} | S_z | \zeta^{(k')} \rangle := \delta^{kr} (k - k') \pi(\zeta^{(k)}), \quad (12.3.76)$$

from which it follows that $T^\dagger S_z T = S_z$, where T is any 2^{N_2} -dimensional representative matrix of $\mathcal{S}_{N_2}$. An important property of the transfer matrix is:

Theorem 12.3.1

$$\langle \zeta^{(k)} | V(v, \eta) | \zeta^{(k')} \rangle \neq 0 \text{ if and only if } \pi(\zeta^{(k)}) \pi(\zeta^{(k')}) = +1. \quad (12.3.77)$$

Proof The easiest way to picture this is using arrows assigned to edges according to the direction rule as shown in Fig. 3.2. The non-zero terms in the trace in (12.3.23) consist of two terms corresponding to reversal of all the horizontal arrows. Around the horizontal lattice ring the arrow direction will change at a vertex with different horizontal, and hence for a non-zero contribution different vertical, edge variables. There can be only an even number of such vertices and, since

$$\pi(\zeta^{(k)}) \pi(\zeta^{(k')}) = \prod_{j=1}^{N_2} (\zeta_j^{(k)} \zeta_j^{(k')}), \quad (12.3.78)$$

the theorem follows. $\square$

From (12.3.77)

$$[S_z, V(v, \eta)] = 0, \quad (12.3.79)$$

and (12.3.66) and this equation imply that $\mathcal{S}_E(\mathbf{S}_x, \mathbf{V})$ and $\mathcal{S}_E(\mathbf{S}_z, \mathbf{V})$ are non-empty.⁵³ However,

$$\mathbf{S}_x \mathbf{S}_z = (-1)^{N_2} \mathbf{S}_z \mathbf{S}_x, \quad (12.3.80)$$

although, from (12.3.66) and (12.3.79),

$$[\mathbf{S}_z \mathbf{S}_x, \mathbf{V}(v, \eta)] = \mathbf{0}. \quad (12.3.81)$$

All the eigenvalues of $\mathbf{S}_x$ and $\mathbf{S}_z$, denoted respectively by θ_x and θ_z , are ± 1 .

- If N_2 is odd $\mathbf{S}_x$ and $\mathbf{S}_z$ have no common eigenvectors and all the eigenvalues of $\mathbf{V}(v, \eta)$ are degenerate (Sect. 17.2.2).
- If N_2 is even every non-degenerate eigenvector of $\mathbf{V}(v, \eta)$ is an eigenvector of $\mathbf{S}_x$ and $\mathbf{S}_z$. The subspace defined by a k -degenerate eigenvalue of $\mathbf{V}(v, \eta)$ contains k independent common eigenvectors of $\mathbf{S}_x$ and $\mathbf{S}_z$.

Now we have a second block-diagonalization provided by Theorem 12.3.1. The permutation matrix $\mathbf{P}$, which collects the diagonal elements of $\mathbf{S}_z$ with the same sign, will shuffled the rows and columns of $\mathbf{V}(v, \eta)$ so that

$$\mathbf{V}_z(v, \eta) := \mathbf{P}^\dagger \mathbf{V}(v, \eta) \mathbf{P}. \quad (12.3.82)$$

The first and second blocks of $\mathbf{V}_z$ are labelled with $\tau_z = +1$ and -1 although, in this case, the variable is not simply related to characters of $\mathcal{S}_{N_2}$.⁵⁴ Since the reduction to block diagonal form is in this case effected by a permutation matrix, $\mathbf{V}$ is reducible, which is a necessary, but not sufficient, condition for the eigenvalue equal to the spectral radius $\mathcal{SR}(\mathbf{V})$ to have a multiplicity of at least two (Sect. 17.2.4); this becomes exactly two if the two blocks of $\mathbf{V}_z$ can be shown to be irreducible.

12.3.6 The VQ Equation

To this point our analysis has been for zone $\mathcal{R}_L(\text{I})$, whereas, as we have already noted, most authors have studied $\mathcal{R}_L(\text{III})$. Further discussion using our formulation exposes the wisdom of their choice⁵⁵ and we bring our variables in line with theirs by using the transformation $\mathcal{T}_2$ to $\mathcal{R}_L(\text{III})$ effected by the change of variables

$$\hat{v} := v + \frac{1}{2}i\mathcal{K}'_1, \quad \hat{\eta} := \eta + \frac{1}{2}i\mathcal{K}'_1. \quad (12.3.83)$$

⁵³ For this notation see Sect. 17.2.

⁵⁴ Although $\mathbf{S}_z$ commutes with $\mathbf{V}(v, \eta)$ it is a representative matrix of $\mathcal{S}_{N_2}$ only when N_2 is even.

⁵⁵ In particular the elliptic roots of unity condition (12.3.128) in our variables takes the cumbersome form $\eta + \frac{1}{2}i\mathcal{K}'_1 = \mathcal{U}(2m, m'|k_1)/L$, where L makes its appearance in Sect. 12.3.6.2 as the dimension of the matrices $\mathbf{T}_{R1}$.

For the expressions for the Boltzmann factors given by (12.3.54) this gives

$$\begin{aligned}
 a(v, \eta|k_1) &= -\mathfrak{E}(-m[i\mathcal{K}'_1 - \hat{v} - 3\hat{\eta}]|k_1)\hat{a}(\hat{v}, \hat{\eta}|k_1), \\
 b(v, \eta|k_1) &= -\mathfrak{E}(-m[i\mathcal{K}'_1 - \hat{v} - 3\hat{\eta}]|k_1)\hat{b}(\hat{v}, \hat{\eta}|k_1), \\
 c(v, \eta|k_1) &= -\mathfrak{E}(-m[i\mathcal{K}'_1 - \hat{v} - 3\hat{\eta}]|k_1)\hat{c}(\hat{v}, \hat{\eta}|k_1), \\
 d(v, \eta|k_1) &= -\mathfrak{E}(-m[i\mathcal{K}'_1 - \hat{v} - 3\hat{\eta}]|k_1)\hat{d}(\hat{v}, \hat{\eta}|k_1),
 \end{aligned} \tag{12.3.84}$$

where

$$\begin{aligned}
 \hat{a}(\hat{v}, \hat{\eta}|k_1) &:= \Psi_{\mathfrak{B}}(-1, +1; \hat{v}, \hat{\eta}|k_1), & \hat{b}(\hat{v}, \hat{\eta}|k_1) &:= \Psi_{\mathfrak{B}}(-1, -1; \hat{v}, \hat{\eta}|k_1), \\
 \hat{c}(\hat{v}, \hat{\eta}|k_1) &:= \Psi_{\mathfrak{B}}(+1, +1; \hat{v}, \hat{\eta}|k_1), & \hat{d}(\hat{v}, \hat{\eta}|k_1) &:= \Psi_{\mathfrak{B}}(+1, -1; \hat{v}, \hat{\eta}|k_1).
 \end{aligned} \tag{12.3.85}$$

We again consider the effect of the Bazhanov–Mangazeev transformation, where (12.3.55) is now replaced by

$$\begin{aligned}
 w_1 &:= W_{\mathfrak{B}}(+1, +1; \hat{v}, \hat{\eta}|k_1) = \hat{c}(\hat{v}, \hat{\eta}|k_1) + \hat{d}(\hat{v}, \hat{\eta}|k_1), \\
 w_2 &:= W_{\mathfrak{B}}(+1, -1; \hat{v}, \hat{\eta}|k_1) = \hat{c}(\hat{v}, \hat{\eta}|k_1) - \hat{d}(\hat{v}, \hat{\eta}|k_1), \\
 w_3 &:= W_{\mathfrak{B}}(-1, +1; \hat{v}, \hat{\eta}|k_1) = \hat{a}(\hat{v}, \hat{\eta}|k_1) + \hat{b}(\hat{v}, \hat{\eta}|k_1), \\
 w_4 &:= W_{\mathfrak{B}}(-1, -1; \hat{v}, \hat{\eta}|k_1) = \hat{a}(\hat{v}, \hat{\eta}|k_1) - \hat{b}(\hat{v}, \hat{\eta}|k_1)
 \end{aligned} \tag{12.3.86}$$

and (12.3.56) by

$$\begin{aligned}
 W_{\mathfrak{B}}(+1, +1; \hat{v}_3, \hat{\eta}_3|k_3) &= \Pi_{\mathfrak{B}}(\hat{v}, \hat{\eta}|k_1)[\hat{a}(\hat{v}, \hat{\eta}|k_1) + \hat{b}(\hat{v}, \hat{\eta}|k_1)], \\
 W_{\mathfrak{B}}(+1, -1; \hat{v}_3, \hat{\eta}_3|k_3) &= \Pi_{\mathfrak{B}}(\hat{v}, \hat{\eta}|k_1)[\hat{c}(\hat{v}, \hat{\eta}|k_1) + \hat{d}(\hat{v}, \hat{\eta}|k_1)], \\
 W_{\mathfrak{B}}(-1, +1; \hat{v}_3, \hat{\eta}_3|k_3) &= \Pi_{\mathfrak{B}}(\hat{v}, \hat{\eta}|k_1)[\hat{a}(\hat{v}, \hat{\eta}|k_1) - \hat{b}(\hat{v}, \hat{\eta}|k_1)], \\
 W_{\mathfrak{B}}(-1, -1; \hat{v}_3, \hat{\eta}_3|k_3) &= \Pi_{\mathfrak{B}}(\hat{v}, \hat{\eta}|k_1)[\hat{c}(\hat{v}, \hat{\eta}|k_1) - \hat{d}(\hat{v}, \hat{\eta}|k_1)],
 \end{aligned} \tag{12.3.87}$$

where

$$\hat{v}_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - \hat{v}]}{\mathcal{K}'_1}, \quad \hat{\eta}_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - \hat{\eta}]}{\mathcal{K}'_1}. \tag{12.3.88}$$

From Table 9.1 it can be seen that this transformation is again $\mathcal{T}_2\mathcal{T}_{\text{wg}}$, but now it maps $\mathcal{R}_{\text{L}}(\text{III})$ into $\mathcal{R}_{\text{O}}(\text{I})$.

We keep track of the prefactors in (12.3.84) by defining

$$\mathbf{V}(v, \eta) := [-\mathfrak{E}(-m[i\mathcal{K}'_1 - \hat{v} - 3\hat{\eta}][k_1])]^{N_2} \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}), \quad (12.3.89)$$

for which

$$\widehat{\mathbf{V}}^\top(\hat{v}, \hat{\eta}) = (-1)^{N_2} \widehat{\mathbf{V}}(-\hat{v}, \hat{\eta}) = \widehat{\mathbf{V}}(-\hat{v}, \hat{\eta}), \quad (12.3.90)$$

$$\widehat{\mathbf{V}}(\hat{v} \pm 2\mathcal{K}_1, \hat{\eta}) = \{-\mathfrak{E}(2m'[\mathcal{K}_1 \pm \hat{v}][k_1])\}^{N_2} \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}), \quad (12.3.91)$$

$$\widehat{\mathbf{V}}(\hat{v} \pm i\mathcal{K}'_1, \hat{\eta}) = \{-\mathfrak{E}(-m[i\mathcal{K}'_1 \pm 2\hat{v}][k_1])\}^{N_2} \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}). \quad (12.3.92)$$

From (12.3.22) and (12.3.23),

$$\langle \mathfrak{S} | \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}) | \mathfrak{S}' \rangle = \text{Trace} \left\{ \widehat{U}_1 \left(\begin{smallmatrix} \mathfrak{S}'_1 \\ \mathfrak{S}_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \widehat{U}_1 \left(\begin{smallmatrix} \mathfrak{S}'_{N_2} \\ \mathfrak{S}_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\}, \quad (12.3.93)$$

where

$$\begin{aligned} \widehat{U}_1 \left(\begin{smallmatrix} +1 \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \begin{pmatrix} \hat{a}(\hat{v}, \hat{\eta}) & 0 \\ 0 & \hat{b}(\hat{v}, \hat{\eta}) \end{pmatrix}, \quad \widehat{U}_1 \left(\begin{smallmatrix} -1 \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) = \begin{pmatrix} 0 & \hat{d}(\hat{v}, \hat{\eta}) \\ \hat{c}(\hat{v}, \hat{\eta}) & 0 \end{pmatrix}, \\ \widehat{U}_1 \left(\begin{smallmatrix} +1 \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \begin{pmatrix} 0 & \hat{c}(\hat{v}, \hat{\eta}) \\ \hat{d}(\hat{v}, \hat{\eta}) & 0 \end{pmatrix}, \quad \widehat{U}_1 \left(\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) = \begin{pmatrix} \hat{b}(\hat{v}, \hat{\eta}) & 0 \\ 0 & \hat{a}(\hat{v}, \hat{\eta}) \end{pmatrix}. \end{aligned} \quad (12.3.94)$$

Baxter's strategy for determining the eigenvalues of the transfer matrix involved the determination of two matrices $\mathbf{Q}_R(\hat{v}, \hat{\eta})$ and $\mathbf{Q}_L(\hat{v}, \hat{\eta})$ such that:

(i) They satisfy

$$\begin{aligned} \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}) \mathbf{Q}_R(\hat{v}, \hat{\eta}) &= [\mathfrak{h}_{\mathfrak{B}}(\hat{v} + \hat{\eta}|k_1)]^{N_2} \mathbf{Q}_R(\hat{v} - 2\hat{\eta}, \hat{\eta}) \\ &\quad + [\mathfrak{h}_{\mathfrak{B}}(\hat{v} - \hat{\eta}|k_1)]^{N_2} \mathbf{Q}_R(\hat{v} + 2\hat{\eta}, \hat{\eta}), \end{aligned} \quad (12.3.95)$$

$$\begin{aligned} \mathbf{Q}_L(\hat{v}, \hat{\eta}) \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}) &= [\mathfrak{h}_{\mathfrak{B}}(\hat{v} + \hat{\eta}|k_1)]^{N_2} \mathbf{Q}_L(\hat{v} - 2\hat{\eta}, \hat{\eta}) \\ &\quad + [\mathfrak{h}_{\mathfrak{B}}(\hat{v} - \hat{\eta}|k_1)]^{N_2} \mathbf{Q}_L(\hat{v} + 2\hat{\eta}, \hat{\eta}), \end{aligned} \quad (12.3.96)$$

where $\mathfrak{h}_{\mathfrak{B}}(u|k)$ is given by (18.3.155).

(ii) For at least one value $\hat{v}_0$ of $\hat{v}$, $\mathbf{Q}_L(\hat{v}_0, \hat{\eta})$ and $\mathbf{Q}_R(\hat{v}_0, \hat{\eta})$ are non-singular for all $\hat{\eta}$.

(iii) That the interchange relationship

$$\mathbf{Q}_L(\hat{v}, \hat{\eta}) \mathbf{E} \mathbf{Q}_R(\hat{v}', \hat{\eta}) = \mathbf{Q}_L(\hat{v}', \hat{\eta}) \mathbf{E} \mathbf{Q}_R(\hat{v}, \hat{\eta}), \quad (12.3.97)$$

is satisfied for all $\hat{v}$, $\hat{v}'$ and $\hat{\eta}$, where $\mathbf{E}$ is a constant self-inverse matrix which commutes with $\hat{\mathbf{V}}(\hat{v}, \hat{\eta})$.

In Sect. 12.3.6.2 a matrix $\mathbf{Q}_R$ satisfying (12.3.95) is derived. Then, from (12.3.90), (12.3.95) and (12.3.96),

$$\mathbf{Q}_L(-\hat{v}, \hat{\eta}) = \mathbf{Q}_R^T(\hat{v}, \hat{\eta}). \quad (12.3.98)$$

Baxter (1972b, p. 218) (see also Baxter 1982b, p. 220) *assumed* (for $m' = 0$) that there exists a $\hat{v}_0$ such that $\mathbf{Q}_R(\hat{v}_0, \hat{\eta})$ is non-singular for all $\hat{\eta}$. More recent work by Fabricius and McCoy (2003) tested this assertion for small values of N_2 , L and m with $m' = 0$. They found that $\mathbf{Q}_R(\hat{v}, \hat{\eta})$ is generically non-singular (singular for only a small number of values of $\hat{v}$) if either L is even or both L and m are odd. When L is odd and m is even singularity depends on the parity of N_2 , with $\mathbf{Q}_R(\hat{v}, \hat{\eta})$ being singular for all L if N_2 is odd and singular if $N_2 \geq L - 1$ if N_2 is even. These investigations were extended to the case $m' \neq 0$ by Fabricius and McCoy (2007). They found the same type of behaviour when m and m' were of different parity and that, when m and m' are both even and N_2 is odd, $\mathbf{Q}_R(\hat{v}, \hat{\eta})$ is singular for all $\hat{v}$. Beginning with the case m and N_2 even and $m' = 0$, a different $\mathbf{Q}_R(\hat{v}, \hat{\eta})$, adapted to the situation where both m and m' are even, was developed by Fabricius (2007) and extended to $m' \neq 0$ and even by Fabricius and McCoy (2007). This involves the use of the cyclic condition, described on p. 455.⁵⁶

Henceforth we shall suppose that N_2 , m , m' and L have values such that $\mathbf{Q}_R(\hat{v}_0, \hat{\eta})$ is non-singular and define

$$\mathbf{Q}(\hat{v}, \hat{\eta}) := \mathbf{Q}_R(\hat{v}, \hat{\eta}) \mathbf{Q}_R^{-1}(\hat{v}_0, \hat{\eta}), \quad (12.3.99)$$

so that, from (12.3.95),

$$\begin{aligned} \hat{\mathbf{V}}(\hat{v}, \hat{\eta}) \mathbf{Q}(\hat{v}, \hat{\eta}) &= [\mathbf{h}_{\mathfrak{B}}(\hat{v} + \hat{\eta} | \mathbf{k}_1)]^{N_2} \mathbf{Q}(\hat{v} - 2\hat{\eta}, \hat{\eta}) \\ &\quad + [\mathbf{h}_{\mathfrak{B}}(\hat{v} - \hat{\eta} | \mathbf{k}_1)]^{N_2} \mathbf{Q}(\hat{v} + 2\hat{\eta}, \hat{\eta}). \end{aligned} \quad (12.3.100)$$

This is the *VQ equation*.⁵⁷ The programme (i)–(iii) is completed by a proof of the interchange relationship (12.3.97) in Sect. 12.3.6.3, from which it follows that

$$[\mathbf{Q}(\hat{v}, \hat{\eta}), \mathbf{Q}(\hat{v}', \hat{\eta})] = \mathbf{0}, \quad (12.3.101)$$

⁵⁶ In the work of Fabricius and McCoy the approach used here and the cyclic case are distinguished by a superscripts (1) and (2) respectively.

⁵⁷ Baxter (1972b) used the symbol $\mathbf{T}$ for the transfer matrix and this convention has been adopted in most subsequent work, with the consequent naming of (12.3.100) as the TQ equation.

$$[\widehat{V}(\hat{v}, \hat{\eta}), \mathbf{Q}(\hat{v}, \hat{\eta})] = \mathbf{0}. \quad (12.3.102)$$

These commutation relationships imply that the set $\mathcal{S}_E(\mathbf{Q})$ of eigenmatrices of $\mathbf{Q}(\hat{v}, \hat{\eta})$ is, like that of $\widehat{V}(\hat{v}, \hat{\eta})$ independent of $\hat{v}$, but dependent on $\hat{\eta}$, and if $X_Q(\hat{\eta}) \in \mathcal{S}_E(\mathbf{Q})$ then $X_Q(\hat{\eta}) \in \mathcal{S}_E(\widehat{V})$, so that $\mathcal{S}_E(\widehat{V}, \mathbf{Q}) = \mathcal{S}_E(\mathbf{Q})$.

12.3.6.1 Commutativity and Degeneracy

It follows from (12.3.102) that $\widehat{V}(\hat{v}, \hat{\eta})$ and $\mathbf{Q}(\hat{v}, \hat{\eta})$ have a complete set of common eigenvectors (Sect. 17.2.2). If for one such the respective eigenvalues are $\hat{\Lambda}_V(\hat{v}, \hat{\eta})$ and $\Lambda_Q(\hat{v}, \hat{\eta})$, from (12.3.100),

$$\begin{aligned} \hat{\Lambda}_V(\hat{v}, \hat{\eta}|k_1) \Lambda_Q(\hat{v}, \hat{\eta}|k_1) &= [\mathbf{h}_{\mathfrak{B}}(\hat{v} + \hat{\eta}|k_1)]^{N_2} \Lambda_Q(\hat{v} - 2\hat{\eta}, \hat{\eta}|k_1) \\ &\quad + [\mathbf{h}_{\mathfrak{B}}(\hat{v} - \hat{\eta}|k_1)]^{N_2} \Lambda_Q(\hat{v} + 2\hat{\eta}, \hat{\eta}|k_1). \end{aligned} \quad (12.3.103)$$

It is clear that if this equation is satisfied by more than one eigenvalue of $\mathbf{Q}(\hat{v}, \hat{\eta})$ for a particular eigenvalue of $\widehat{V}(\hat{v}, \hat{\eta})$ then that eigenvalue is degenerate. If $\hat{\Lambda}_V(\hat{v}, \hat{\eta})$ is non-degenerate then the eigenvalue of $\mathbf{Q}(\hat{v}, \hat{\eta})$ satisfying (12.3.103) for this $\hat{\Lambda}_V(\hat{v}, \hat{\eta})$ is unique.

In Sect. 12.3.5 we discussed the commutation relationships of $\widehat{V}(\hat{v}, \hat{\eta})$ with S_x and S_z . Consider now the possible commutation relations

$$[S_x, \mathbf{Q}(\hat{v}, \hat{\eta})] = \mathbf{0}, \quad (12.3.104)$$

$$[S_z, \mathbf{Q}(\hat{v}, \hat{\eta})] = \mathbf{0}, \quad (12.3.105)$$

$$[S_x S_z, \mathbf{Q}(\hat{v}, \hat{\eta})] = \mathbf{0}. \quad (12.3.106)$$

If none of the eigenvalues of $\widehat{V}(\hat{v}, \hat{\eta})$ is degenerate (for which it must be the case that N_2 is even) then its unique eigenmatrix is also an eigenmatrix of S_x , S_z , $S_x S_z$ and $\widehat{Q}(\hat{v}, \hat{\eta})$, which thus satisfies all of (12.3.104)–(12.3.106). If $\mathbf{Q}(\hat{v}, \hat{\eta})$ does not satisfy all these three equations then it will satisfy either none or exactly one of them (Sect. 17.2.2). In those two situations it follows that $\widehat{V}(\hat{v}, \hat{\eta})$ will have degenerate eigenvalues and $\mathbf{Q}(\hat{v}, \hat{\eta})$ will not be unique, since although $\Lambda_Q(\hat{v}, \hat{\eta})$ is unique $X^\dagger(\hat{\eta}) \Lambda_Q(\hat{v}, \hat{\eta}) X(\hat{\eta})$ is a possible $\mathbf{Q}(\hat{v}, \hat{\eta})$ for all $X(\hat{\eta}) \in \mathcal{S}_E(\widehat{V})$.

A relationship between the interchange condition (12.3.97) and the commutation relations (12.3.104)–(12.3.106) is encapsulated in the following result:

Theorem 12.3.2 *If E_1 are E_2 are two of the four matrices I , S_x , S_z and $S_x S_z$, then $[E_1 E_2, \mathbf{Q}(\hat{v}, \hat{\eta})] = \mathbf{0}$ if E_1 and E_2 each satisfy (12.3.97).*

Proof If E_1 and E_2 satisfy (12.3.97), from (12.3.99),

$$\mathcal{Q}(\hat{v}, \hat{\eta}) = \mathcal{Q}_R(\hat{v}, \hat{\eta}) \mathcal{Q}_R^{-1}(\hat{v}_0, \hat{\eta}) = E_i \mathcal{Q}_L^{-1}(\hat{v}_0, \hat{\eta}) \mathcal{Q}_L(\hat{v}, \hat{\eta}) E_i, \quad i = 1, 2, \quad (12.3.107)$$

and

$$\begin{aligned} (E_1 E_2) \mathcal{Q}(\hat{v}, \hat{\eta}) &= E_1 \mathcal{Q}_L^{-1}(\hat{v}_0, \hat{\eta}) \mathcal{Q}_L(\hat{v}, \hat{\eta}) E_2 \\ &= E_1 \mathcal{Q}_L^{-1}(\hat{v}_0, \hat{\eta}) \mathcal{Q}_L(\hat{v}, \hat{\eta}) E_1 (E_1 E_2) \\ &= \mathcal{Q}(\hat{v}, \hat{\eta}) (E_1 E_2). \end{aligned} \quad (12.3.108)$$

□

The matrix $\mathcal{Q}(\hat{v}, \hat{\eta})$ is a function of the parameters m and m' , which enter through the Baxter's modified theta functions. We show below, in Sect. 12.3.6.3, that the interchange condition holds for $E = I$ unless both m and m' are odd. Fabricius and McCoy (2007) showed that the interchange condition holds for $E = S_x$ except if m is odd and m' even, for $E = S_z$ except if m is even and m' odd and $E = S_x S_z$ only when both m and m' are even. It can, therefore, be concluded from Theorem 12.3.2 that:

- (i) If m is odd and m' is even $\mathcal{Q}(\hat{v}, \hat{\eta})$ commutes with S_z .
- (ii) If m is even and m' is odd $\mathcal{Q}(\hat{v}, \hat{\eta})$ commutes with S_x .
- (iii) If m and m' are odd $\mathcal{Q}(\hat{v}, \hat{\eta})$ commutes with $S_x S_z$.
- (iv) If m and m' are even $\mathcal{Q}(\hat{v}, \hat{\eta})$ commutes with S_z , S_x and $S_x S_z$.

A key element of the derivation of $\mathcal{Q}(\hat{v}, \hat{\eta})$ is whether or not the value of $\hat{\eta}$ is an elliptic root of unity, as specified by (12.3.128). As a digression from the main aim of the paper Baxter (1973a, Sect. 6) presented the derivation of a $\mathcal{Q}(\hat{v}, \hat{\eta})$ satisfying (12.3.104)–(12.3.106). This matrix is unique, for specified values of $\hat{v}$, k_0 and $\hat{\eta}$, and all the eigenvalues of $\widehat{V}(\hat{v}, \hat{\eta})$ are non-degenerate. This case was presented in greater detail by Baxter (1982b, Chap. 10).

In contrast the first solution of the model by Baxter (1972b) adopted the elliptic roots of unity condition for $\hat{\eta}$. Then the matrix $\mathcal{Q}(\hat{v}, \hat{\eta})$ is not unique, not all of (12.3.104)–(12.3.106) are satisfied and some of the eigenvalues of $\widehat{V}(\hat{v}, \hat{\eta})$ have degeneracies. A comprehensive list of all the known $\mathcal{Q}$ matrices is given by Fabricius and McCoy (2009) together with conjectured functional equations which they satisfy.

It should be emphasized that all the methods must give the same eigenvalues for $\widehat{V}(\hat{v}, \hat{\eta})$, otherwise they would yield different results for the free energy; the differences are evidenced by different degeneracies.

12.3.6.2 The Derivation of $\mathcal{Q}_R$ and $\mathcal{Q}_L$

The matrices $\mathcal{Q}_R(\hat{v}, \hat{\eta})$ and $\mathcal{Q}_L(\hat{v}, \hat{\eta})$ can be derived independently, but the relationship (12.3.98) allows us to consider only $\mathcal{Q}_R(\hat{v}, \hat{\eta})$. Suppose that

$$\langle \mathcal{S} | \mathcal{Q}_R(\hat{v}, \hat{\eta}) | \mathcal{S}' \rangle = \text{Trace} \left\{ \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S}'_1 \\ \mathcal{S}_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S}'_{N_2} \\ \mathcal{S}_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\}, \quad (12.3.109)$$

where the matrices $\mathbf{T}_{R1}$ are of dimension L . As in (12.3.24) and (12.3.25),

$$\langle \mathcal{S} | \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}) \mathcal{Q}_R(\hat{v}, \hat{\eta}) | \mathcal{S}' \rangle = \text{Trace} \left\{ \mathbf{Y}_R \left(\begin{smallmatrix} \mathcal{S}'_1 \\ \mathcal{S}_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \mathbf{Y}_R \left(\begin{smallmatrix} \mathcal{S}'_{N_2} \\ \mathcal{S}_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\}, \quad (12.3.110)$$

where

$$\mathbf{Y}_R \left(\begin{smallmatrix} \mathcal{S}' \\ \mathcal{S} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) = \sum_{\mathcal{S}''=\pm 1} \widehat{\mathbf{U}}_1 \left(\begin{smallmatrix} \mathcal{S}'' \\ \mathcal{S} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \otimes \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S}' \\ \mathcal{S}'' \end{smallmatrix}; \hat{v}, \hat{\eta} \right). \quad (12.3.111)$$

From (12.3.94)

$$\mathbf{Y}_R \left(\begin{smallmatrix} \mathcal{S} \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) = \begin{pmatrix} \hat{\mathbf{a}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \hat{\mathbf{d}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \\ \hat{\mathbf{c}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \hat{\mathbf{b}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \end{pmatrix}, \quad (12.3.112)$$

$$\mathbf{Y}_R \left(\begin{smallmatrix} \mathcal{S} \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) = \begin{pmatrix} \hat{\mathbf{b}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \hat{\mathbf{c}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \\ \hat{\mathbf{d}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \hat{\mathbf{a}}(\hat{v}, \hat{\eta} | \mathbf{k}_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \mathcal{S} \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \end{pmatrix}. \quad (12.3.113)$$

Suppose that a matrix $\mathbf{W}(\hat{\eta})$ can be found so that

$$\mathbf{W}^{-1}(\hat{\eta}) \mathbf{Y}_{\mathbf{R}} \left(\begin{smallmatrix} \varsigma \\ \varsigma' \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}(\hat{\eta}) = \begin{pmatrix} \mathbf{A} \left(\begin{smallmatrix} \varsigma \\ \varsigma' \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \mathbf{0} \\ \mathbf{O} \left(\begin{smallmatrix} \varsigma \\ \varsigma' \end{smallmatrix}; \hat{v}, \hat{\eta} \right) & \mathbf{B} \left(\begin{smallmatrix} \varsigma \\ \varsigma' \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \end{pmatrix}. \quad (12.3.114)$$

Then (12.3.110) decomposes into the form

$$\begin{aligned} \langle \varsigma | \widehat{\mathbf{V}}(\hat{v}, \hat{\eta}) \mathbf{Q}_{\mathbf{R}}(\hat{v}, \hat{\eta}) | \varsigma' \rangle &= \text{Trace} \left\{ \mathbf{A} \left(\begin{smallmatrix} \varsigma'_1 \\ \varsigma_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \mathbf{A} \left(\begin{smallmatrix} \varsigma'_{N_2} \\ \varsigma_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\} \\ &+ \text{Trace} \left\{ \mathbf{B} \left(\begin{smallmatrix} \varsigma'_1 \\ \varsigma_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \mathbf{B} \left(\begin{smallmatrix} \varsigma'_{N_2} \\ \varsigma_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\}. \end{aligned} \quad (12.3.115)$$

Suppose that

$$\mathbf{W}(\hat{\eta}) := \begin{pmatrix} \mathbf{I}_L & \mathbf{W}_L(\hat{\eta}) \\ \mathbf{0} & \mathbf{I}_L \end{pmatrix}, \quad (12.3.116)$$

where $\mathbf{I}_L$ is the L -dimensional unit matrix and

$$\langle k | \mathbf{W}_L(\hat{\eta}) | k' \rangle = \left\langle \begin{smallmatrix} +1 \\ k \end{smallmatrix} \left| \mathbf{W}(\hat{\eta}) \right| \begin{smallmatrix} -1 \\ k' \end{smallmatrix} \right\rangle = \delta^{\text{Kr}}(k - k') W(\hat{\eta}; k | k_1). \quad (12.3.117)$$

Then

$$\mathbf{W}^{-1}(\hat{\eta}) := \begin{pmatrix} \mathbf{I}_L - \mathbf{W}_L(\hat{\eta}) \\ \mathbf{0} & \mathbf{I}_L \end{pmatrix}, \quad (12.3.118)$$

giving

$$\begin{aligned} \mathbf{A} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \hat{\mathbf{a}}(\hat{v}, \hat{\eta} | k_1) \mathbf{T}_{\mathbf{R}1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) - \hat{\mathbf{c}}(\hat{v}, \hat{\eta} | k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{\mathbf{R}1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right), \\ \mathbf{A} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \hat{\mathbf{b}}(\hat{v}, \hat{\eta} | k_1) \mathbf{T}_{\mathbf{R}1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) - \hat{\mathbf{d}}(\hat{v}, \hat{\eta} | k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{\mathbf{R}1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right), \end{aligned} \quad (12.3.119)$$

$$\begin{aligned}
\mathbf{B} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \hat{\mathbf{b}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) + \hat{\mathbf{c}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}), \\
\mathbf{B} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \hat{\mathbf{a}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) + \hat{\mathbf{d}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}),
\end{aligned}
\tag{12.3.120}$$

$$\begin{aligned}
&\hat{\mathbf{a}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}) - \hat{\mathbf{b}}(\hat{v}, \hat{\eta}|k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \\
&+ \hat{\mathbf{d}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) - \hat{\mathbf{c}}(\hat{v}, \hat{\eta}|k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}) = \mathbf{0}, \\
&\hat{\mathbf{b}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}) - \hat{\mathbf{a}}(\hat{v}, \hat{\eta}|k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \\
&+ \hat{\mathbf{c}}(\hat{v}, \hat{\eta}|k_1) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) - \hat{\mathbf{d}}(\hat{v}, \hat{\eta}|k_1) \mathbf{W}_L(\hat{\eta}) \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \mathbf{W}_L(\hat{\eta}) = \mathbf{0}.
\end{aligned}
\tag{12.3.121}$$

To satisfy (12.3.121) for each matrix element either

$$\langle k | \mathbf{T}_{R1} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle = 0
\tag{12.3.122}$$

or

$$\begin{aligned}
8\hat{\Delta}(\hat{\eta})W(\hat{\eta}; k|k_1)W(\hat{\eta}; k'|k_1) &= [\hat{\Gamma}(\hat{\eta}) + 1][W(\hat{\eta}; k|k_1)^2 + W(\hat{\eta}; k'|k_1)^2] \\
&+ [\hat{\Gamma}(\hat{\eta}) - 1][1 + W(\hat{\eta}; k|k_1)^2W(\hat{\eta}; k'|k_1)^2],
\end{aligned}
\tag{12.3.123}$$

where $\hat{\Delta}(\hat{\eta})$ and $\hat{\Gamma}(\hat{\eta})$ are given by (9.3.1) with the Boltzmann weights $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ replaced by $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$ and $\hat{\mathbf{d}}$. In zone $\mathcal{R}_L(\mathbf{I})$, $\Delta = \Delta^*$ and $\Gamma = \Gamma^*$. So, using the substitution (12.3.83) in Table 12.2,

$$\begin{aligned}
2\text{cn}(2\hat{\eta}|k_1)\text{dn}(2\hat{\eta}|k_1)W(\hat{\eta}; k|k_1)W(\hat{\eta}; k'|k_1) &- [W(\hat{\eta}; k|k_1)^2 + W(\hat{\eta}; k'|k_1)^2] \\
&+ k_1 \text{sn}^2(2\hat{\eta}|k_1)[1 + W(\hat{\eta}; k|k_1)^2W(\hat{\eta}; k'|k_1)^2] = 0.
\end{aligned}
\tag{12.3.124}$$

This equation is satisfied by

$$\begin{aligned}
W(\hat{\eta}; k|k_1) &= \sqrt{k_1} \text{sn}(x \pm \hat{\eta}|k_1), \\
W(\hat{\eta}; k'|k_1) &= \sqrt{k_1} \text{sn}(x \mp \hat{\eta}|k_1),
\end{aligned}
\tag{12.3.125}$$

for any x , with the two solutions coinciding when $x = \mathcal{K}_1 + \mathcal{U}(2\ell, \ell'|k_1)$ to give

$$W(\hat{\eta}; k|k_1) = W(\hat{\eta}; k'|k_1) = \sqrt{k_1} \operatorname{sn}(\mathcal{K}_1 + \mathfrak{U}(2\ell, \ell'|k_1) \pm \hat{\eta}|k_1), \quad (12.3.126)$$

where ℓ and ℓ' are any integers and, as defined in (18.3.21), $\mathfrak{U}(2\ell, \ell'|k_1) := 2\ell \mathcal{K}_1 + \ell' i \mathcal{K}_1$. The Eq. (12.3.125) cannot be satisfied for all pairs of integers k and k' in the range $[1, L]$. So (12.3.122) is imposed except for a sparse set of matrix elements. Two choices have been shown to satisfy conditions (i)–(iii) of p. 448. Matrix elements are non-zero only when either:

$$(1) \quad (a) k = k' = 1, (b) k' = k \pm 1, (c) k = k' = L,$$

or

$$(2) \quad k' = k \pm 1, \text{ with } k', k = 1, 2, \dots, L \pmod{L}.$$

In this chapter attention is confined to case (1). For the cyclic case (2), which was proposed by Fabricius (2007), the reader is referred to Fabricius and McCoy (2007) and McCoy (2010).

Conditions (a) and (b) are satisfied by

$$W(\hat{\eta}; k|k_1) := \sqrt{k_1} \operatorname{sn}(\mathcal{K}_1 + (2k - 1)\hat{\eta}|k_1) = W(-\hat{\eta}; k|k_1). \quad (12.3.127)$$

To satisfy (c), from (18.3.48),

$$\hat{\eta} = \frac{\mathfrak{U}(2m, m'|k_1)}{2L} = \frac{m}{L} \mathcal{K}_1 + \frac{m'}{2L} i \mathcal{K}'_1, \quad (12.3.128)$$

for some integers m and m' ; this is called the *elliptic roots of unity condition*.

We pause to note the remarkable nature of this result. A parameter $\hat{\eta}$, to which our initial discussions attributed only imaginary values, must now take one of a set of discrete complex values, called the elliptic roots of unity. Further it is the case that the original solution of Baxter (1972b) corresponds to $m' = 0$, when $\hat{\eta}$ becomes real. The effect of $\hat{\eta}$ changing from a purely imaginary to a complex and more particularly to a real value is to shift the zone of validity from $\mathcal{R}_L(\text{III})$. The zone into which it is transformed will become clear when the calculation yields a form for the free-energy density.

Given that the same integer pair (m, m') is used to define $\hat{\eta}$ and the Baxter modified theta functions, so that $\mathfrak{U}(2m, m'|k_1) := \mathfrak{U}_{\mathfrak{B}1}$, from (18.3.131),

$$\begin{aligned} W(\hat{\eta}; k|k_1) &= \frac{H_{\mathfrak{B}}(\mathcal{K}_1 + (2k - 1)\hat{\eta}|k_1)}{\Theta_{\mathfrak{B}}(\mathcal{K}_1 + (2k - 1)\hat{\eta}|k_1)}, \quad k = 1, 2, \dots, L - 1, \\ W(\hat{\eta}; L|k_1) &= \frac{H_{\mathfrak{B}}(\mathcal{K}_1 + \mathfrak{U}_{\mathfrak{B}1} - \hat{\eta}|k_1)}{\Theta_{\mathfrak{B}}(\mathcal{K}_1 + \mathfrak{U}_{\mathfrak{B}1} - \hat{\eta}|k_1)}. \end{aligned} \quad (12.3.129)$$

From (12.3.121),

$$\begin{aligned}
 \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle F(-1, +1; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1) \\
 = \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle F(+1, -1; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1), \\
 \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle F(+1, +1; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1) \\
 = \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle F(-1, -1; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1),
 \end{aligned} \tag{12.3.130}$$

where, from (12.3.127), the functions

$$\begin{aligned}
 F(+1, \beta; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1) &:= \Psi_{\mathfrak{B}}(+1, \beta; \hat{v}, \hat{\eta} | \mathbf{k}_1) \\
 &\quad - \Psi_{\mathfrak{B}}(+1, -\beta; \hat{v}, \hat{\eta} | \mathbf{k}_1) W(\hat{\eta}; k | \mathbf{k}_1) W(\eta; k' | \mathbf{k}_1), \\
 F(-1, \beta; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1) &:= \Psi_{\mathfrak{B}}(-1, -\beta; \hat{v}, \hat{\eta} | \mathbf{k}_1) W(\hat{\eta}; k | \mathbf{k}_1) \\
 &\quad - \Psi_{\mathfrak{B}}(-1, \beta; \hat{v}, \hat{\eta} | \mathbf{k}_1) W(\hat{\eta}; k' | \mathbf{k}_1),
 \end{aligned} \tag{12.3.131}$$

have the properties

$$F(\alpha, \beta; \hat{v}, \hat{\eta}; k', k | \mathbf{k}_1) = \alpha F(\alpha, \beta\alpha; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1), \tag{12.3.132}$$

$$F(\alpha, \beta; -\hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1) = \alpha F(\alpha, \beta\alpha; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1), \tag{12.3.133}$$

$$F(\alpha, \beta; \hat{v}, -\hat{\eta}; k, k' | \mathbf{k}_1) = -\alpha F(\alpha, \beta\alpha; \hat{v}, \hat{\eta}; k, k' | \mathbf{k}_1), \tag{12.3.134}$$

from which it follows that the conditions

$$\langle k' | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ \nu \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k \rangle = \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ \nu \end{smallmatrix}; -\hat{v}, \hat{\eta} \right) | k' \rangle = \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ \nu \end{smallmatrix}; \hat{v}, -\hat{\eta} \right) | k' \rangle \tag{12.3.135}$$

can be consistently imposed. From (12.3.85), (12.3.130), (12.3.131), (12.3.135), (18.3.33) and (18.3.34),

$$\begin{aligned}
 \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k+1 \rangle &= H_{\mathfrak{B}}(\mathcal{K}_1 - \hat{v} + 2k\hat{\eta}) | \mathbf{k}_1 \rangle G(\varsigma; k), \\
 \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k+1 \rangle &= \Theta_{\mathfrak{B}}(\mathcal{K}_1 - \hat{v} + 2k\hat{\eta}) | \mathbf{k}_1 \rangle G(\varsigma; k),
 \end{aligned} \tag{12.3.136}$$

where the pair $(k, k + 1)$ is in the range $[1, L]$, and

$$\begin{aligned} \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | 1 \rangle &= \mathbf{H}_{\mathfrak{B}}(\mathcal{K}_1 - \hat{v} | \mathbf{k}_1) G_0(\varsigma; 1), \\ \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | 1 \rangle &= \mathbf{\Theta}_{\mathfrak{B}}(\mathcal{K}_1 - \hat{v} | \mathbf{k}_1) G_0(\varsigma; 1), \end{aligned} \quad (12.3.137)$$

$$\begin{aligned} \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ +1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | L \rangle &= \mathbf{H}_{\mathfrak{B}}(\mathcal{K}_1 + \mathcal{U}_{\mathfrak{B}1} - \hat{v} | \mathbf{k}_1) G_0(\varsigma; L), \\ \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ -1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | L \rangle &= \mathbf{\Theta}_{\mathfrak{B}}(\mathcal{K}_1 + \mathcal{U}_{\mathfrak{B}1} - \hat{v} | \mathbf{k}_1) G_0(\varsigma; L), \end{aligned} \quad (12.3.138)$$

where $G(\varsigma; k)$, $G_0(\varsigma; 1)$ and $G_0(\varsigma; L)$ are arbitrary constants.⁵⁸

It follows from (12.3.119) and (12.3.120) the $\mathbf{A}$ and $\mathbf{B}$ matrices have the same zero elements as $\mathbf{T}_{\text{r1}}$ with

$$\begin{aligned} \mathbf{A} \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \mathbf{h}_{\mathfrak{B}}(\hat{v} - \hat{\eta} | \mathbf{k}) \mathbf{\Theta}_{\mathfrak{B}}^{-1} \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \hat{v} + 2\hat{\eta}, \hat{\eta} \right) \mathbf{\Theta}_{\mathfrak{B}}, \\ \mathbf{B} \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \hat{v}, \hat{\eta} \right) &= \mathbf{h}_{\mathfrak{B}}(\hat{v} + \hat{\eta} | \mathbf{k}) \mathbf{\Theta}_{\mathfrak{B}} \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma' \\ \varsigma \end{smallmatrix}; \hat{v} - 2\hat{\eta}, \hat{\eta} \right) \mathbf{\Theta}_{\mathfrak{B}}^{-1}, \end{aligned} \quad (12.3.139)$$

where $\mathbf{\Theta}_{\mathfrak{B}}$ is the diagonal matrix with elements

$$\langle k | \mathbf{\Theta}_{\mathfrak{B}} | k' \rangle := \delta^{\text{Kr}}(k - k') \mathbf{\Theta}_{\mathfrak{B}}(\mathcal{K}_1 + (2k - 1)\hat{\eta} | \mathbf{k}_1), \quad (12.3.140)$$

confirming, from (12.3.115), that the matrix $\mathbf{Q}_{\text{r}}(\hat{v}, \hat{\eta})$, defined by (12.3.109), satisfies (12.3.95).

12.3.6.3 The Interchange Relationship for $\mathbf{Q}_{\text{r}}$ and $\mathbf{Q}_{\text{L}}$

Here we need the companion formula

$$\langle \varsigma | \mathbf{Q}_{\text{L}}(\hat{v}, \hat{\eta}) | \varsigma' \rangle = \text{Trace} \left\{ \mathbf{T}_{\text{L1}} \left(\begin{smallmatrix} \varsigma'_1 \\ \varsigma_1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \cdots \mathbf{T}_{\text{L1}} \left(\begin{smallmatrix} \varsigma'_{N_2} \\ \varsigma_{N_2} \end{smallmatrix}; \hat{v}, \hat{\eta} \right) \right\} \quad (12.3.141)$$

to (12.3.109), where, from (12.3.98),

⁵⁸ Of course, this is another *assumption* built into the derived $\mathbf{Q}$ matrix; in general there would be nothing to prevent the presence of arbitrary functions.

$$T_{L1} \left(\frac{\varsigma'}{\varsigma}; \hat{v}, \hat{\eta} \right) = T_{R1} \left(\frac{\varsigma}{\varsigma'}; -\hat{v}, \hat{\eta} \right). \quad (12.3.142)$$

We consider only the case where $E = I$ in (12.3.97). As in (12.3.110) and (12.3.111)

$$\langle \varsigma | Q_L(\hat{v}, \hat{\eta}) Q_R(\hat{v}', \hat{\eta}) | \varsigma' \rangle = \text{Trace} \left\{ Z \left(\frac{\varsigma'_1}{\varsigma_1}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) \cdots Z \left(\frac{\varsigma'_{N_2}}{\varsigma_{N_2}}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) \right\}, \quad (12.3.143)$$

where

$$Z \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) = Z^{(+)} \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) + Z^{(-)} \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right), \quad (12.3.144)$$

$$Z^{(\pm)} \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) = T_{L1} \left(\frac{\pm 1}{\varsigma}; \hat{v}, \hat{\eta} \right) \otimes T_{R1} \left(\frac{\varsigma'}{\pm 1}; \hat{v}', \hat{\eta} \right), \quad (12.3.145)$$

are L^2 -dimensional matrices. It is convenient to index their rows and columns by the pair $\begin{pmatrix} k \\ j \end{pmatrix}, j, k = 1, 2, \dots, L$, so that

$$\begin{aligned} \left\langle \begin{pmatrix} k \\ j \end{pmatrix} \middle| Z^{(\pm)} \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) \middle| \begin{pmatrix} k' \\ j' \end{pmatrix} \right\rangle &= \langle j | T_{L1} \left(\frac{\pm 1}{\varsigma}; \hat{v}, \hat{\eta} \right) | j' \rangle \langle k | T_{R1} \left(\frac{\varsigma'}{\pm 1}; \hat{v}', \hat{\eta} \right) | k' \rangle \\ &= \langle j | T_{R1} \left(\frac{\varsigma}{\pm 1}; -\hat{v}, \hat{\eta} \right) | j' \rangle \langle k | T_{R1} \left(\frac{\varsigma'}{\pm 1}; \hat{v}', \hat{\eta} \right) | k' \rangle. \end{aligned} \quad (12.3.146)$$

A sufficient condition for the interchange relationship to hold is that there exists, for some matrix D , the similarity transformation

$$Z \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) = D^{-1} \left(\frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) Z \left(\frac{\varsigma'}{\varsigma}; \frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right) D \left(\frac{\hat{v}'}{\hat{v}}, \hat{\eta} \right). \quad (12.3.147)$$

In general the task of finding such a matrix is difficult and Baxter (1972b) and later authors (see e.g. McCoy 2010) have confined their attention to a diagonal transformation.⁵⁹

⁵⁹ Of course, in cases (specifically for different parities of m and m') where the interchange condition is established using a diagonal transformation this is sufficient. However, when a diagonal transformation does not exist it leave open the question as to whether a more general transformation may perform the task.

We first consider the four diagonal elements of $\mathbf{Z}$, where the contributions from the elements of the diagonal matrix $\mathbf{D}$ cancel,

$$\left\langle \frac{1}{1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{1}{1} \right\rangle = \mathfrak{L}(\hat{v}, \hat{v}' | k_1) G_0(\varsigma; 1) G_0(\varsigma'; 1), \quad (12.3.148)$$

$$\left\langle \frac{L}{L} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{L}{L} \right\rangle = \mathfrak{L}(\hat{v} - \mathfrak{U}_{\mathfrak{B}1}, \hat{v}' - \mathfrak{U}_{\mathfrak{B}1} | k_1) G_0(\varsigma; L) G_0(\varsigma'; L), \quad (12.3.149)$$

$$\left\langle \frac{L}{1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{L}{1} \right\rangle = \mathfrak{L}(\hat{v}, \hat{v}' - \mathfrak{U}_{\mathfrak{B}1} | k_1) G_0(\varsigma; 1) G_0(\varsigma'; L), \quad (12.3.150)$$

$$\left\langle \frac{1}{L} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{1}{L} \right\rangle = \mathfrak{L}(\hat{v} - \mathfrak{U}_{\mathfrak{B}1}, \hat{v}' | k_1) G_0(\varsigma; L) G_0(\varsigma'; 1), \quad (12.3.151)$$

where $\mathfrak{L}(v, u | k)$ is given by (18.3.143). It is clear, from (18.3.147), that the matrix elements given by (12.3.148) and (12.3.149) are invariant under the interchange of $\hat{v}$ and $\hat{v}'$, but that, from (18.3.148),

$$\left\langle \frac{L}{1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{L}{1} \right\rangle = (-1)^{mm'} \left\langle \frac{L}{1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}, \hat{\eta} \right) \right| \frac{L}{1} \right\rangle, \quad (12.3.152)$$

$$\left\langle \frac{1}{L} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{1}{L} \right\rangle = (-1)^{mm'} \left\langle \frac{1}{L} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}, \hat{\eta} \right) \right| \frac{1}{L} \right\rangle. \quad (12.3.153)$$

The interchange condition is not satisfied if both m and m' are odd. The non-diagonal elements are

$$\begin{aligned} \left\langle \frac{k}{j} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{k+1}{j+1} \right\rangle &= \mathfrak{L}(\hat{v} + 2j\hat{\eta}, \hat{v}' - 2k\hat{\eta} | k_1) G(\varsigma; j) G(\varsigma'; k), \\ \left\langle \frac{k}{j+1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{k+1}{j} \right\rangle &= \mathfrak{L}(\hat{v} - 2j\hat{\eta}, \hat{v}' - 2k\hat{\eta} | k_1) G(\varsigma; j) G(\varsigma'; k), \\ \left\langle \frac{k+1}{j} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{k}{j+1} \right\rangle &= \mathfrak{L}(\hat{v} + 2j\hat{\eta}, \hat{v}' + 2k\hat{\eta} | k_1) G(\varsigma; j) G(\varsigma'; k), \\ \left\langle \frac{k+1}{j+1} \left| \mathbf{Z} \left(\frac{\varsigma'}{\varsigma}; \hat{v}', \hat{\eta} \right) \right| \frac{k}{j} \right\rangle &= \mathfrak{L}(\hat{v} - 2j\hat{\eta}, \hat{v}' + 2k\hat{\eta} | k_1) G(\varsigma; j) G(\varsigma'; k), \end{aligned} \quad (12.3.154)$$

$$\begin{aligned}
\left\langle \begin{matrix} k \\ 1 \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} k+1 \\ 1 \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v}, \hat{v}' - 2k\hat{\eta}|\mathbf{k}_1)G_0(\varsigma; 1)G(\varsigma'; k), \\
\left\langle \begin{matrix} k+1 \\ 1 \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} k \\ 1 \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v}, \hat{v}' + 2k\hat{\eta}|\mathbf{k}_1)G_0(\varsigma; 1)G(\varsigma'; k), \\
\left\langle \begin{matrix} 1 \\ j \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} 1 \\ j+1 \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} + 2j\hat{\eta}, \hat{v}'|\mathbf{k}_1)G(\varsigma; j)G_0(\varsigma'; 1), \\
\left\langle \begin{matrix} 1 \\ j+1 \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} 1 \\ j \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} - 2j\hat{\eta}, \hat{v}'|\mathbf{k}_1)G(\varsigma; j)G_0(\varsigma'; 1),
\end{aligned}
\tag{12.3.155}$$

$$\begin{aligned}
\left\langle \begin{matrix} k \\ L \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} k+1 \\ L \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} - \mathfrak{U}_{\mathfrak{B}1}, \hat{v}' - 2k\hat{\eta}|\mathbf{k}_1)G_0(\varsigma; L)G(\varsigma'; k), \\
\left\langle \begin{matrix} k+1 \\ L \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} k \\ L \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} - \mathfrak{U}_{\mathfrak{B}1}, \hat{v}' + 2k\hat{\eta}|\mathbf{k}_1)G_0(\varsigma; L)G(\varsigma'; k), \\
\left\langle \begin{matrix} L \\ j \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} L \\ j+1 \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} + 2j\hat{\eta}, \hat{v}' - \mathfrak{U}_{\mathfrak{B}1}|\mathbf{k}_1)G(\varsigma; j)G_0(\varsigma'; L), \\
\left\langle \begin{matrix} L \\ j+1 \end{matrix} \middle| \mathbf{Z} \left(\begin{matrix} \varsigma' \\ \varsigma \end{matrix}; \begin{matrix} \hat{v}' \\ \hat{v} \end{matrix}, \hat{\eta} \right) \middle| \begin{matrix} L \\ j \end{matrix} \right\rangle &= \mathfrak{L}(\hat{v} - 2j\hat{\eta}, \hat{v}' - \mathfrak{U}_{\mathfrak{B}1}|\mathbf{k}_1)G(\varsigma; j)G_0(\varsigma'; L).
\end{aligned}
\tag{12.3.156}$$

Let the diagonal elements of $\mathbf{D}$ be $D(\hat{v}, \hat{v}', \hat{\eta}; j, k|\mathbf{k}_1)$ then, using (18.3.149), for the twelve elements in (12.3.154)–(12.3.156) the interchange conditions are satisfied if

$$\begin{aligned}
\frac{D(\hat{v}, \hat{v}', \hat{\eta}; j+1, k+1|\mathbf{k}_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; j, k|\mathbf{k}_1)} &= \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}[j+k]|\mathbf{k}_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}[j+k]|\mathbf{k}_1)}, \\
\frac{D(\hat{v}, \hat{v}', \hat{\eta}; j, k+1|\mathbf{k}_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; j+1, k|\mathbf{k}_1)} &= \frac{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}[j-k]|\mathbf{k}_1)}{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}[j-k]|\mathbf{k}_1)},
\end{aligned}
\tag{12.3.157}$$

$$\begin{aligned}
\frac{D(\hat{v}, \hat{v}', \hat{\eta}; 1, k+1|\mathbf{k}_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; 1, k|\mathbf{k}_1)} &= \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}k|\mathbf{k}_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}k|\mathbf{k}_1)}, \\
\frac{D(\hat{v}, \hat{v}', \hat{\eta}; j+1, 1|\mathbf{k}_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; j, 1|\mathbf{k}_1)} &= \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}j|\mathbf{k}_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}j|\mathbf{k}_1)},
\end{aligned}
\tag{12.3.158}$$

$$\frac{D(\hat{v}, \hat{v}', \hat{\eta}; L, k+1|k_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; L, k|k_1)} = (-1)^{mm'} \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}[k+L]|k_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}[k+L]|k_1)}, \quad (12.3.159)$$

$$\frac{D(\hat{v}, \hat{v}', \hat{\eta}; j+1, L|k_1)}{D(\hat{v}, \hat{v}', \hat{\eta}; j, L|k_1)} = (-1)^{mm'} \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}[j+L]|k_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}[j+L]|k_1)},$$

where $\mathfrak{H}(u|k)$ is given by (18.3.146). Let

$$\frac{\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; k+2|k_1)}{\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; k|k_1)} := \frac{\mathfrak{H}(\hat{v}' - \hat{v} - 2\hat{\eta}k|k_1)}{\mathfrak{H}(\hat{v}' - \hat{v} + 2\hat{\eta}k|k_1)} = \frac{\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; -k|k_1)}{\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; 2-k|k_1)}. \quad (12.3.160)$$

Then Eqs. (12.3.157) are satisfied by

$$D(\hat{v}, \hat{v}', \hat{\eta}; j, k|k_1) = \mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; j+k|k_1) \mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; j-k+1|k_1). \quad (12.3.161)$$

It follows, by induction on (12.3.160), that $\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; 1+k|k_1) = \mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; 1-k|k_1)$ and thus that (12.3.161) satisfies (12.3.158). With $\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; 1|k_1)$ and $\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; 2|k_1)$ set to unity the recurrence relationship (12.3.161) generates independent sequences of terms for $\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; k|k_1)$ with k odd and even and the elements of $\mathbf{D}$ are formed as products of one term from each sequence. Since $\mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; k + \mathcal{U}_{\mathfrak{B}1}|k_1) = \mathfrak{T}(\hat{v}, \hat{v}', \hat{\eta}; k - \mathcal{U}_{\mathfrak{B}1}|k_1)$, (12.3.161) satisfies (12.3.159) if not both of m and m' are odd and this is then the condition for the interchange condition (12.3.97) to hold when $\mathbf{E} = \mathbf{I}$.

12.3.6.4 Periodicity Properties of $\mathcal{Q}$

Baxter's modified theta functions satisfy the periodicity conditions (18.3.127), (18.3.128) and (18.3.130)–(18.3.136), which can then be applied to (12.3.136)–(12.3.138) to give⁶⁰

$$\langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \mathcal{S} \\ \pm 1 \end{smallmatrix}; \hat{v} + 2\mathcal{O}_{\mathfrak{B}1}, \hat{\eta} \right) | k' \rangle = (\mp 1)^{\bar{m}} (-1)^{\bar{m}\bar{m}'} \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \mathcal{S} \\ \pm 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle, \quad (12.3.162)$$

$$\begin{aligned} \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \mathcal{S} \\ \pm 1 \end{smallmatrix}; \hat{v} + 2\mathcal{O}'_{\mathfrak{B}1}, \hat{\eta} \right) | k+1 \rangle &= (\mp 1)^n (-1)^{nm'} \\ &\times \exp \left(-\frac{\pi i \{ \mathcal{O}'_{\mathfrak{B}1} + \hat{v} - 2k\hat{\eta} \}}{\mathcal{O}_{\mathfrak{B}1}} \right) \langle k | \mathbf{T}_{\text{rl}} \left(\begin{smallmatrix} \mathcal{S} \\ \pm 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k+1 \rangle, \end{aligned} \quad (12.3.163)$$

⁶⁰ With $\mathcal{O}_{\mathfrak{B}1} := \mathcal{U}_{\mathfrak{B}1}/m_0 = \bar{m}\mathcal{K}_1 + i\bar{m}'\mathcal{K}'_1$.

$$\begin{aligned} \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + 2\mathcal{O}'_{\mathfrak{B}1}, \hat{\eta} \right) | 1 \rangle &= (\mp 1)^n (-1)^{nm'} \\ &\times \exp \left(-\frac{\pi i \{ \mathcal{O}'_{\mathfrak{B}1} + \hat{v} \}}{\mathcal{O}_{\mathfrak{B}1}} \right) \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | 1 \rangle, \end{aligned} \quad (12.3.164)$$

$$\begin{aligned} \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + 2\mathcal{O}'_{\mathfrak{B}1}, \hat{\eta} \right) | L \rangle &= (\mp 1)^n (-1)^{nm'} (-1)^{m_0} \\ &\times \exp \left(-\frac{\pi i \{ \mathcal{O}'_{\mathfrak{B}1} + \hat{v} \}}{\mathcal{O}_{\mathfrak{B}1}} \right) \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | L \rangle, \end{aligned} \quad (12.3.165)$$

$$\begin{aligned} \langle k | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + i\mathcal{K}'_1, \hat{\eta} \right) | k+1 \rangle &= \mp \exp \left(\frac{\pi i m [4\mathcal{K}_1 - i\mathcal{K}'_1 - 2\hat{v} + 4k\hat{\eta}]}{2\mathcal{U}_{\mathfrak{B}1}} \right) \\ &\times \langle k | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \mp 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k+1 \rangle, \end{aligned} \quad (12.3.166)$$

$$\begin{aligned} \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + i\mathcal{K}'_1, \hat{\eta} \right) | 1 \rangle &= \mp \exp \left(\frac{\pi i m [4\mathcal{K}_1 - i\mathcal{K}'_1 - 2\hat{v}]}{2\mathcal{U}_{\mathfrak{B}1}} \right) \\ &\times \langle 1 | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \mp 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | 1 \rangle, \end{aligned} \quad (12.3.167)$$

$$\begin{aligned} \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + i\mathcal{K}'_1, \hat{\eta} \right) | L \rangle &= \mp (-1)^m \exp \left(\frac{\pi i m [4\mathcal{K}_1 - i\mathcal{K}'_1 - 2\hat{v}]}{2\mathcal{U}_{\mathfrak{B}1}} \right) \\ &\times \langle L | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \mp 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | L \rangle. \end{aligned} \quad (12.3.168)$$

We also have, in the case that $\bar{m}$ is even and $\bar{m}'$ odd,

$$\begin{aligned} \langle k | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \pm 1 \end{smallmatrix}; \hat{v} + \mathcal{O}_{\mathfrak{B}1}, \hat{\eta} \right) | k' \rangle &= (\mp 1)^{\frac{1}{2}\bar{m}} \exp \left(\frac{1}{4} i \pi \bar{m} \bar{m}' \right) \\ &\times \langle k | \mathbf{T}_{\text{r1}} \left(\begin{smallmatrix} \varsigma \\ \mp 1 \end{smallmatrix}; \hat{v}, \hat{\eta} \right) | k' \rangle. \end{aligned} \quad (12.3.169)$$

A term in the trace in (12.3.109) is indexed by a sequence $k_1, k_2, \dots, k_{N_2-1}, k_{N_2}, k_{N_2+1}$, where k_{j+1} is one of $k_j - 1$, k_j or $k_j + 1$ and $k_{N_2+1} = k_1$. Every pair (k_j, k_{j+1}) , where $k_j \neq k_{j+1}$, is matched by another pair (k_i, k_{i+1}) where $k_j = k_{i+1}$ and $k_i = k_{j+1}$. So the pre-factor contributions from the index k for the traces constructed from these matrix elements all cancel. Consider now the different choices for the parities of m , m' and N_2 .

m odd and m' even:

It follows from (18.3.129) that m_0 is even and $\bar{m}$ is odd, but $\bar{m}'$ can be of either parity. From (12.3.76), (12.3.99) and (12.3.162)

$$\mathcal{Q}_R(\hat{v} + 2\mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) = (-1)^{N_2\bar{m}'+1} S_z \mathcal{Q}_R(\hat{v}, \hat{\eta}), \quad (12.3.170)$$

and, from (12.3.98),

$$\mathcal{Q}_L(\hat{v} + 2\mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) = (-1)^{N_2\bar{m}'+1} \mathcal{Q}_L(\hat{v}, \hat{\eta}) S_z. \quad (12.3.171)$$

Applying (12.3.99) to (12.3.170) and (12.3.107), with $E = I$, to (12.3.171)⁶¹ gives

$$\begin{aligned} \mathcal{Q}(\hat{v} + 2\mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) &= (-1)^{N_2\bar{m}'+1} S_z \mathcal{Q}(\hat{v}, \hat{\eta}) \\ &= (-1)^{N_2\bar{m}'+1} \mathcal{Q}(\hat{v}, \hat{\eta}) S_z, \end{aligned} \quad (12.3.172)$$

establishing the commutation relation (12.3.105). From (12.3.163)–(12.3.165),

$$\mathcal{Q}(\hat{v} + 2\mathcal{O}'_{\mathfrak{B}1}, \hat{\eta}) = (-1)^{N_2nm'} \exp\left(-\frac{\pi i N_2\{\mathcal{O}'_{\mathfrak{B}1} + \hat{v}\}}{\mathcal{O}_{\mathfrak{B}1}}\right) (-S_z)^n \mathcal{Q}(\hat{v}, \hat{\eta}). \quad (12.3.173)$$

When $m' = 0$ the modified theta functions reduce to their standard equivalents and, from (18.3.129), (18.3.130), (18.3.133), (18.3.135), (12.3.89) and (12.3.128)

$$m_0 = 2m, \quad \bar{m} = 1, \quad \bar{m}' = 0, \quad n = 0, \quad n' = 1, \quad (12.3.174)$$

$$\begin{aligned} \mathcal{O}_{\mathfrak{B}1} &= \mathcal{K}_1, \quad \mathcal{O}'_{\mathfrak{B}1} = i\mathcal{K}'_1, \quad \mathcal{U}_{\mathfrak{B}1} = 2m\mathcal{K}_1, \\ \hat{\eta} &= m\mathcal{K}_1/L. \end{aligned} \quad (12.3.175)$$

The Eqs. (12.3.172) and (12.3.173) simplify to

$$\mathcal{Q}(\hat{v} + 2\mathcal{K}_1, \hat{\eta}) = -S_z \mathcal{Q}(\hat{v}, \hat{\eta}) = -\mathcal{Q}(\hat{v}, \hat{\eta}) S_z, \quad (12.3.176)$$

$$\mathcal{Q}(\hat{v} + 2i\mathcal{K}'_1, \hat{\eta}) = \frac{1}{\mathcal{Q}_1^{N_2}} \exp\left(-\frac{\pi i N_2 \hat{v}}{\mathcal{K}_1}\right) \mathcal{Q}(\hat{v}, \hat{\eta}) \quad (12.3.177)$$

and if, in addition, N_2 is even, from (12.3.166)–(12.3.168),

$$\mathcal{Q}(\hat{v} + i\mathcal{K}'_1, \hat{\eta}) = -\frac{1}{\mathcal{Q}_1^{\frac{1}{4}N_2}} \exp\left(-\frac{\pi i N_2 \hat{v}}{2\mathcal{K}_1}\right) S_z S_x \mathcal{Q}(\hat{v}, \hat{\eta}). \quad (12.3.178)$$

m even and m' odd:

It follows from (18.3.129) that m_0 and $\bar{m}'$ are odd and $\bar{m}$ is divisible by four. The analysis is the same as that when m is odd and m' even except that (12.3.169) is used in place of (12.3.162) to give

⁶¹ Valid since the interchange relationship holds with $E = I$, when m is odd and m' even.

$$\begin{aligned}
\mathcal{Q}(\hat{v} + \mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) &= (-1)^{\frac{1}{4}N_2\bar{m}\bar{m}'} S_x \mathcal{Q}(\hat{v}, \hat{\eta}) \\
&= (-1)^{\frac{1}{4}N_2\bar{m}\bar{m}'} \mathcal{Q}(\hat{v}, \hat{\eta}) S_x,
\end{aligned} \tag{12.3.179}$$

establishing the commutation relation (12.3.104). Because of the factor $(-1)^{m_0}$ in (12.3.165) the period of quasi-periodicity is now doubled with respect to $\mathcal{O}'_{\mathfrak{B}1}$ giving

$$\mathcal{Q}(\hat{v} + 4\mathcal{O}'_{\mathfrak{B}1}, \hat{\eta}) = \exp\left(\frac{-2i\pi N_2\{\mathcal{O}'_{\mathfrak{B}1} + \hat{v}\}}{\mathcal{O}_{\mathfrak{B}1}}\right) \mathcal{Q}(\hat{v}, \hat{\eta}). \tag{12.3.180}$$

m and m' odd:

It follows from (18.3.129) that m_0 and $\bar{m}'$ are odd and $\bar{m}$ is even but not divisible by four, so (12.3.180) remains valid. In place of (12.3.179) we first have

$$\mathcal{Q}_R(\hat{v} + \mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) = \exp\left(\frac{1}{4}N_2i\pi\bar{m}\bar{m}'\right) S_x S_z \mathcal{Q}_R(\hat{v}, \hat{\eta}), \tag{12.3.181}$$

$$\mathcal{Q}_L(\hat{v} + \mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) = \exp\left(\frac{1}{4}N_2i\pi\bar{m}\bar{m}'\right) \mathcal{Q}_L(\hat{v}, \hat{\eta}) S_z S_x. \tag{12.3.182}$$

The application of (12.3.99) to (12.3.181) is as in the previous cases, but now, according to the results of Fabricius and McCoy (2007), (12.3.107) becomes

$$S_x \mathcal{Q}(\hat{v}, \hat{\eta}) S_x = S_z \mathcal{Q}(\hat{v}, \hat{\eta}) S_z = \mathcal{Q}_L^{-1}(\hat{v}_0, \hat{\eta}) \mathcal{Q}_L(\hat{v}, \hat{\eta}). \tag{12.3.183}$$

Applying this to (12.3.182) gives

$$\begin{aligned}
\mathcal{Q}(\hat{v} + \mathcal{O}_{\mathfrak{B}1}, \hat{\eta}) &= \exp\left(\frac{1}{4}N_2i\pi\bar{m}\bar{m}'\right) S_x S_z \mathcal{Q}(\hat{v}, \hat{\eta}) \\
&= \exp\left(\frac{1}{4}N_2i\pi\bar{m}\bar{m}'\right) \mathcal{Q}(\hat{v}, \hat{\eta}) S_x S_z,
\end{aligned} \tag{12.3.184}$$

establishing the commutation relation (12.3.106).

12.3.6.5 The Eigenvalues of $\mathcal{Q}$ and $\hat{V}$

For simplicity we concentrate on the case m odd and $m' = 0$ when (12.3.174) and (12.3.175) apply. For accounts of the other parity choices, including the case $\hat{\eta}$ complex ($m' \neq 0$), the reader is referred to Fabricius and McCoy (2007) or McCoy (2010, Sect. 14.3).

The matrix $\mathcal{Q}(\hat{v}, \hat{\eta})$ with periodicity properties (12.3.170) and (12.3.173) is not unique, since it commutes with S_z but not with S_x and $S_x S_z$. Of course, since S_z is diagonal, with elements $\pi(\zeta^{(k)}) = \pm 1$, we know its eigenvalues. However, we need to associate these values with the eigenvalues of $\mathcal{Q}(\hat{v}, \hat{\eta})$. Suppose that $\Lambda_Q(\hat{v}, \hat{\eta}|k_1)$

and θ_z are associated eigenvalues of $\mathcal{Q}(\hat{v}, \hat{\eta})$ and $\mathcal{S}_z$.⁶² Then, from (12.3.172) and (12.3.173),

$$\Lambda_Q(\hat{v} + 2\mathcal{K}_1, \hat{\eta}|\mathbf{k}_1) = -\theta_z \Lambda_Q(\hat{v}, \hat{\eta}|\mathbf{k}_1), \quad (12.3.185)$$

$$\Lambda_Q(\hat{v} + 2i\mathcal{K}'_1, \hat{\eta}|\mathbf{k}_1) = \frac{1}{Q_1^{N_2}} \exp\left(\frac{-i\pi N_2 \hat{v}}{\mathcal{K}_1}\right) \Lambda_Q(\hat{v}, \hat{\eta}|\mathbf{k}_1). \quad (12.3.186)$$

According to the definition in Sect. 18.2, $\Lambda_Q(\hat{v}, \hat{\eta}|\mathbf{k}_1)$ is a quasi-periodic function in the class $\text{QP}(2\mathcal{K}_1, 2i\mathcal{K}'_1; -\theta_z, 1; N_2, N_2)$. From Theorem 18.2.5 it has N_2 zeros in the period parallelogram $\mathbb{W}(2\mathcal{K}_1, 2i\mathcal{K}'_1)$. Let these be $\hat{v}_j, j = 1, 2, \dots, N_2$. Then, from Theorem 18.3.2, an eigenvalue of $\mathcal{Q}(\hat{v}, \hat{\eta})$, identified by a constant $\aleph$ and the Q -zero set $\{\hat{v}_j\}$, is given, to within a multiplicative constant, by

$$\Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) = \exp\left(-\frac{\pi i \aleph \hat{v}}{2\mathcal{K}_1}\right) \prod_{j=1}^{N_2} H(\hat{v} - \hat{v}_j|\mathbf{k}_1), \quad (12.3.187)$$

where

$$\aleph = 2\ell' + \frac{1}{2}(1 - \theta_z) + 1 + N_2, \quad (12.3.188)$$

$$\sum_{j=1}^{N_2} \hat{v}_j = \aleph i \mathcal{K}'_1 + (N_2 - 2\ell)\mathcal{K}_1, \quad (12.3.189)$$

for some integers ℓ and ℓ' . There are thus four parameters determining the Q -zero set, $N_2, \aleph, \mathbf{k}_1$ and m/L .

The corresponding eigenvalue $\hat{\Lambda}_V(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1)$ is now obtained by substituting from (12.3.187) into (12.3.103) to give

$$\hat{\Lambda}_V(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}) = F(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) G(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1), \quad (12.3.190)$$

where

$$\begin{aligned} F(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) &:= \exp\left(-\frac{\pi i \aleph \hat{\eta}}{\mathcal{K}_1}\right) [h(\hat{v} + \hat{\eta}|\mathbf{k}_1)]^{N_2} \\ &\times \prod_{j=1}^{N_2} \frac{H(\hat{v} + 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}{H(\hat{v} - \hat{v}_j|\mathbf{k}_1)}, \end{aligned} \quad (12.3.191)$$

⁶² Note that θ_z satisfies the same relationship (18.2.5) as τ .

$$\begin{aligned}
G(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) &:= \left[\frac{h(\hat{v} - \hat{\eta}|\mathbf{k}_1)}{h(\hat{v} + \hat{\eta}|\mathbf{k}_1)} \right]^{N_2} + \exp \left(\frac{2\pi i \aleph \hat{\eta}}{\mathcal{K}_1} \right) \\
&\times \prod_{j=1}^{N_2} \frac{H(\hat{v} - 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}{H(\hat{v} + 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}. \tag{12.3.192}
\end{aligned}$$

$F(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1)$ is a meromorphic function with poles at all members of the Q -zero set not eliminated by cancellation in the numerator and denominator of the product. Since $\hat{\Lambda}_v(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1)$ is an entire function every pole $\hat{v}_j$ of $F(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1)$ must satisfy the equation⁶³

$$G(\hat{v}_j, \hat{\eta}; \aleph, \{\hat{v}_i\}|\mathbf{k}_1) = 0, \tag{12.3.193}$$

which, following Fabricius and McCoy, we refer to as the *Bethe equation*. Apart from eliminations arising from complete L -strings as described below, the solutions, for fixed $\aleph$, of this set of N_2 variables would allow the determination of the complete set of eigenvalues of $\hat{V}(\hat{v}, \hat{\eta})$. In fact, since we are concerned here only with determining the free-energy density, which is given by the largest eigenvalue, the task is simplified by an investigation of the properties of the Q -zero set $\{\hat{v}_j\}$.

12.3.6.6 Properties of the Q -Zero Set: Strings

Suppose that, for some $\hat{v}_\mu$,

$$\begin{aligned}
\hat{v}_{\mu+k} &= \hat{v}_\mu - 2k\hat{\eta} + \varepsilon_k = \hat{v}_{\mu+k-1} - 2\hat{\eta} + \varepsilon_k - \varepsilon_{k-1}, \\
k &= 1, 2, \dots, \nu - 1, \quad \text{for } \nu < L, \tag{12.3.194}
\end{aligned}$$

where the ε_k are small. Such a set $\hat{v}_\mu, \dots, \hat{v}_{\mu+\nu-1}$ is called a ν -string and

$$\hat{v}_{\mu+k} + 2\hat{\eta} = \hat{v}_{\mu+k-2} - 2\hat{\eta} + \varepsilon_k - \varepsilon_{k-2}, \quad k = 2, 3, \dots, \nu - 1. \tag{12.3.195}$$

The product in (12.3.191) contains terms

⁶³ If the root $\hat{v}_j$ is of multiplicity s then

$$\left(\frac{\partial^k G(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_i\}|\mathbf{k}_1)}{\partial \hat{v}^k} \right)_{v=\hat{v}_j} = 0, \quad k = 0, 1, \dots, s-1.$$

Such solutions would exist, if at all, only for special values of $\hat{\eta}$. They have not been found for small values of N_2 , when $m' = 0$ (McCoy 2010, Sect. 14.3.2) and we shall not explore this possibility.

$$\frac{H(\hat{v} + 2\hat{\eta} - \hat{v}_{\mu+k-1}|\mathbf{k}_1)}{H(\hat{v} - \hat{v}_{\mu+k}|\mathbf{k}_1)} = \frac{H(\hat{v} - \hat{v}_{\mu+k} - \varepsilon_k + \varepsilon_{k-1}|\mathbf{k}_1)}{H(\hat{v} - \hat{v}_{\mu+k}|\mathbf{k}_1)},$$

$$k = 1, 2, \dots, \nu - 1, \quad (12.3.196)$$

and that in (12.3.192) the terms

$$\frac{H(\hat{v} - 2\hat{\eta} - \hat{v}_{\mu+k}|\mathbf{k}_1)}{H(\hat{v} + 2\hat{\eta} - \hat{v}_{\mu+k-2}|\mathbf{k}_1)} = \frac{H(\hat{v} - 2\hat{\eta} - \hat{v}_{\mu+k}|\mathbf{k}_1)}{H(\hat{v} - 2\hat{\eta} - \hat{v}_{\mu+k} - \varepsilon_k + \varepsilon_{k-2}|\mathbf{k}_1)},$$

$$k = 2, 3, \dots, \nu - 1. \quad (12.3.197)$$

In each case these terms ‘almost’ cancel. Although, of course, this would not, even if all the ε_k were zero, eliminate the string-zeros from the product. In (12.3.191) the denominator retains the term in $\hat{v}_\mu$ and the numerator the term in $\hat{v}_{\mu+\nu-1}$. In (12.3.192) the denominator retains the terms in $\hat{v}_{\mu+\nu-1}$ and $\hat{v}_{\mu+\nu-2}$ and the numerator the terms in $\hat{v}_\mu$ and $\hat{v}_{\mu+1}$.

Consider, however, the case where all the ε_k are zero and $\nu = L$ with

$$\hat{v}_{\mu+k} = \hat{v}_\mu - 2k\hat{\eta}, \quad k = 0, 1, \dots, L - 1. \quad (12.3.198)$$

This is called a *complete L-string* and $\hat{v}_\mu$ is its *centre*.⁶⁴ Suppose that there are N_{21} complete L -strings with centres, $\hat{v}_{\mu_s}$ $s = 1, 2, \dots, N_{21}$ with $N_{22} := N_2 - LN_{21}$. Then, from (12.3.189), (12.3.191) and (12.3.192),

$$\sum_{j=1}^{N_{22}} \hat{v}_j = \aleph i \mathcal{K}'_1 + [N_2 - 2\ell - mN_{21}(L - 1)]\mathcal{K}_1 - L \sum_{s=1}^{N_{21}} \hat{v}_{\mu_s}, \quad (12.3.199)$$

$$F(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) = (-1)^{N_{21}} \exp\left(-\frac{\pi i \aleph \hat{\eta}}{\mathcal{K}_1}\right) [h(\hat{v} + \hat{\eta}|\mathbf{k}_1)]^{N_2}$$

$$\times \prod_{j=1}^{N_{22}} \frac{H(\hat{v} + 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}{H(\hat{v} - \hat{v}_j|\mathbf{k}_1)}, \quad (12.3.200)$$

$$G(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|\mathbf{k}_1) = \left[\frac{h(\hat{v} - \hat{\eta}|\mathbf{k}_1)}{h(\hat{v} + \hat{\eta}|\mathbf{k}_1)}\right]^{N_2} + \exp\left(\frac{2\pi i \aleph \hat{\eta}}{\mathcal{K}_1}\right)$$

$$\times \prod_{j=1}^{N_{22}} \frac{H(\hat{v} - 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}{H(\hat{v} + 2\hat{\eta} - \hat{v}_j|\mathbf{k}_1)}, \quad (12.3.201)$$

where $\{\hat{v}_j\}$ is the reduced Q -zero set $\hat{v}_j, j = 1, 2, \dots, N_{22}$.

⁶⁴ Although, since, from (12.3.174), $2\hat{\eta}L = 2m\mathcal{K}_1$, it is clear that any member of the string can equally well be taken as its centre.

It is clear that the centres of the L -strings cannot be determined from (12.3.193), which is now reduced to

$$G(\hat{v}_j, \hat{\eta}; \aleph, \overline{\{\hat{v}_i\}}|k_1) = 0. \quad (12.3.202)$$

A method for doing this has been conjectured by Fabricius and McCoy (2003) and proved for $L = 2$ by Fabricius and McCoy (2004). These authors also derived numerical results for $L = 3$ and various values of N_2 (see Fig. 14.10, p. 522, McCoy 2010, for the case $N_2 = 8$). It was shown that:

- (a) The number of L -strings N_{21} is even, so that the factors $mN_{21}(L-1)$ and $(-1)^{N_{21}}$ are removed from (12.3.199) and (12.3.200) respectively.
- (b) For every configuration of string centres $\{\hat{v}_{\mu_s}\}$ there is another obtained by adding or not adding $i\mathcal{K}_1$ to a member of this set, giving $2^{N_{21}}$ solutions in all.

The transformation $\aleph \rightarrow \aleph + L$ and $\hat{v}_{\mu_s} \rightarrow \hat{v}_{\mu_s} + i\mathcal{K}'_1$, for one $\hat{v}_{\mu_s}$, leaves (12.3.199) and (12.3.201) unaffected but changes the sign in (12.3.200) and hence of $\hat{\Lambda}_V(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\})$. The presence of the L -strings implies the existence of a pair of eigenvalues of equal magnitude, opposite sign and degeneracy $2^{N_{21}-1}$. From (12.3.188) it follows that, the two degenerate eigenvalues have the same and different parities with respect to S_z according as L is even or odd. From (12.3.187)

$$\begin{aligned} \Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_{\mu_s}\}, \overline{\{\hat{v}_j\}}|k_1) &= \exp\left(-\frac{\pi i \aleph \hat{v}}{2\mathcal{K}_1}\right) L(\hat{v}, \hat{\eta}; \{\hat{v}_{\mu_s}\}|k_1) \\ &\times \prod_{j=1}^{N_{22}} H(\hat{v} - \hat{v}_j|k_1), \end{aligned} \quad (12.3.203)$$

where

$$L(\hat{v}, \hat{\eta}; \{\hat{v}_{\mu_s}\}|k_1) := \prod_{s=1}^{N_{21}} \prod_{k=0}^{L-1} H(\hat{v} - \hat{v}_{\mu_s+k}|k_1) = \prod_{s=1}^{N_{21}} \prod_{k=1}^{L-1} H(\hat{v} - \hat{v}_{\mu_s} + 2k\hat{\eta}|k_1). \quad (12.3.204)$$

Since

$$L(\hat{v}, \hat{\eta}; \{\hat{v}_{\mu_s}\}|k_1) = L(\hat{v} \pm 2\hat{\eta}, \hat{\eta}; \{\hat{v}_{\mu_s}\}|k_1) = L(\hat{v}, \hat{\eta}; \{\hat{v}_{\mu_s} \pm 2\hat{\eta}\}|k_1), \quad (12.3.205)$$

this term cancels in (12.3.203), where it represents the degeneracy of the eigenvalue of $\hat{V}(\hat{v}, \hat{\eta})$.

12.3.6.7 Properties of the Q -Zero Set: Bethe Pairs

The analysis in this section is independent of the presence or absence of complete L -strings. In the latter case $\{\hat{v}_{\mu_s}\} = \emptyset$, $\{\hat{v}_j\} = \{\hat{v}_j\}$ and $N_{22} = N_2$. So, supposing the possible presence of complete L -strings, (12.3.190) is replaced by

$$\hat{\Lambda}_v(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) = F(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) G(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1). \quad (12.3.206)$$

From (18.3.25), (18.3.158), (12.3.199)–(12.3.201),⁶⁵

$$h(\hat{v} + i\mathcal{K}'_1|k_1) = -\frac{1}{\sqrt{Q_1}} \exp\left(-\frac{\pi i \hat{v}}{\mathcal{K}_1}\right) h(\hat{v}|k_1), \quad (12.3.207)$$

$$\begin{aligned} F(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) &= \left(-\frac{1}{\sqrt{Q_1}}\right)^{N_2} \exp\left(-\frac{\pi i N_2[\hat{v} + 2\hat{\eta}]}{\mathcal{K}_1}\right) \\ &\times F(\hat{v}, \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_j + i\mathcal{K}'_1\}|k_1), \end{aligned} \quad (12.3.208)$$

$$\begin{aligned} G(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) &= \exp\left(\frac{2\pi i N_2 \hat{\eta}}{\mathcal{K}_1}\right) \\ &\times G(\hat{v}, \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_j + i\mathcal{K}'_1\}|k_1), \end{aligned} \quad (12.3.209)$$

$$\begin{aligned} \hat{\Lambda}_v(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) &= \left(-\frac{1}{\sqrt{Q_1}}\right)^{N_2} \exp\left(-\frac{\pi i N_2 \hat{v}}{\mathcal{K}_1}\right) \\ &\times \hat{\Lambda}_v(\hat{v}, \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_j + i\mathcal{K}'_1\}|k_1). \end{aligned} \quad (12.3.210)$$

But, directly from (12.3.92),

$$\begin{aligned} \hat{\Lambda}_v(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1) &= \left(-\frac{1}{\sqrt{Q_1}}\right)^{N_2} \exp\left(-\frac{\pi i N_2 \hat{v}}{\mathcal{K}_1}\right) \\ &\times \hat{\Lambda}_v(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1). \end{aligned} \quad (12.3.211)$$

So

$$\hat{\Lambda}_v(\hat{v}, \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_j + i\mathcal{K}'_1\}|k_1) = \hat{\Lambda}_v(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|k_1). \quad (12.3.212)$$

⁶⁵ Since

$$\exp\left(\frac{2\pi i N_2 \hat{\eta}}{\mathcal{K}_1}\right) = \exp\left(\frac{2\pi i N_{22} \hat{\eta}}{\mathcal{K}_1}\right)$$

the form of these equations is unaffected by the presence of complete L -strings. In particular N_2 could be replaced by N_{22} in (12.3.214). Incomplete v -strings, $v < L$, are included in the set $\{\hat{v}_j\}$.

It follows that

$$\mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) := C \Lambda_Q(\hat{v}; \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_{\mu_s}\}, \overline{\{\hat{v}_j + i\mathcal{K}'_1\}}|\mathbf{k}_1) \quad (12.3.213)$$

satisfies (12.3.103) with the same $\hat{\Lambda}_V(\hat{v}, \hat{\eta}; \mathfrak{N}, \overline{\{\hat{v}_j\}}|\mathbf{k}_1)$, for any $\hat{v}$ -independent quantity C . From (12.3.203)

$$\begin{aligned} \exp\left(\frac{\pi i N_2 \hat{v}}{2\mathcal{K}_1}\right) \Lambda_Q(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) &= (-1)^{N_2} Q_1^{\frac{1}{2}\mathfrak{N}} \\ &\times \Lambda_Q(\hat{v}, \hat{\eta}; \mathfrak{N} + N_2, \{\hat{v}_{\mu_s}\}, \overline{\{\hat{v}_j + i\mathcal{K}'_1\}}|\mathbf{k}_1) \end{aligned} \quad (12.3.214)$$

and, from (12.3.185) and (12.3.186),

$$\Lambda_Q(\hat{v} + 2\mathcal{K}_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) = -\theta_z \Lambda_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1), \quad (12.3.215)$$

$$\Lambda_Q(\hat{v} + 2i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) = \frac{1}{Q_1^{N_2}} \exp\left(\frac{-\pi i N_2 \hat{v}}{\mathcal{K}_1}\right) \Lambda_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1), \quad (12.3.216)$$

giving, from (12.3.213) and (12.3.214),

$$\mathring{\Lambda}_Q(\hat{v} + 2\mathcal{K}_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) = -\theta_z (-1)^{N_2} \mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1), \quad (12.3.217)$$

$$\mathring{\Lambda}_Q(\hat{v} + 2i\mathcal{K}'_1, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1) = \frac{1}{Q_1^{N_2}} \exp\left(\frac{-\pi i N_2 \hat{v}}{\mathcal{K}_1}\right) \mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1). \quad (12.3.218)$$

Remembering that N_2 and N_{22} have the same parity:

For N_2 odd:

$\Lambda_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1)$ and $\mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \mathfrak{N}, \{\hat{v}_j\}|\mathbf{k}_1)$ with the same complete L -strings and remaining zeros $\overline{\{\hat{v}_j\}}$ and $\overline{\{\hat{v}_j + i\mathcal{K}'_1\}}$, respectively, are distinct since they are associated with opposite eigenvalues of S_z . They are, however, associated with the same eigenvalue of $\widehat{V}(\hat{v}, \hat{\eta})$, which must be (at least) doubly degenerate, in addition to the degeneracy associated with complete L -strings.

For N_2 even:

Either

- (a) $\Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1)$ and $\mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1)$ are distinct eigenvalues of $\mathcal{Q}(\hat{v}, \hat{\eta})$ both associated with the same eigenvalue $\hat{\Lambda}_V(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1)$ of $\hat{V}(\hat{v}, \hat{\eta})$.
- (b) $\Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1) = \mathring{\Lambda}_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1)$. Then, from (12.3.213) and (12.3.214),

$$\begin{aligned} \Lambda_Q(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1) &= C \exp\left(-\frac{\pi i N_2 \hat{v}}{2\mathcal{K}_1}\right) Q_1^{\frac{1}{2}\aleph} \\ &\times \Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1). \end{aligned} \quad (12.3.219)$$

From (12.3.178) we see that an eigenvalue relationship of this form applies only when the corresponding eigenvector is an eigenvector of $\mathbf{S}_x$, when it take the form

$$\begin{aligned} \Lambda_Q(\hat{v} + i\mathcal{K}'_1, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1) &= -\exp\left(-\frac{\pi i N_2 \hat{v}}{2\mathcal{K}_1}\right) \frac{\theta_z \theta_x}{Q_1^{\frac{1}{4}N_2}} \\ &\times \Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1), \end{aligned} \quad (12.3.220)$$

allowing the identification of C .

The zeros of $\Lambda_Q(\hat{v}, \hat{\eta}; \aleph, \{\hat{v}_j\}|k_1)$ are in pairs. These *Bethe pairs* $\{\hat{v}_j, \hat{v}_j + i\mathcal{K}'_1\}$ are solutions of (12.3.202) and numerical studies indicate (see, McCoy 2010, Sect. 14.3.2) that a pair either belongs to two ν -strings ($\nu < L$) or has $\Re\{\hat{v}_j\} = 0$ or $\mathcal{K}_1$. In the case of the largest eigenvalue with N_2 even, the Q -zero set consists entirely of Bethe pairs with $\Re\{\hat{v}_j\} = \mathcal{K}_1$, so we set

$$\hat{v} := \mathcal{K}_1 + i q. \quad (12.3.221)$$

12.3.6.8 The Largest Eigenvalue of $\hat{V}(\hat{v}, \hat{\eta})$ for N_2 Even

The analysis of Sect. 12.1.5 needed the condition that N_2 is even. Although this is no longer necessary we confine discussion to this case.⁶⁶

Let $N_{23} := \frac{1}{2}N_2$ and, with a suitable numbering of the roots, $q_j \in [-\frac{1}{2}\mathcal{K}'_1, \frac{1}{2}\mathcal{K}'_1)$, for $j = 1, 2, \dots, N_{23}$. Of these values suppose that there are $N_{23}^{(-)}$ in the interval

⁶⁶ For treatments of N_2 odd the reader is referred to Fabricius and McCoy (2005) and McCoy (2010, Chap. 14).

$[-\frac{1}{2}\mathcal{K}'_1, 0)$ and $N_{23}^{(+)}$ in the interval $[0, \frac{1}{2}\mathcal{K}'_1)$. Then there will be $N_{23}^{(+)}$ roots of the form $q_j - i\mathcal{K}'_1$ in the interval $[-\mathcal{K}'_1, -\frac{1}{2}\mathcal{K}'_1)$ and $N_{23}^{(-)}$ roots of the form $q_j + i\mathcal{K}'_1$ in the interval $[\frac{1}{2}\mathcal{K}'_1, \mathcal{K}'_1)$. From (12.3.189), with $\ell = 0$,

$$\sum_{j=1}^{N_{23}} q_j = \mathcal{K}'_1 [\aleph + N_{23}^{(+)} - N_{23}^{(-)}]. \quad (12.3.222)$$

Since the largest eigenvalue is real the roots must be in conjugate-complex pairs or be real. Numerical studies indicate that real roots occur only when N_{23} is odd when there is exactly one such. We shall assume that this is generally true, giving for N_{23}

(i) even $N_{23}^{(+)} = N_{23}^{(-)}$ and $\aleph = 0$,

(ii) odd $N_{23}^{(+)} = N_{23}^{(-)} + 1$ and $\aleph = -1$.

In each case, from (18.3.156)–(18.3.158), (12.3.191) and (12.3.192) become

$$F(q, \hat{\eta}; \{q_i\} | k_1) = [\hbar(\hat{\eta} + \mathcal{K}_1 + i q | k_1)]^{2N_{23}} \prod_{i=1}^{N_{23}} \frac{\hbar(2\hat{\eta} + i q - i q_i | k_1)}{\hbar(i q - i q_i | k_1)}, \quad (12.3.223)$$

$$G(q, \hat{\eta}; \{q_i\} | k_1) = \left[\frac{\hbar(\hat{\eta} + \mathcal{K}_1 - i q | k_1)}{\hbar(\hat{\eta} + \mathcal{K}_1 + i q | k_1)} \right]^{2N_{23}} + (-1)^{N_{23}} \prod_{i=1}^{N_{23}} \frac{\hbar(2\hat{\eta} - i q + i q_i | k_1)}{\hbar(2\hat{\eta} + i q - i q_i | k_1)}. \quad (12.3.224)$$

The members of the Q -zero set $\{q_j\}$ are solutions of the N_{23} Bethe equations $G_1(q_j, \hat{\eta}; \{q_i\}) = 0$, $j = 1, 2, \dots, N_{23}$. The next step is to take the logarithms of these equations. This will involve

$$\ln[(-1)^{N_{23}+1}] = 2\pi i [N_{23} + 1] \left[p + \frac{1}{2} \right], \quad (12.3.225)$$

where the *branch number* p can take any integer value. Let p_j be the branch number associated with q_j , that is p is a function of q with

$$I(q) := [N_{23} + 1] \left[p(q) + \frac{1}{2} \right] \quad (12.3.226)$$

an integer if N_{23} is odd and an integer plus a half if N_{23} is even. The Bethe equations take the form

$$2N_{23} f(\hat{\eta}, q_j | k_1) = 2\pi i I(q_j) + \sum_{i=1}^{N_{23}} f(2\hat{\eta} - \mathcal{K}_1, q_j - q_i | k_1), \quad (12.3.227)$$

where $f(v, u|k)$, given by (18.3.160), consists of logarithms which are purely imaginary, lie in the interval $[-\pi i, \pi i]$ and are zero when $v = 0$.

In the limit $N_{23} \rightarrow \infty$ the values $q_j, j = 1, 2, \dots, N_{23}$ are dense on the interval $[-\frac{1}{2}\mathcal{K}'_1, \frac{1}{2}\mathcal{K}'_1]$ with a normalized density function $\rho(q)$. The aim now is to use (12.3.226) and $f(\hat{\eta}, q_j|k_1)$ to determine this density function and for this it is necessary to suppose a distribution for the members of the Q -zero set. A form for this has not been obtained in general, but it was shown by Yang and Yang (1966b), for the XXZ Heisenberg chain, and hence for the six-vertex model, that, in the limit of large N_{23} the number of members of the Q -zero set in the interval $[q, q + dq]$ is $I(q + dq) - I(q)$. We suppose that this result also applies to the eight-vertex model, and, from (12.3.226),⁶⁷

$$\frac{1}{N_{23}} \int_{-\frac{1}{2}\mathcal{K}'_1}^{\frac{1}{2}\mathcal{K}'_1} \dot{I}(q) dq = \int_{-\frac{1}{2}\mathcal{K}'_1}^{\frac{1}{2}\mathcal{K}'_1} \dot{\rho}(q) dq. \quad (12.3.228)$$

So $\dot{\rho}(q)$ is the normalized density function. Differentiating (12.3.227) and taking the limit $N_{23} \rightarrow \infty$, when the sum becomes an integral, gives

$$2\dot{f}(\hat{\eta}, q|k_1) = 2\pi i \rho(q) + \int_{-\frac{1}{2}\mathcal{K}'_1}^{\frac{1}{2}\mathcal{K}'_1} \rho(x) \dot{f}(2\hat{\eta} - \mathcal{K}_1, q - x|k_1) dx. \quad (12.3.229)$$

Taking the Fourier transform of (12.3.228) and using the convolution formula (18.1.22)

$$2\dot{f}^*(\hat{\eta}, m|k_1) = 2\pi i \rho^*(m) + \rho^*(m) \mathcal{K}'_1 \dot{f}^*(2\hat{\eta} - \mathcal{K}_1, m|k_1) \quad (12.3.230)$$

and, substituting from (18.3.162), gives

$$\rho^*(m) = \frac{1}{\mathcal{K}'_1 \cosh(2m\pi[\hat{\eta} - \mathcal{K}_1]/\mathcal{K}'_1)}. \quad (12.3.231)$$

Reverting to the variable $\hat{v}$, from (12.3.190), (12.3.223) and (12.3.224),

$$\hat{\Lambda}_v(\hat{v}, \hat{\eta}; \{q_j\}|k_1) = X(\hat{v}, \hat{\eta}; \{q_i\}|k_1) + X(\hat{v}, -\hat{\eta}; \{q_i\}|k_1), \quad (12.3.232)$$

where

$$X(\hat{v}, \hat{\eta}; \{q_i\}|k_1) = [h(\hat{\eta} + \hat{v}|k_1)]^{2N_{23}} \prod_{i=1}^{N_{23}} \frac{h(i q_i + 2\hat{\eta} - \hat{v} + \mathcal{K}_1|k_1)}{h(i q_i - \hat{v} + \mathcal{K}_1|k_1)}. \quad (12.3.233)$$

In the limit $N_2 \rightarrow \infty$, using (18.1.8),

⁶⁷ Here a dot is used to denote differentiation with respect to q .

$$\frac{\ln[X(\hat{v}, \hat{\eta}|\mathbf{k}_1)]}{N_2} \simeq Y(\hat{v}, \hat{\eta}|\mathbf{k}_1), \quad (12.3.234)$$

where

$$\begin{aligned} Y(\hat{v}, \hat{\eta}|\mathbf{k}_1) &= \ln[h(\hat{\eta} + \hat{v}|\mathbf{k}_1)] + \frac{1}{2} \int_{-\frac{1}{2}\mathcal{K}'_1}^{\frac{1}{2}\mathcal{K}'_1} \rho(x) f(ix + \hat{\eta} - \hat{v}; i\hat{\eta}|\mathbf{k}_1) dx \\ &= \ln[h(\hat{\eta} + \hat{v}|\mathbf{k}_1)] \\ &\quad + \frac{1}{2} \sum_{m=-\infty}^{\infty} \rho^*(m) \int_{-\frac{1}{2}\mathcal{K}'_1}^{\frac{1}{2}\mathcal{K}'_1} \exp\left(\frac{2m\pi i x}{\mathcal{K}'_1}\right) f(ix + \hat{\eta} - \hat{v}; i\hat{\eta}|\mathbf{k}_1) dx. \end{aligned} \quad (12.3.235)$$

Inserting (12.3.231) and the series expansion from (18.3.160) gives

$$\begin{aligned} Y(\hat{v}, \hat{\eta}|\mathbf{k}_1) &= \ln[h(\hat{\eta} + \hat{v}|\mathbf{k}_1)] + \frac{\pi \hat{\eta}(\hat{v} - \hat{\eta})}{\mathcal{K}_1 \mathcal{K}'_1} \\ &\quad + \sum_{n=1}^{\infty} \frac{\sinh(2n\pi \hat{\eta}/\mathcal{K}'_1) \sinh(2n\pi [\hat{v} - \hat{\eta}]/\mathcal{K}'_1)}{n \cosh(2n\pi [\mathcal{K}_1 - \hat{\eta}]/\mathcal{K}'_1) \sinh(2n\pi \mathcal{K}_1/\mathcal{K}'_1)}. \end{aligned} \quad (12.3.236)$$

So

$$\hat{\Lambda}_V(\hat{v}, \hat{\eta}|\mathbf{k}_1) \simeq \exp[N_2 Y(\hat{v}, \hat{\eta}|\mathbf{k}_1)] + \exp[N_2 Y(\hat{v}, -\hat{\eta}|\mathbf{k}_1|\mathbf{k}_1)]. \quad (12.3.237)$$

12.3.7 The Free Energy and Magnetization

For $0 < \hat{\eta} \leq \hat{v} \leq 2\mathcal{K}_1$ it can be verified that $Y(\hat{v}, -\hat{\eta}|\mathbf{k}_1) < Y(\hat{v}, \hat{\eta}|\mathbf{k}_1)$ and thus, in the limit $N_2 \rightarrow \infty$, the first term in (12.3.237) is dominant giving, from (11.1.14), the free-energy density

$$\phi^{(8\text{vm})}(\mathcal{N}^{\otimes}, \hat{v}, \hat{\eta}|\mathbf{k}_1) = -\ln[h(\hat{\eta} + \hat{v}|\mathbf{k}_1)] - \frac{\pi \hat{\eta}(\hat{v} - \hat{\eta})}{\mathcal{K}_1 \mathcal{K}'_1} - \sum_{n=1}^{\infty} S_n^{(1)}(\hat{v}, \hat{\eta}|\mathbf{k}_1), \quad (12.3.238)$$

where

$$S_n^{(1)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) := \frac{\sinh\left(\frac{2n\pi \hat{\eta}}{\mathcal{K}'_1}\right) \sinh\left(\frac{2n\pi [\hat{v} - \hat{\eta}]}{\mathcal{K}'_1}\right)}{n \cosh\left(\frac{2n\pi [\mathcal{K}_1 - \hat{\eta}]}{\mathcal{K}'_1}\right) \sinh\left(\frac{2n\pi \mathcal{K}_1}{\mathcal{K}'_1}\right)}. \quad (12.3.239)$$

Translated into our notation the form for the free-energy density derived by Baxter (1972b, p. 202) is

$$\phi_{\text{Bx}}^{(8\text{vM})}(\mathcal{N}^{\otimes}, \hat{v}, \hat{\eta}|\mathbf{k}_1) = -\ln[\hat{c}(\hat{v}, \hat{\eta}|\mathbf{k}_1)] - \sum_{n=1}^{\infty} S_n^{(2)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) \quad (12.3.240)$$

where

$$S_n^{(2)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) := \frac{2 \sinh^2 \left(\frac{n\pi i[2\hat{\eta} - i\mathcal{K}'_1]}{2\mathcal{K}_1} \right) \left[\cosh \left(\frac{n\pi i\hat{\eta}}{\mathcal{K}_1} \right) - \cosh \left(\frac{n\pi i\hat{v}}{\mathcal{K}_1} \right) \right]}{n \sinh \left(\frac{n\pi \mathcal{K}'_1}{\mathcal{K}_1} \right) \cosh \left(\frac{n\pi i\hat{\eta}}{\mathcal{K}_1} \right)}. \quad (12.3.241)$$

In the Bazhanov–Mangazeev transformation, defined in Sect. 18.3.4, the elliptic modulus $\mathbf{k}_1$ is replaced by $\mathbf{k}_3$ and the variables $\hat{v}$ and $\hat{\eta}$ are replaced by $\hat{v}_3$ and $\hat{\eta}_3$ given by (12.3.87). For this it may be shown that

$$S_n^{(1)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) - S_n^{(2)}(\hat{v}_3, \hat{\eta}_3|\mathbf{k}_3) = S_n^{(3)}(\hat{v}, \hat{\eta}|\mathbf{k}_1), \quad (12.3.242)$$

where

$$S_n^{(3)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) := \frac{2 \sinh \left(\frac{2n\pi \hat{\eta}}{\mathcal{K}'_1} \right) \left[\sinh \left(\frac{2n\pi \hat{v}}{\mathcal{K}'_1} \right) - \sinh \left(\frac{2n\pi \hat{\eta}}{\mathcal{K}'_1} \right) \right]}{n \sinh \left(\frac{4\pi n \mathcal{K}_1}{\mathcal{K}'_1} \right)} \quad (12.3.243)$$

and

$$\begin{aligned} \sum_{n=1}^{\infty} S_n^{(3)}(\hat{v}, \hat{\eta}|\mathbf{k}_1) &= \ln[\hat{c}(\hat{v}_3, \hat{\eta}_3|\mathbf{k}_3)] - \ln[h(\hat{v} + \hat{\eta}|\mathbf{k}_1)] - \frac{\pi \hat{\eta}(\hat{v} - \hat{\eta})}{\mathcal{K}_1 \mathcal{K}'_1} \\ &\quad + \ln[\Pi(\hat{v}, \hat{\eta}|\mathbf{k}_1)], \end{aligned} \quad (12.3.244)$$

where the last term is given by (18.3.167). Thus we have

$$\phi_{\text{Bx}}^{(8\text{vM})}(\mathcal{N}^{\otimes}, \hat{v}_3, \hat{\eta}_3|\mathbf{k}_3) = \phi_{\text{Bx}}^{(8\text{vM})}(\mathcal{N}^{\otimes}, \hat{v}, \hat{\eta}|\mathbf{k}_1) + \ln[\Pi(\hat{v}, \hat{\eta}|\mathbf{k}_1)]. \quad (12.3.245)$$

The zone into which the choice of a non-imaginary value for $\hat{\eta}$ projected the solution is now identified as $\mathcal{R}_o(\text{I})$. The method of solution of the eight-vertex model which has been developed in this text, which follows the work of Fabricius and McCoy, has been shown to recover, by means of the Bazhanov–Mangazeev transformation Baxter’s 1972 expression for the free energy density in $\mathcal{R}_L(\text{III})$. Of course, as we

have already observed, all methods of solution which use the elliptic roots of unity condition, which for $m' = 0$ reduces to (12.3.175), impose a restriction on the values of $\hat{\eta}$. But as Baxter remarked Baxter (1972b, p. 215) “as the positive integer L is arbitrary, we can approach arbitrarily close to any desired value.”

In Sect. 12.3.2 it was shown that the transformation $\mathcal{T}_2$ from $\mathcal{R}_L(\text{I})$ to $\mathcal{R}_L(\text{III})$ was achieved by the variable change $(z, \xi) \rightarrow (\hat{z}, \hat{\xi})$, given by (12.3.9), and this was mirrored for $(v, \eta) \rightarrow (\hat{v}, \hat{\eta})$, with these latter variables given by (12.3.83), in Sect. 12.3.6. The ground state for $\mathcal{R}_L(\text{I})$ has all vertices of type 1 or 2 with Boltzmann weight a and the ground state for $\mathcal{R}_L(\text{III})$ has all vertices of type 5 or 6 with Boltzmann weight c . The full free energy of the system involving all the eigenvalues of the transfer matrix is invariant under $\mathcal{T}_2$ and the change of variables $(v, \eta) \rightarrow (\hat{v}, \hat{\eta})$. However, this symmetry is broken by the process of selecting the largest eigenvalue which is involved in the direct derivation of (12.3.240) by Baxter (1972b) and in the more indirect derivation given above. Notwithstanding this, the formula (12.3.240) with the $\hat{c}(\hat{v}, \hat{\eta}|\mathbf{k}_1)$ replaced, according to Table 9.1, by the ‘zone-neutral’ expression $w_1 + w_2$ and $(\hat{v}, \hat{\eta})$ by (v, η) applies, with an appropriate choice of v and η , to all low-temperature zones $\mathcal{R}_L(\text{J})$, $\text{J} = \text{I}, \text{II}, \text{III}, \text{IV}$ (Baxter 1972b, p. 202). This can be expressed in the form

$$\phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, w_1, w_2, w_3, w_4) = -\ln[w_1 + w_2] - \sum_{n=1}^{\infty} \frac{(\mathcal{X}^{2n} - \mathcal{Q}^n)^2 (\mathcal{X}^n + \mathcal{X}^{-n} - \mathcal{Y}^n - \mathcal{Y}^{-n})}{n\mathcal{X}^n (1 - \mathcal{Q}^{2n})(1 + \mathcal{X}^{2n})}, \quad (12.3.246)$$

where

$$\begin{aligned} \mathcal{X} &:= \exp(-\pi \xi / \mathcal{K}_0) = \exp(i\pi \eta / \mathcal{K}_1), \\ \mathcal{Y} &:= \exp(-\pi z / \mathcal{K}_0) = \exp(i\pi v / \mathcal{K}_1), \\ \mathcal{Q} &:= \exp(-2\pi \mathcal{K}'_0 / \mathcal{K}_0) = \exp(-\pi \mathcal{K}'_1 / \mathcal{K}_1), \end{aligned} \quad (12.3.247)$$

$\mathcal{Q}$ being the nome $\mathcal{Q}_1$ of $\mathbf{k}_1$ and the square $\mathcal{Q}_0^2$ of the nome of $\mathbf{k}_0$.

The spontaneous relative magnetization m_s in the spin interpretation of the eight-vertex model can be derived using ‘corner’ transfer matrices (Baxter 1982b) instead of ‘row’ transfer matrices like those used here and in Chap. 11. It is found that

$$m_s = \prod_{n=1}^{\infty} \frac{1 - \mathcal{X}^{4n-2}}{1 + \mathcal{X}^{4n-2}}. \quad (12.3.248)$$

If $\mathbf{k}_m$ is defined by the relation

$$\frac{\mathcal{K}(\mathbf{k}'_m)}{\mathcal{K}(\mathbf{k}_m)} := \frac{2\xi}{\mathcal{K}_0}, \quad (12.3.249)$$

then it follows from (12.3.247) and (18.3.11) that $\mathfrak{X}^2$ is the nome corresponding to k_m and, from (18.3.13), that

$$m_s = (k'_m)^{\frac{1}{4}} = (1 - k_m^2)^{\frac{1}{8}}. \quad (12.3.250)$$

12.3.8 Critical Behaviour

The phase space of the eight-vertex model is three-dimensional, with axes corresponding, as in Sect. 9.3.2, to ratios of Boltzmann weights or to ratios of (w_1, w_2, w_3, w_4) . In the low-temperature zone $\mathcal{R}_L(J)$ of any of the regions $J = \text{I, II, III, IV}$ of phase space, the free-energy density given by (12.3.246) is an entire function of (w_1, w_2, w_3, w_4) which satisfy (9.3.24). Variations of these quantities can be achieved by varying the members of the set (z, ξ, k_0) , using the first column of Table 12.2. Alternatively, through (3.8.19), variations of the Boltzmann weights can be understood as a variation of the temperature with fixed values of the parameters $\varepsilon_a, \varepsilon_b, \varepsilon_c$ and ε_d .

Singular behaviour will occur (if at all), as T varies, only at points on a boundary of $\mathcal{R}_L(J)$. As we have seen, the weak-graph transformation $\mathcal{T}_{\text{wg}}$ maps from the low-temperature zone $\mathcal{R}_L(J)$ to the high-temperature zone $\mathcal{R}_o(J)$ with the zones separated by the surface of fixed points $\mathcal{C}(J)$, given by $w_2 = w_3$ and $k_0 = 1$. In Sect. 9.3.2.2 this was proposed as a surface of phase transitions, a claim that can now be investigated using the expression (12.3.246) for the free energy. Let

$$\varphi := \pi z / \mathcal{K}'_0, \quad \vartheta := \pi \xi / \mathcal{K}'_0, \quad \varpi := \mathcal{K}'_0 / \mathcal{K}_0. \quad (12.3.251)$$

Then (12.3.246) becomes

$$\phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) = -\ln(w_1 + w_2) - \varpi \sum_{n=1}^{\infty} \psi(n\varpi; \varphi, \vartheta), \quad (12.3.252)$$

where

$$\psi(x; \varphi, \vartheta) := \frac{2 \sinh^2[(\pi - \vartheta)x] \{ \cosh(\vartheta x) - \cosh(\varphi x) \}}{u \sinh(2\pi x) \cosh(\vartheta x)}. \quad (12.3.253)$$

Using the Poisson summation formula (18.1.27),

$$\phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) = -\ln(w_1 + w_2) - \frac{1}{2} \psi^*(0; \varphi, \vartheta) - \sum_{n=1}^{\infty} \psi^*(2\pi n / \varpi; \varphi, \vartheta), \quad (12.3.254)$$

where

$$\psi^*(s; \varphi, \vartheta) := \int_{-\infty}^{\infty} \exp(i x s) \psi(x; \varphi, \vartheta) dx \quad (12.3.255)$$

is an even function of s with $\psi(0; \varphi, \vartheta) = 0$. It has simple poles at $x = i m/2$ and $x = i \pi(2m-1)/(2\vartheta)$, for $m = 1, 2, \dots$, in the upper half of the complex x -plane. The integral (12.3.255) can, therefore, be evaluated using the residue method to give

$$\begin{aligned} \psi^*(s; \varphi, \vartheta) = & 2 \sum_{m=1}^{\infty} \frac{\exp(-m s) [\cos(m \vartheta) - (-1)^m] \left[\cos\left(\frac{m \vartheta}{2}\right) - \cos\left(\frac{m \varphi}{2}\right) \right]}{m \cos\left(\frac{m \vartheta}{2}\right)} \\ & + 2 \sum_{m=1}^{\infty} \frac{(-1)^m \exp\left[-\frac{(2m-1)\pi s}{2\vartheta}\right] \left\{ \cos\left[\frac{(2m-1)\pi^2}{\vartheta}\right] + 1 \right\} \cos\left[\frac{(2m-1)\pi \varphi}{2\vartheta}\right]}{\left(m - \frac{1}{2}\right) \sin\left[\frac{(2m-1)\pi^2}{\vartheta}\right]}. \end{aligned} \quad (12.3.256)$$

On the critical surface $k_0 = 1$, $k'_0 = 0$, $\vartheta = 2\xi$ and $\varphi = 2z$, giving, from the second column of Table 12.2 and (18.3.41),

$$w_1 : w_2 : w_3 : w_4 = \frac{\cos\left(\frac{1}{2}\varphi\right)}{\cos\left(\frac{1}{2}\vartheta\right)} : 1 : 1 : \frac{\sin\left(\frac{1}{2}\varphi\right)}{\sin\left(\frac{1}{2}\vartheta\right)}. \quad (12.3.257)$$

From (12.3.251) and (18.3.10), $\varpi \rightarrow 0$, as $k'_0 \rightarrow 0$, and $\psi^*(2\pi n/\varpi; \varphi, \vartheta) \rightarrow 0$ giving, from (12.3.253)–(12.3.255),

$$\begin{aligned} \phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) &= \phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T_c) = -\ln(w_1 + w_2) - \frac{1}{2} \psi^*(0; \varphi, \vartheta) \\ &= -\ln(w_1 + w_2) - \int_{-\infty}^{\infty} \frac{\sinh^2[(\pi - \vartheta)x] [\cosh(\vartheta x) - \cosh(\varphi x)]}{\omega \sinh(2\pi x) \cosh(\vartheta x)} dx, \end{aligned} \quad (12.3.258)$$

with ϑ and φ taking the critical values ϑ_c and φ_c , obtained by solving (12.3.257) for some fixed values of ε_a , ε_b , ε_c and ε_d . For such a set of fixed values, as the temperature is increased, the system describes a trajectory in parameter space which crosses the critical surface as some critical temperature T_c . From (9.3.31)–(9.3.33), $\Delta = \Delta^* > 1$ in $\mathcal{R}_L(\text{I})$ and $\mathcal{R}_L(\text{II})$ and $\Delta = -\Delta^* < -1$ in $\mathcal{R}_L(\text{III})$ and $\mathcal{R}_L(\text{IV})$. From column 2 of Table 12.2, on $\mathcal{C}(\text{J})$,

$$\Delta = \Delta_c := \pm 1, \quad \Gamma = \Gamma_c := \mp \cos(\vartheta_c), \quad (12.3.259)$$

with the upper and lower signs corresponding, respectively, to the pairs of regions (I, II) and (III, IV). At a temperature $T = T_c - \delta T$,

$$\Delta \simeq \Delta_c + \Delta_0 |\delta T|, \quad \Gamma \simeq \Gamma_c + \Gamma_0 \delta T, \quad (12.3.260)$$

where Δ_0 and Γ_0 are some functions of $\varepsilon_a, \varepsilon_b, \varepsilon_c$ and ε_d , with $\Delta_0 \geq 0$ in (I, II) and $\Delta_0 \leq 0$ in (III, IV). From (12.1.18),

$$k_0'^2 = \frac{\Delta^2 - 1}{\Delta^2 - \Gamma^2} \quad (12.3.261)$$

and, substituting from (12.3.260),

$$k_0'^2 \simeq \frac{2|\Delta_0||\delta T|}{\sin^2(\vartheta_c)}, \quad (12.3.262)$$

giving, from (12.3.251) and (18.3.10),

$$\exp(-\pi/\varpi) \simeq \exp[-2\mathcal{K}_0] \simeq \frac{1}{16} k_0'^2. \quad (12.3.263)$$

Combining these results with (12.3.254) and (12.3.256) we see that, at the temperature $T = T_c - \delta T$, near to the critical surface, $\phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T)$ can, as in (4.2.8), be divided into a sum of smooth and singular parts with

$$\begin{aligned} \phi_{\text{smth}}^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) &\simeq \phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T_c) \\ &- 4 \left[\frac{|\Delta_0||\delta T|}{8 \sin^2(\vartheta_c)} \right]^2 \cos\left(\frac{\vartheta_c}{2}\right) \left[\cos\left(\frac{\vartheta_c}{2}\right) + \cos\left(\frac{\varphi_c}{2}\right) \right], \end{aligned} \quad (12.3.264)$$

$$\phi_{\text{sing}}^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) \simeq 4 \left[\frac{|\Delta_0||\delta T|}{8 \sin^2(\vartheta_c)} \right]^{\pi/\vartheta_c} \cot\left(\frac{\pi^2}{2\vartheta_c}\right) \cos\left(\frac{\pi\varphi_c}{2\vartheta_c}\right). \quad (12.3.265)$$

Unless π/ϑ_c is an integer

$$\phi_{\text{sing}}^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) \sim |\delta T|^{\pi/\vartheta_c}. \quad (12.3.266)$$

Exceptional cases occur if $\vartheta_c = \pi/p$, where p is an integer. If p is even the factor $\cot(\pi^2/2\vartheta_c)$ in (12.3.265) is infinite. This is due to the coincidence of two poles of $\psi(x; \varphi, \vartheta)$. If the residue of the resulting double pole is calculated, (12.3.266) is replaced by

$$\phi_{\text{sing}}^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, T) \sim |\delta T|^{\pi/\vartheta_c} \ln |\delta T|. \quad (12.3.267)$$

If p is odd $\cot(\pi^2/2\vartheta_c) = 0$ and it is necessary to consider the dependence of ϑ on T .

When π/ϑ_c is non-integer (12.3.266) implies that the heat capacity exponents α and α' defined in (4.4.12) are given by

$$\alpha = \alpha' = 2 - \frac{\pi}{\vartheta_c}. \quad (12.3.268)$$

The critical value ϑ_c of ϑ is finite and non-zero and, from (12.3.251), $\xi_c = \frac{1}{2}\vartheta_c$. It follows from (12.3.249) that, as $k'_0 \rightarrow 0$, $k'_m \rightarrow 0$ and, using (18.3.10),

$$k'_m \simeq 4 \left(\frac{k'_0}{4} \right)^{\pi/2\vartheta_c}. \quad (12.3.269)$$

From (12.3.250), (12.3.262) and (12.3.269) the magnetic exponent β , defined in (4.4.13) (with $\rho = m_s$, $\rho_c = 0$) is given by

$$\beta = \frac{\pi}{16\vartheta_c}. \quad (12.3.270)$$

If the Essam–Fisher and Widom scaling laws, given by (4.5.44) and (4.5.45) respectively, are assumed, then it follows from (12.3.269) and (12.3.270) that

$$\gamma = \frac{7\pi}{8\vartheta_c}, \quad \delta = 15. \quad (12.3.271)$$

To complete the picture of the critical properties of the eight-vertex model it should be noted that the correlation length τ can be shown to take the form

$$\tau = \frac{2}{\ln(1/k_m)}, \quad (12.3.272)$$

(Baxter 1982b). From (12.3.262) and (12.3.269) it then follows that the critical exponents ν and ν' , defined in (4.5.52), are given by

$$\nu = \nu' = \frac{\pi}{2\vartheta_c}. \quad (12.3.273)$$

From (12.3.268) and (12.3.273) the Josephson hyper-scaling law (4.5.54), with $d = 2$ is satisfied. If the Fisher scaling law (4.5.58) is assumed then, from (12.3.271) and (12.3.273), the exponent η , defined in (4.5.55), is given by

$$\eta = \frac{1}{4}. \quad (12.3.274)$$

Since ϑ_c is given by (12.3.259) it varies continuously as a function of ε_a , ε_b , ε_c and ε_d . This is a situation where the critical exponents, except δ and η , vary as continuous functions of the energy parameters of the model. The implication of this for the concept of universality is discussed in Sect. 12.3.11.

12.3.9 The Coupling Form and the Ising Model Limit

In terms of the couplings $K^{(f)}$, $K^{(b)}$ and $K^{(4)}$ the critical surfaces are given by (9.3.34) with the critical exponents varying as functions of ϑ_c , which is given, from (12.1.13) (dropping the subscript j) and (12.3.259), by

$$\vartheta_c = \begin{cases} \pi - \arccos[\tanh(2K^{(4)})], & \text{for } \mathcal{C}(\text{I}) \text{ and } \mathcal{C}(\text{II}), \\ \arccos[\tanh(2K^{(4)})], & \text{for } \mathcal{C}(\text{III}) \text{ and } \mathcal{C}(\text{IV}). \end{cases} \quad (12.3.275)$$

It has been shown by Kadanoff and Wegner (1971) that the four-spin interaction in the eight-vertex model scales as $1/r^2$. This means, in terms of our discussion in Sect. 4.3.8, that $K^{(4)}$ is a marginal coupling, and the variation of the critical exponents α , β and γ , given by (12.3.268), (12.3.270) and (12.3.271) respectively, as functions of $K^{(4)}$ does not conflict with scaling theory.

Again using upper and lower signs for the pairs of regions (I, II) and (III, IV), respectively, the critical surfaces given by (9.3.34) cut the planes $K^{(f)} = \pm K^{(b)} = K$ with $K^{(4)} \geq 0$ in a curve on which⁶⁸

$$K = \frac{1}{2} \operatorname{arcsinh} \left(\sqrt{\frac{1 - \tanh(2|K^{(4)}|)}{1 + \tanh(2|K^{(4)}|)}} \right). \quad (12.3.276)$$

The plot of this, shown in Fig. 12.8, meets the $K^{(4)} = 0$ axis at $K = K_c$ given by $\sinh(2K_c) = 1$, which is the standard formula for the critical temperature of the square-lattice Ising model given in (9.2.36). We also note that, as $K \rightarrow 0$ on the curve, $K^{(4)} \rightarrow \infty$, showing that there is no phase transition in a spin- $\frac{1}{2}$ model on a square lattice with a purely four-spin interaction.

The case $K^{(4)} = 0$, where the eight-vertex model, in the coupling formulation, reduces to two disconnected Ising models, has been explored in great detail in Sect. 12.2. Here we simply exploit the information acquired from the derivation of the free energy in this special case, restricting attention to the situation $K^{(f)} = K^{(b)} = K$ (in region I or II). Now, from (3.8.12) and (3.8.13),

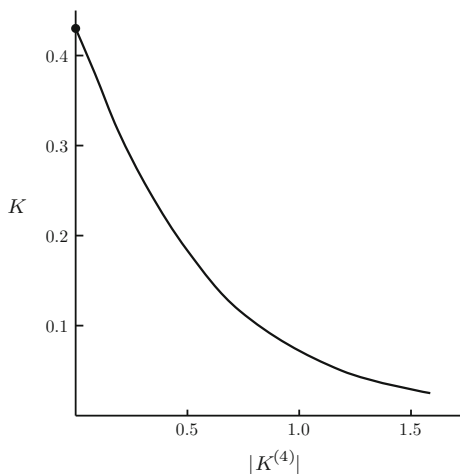
$$\Delta = \sinh^2(2K), \quad \Gamma = 0, \quad (12.3.277)$$

from (12.3.1), (12.3.4) and (12.3.251),

$$\begin{aligned} \lambda &= \mathcal{K}'_0, & \xi &= \frac{1}{2} \mathcal{K}'_0, & \vartheta &= \frac{1}{2} \pi, \\ u &= \frac{1}{2} \lambda, & z &= \varphi = 0, \end{aligned} \quad (12.3.278)$$

⁶⁸ Of course, there is also a choice of sign for K but this applies to all the regions.

Fig. 12.8 The critical curve for the eight-vertex model, where the critical point of the isotropic Ising model lies on the $K^{(4)} = 0$ axis at $K = 0.4407$



from (12.3.261),

$$\Delta = \begin{cases} 1/k_0, & K \geq K_c, (T \leq T_c), \\ k_0, & K \leq K_c, (T \geq T_c) \end{cases} \quad (12.3.279)$$

and, from (12.3.268), (12.3.270)–(12.3.274) and (12.3.278),

$$\begin{aligned} \alpha &= 0, & \beta &= \frac{1}{8}, & \gamma &= \frac{7}{4}, \\ \delta &= 15, & \nu &= 1, & \eta &= \frac{1}{4}, \end{aligned} \quad (12.3.280)$$

which are the well-known results for the two-dimensional Ising model. The fact that $\pi/\vartheta_c = 2$ shows, from the discussion of Sect. 12.3.8, that the heat capacity has a logarithmic singularity at the critical point. From (12.3.249) and (12.3.278),

$$\frac{\mathcal{K}(k'_m)}{\mathcal{K}(k_m)} = \frac{\mathcal{K}'_0}{\mathcal{K}_0}, \quad (12.3.281)$$

from which it follows that $k_m = k'_0$. Then, from (12.3.250),⁶⁹

$$m_s = (k'_m)^{\frac{1}{4}} = (1 - k_0'^2)^{\frac{1}{8}}, \quad (12.3.282)$$

⁶⁹ This is a result which applies to the square, triangular and honeycomb lattices (McCoy and Wu 1973).

yielding, from (12.3.277) and (12.3.278).

$$m_s = [1 - \sinh^{-4}(2K)]^{\frac{1}{8}}, \quad T < T_c. \quad (12.3.283)$$

This is show, in comparison with approximate results, in Fig. 3.1.

The series expansion (12.3.246) will now be used to obtain an expression for the internal energy of the Ising model following the procedure of Lavis (1996). From (12.3.278), Eqs. (12.3.247) reduce to

$$\mathcal{Y} = 1, \quad \mathcal{X} = Q^{\frac{1}{4}} = \exp(-\pi \mathcal{K}'_0 / 2\mathcal{K}_0). \quad (12.3.284)$$

Using the Landen modulus k_2 defined in Sect. 18.3.4,

$$\mathcal{X} = \exp(-\pi \mathcal{K}'_2 / \mathcal{K}_2) \quad (12.3.285)$$

is the nome of k_2 . From (3.8.11), Table 9.1, (12.3.246), (12.3.284) and (12.3.285),

$$\phi^{(8_{\text{VM}})}(\mathcal{N}^{\otimes}, K) = \begin{cases} -2|K| - \sum_{n=1}^{\infty} S_n(\mathcal{X}), & T < T_c, \\ -\ln[\cosh(2K) + 1] - \sum_{n=1}^{\infty} S_n(\mathcal{X}), & T > T_c, \end{cases} \quad (12.3.286)$$

where

$$S_n(\mathcal{X}) = \frac{\mathcal{X}^{2n}(1 - \mathcal{X}^{2n})^2(1 - \mathcal{X}^n)^2}{n(1 - \mathcal{X}^{8n})(1 + \mathcal{X}^{2n})}. \quad (12.3.287)$$

The dimensionless internal-energy density u is given by (2.3.27)⁷⁰ and, from (12.3.277), (12.3.279), (12.3.286) and (18.3.12),

$$u = \begin{cases} -2 + \frac{B(K)}{\mathcal{K}_2^2} \sum_{n=1}^{\infty} \mathcal{X} \frac{dS_n(\mathcal{X})}{d\mathcal{X}}, & T < T_c, \\ -\frac{2 \sinh(2K)}{1 + \cosh(2K)} + \frac{B(K)}{\mathcal{K}_2^2} \sum_{n=1}^{\infty} \mathcal{X} \frac{dS_n(\mathcal{X})}{d\mathcal{X}}, & T > T_c, \end{cases} \quad (12.3.288)$$

⁷⁰ In the notation of Sect. 2.3 ϕ is here $\phi_1(K, 0)$, the free-energy density with one external coupling, $L := \mathcal{H}/T$ set equal to zero.

where

$$B(K) = \frac{\pi^2 \cosh^3(2K)}{\sinh(2K)[\sinh^2(2K) - 1]}. \quad (12.3.289)$$

From (12.3.287)

$$\begin{aligned} \mathcal{X} \frac{dS_n(\mathcal{X})}{d\mathcal{X}} &= \frac{\mathcal{X}^n}{1 + \mathcal{X}^{2n}} + \frac{2\mathcal{X}^{2n}}{(1 + \mathcal{X}^{2n})^2} - \frac{8\mathcal{X}^{3n}}{(1 + \mathcal{X}^{2n})^3} \\ &\quad - \frac{\mathcal{X}^n(1 + \mathcal{X}^{2n})}{1 + \mathcal{X}^{4n}} - \frac{4\mathcal{X}^{4n}}{(1 + \mathcal{X}^{4n})^2} + \frac{4\mathcal{X}^{3n}(1 + \mathcal{X}^{2n})}{(1 + \mathcal{X}^{2n})^3}, \end{aligned} \quad (12.3.290)$$

and hence, from Eqs. (18.3.91)–(18.3.93),

$$\sum_{n=1}^{\infty} \mathcal{X} \frac{dS_n(\mathcal{X})}{d\mathcal{X}} = \frac{k'_2 \mathcal{K}_2^2}{\pi^2} \left[\sqrt{2(1 + k'_2)} - 1 \right] - \frac{2k'_2{}^2 \mathcal{K}_2^3}{\pi^3}. \quad (12.3.291)$$

Since, from (12.3.277), (12.3.279) and (18.3.78),

$$k_2(K) = \frac{2 \sinh(2K)}{\cosh^2(2K)}, \quad (12.3.292)$$

$$k'_2(K) = \pm \left[\frac{1 - \sinh^2(2K)}{\cosh^2(2K)} \right], \quad T \geq T_c, \quad (12.3.293)$$

by substituting from (12.3.292) and (12.3.293) into (12.3.291) and then into (12.3.288) the formula (12.2.178) for the internal energy per lattice site of the Ising model on the square lattice at all temperatures is obtained.

12.3.10 The Six-Vertex Model

The six-vertex model corresponds to the limiting case $\varepsilon_d \rightarrow \infty$ ($d \rightarrow 0$) of the eight-vertex model with, from (9.3.1), $\Gamma = 1$. It can fall in region I, II or III, but not IV.

12.3.10.1 Low-Temperature Zones $\mathcal{R}_L(\text{I})$ and $\mathcal{R}_L(\text{II})$

In these zones, from (9.3.31), (9.3.32), Table 12.2 and (18.3.43),

$$\Delta = \Delta^* = nd(2\xi|k'_0) > 0, \quad \Gamma = \Gamma^* = -cd(2\xi|k'_0) = 1. \quad (12.3.294)$$

From (18.3.39) these relationships hold in the limit $k_0 \rightarrow 0$ with

$$\Delta^* = \cosh(2\xi). \quad (12.3.295)$$

This corresponds to a case of the XYX (in standard notation XXZ) linear Heisenberg model (Sects. 3.2.1 and 12.3.4.2, case (vi)). From Table 9.1, for small d , w_4 is positive in $\mathcal{R}_L(I)$ and negative in $\mathcal{R}_L(II)$. So, from Table 12.2, z is positive in $\mathcal{R}_L(I)$ and negative in $\mathcal{R}_L(II)$ with $|z| \leq \xi$. From (12.3.247), in the limit $k_0 \rightarrow 0$,

$$\begin{aligned} \mathcal{Q} &= \exp(-2\pi\mathcal{K}'_0/\mathcal{K}_0) \rightarrow \mathcal{X}^2 \rightarrow 0, \\ \mathcal{X} &= \exp(-\pi\xi/\mathcal{K}_0) \rightarrow 0, \\ \mathcal{Y} &= \begin{cases} \exp(-\pi|z|/\mathcal{K}_0) \rightarrow \mathcal{X}^2 \rightarrow 0 & \text{in } \mathcal{R}_L(I), \\ \exp(\pi|z|/\mathcal{K}_0) \rightarrow \mathcal{X}^{-2} \rightarrow \infty & \text{in } \mathcal{R}_L(II). \end{cases} \end{aligned} \quad (12.3.296)$$

From (12.3.246) we have, therefore,

$$\begin{aligned} \phi^{(6_{\text{VM}})}(\mathcal{N}^{\oplus}, a, b, c) &= -\ln(w_1 + w_2) \\ &= -\ln\left[\frac{1}{2}(a + b) + \frac{1}{2}|a - b|\right], \end{aligned} \quad (12.3.297)$$

which is equivalent to (9.3.20), in $\mathcal{R}_L(I)$, where $a > b$. From (12.3.259) $\vartheta_c = \pi$ in this case and, from (12.3.268), $\alpha = \alpha' = 1$. It thus follows that the scaling exponent y_2 of (4.5.39) is equal to $d = 2$, indicative of a first-order transition. This is of course as it should be since the transition is to the frozen completely-ordered ferroelectric state.

12.3.10.2 Low-Temperature Zone $\mathcal{R}_L(III)$

Here, from (9.3.31), (9.3.32), Table 12.2 and (18.3.43),

$$\begin{aligned} \Delta &= -\Delta^* = -nd(2\xi|k'_0|) < 0, \\ \Gamma &= -\Gamma^* = cd(2\xi|k'_0|) = 1. \end{aligned} \quad (12.3.298)$$

These relationships again hold in the limit $k_0 \rightarrow 0$, with Δ^* given by (12.3.295), and the same equivalence to the XXZ model. From Table 12.2,

$$\frac{a}{c} = \frac{\sinh(\xi + z)}{\sinh(2\xi)}, \quad \frac{b}{c} = \frac{\sinh(\xi - z)}{\sinh(2\xi)} \quad (12.3.299)$$

and, from (12.3.247),

$$\Omega = 0, \quad \mathcal{X} = \exp(-2\xi), \quad \mathcal{Y} = \exp(-2z). \quad (12.3.300)$$

Substituting into (12.3.246) gives

$$\phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c) = -\ln(c) - \sum_{n=1}^{\infty} \frac{\exp(-4n\xi)[\cosh(2n\xi) - \cosh(2nz)]}{n \cosh(2n\xi)} \quad (12.3.301)$$

and, substituting for c from the first of Eqs. (12.3.299) and expanding the resulting logarithmic series, gives the alternative form

$$\phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c) = -(\xi - z) - \sum_{n=1}^{\infty} \frac{\exp(-2n\xi) \sinh[2n(\xi - z)]}{n \cosh(2n\xi)}. \quad (12.3.302)$$

12.3.10.3 High-Temperature Zones $\mathcal{R}_d(\text{I})$, $\mathcal{R}_d(\text{II})$ and $\mathcal{R}_d(\text{III})$

In these zones, from Tables 9.1 and 12.2,

$$\begin{aligned} \Delta &= \Gamma^*/\Delta^* = -\text{cn}(2\xi|k'_0), \\ \Gamma &= 1/\Delta^* = \text{dn}(2\xi|k'_0) = 1, \end{aligned} \quad (12.3.303)$$

which are satisfied, in the limit $k_0 \rightarrow 1$, when $\vartheta = 2\xi$ and

$$\Delta = -\cos(\vartheta). \quad (12.3.304)$$

Since, in these zones, from Table 9.1 with $d = 0$, $w_2 = w_3$, this is the critical surface with

$$\frac{a}{c} = \frac{\sin\left[\frac{1}{2}(\vartheta + \varphi)\right]}{\sin(\vartheta)}, \quad \frac{b}{c} = \frac{\sin\left[\frac{1}{2}(\vartheta - \varphi)\right]}{\sin(\vartheta)}, \quad (12.3.305)$$

and the free-energy density is, from (12.3.258),

$$\begin{aligned} \phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c) &= -\ln\left[\frac{1}{2}(a + b + c)\right] \\ &\quad - \int_{-\infty}^{\infty} \frac{\sinh^2[(\pi - \vartheta)x][\cosh(\vartheta x) - \cosh(\varphi x)]}{x \sinh(2\pi x) \cosh(\vartheta x)} dx. \end{aligned} \quad (12.3.306)$$

This is the expression given by Baxter (1972b). With some manipulation it can be re-expressed in the form

$$\phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, c) = -\ln(a) - \frac{1}{2} \int_{-\infty}^{\infty} \frac{\sinh[(\pi - \vartheta)x] \sinh[(\vartheta - \varphi)x]}{x \sinh(\pi x) \cosh(\vartheta x)} dx, \quad (12.3.307)$$

given by Lieb and Wu (1972).

12.3.10.4 Square Ice

Here $a = b = c = 1$ imposing, in addition to (12.3.309), the condition that $\sin(\vartheta) = -\sin(2\vartheta)$. This means that $\vartheta = \frac{2}{3}\pi$ in (12.3.310), giving, for the entropy of square ice,

$$\begin{aligned} \ln[\Omega(N^{\otimes})] &:= -N^{\otimes} \phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, 1, 1, 1) = N^{\otimes} \ln \left| \frac{4 \cot(3\pi/4)}{3 \sin(2\pi/3)} \right| \\ &= \frac{3}{2} N^{\otimes} \ln \left(\frac{4}{3} \right) = N^{\otimes} \ln(1.5396), \end{aligned} \quad (12.3.308)$$

This is the result obtained by Lieb (1967a, b). It is of interest that it is very close to an estimate obtained by Pauling (1935).⁷¹ He simply took the total unrestricted number of arrow arrangements $2^{2N^{\otimes}}$ and scaled by $(6/16)^{N^{\otimes}}$ to compensate for the fact that only six of the sixteen arrangements around a vertex are permitted; thus giving $\Omega(N^{\otimes}) = (1.5)^{N^{\otimes}}$.

12.3.10.5 The KDP Model

From Sect. 3.8.2, $a > b = c$, giving, from (9.3.1), (12.3.304) and (12.3.305),

$$\sin \left[\frac{1}{2}(\vartheta + \varphi) \right] = -\sin(2\vartheta), \quad \sin \left[\frac{1}{2}(\vartheta - \varphi) \right] = \sin(\vartheta), \quad (12.3.309)$$

which have the solutions $(\varphi, \vartheta) = (0, \frac{2}{3}\pi)$ and (π, π) . Glasser (1969) showed that when these conditions are satisfied the integral in (12.3.307) can be expressed in terms of elementary functions to give

$$\phi^{(6_{\text{VM}})}(\mathcal{N}^{\otimes}, a, b, b) = -\ln(a) - \ln \left| \frac{2\vartheta}{\pi \sin(\vartheta)} \cot \left(\frac{\pi^2}{2\vartheta} \right) \right|. \quad (12.3.310)$$

⁷¹ And, as noted above, even closer to $N^{\otimes} \ln(1.543)$ the result of the series approximation by Nagle (1966a).

12.3.10.6 The F Model and Critical Potts Model

From Sect. 3.8.2, $a = b < c$ and, as we saw in Sect. 3.8.2.1, setting the Boltzmann weights to $a = b = 1$, $c = (2 + \sqrt{\nu})^{\frac{1}{2}}$ gives equivalence to the square-lattice ferromagnetic ν -state Potts model at its critical temperature. From (9.3.1),

$$\Delta = -\frac{1}{2}\sqrt{\nu} \quad (12.3.311)$$

and variations of ν now play the role of temperature variation.

- For $0 < \nu < 4$ $-1 < \Delta < 0$, placing the model in the disordered (high-temperature in the six-vertex sense) zone with, from (12.3.295) and (12.3.305), $\varphi = 0$ and

$$\sqrt{\nu} = 2 \cos(\vartheta), \quad 0 < \vartheta < \frac{1}{2}\pi. \quad (12.3.312)$$

Then, from (3.8.41) and (12.3.307),

$$\phi_c^{(\text{PM})}(\mathcal{N}, \nu) = -\frac{1}{2} \ln(\nu) - 2 \int_0^\infty \frac{\sinh[(\pi - \vartheta)x] \tanh(\vartheta x)}{x \sinh(\pi x)} dx. \quad (12.3.313)$$

- For $\nu \geq 4$ $\Delta < -1$, placing the model in $\mathcal{R}_L(\text{III})$. From Table 9.1 $w_4 = 0$ and hence, from Table 12.2, $z = 0$. From (12.3.295),

$$\sqrt{\nu} = 2 \cosh(2\xi), \quad (12.3.314)$$

and, from (12.3.302),

$$\phi_c^{(\text{PM})}(\mathcal{N}, \nu) = -\frac{1}{2} \ln(\nu) - \xi - \sum_{n=1}^{\infty} \frac{\exp(-2n\xi) \tanh(2n\xi)}{n}. \quad (12.3.315)$$

- For the boundary $\nu = 4$ the free-energy density can be obtained by taking the limit $\vartheta \rightarrow 0$ in (12.3.313), or $\xi \rightarrow 0$ in (12.3.315), to give

$$\phi_c^{(\text{PM})}(\mathcal{N}, \nu) = -\ln(2) - 2 \int_0^\infty \exp(-x) \tanh(x) dx = -8 \ln \left[\frac{\Gamma(\frac{1}{4})}{2^{\frac{5}{8}} \pi^{\frac{1}{2}}} \right]. \quad (12.3.316)$$

It is clear from these results that the model differs in its behaviour according to whether ν is greater or less than four. One way to understand more clearly the effect on the nature of the phase transition is to calculate the internal energy as the critical point is approached from above and below. For the ferromagnetic case this has been achieved by Baxter (1973d). In our present notation

$$u = -R(1 + \nu^{-\frac{1}{2}})[1 \pm p_0 \tanh(\xi)], \quad (12.3.317)$$

where p_0 is the spontaneous polarization with the $+$ and $-$ signs applying according to whether the critical point is approached from below or above. Baxter (1973e, f) showed that

$$p_0 = \begin{cases} 0, & 0 < \nu \leq 4, \\ \prod_{n=1}^{\infty} \tanh^2(2n\xi), & \nu > 4. \end{cases} \quad (12.3.318)$$

Twice the second term in (12.3.317), therefore, represents the latent heat at the transition for $\nu > 4$, the transition being first-order. In all cases the first term in (12.3.317) gives the internal energy per site averaged over the values below and above the transition. This was first calculated by Potts (1952).

For $0 < \nu \leq 4$ it has been conjectured by den Nijs (1979) that

$$\alpha = \frac{2\pi - 8\vartheta}{3\pi - 6\vartheta} \quad (12.3.319)$$

and independently by Nienhuis et al. (1980) and Pearson (1980) that

$$\beta = \frac{\pi + 2\vartheta}{12\pi}, \quad (12.3.320)$$

where ϑ is given by (12.3.312). Using the scaling laws (4.5.44), (4.5.45) and (4.5.47) this gives

$$\gamma = \frac{7\pi^2 - 8\vartheta\pi + 4\vartheta^2}{6\pi(\pi - 2\vartheta)}, \quad \delta = \frac{15\pi^2 - 16\vartheta\pi + 4\vartheta^2}{\pi^2 - 4\vartheta^2}. \quad (12.3.321)$$

The formula (12.3.319) for α has now been established by Black and Emery (1981) by further exploiting the relationship to the staggered F model. From (12.3.312), with $\nu = 2$, $\vartheta = \pi/4$ yielding the exponent values (12.3.280) for the Ising model, When $\nu = 3$, $\vartheta = \pi/6$,

$$\alpha = \frac{1}{3}, \quad \beta = \frac{1}{9}, \quad \gamma = \frac{13}{9}, \quad \delta = 14. \quad (12.3.322)$$

These values are strongly supported by the finite-size scaling analysis of Blöte and Nightingale (1982) (Sect. 4.10) and by conformal invariance arguments of Friedan et al. (1984) and Cardy (1996) (Sect. 4.11). They are the same as those of the hard-hexagon model (Baxter 1982b) and support the argument of Alexander (1975) that the 3-state Potts model and hard-hexagon model lie in the same universality class. When $\nu = 4$, $\vartheta = 0$,

$$\alpha = \frac{2}{3}, \quad \beta = \frac{1}{12}, \quad \gamma = \frac{7}{6}, \quad \delta = 15, \quad (12.3.323)$$

but this gives only a partial picture of the asymptotic behaviour because of the presence of logarithmic corrections (Cardy et al. 1980; Nauenberg and Scalapino 1980; Salas and Sokal 1997).

12.3.11 The Eight-Vertex Model and Universality

The concept of universality classes was discussed in Sect. 4.2, where we listed Kadanoff's three criteria for the division of phase transitions into classes. It was also pointed out that any particular model may well exhibit phase transitions in different universality classes within different parts of its phase space. We saw in Sect. 5.4.1 that the symmetry group of the zero-field spin-1 Ising model is in general $\mathcal{S}_2$, the same as that of the standard zero-field spin- $\frac{1}{2}$ Ising model and thus, for any particular value of the dimension d , it will exhibit a phase transition in the universality class of the latter. However, when the interaction energy parameters attain certain special values there is a change of symmetry group to $\mathcal{S}_3$, that of the 3-state Potts model (Sect. 5.4.2), for which the transition lies in a different universality class. Examples can also be given when a change in the energy parameters will lead to a change in universality class by means of an effective change in the dimension of the system rather than the symmetry group. We could, for example, consider an anisotropic lattice model on a d -dimensional hypercubic lattice. If the energy parameters connecting microsystems in the direction of one axis tend to zero there will, from the point of view of cooperative phenomena, be an effective change of dimension of the system from d to $d - 1$, with a resulting change of universality class for the phase transition.

At this point it is useful to distinguish between the *universality hypothesis*, in whatever form it is represented, and the *scaling hypothesis*, introduced in Sect. 4.3. It is a consequence of the latter and the corresponding hypothesis for correlations of Sect. 4.5.7 that the critical exponents satisfy the four scaling laws (4.5.44), (4.5.45), (4.5.54) and (4.5.58). It is sometimes asserted that the eight-vertex model contradicts the universality hypothesis because its critical exponents vary as a function of a parameter. We have shown in Sect. 4.3.8 that the scaling hypothesis encompasses the possibility of exponents varying as functions of a marginal coupling and in Sect. 12.3.9 it was shown that the four-spin coupling in the spin representation of the eight-vertex model is of this type. The interesting question in relation to universality is whether the variation of the four-spin coupling corresponds to a change of symmetry and thus a change of universal class according to Kadanoff's criterion (b), or whether the symmetry group remains the same. As was pointed out by Kadanoff and Wegner (1971) and Baxter (1982b) there is certainly a change of symmetry when the first two conditions $\mathfrak{X}_1 = \mathfrak{X}_2$, $\mathfrak{X}_3 = \mathfrak{X}_4$ in (3.8.4) are applied. These conditions

would be violated if horizontal and vertical spin couplings $K^{(h)}$ and $K^{(v)}$ were added⁷² to the Boltzmann weights (3.8.9). With $K^{(h)} = K^{(v)} = 0$ the symmetry group is the direct product $\mathcal{S}_2^{(m)} \times \mathcal{S}_2^{(m)}$, of spin-reversal symmetry applied to the two interpenetrating sublattices independently. With one or both of $K^{(h)}$ and $K^{(v)}$ non-zero the symmetry group is reduced to simply the spin-reversal symmetry group $\mathcal{S}_2^{(m)}$ of the Ising model (on the whole lattice). Although the general eight-vertex model, without the conditions $\mathfrak{X}_1 = \mathfrak{X}_2$, $\mathfrak{X}_3 = \mathfrak{X}_4$ has not been solved, the solution of the special free-fermion case by Fan and Wu (1970) (Sect. 3.8.1.5) gives a strong indication that the critical exponents are those of the Ising model. According to Kadanoff's criteria (a) and (b) for universality classes this is to be expected, since the dimensions and symmetry groups of the models are the same. By the same token we expect the exponents of the eight-vertex model with conditions (3.8.4) to be different. That they vary is understandable in terms of scaling. To set them in the context of Kadanoff's criteria of universality classes two approaches are possible:

- (i) The model with varying exponents can be taken as one universality class.
- (ii) Criterion (c) can be invoked with the marginal coupling taken as another criterion differentiating between classes.

The latter has some advantages since the concept of constant exponents *within* classes is preserved. Also since, in renormalization group theory (Chap. 15) a marginal exponent is associated with a line of fixed points and a fixed point is usually associated with a universality class, to associate each value of K_4 with a different class has some value.

Finally, the proposal of *weak universality* by Suzuki (1974) should be mentioned. He observed that the exponents in the eight-vertex model which vary are α , β , γ and ν (see Eqs. (12.3.268), (12.3.270), (12.3.271) and (12.3.273) respectively). These are precisely those which represent asymptotic behaviour in terms of $|T - T_c|$ (see Eqs. (4.4.12)–(4.4.14) and (4.5.52)). If, however, asymptotic behaviour is measured in terms of the inverse correlation length τ^{-1} then this leads to the use of the exponents

$$\hat{\phi} = (2 - \alpha)/\nu, \quad \hat{\beta} = \beta/\nu, \quad \hat{\gamma} = \gamma/\nu. \quad (12.3.324)$$

The four scaling laws (4.5.45), (4.5.44), (4.5.54) and (4.5.58) are replaced by the *weak scaling laws*

$$\begin{aligned} \hat{\gamma} &= \hat{\beta}(\delta - 1), & 2\hat{\beta} + \hat{\gamma} &= \hat{\phi}, \\ d &= \hat{\phi}, & 2 - \eta &= \hat{\gamma}. \end{aligned} \quad (12.3.325)$$

⁷² As in (3.8.20) but now with equal values to left and right and above and below.

For the eight-vertex model the complete set of exponents is now

$$\hat{\phi} = 2, \quad \hat{\beta} = \frac{1}{8}, \quad \hat{\gamma} = \frac{7}{4}, \quad \delta = 15, \quad \eta = \frac{1}{4}. \quad (12.3.326)$$

These are, of course, the same for the two-dimensional Ising model (Eq. (12.3.280)), which unfortunately obscures the important symmetry difference between the two models.

Chapter 13

Dimer Models

13.1 The Dimer Partition Function

Here we consider a lattice $\mathcal{N}$ with pure dimer coverings, as defined in Sect. 3.9. Let $\Pi_0 := (1, 2, \dots, N)$ be some labelling of the sites of $\mathcal{N}$ so that the partition into pairs $(1, 2), (3, 4), \dots, (N-1, N)$, with (i, j) being a first-neighbour pair occupied by a dimer, represents such a covering.¹ Then

$$Z(\mathcal{N}; \mathbf{T}) := \sum_{\{\Pi\}} |t(\pi_1, \pi_2)| |t(\pi_3, \pi_4)| \cdots |t(\pi_{N-1}, \pi_N)|, \quad (13.1.1)$$

where, as in the dimer partition function (3.9.6), the sum is over all permutations Π of Π_0 satisfying (3.9.2) and $\mathbf{T}$ is any matrix with elements $t(i, j) := \langle i | \mathbf{T} | j \rangle$ satisfying

$$|t(i, j)| = \begin{cases} |t(j, i)| \neq 0, & i \text{ and } j \text{ a first-neighbour pair,} \\ 0, & i \text{ and } j \text{ not a first-neighbour pair,} \end{cases} \quad (13.1.2)$$

$$t(i, j) = \overline{t(j, i)}, \quad (13.1.3)$$

is a generalization of the formula (3.9.6) for the dimer partition function $Z(\mathcal{N}; \mathbf{\Lambda})$. For physical reasons associated with the problem that concerned them, Lieb and Loss (1993) called $\mathbf{T}$ a *hopping matrix* and showed that²:

Theorem 13.1.1 *If the hopping matrix $\mathbf{T}$ satisfies the canonical flux distribution condition (16.4.8) then*

¹ As observed in Sect. 3.9, for this to be the case N must be even.

² Parts (i) and (ii) of this theorem correspond respectively to their Theorems 3.1 and 16.1.

$$(i) \text{Det}\{T\} = (-1)^{N/2} [Z(\mathcal{N}; T)]^2, \quad (13.1.4)$$

(ii) *there exists a gauge transformation (16.4.5) with*

$$U^\dagger T U = iZ, \quad (13.1.5)$$

where Z is a real antisymmetric matrix.

It follows from Theorem 16.4.2, that a hopping matrix T , satisfying the canonical flux distribution, can be constructed from any symmetric matrix Λ , which satisfies (3.9.3), by taking $t(i, j) := \lambda(i, j) \exp[i\theta(i, j)]$ with a suitable choice of the antisymmetric arguments $\theta(i, j)$. Then, from (13.1.1)–(13.1.3),

$$Z(\mathcal{N}; T) = Z(\mathcal{N}; \Lambda). \quad (13.1.6)$$

It then follows from Theorem 13.1.1 that there is a matrix Z related to Λ by

$$z(i, j) = \pm \lambda(i, j), \quad z(j, i) = \mp \lambda(i, j), \quad (13.1.7)$$

for a particular choice of signs, such that

$$\text{Det}\{Z\} = [Z(\mathcal{N}; \Lambda)]^2. \quad (13.1.8)$$

The evaluation of the dimer partition function (3.9.6) for a particular choice of fugacities $\lambda(i, j)$ is thus reduced to:

- (i) the determination of the choice of signs for the elements of Z ;
- (ii) the evaluation of the determinant of Z .

13.2 Superposition Polynomials and Pfaffians

The method used for task (i), the determination of the signs in (13.1.7) for the antisymmetric matrix Z , is that of Kasteleyn (1961), which employs the *Pfaffian*

$$\text{Pf}\{Z\} := \sum_{\{\Pi\}} \epsilon(\Pi) z(\pi_1, \pi_2) z(\pi_3, \pi_4) \cdots z(\pi_{N-1}, \pi_N) \quad (13.2.1)$$

of Z , where, as in the dimer partition function (3.9.6) and in (13.1.1), the sum is over all permutations Π of Π_0 satisfying (3.9.2), with $\epsilon(\Pi)$ being the *signature*³ of Π . *Pfaff's theorem* states that:

³ $\epsilon(\Pi) = 1$ or -1 according to whether Π is of even or odd parity, that is whether Π is obtained from $\Pi_0 = (1, 2, \dots, N)$ by an even or odd number of transpositions. A transposition is an exchange of the position of two indices, with the order of the remaining indices unchanged. Any transposition can be effected by an odd number of transpositions of adjacent indices.

Theorem 13.2.1 *The determinant and Pfaffian of the $N \times N$ antisymmetric matrix $\mathbf{Z}$ satisfy the relationship*

$$\text{Det}\{\mathbf{Z}\} = [\text{Pf}\{\mathbf{Z}\}]^2. \quad (13.2.2)$$

This result was proved in some special cases by Temperley and Fisher (1961), by induction on the order N by Green and Hurst (1964) and directly by Kasteleyn (1967). It follows from (13.1.7), (13.1.8) and (13.2.2) that⁴

$$\text{Pf}\{\mathbf{Z}\} = Z(\mathcal{N}; \mathbf{Z}). \quad (13.2.3)$$

This result is known as *Kasteleyn's theorem* (Kasteleyn 1961). Now let

$$\mathcal{P}(\mathbf{Z}; \Pi) := \epsilon(\Pi) z(\pi_1, \pi_2) z(\pi_3, \pi_4) \cdots z(\pi_{N-1}, \pi_N). \quad (13.2.4)$$

This is an arbitrary term of the Pfaffian (13.2.1). The correct choice of signs for the elements of $\mathbf{Z}$ to satisfy (13.1.8) is, from (13.2.3), the one for which $\mathcal{P}(\mathbf{Z}; \Pi) > 0$ for all permutations Π . We have introduced a standard dimer covering of $\mathcal{N}$, labelled so that the dimers are the lines $(1, 2), (3, 4), \dots, (N-1, N)$ with $\Pi_0 = (1, 2, \dots, N)$ being the identity permutation. Whatever the eventual choice of signs for the elements of $\mathbf{Z}$ we can choose $\epsilon(\Pi_0)$, so that $\mathcal{P}(\mathbf{Z}; \Pi_0) > 0$ and thus we need to ensure that

$$\text{Sign}\{\mathcal{P}(\mathbf{Z}; \Pi)\} = \text{Sign}\{\mathcal{P}(\mathbf{Z}; \Pi_0)\}, \quad \forall \mathcal{P}(\mathbf{Z}; \Pi) \neq 0. \quad (13.2.5)$$

Now consider the product

$$\epsilon(\mathsf{T}) z(\tau_1, \tau_2) z(\tau_3, \tau_4) \cdots z(\tau_{N-1}, \tau_N), \quad (13.2.6)$$

where $(\tau_1, \tau_2), (\tau_3, \tau_4), \dots, (\tau_{N-1}, \tau_N)$ is a dimer covering, but the indices in the permutation $\mathsf{T} := (\tau_1, \tau_2, \tau_3, \tau_4, \dots)$ are otherwise unordered. Suppose that for one of the pairs $\tau_{2k} < \tau_{2k-1}$. The transposition of these indices multiplies $\epsilon(\mathsf{T})$ by -1 ; but, from (13.1.7), $z(\tau_{2k}, \tau_{2k-1}) = -z(\tau_{2k-1}, \tau_{2k})$, so the product (13.2.6) remains unchanged. Again (τ_{2k-1}, τ_{2k}) can be exchanged with (τ_{2j-1}, τ_{2j}) for any k and j without affecting the sign of the product (13.2.6) since it involves two transpositions. It follows that the indices can be rearranged to satisfy (3.9.2) without altering the value of (13.2.6), which is thus equal to the Pfaffian term $\mathcal{P}(\mathbf{Z}; \Pi)$ corresponding to the same dimer configuration on the lattice. This degree of flexibility in the order of the indices is helpful when the assignment of signs to edges is discussed.

The relationship between an arbitrary dimer configuration and the standard configuration can be represented by a *superposition diagram*. In such a diagram full and broken lines respectively are used to denote the dimers of the arbitrary and standard configurations. There will thus be exactly one full and one broken line

⁴ In order to achieve the positive sign it may be necessary to change the definition of $\mathbf{Z}$ by multiplying the first row and column by -1 .

of the sites of the small polygon in Fig. 13.1, the transpositions from M_0 to M are

$$(9, 10, 16, 15) \rightarrow (9, 10, 15, 16) \rightarrow (9, 15, 10, 16) \rightarrow (15, 9, 10, 16),$$

yielding a factor $(-1)^3 = -1$. Thus

$$\epsilon(M) = (-1)^s \epsilon(M_0), \quad (13.2.9)$$

where s is the number of superposition polygons in the diagram, which can be even or odd. Equation (13.2.5) applies if the product (13.2.6) and the similar product for M_0 have the same sign. For this, a sufficient condition is that, for each superposition polygon,

$$\begin{aligned} & \text{Sign}\{z(\mu_1, \mu_2)z(\mu_3, \mu_4) \cdots z(\mu_{2n-1}, \mu_{2n})\} \\ &= -\text{Sign}\{z(\mu_{2n}, \mu_1)z(\mu_2, \mu_3) \cdots z(\mu_{2n-2}, \mu_{2n-1})\}, \end{aligned} \quad (13.2.10)$$

since each factor (-1) in (13.2.9) is then neutralized by another factor (-1) . Condition (13.2.10) can be re-expressed in the form

$$\text{Sign}\{z(\mu_1, \mu_2)z(\mu_2, \mu_3)z(\mu_3, \mu_4) \cdots z(\mu_{2n-1}, \mu_{2n})z(\mu_{2n}, \mu_1)\} = -1, \quad (13.2.11)$$

where $(\mu_1, \mu_2), (\mu_2, \mu_3), \dots, (\mu_{2n}, \mu_1)$ specify the edges, taken in anticlockwise order, of any lattice polygon which can be a superposition polygon.⁵ Condition (13.2.11) is sufficient to ensure that all the terms $\mathcal{P}(\mathbf{Z}; \Pi)$ in (13.2.1) are positive and (13.1.8) is then valid.

We can now develop a topological criterion for ascribing appropriate signs to the non-zero $z(i, j) = \pm \lambda(i, j)$. Put an arrow on the edge connecting sites i and j which indicates the order of the indices for a positive factor. Equation (13.2.11) is equivalent to the condition that any superposition polygon is clockwise odd (Sect. 16.1.4). Suppose that an elementary cycle c encloses a subgraph g with $f(g)$ faces, bounded by face cycles c_k , and $n_i(g)$ interior sites. The number ω_k of clockwise arrows on c_k and the number ω of clockwise arrows on c are related by (16.1.10). For a superposition polygon $n_i(g)$ is an even number since it encloses only other superposition polynomials and doubly-connected lines. It follows that ω is odd if all ω_k are odd. When the lattice $\mathcal{N}$ has arrows indicating the signs of the elements of $\mathbf{Z}$ oriented in such a way that all elementary polygons are clockwise odd, then any superposition polygon is clockwise odd and the consequential terms of the Pfaffian are all positive. Thus:

Theorem 13.2.2 *If arrows are assigned to the first-neighbour pairs of $\mathcal{N}$ in such a way that every face cycle is clockwise odd then the assignment of signs to the elements of $z(i, j)$ such that $z(i, j) > 0$ or $z(i, j) < 0$, according as the arrow along (i, j) is in the direction $i \rightarrow j$ or $j \rightarrow i$, is that which satisfies (13.1.8).*

⁵ Provided consistency is maintained, either anticlockwise or clockwise order can be used.

Step (ii) is now the evaluation of the determinant of the $N \times N$ matrix $\mathbf{Z}$. Since N is large this is not an obviously simple task. However, in the case where, as in (3.9.1), the fugacities are the same along each principle lattice direction and periodic boundary conditions are applied, the signs of the elements of $\mathbf{Z}$ can be chosen periodically. Then $\mathbf{Z}$ becomes a cyclic matrix and in Sect. 17.7 a procedure is described for evaluating the determinant of a cyclic $N \times N$ matrix in the limit $N \rightarrow \infty$. We now consider the cases of the square and honeycomb lattices.

13.2.1 The Square-Lattice Case

On the square lattice $\mathcal{N}_s$, $\lambda(i, j) = \mathfrak{z}_1$ and $\lambda(i, j) = \mathfrak{z}_2$, for all horizontal and vertical first-neighbour pairs respectively, with corresponding elements of $\mathbf{Z}$, $z(i, j) = \pm \mathfrak{z}_1$ and $z(i, j) = \pm \mathfrak{z}_2$. The signs are chosen to match an arrow configuration which makes every face cycle clockwise odd. For $\text{Det}\{\mathbf{Z}\}$ to be evaluated the arrow assignment must also be periodic and a suitable one for the square lattice is shown in Fig. 13.2a. With lattice spacing a , and disregarding boundary effects, a translation by a multiple of a in the vertical direction takes the oriented lattice into itself. However, because of the alternating arrow sense on the vertical edges, a translation in the horizontal direction must be by an even multiple of a to achieve this result. The unit cell of the oriented lattice is taken to be a $2a \times a$ rectangle, containing a horizontal first-neighbour pair of sites; such a cell is shown in Fig. 13.2b. The aim, following the general lines of Montroll (1964), is to express $\mathbf{Z}$ as a cyclic block matrix, as defined in Sect. 17.7. Suppose that the lattice consists of N_2 rows each of $2N_1$ sites, so that there are N_2 rows each of N_1 unit cells. The cells are now specified using cartesian vectors of the type $\mathbf{j} := (j_x, j_y)$, where j_x and j_y are integers and the position of

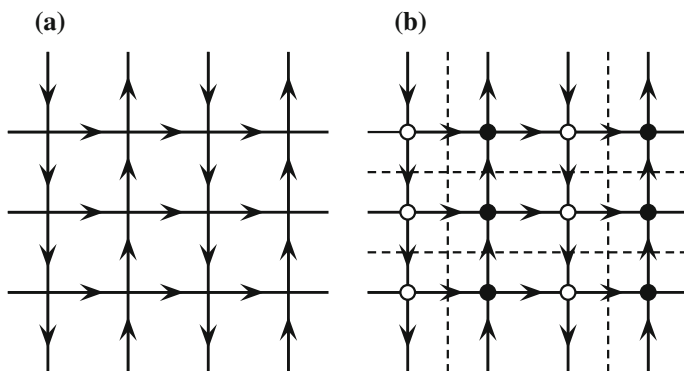


Fig. 13.2 **a** An arrow assignment for the square lattice $\mathcal{N}_s$. **b** A unit cell and sites in the neighbouring unit cells for $\mathcal{N}_s$. The cell boundaries are marked by broken lines and L and R sites are represented by ● and ○ respectively

the centre of cell j is $(2aj_x, aj_y)$. The (scalar) labels of the left and right sites of cell j are j_L and j_R respectively. Let

$$z(j, \ell) := \begin{pmatrix} z(j_L, \ell_L) & z(j_L, \ell_R) \\ z(j_R, \ell_L) & z(j_R, \ell_R) \end{pmatrix}. \quad (13.2.12)$$

Since all cells are identical, $z(j, \ell)$ depends only on the relative positions of j and ℓ and thus

$$z(j, \ell) = z(\ell - j). \quad (13.2.13)$$

From (3.9.3) and (13.1.7), $z(j, \ell) \neq 0$ only if there are first-neighbour links between sites in cells j and ℓ . From Fig. 13.2b it can be seen that the only non-zero blocks are

$$\begin{aligned} z(0) &= \begin{pmatrix} 0 & \mathfrak{z}_1 \\ -\mathfrak{z}_1 & 0 \end{pmatrix}, \\ z(\hat{x}) &= \begin{pmatrix} 0 & 0 \\ \mathfrak{z}_1 & 0 \end{pmatrix}, \quad z(-\hat{x}) = \begin{pmatrix} 0 & -\mathfrak{z}_1 \\ 0 & 0 \end{pmatrix}, \\ z(\hat{y}) &= \begin{pmatrix} \mathfrak{z}_2 & 0 \\ 0 & -\mathfrak{z}_2 \end{pmatrix}, \quad z(-\hat{y}) = \begin{pmatrix} -\mathfrak{z}_2 & 0 \\ 0 & \mathfrak{z}_2 \end{pmatrix}, \end{aligned} \quad (13.2.14)$$

where $\hat{x}$ and $\hat{y}$ are respectively the unit horizontal and vertical vectors. Periodic boundary conditions are now imposed on the right and left and top and bottom of the lattice so that

$$z(j + n) = z(j), \quad n := (n_1 N_1, n_2 N_2), \quad (13.2.15)$$

where n_1 and n_2 are any integers. Equations (13.2.13) and (13.2.15) imply that conditions (17.7.8) are satisfied and that Z is a cyclic block matrix with blocks of dimension 2×2 . The determinant of Z can thus be evaluated using the method of Sect. 17.7. The disadvantage is that extra ‘torus-winding’ terms have been introduced into Z and, since the discussion in Sect. 13.2 depends on the lattice being planar, this approach cannot be used for finite lattices. However, it may be assumed that the boundary conditions will not affect densities in the thermodynamic limit $N_1 \rightarrow \infty$ and $N_2 \rightarrow \infty$.⁶ Using (17.7.10) with $\theta_1 := 2\pi k_1/N_1$ and $\theta_2 := 2\pi k_2/N_2$,

⁶ This is proved by Montroll (1964), Sect. 4.8, using Ledermann’s theorem.

$$\begin{aligned}
\zeta(\theta_1, \theta_2) &= z(\mathbf{0}) + z(\hat{x}) \exp(i\theta_1) + z(-\hat{x}) \exp(-i\theta_1) \\
&\quad + z(\hat{y}) \exp(i\theta_2) + z(-\hat{y}) \exp(-i\theta_2) \\
&= \begin{pmatrix} \mathfrak{Z}_2 \exp(i\theta_2) - \mathfrak{Z}_2 \exp(-i\theta_2) & \mathfrak{Z}_1 - \mathfrak{Z}_1 \exp(-i\theta_1) \\ -\mathfrak{Z}_1 + \mathfrak{Z}_1 \exp(i\theta_1) & \mathfrak{Z}_2 \exp(-i\theta_2) - \mathfrak{Z}_2 \exp(i\theta_2) \end{pmatrix}
\end{aligned} \tag{13.2.16}$$

and thus

$$\text{Det}\{\zeta(\theta_1, \theta_2)\} = 4 \left[\mathfrak{Z}_1^2 \sin^2 \left(\frac{1}{2} \theta_1 \right) + \mathfrak{Z}_2^2 \sin^2 (\theta_2) \right]. \tag{13.2.17}$$

From (13.1.8), (13.2.17) and (17.7.12) the dimensionless free energy per lattice site in the limit $N = 2N_1N_2 \rightarrow \infty$, when the finite square lattice $\mathcal{N}_s \rightarrow \mathcal{L}_s$, is given by

$$\begin{aligned}
\phi(\mathcal{L}_s, \mathfrak{Z}_1, \mathfrak{Z}_2) &= - \lim_{N_1 \rightarrow \infty} \lim_{N_2 \rightarrow \infty} \frac{\ln[Z(\mathcal{N}_s, \mathfrak{Z}_1, \mathfrak{Z}_2)]}{2N_1N_2} \\
&= -\frac{1}{4} \lim_{N_1 \rightarrow \infty} \lim_{N_2 \rightarrow \infty} \frac{\ln[\text{Det}\{\mathbf{Z}\}]}{N_1N_2} \\
&= -\frac{1}{(4\pi)^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln \left\{ 4 \left[\mathfrak{Z}_1^2 \sin^2 \left(\frac{1}{2} \theta_1 \right) + \mathfrak{Z}_2^2 \sin^2 (\theta_2) \right] \right\}.
\end{aligned} \tag{13.2.18}$$

With the change of variable $\theta_2 \rightarrow \frac{1}{2}\theta_2$, this result can be expressed in the more symmetric form

$$\begin{aligned}
\phi(\mathcal{L}_s, \mathfrak{Z}_1, \mathfrak{Z}_2) &= -\frac{1}{2} \ln(2) \\
&\quad + \frac{1}{(4\pi)^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[g(\mathfrak{Z}_1, \mathfrak{Z}_2; \cos(\theta_1), \cos(\theta_2))],
\end{aligned} \tag{13.2.19}$$

where

$$g(\mathfrak{Z}_1, \mathfrak{Z}_2; x_1, x_2) := [\mathfrak{Z}_1^2 + \mathfrak{Z}_2^2 - x_1\mathfrak{Z}_1^2 - x_2\mathfrak{Z}_2^2]^{-1}. \tag{13.2.20}$$

In the thermodynamic limit the numbers ρ_1 and ρ_2 of horizontal and vertical dimers respectively, per lattice site, are given by

$$\rho_i := -\mathfrak{Z}_i \frac{\partial \phi}{\partial \mathfrak{Z}_i} \tag{13.2.21}$$

$$= \frac{1}{2\pi^2} \int_0^\pi d\theta_1 \int_0^\pi d\theta_2 g(\mathfrak{Z}_1, \mathfrak{Z}_2; \cos(\theta_1), \cos(\theta_2)), \quad i = 1, 2. \tag{13.2.22}$$

It can be seen, from these formulae, that $\rho_1 + \rho_2 = \frac{1}{2}$ is necessary for consistency. Using the standard result

$$\int_0^\pi \frac{d\theta}{y - x \cos(\theta)} = \frac{\pi}{\sqrt{y^2 - x^2}},$$

this gives

$$\begin{aligned} \rho_i &= \frac{1}{2\pi} \int_0^\pi \frac{\eta_i^2 [1 - \cos(\theta)]}{\sqrt{[\eta_i^2 + 1 - \eta_i^2 \cos(\theta)]^2 - 1}} d\theta \\ &= \int_0^\pi \frac{\eta_i \sqrt{1 - \cos(\theta)}}{\sqrt{\eta_i^2 + 2 - \eta_i^2 \cos(\theta)}} d\theta, \quad i = 1, 2, \end{aligned} \quad (13.2.23)$$

where $\eta_1 := \mathfrak{z}_1/\mathfrak{z}_2$ and $\eta_2 := 1/\eta_1 := \mathfrak{z}_2/\mathfrak{z}_1$. Using the substitution $u = 1 - \cos \theta$

$$\rho_i = \frac{1}{2\pi} \int_0^2 \frac{du}{\sqrt{(2-u)(2\eta_i^{-2} + u)}} = \frac{\arctan(\eta_i)}{\pi}. \quad (13.2.24)$$

This remarkably simple result is due to Fisher (1961). When $\mathfrak{z}_1 = 0$ all the dimers are vertical $Z(\mathcal{N}_s, 0, \mathfrak{z}_2) = \mathfrak{z}_2^{N/2}$, $\phi(\mathcal{L}_s, 0, \mathfrak{z}_2) = -\frac{1}{2} \ln(\mathfrak{z}_2)$ and similarly, when $\mathfrak{z}_2 = 0$, all the dimers are horizontal and $\phi(\mathcal{L}_s, \mathfrak{z}_1, 0) = -\frac{1}{2} \ln(\mathfrak{z}_1)$. Using these results with (13.2.21) and (13.2.24), we obtain the alternative forms

$$\phi(\mathcal{L}_s, \mathfrak{z}_1, \mathfrak{z}_2) = -\frac{1}{2} \ln(\mathfrak{z}_2) - \frac{1}{\pi} \int_0^{\mathfrak{z}_1/\mathfrak{z}_2} \frac{\arctan(\eta) d\eta}{\eta}, \quad (13.2.25)$$

$$\phi(\mathcal{L}_s, \mathfrak{z}_1, \mathfrak{z}_2) = -\frac{1}{2} \ln(\mathfrak{z}_1) - \frac{1}{\pi} \int_0^{\mathfrak{z}_2/\mathfrak{z}_1} \frac{\arctan(\eta) d\eta}{\eta}. \quad (13.2.26)$$

It follows from these equations that $\phi(\mathcal{L}_s, \mathfrak{z}_1, \mathfrak{z}_2)$ is a continuous function of $\mathfrak{z}_1$ and $\mathfrak{z}_2$ with continuous derivatives of all orders except at $\mathfrak{z}_1 = \mathfrak{z}_2 = 0$. There can, therefore, be no phase transitions. From (3.9.1) the number of ways $\Omega(N)$ of distributing the $\frac{1}{2}N$ dimers on the lattice $\mathcal{N}_s$ of N sites is

$$\Omega(N) = \sum_{\{\mathbf{m}_1 + \mathbf{m}_2 = \frac{1}{2}N\}} \Omega(\mathcal{N}_s; \mathbf{m}_1, \mathbf{m}_2) = Z(\mathcal{N}_s, 1, 1). \quad (13.2.27)$$

For large N , from (13.2.25) or (13.2.26),

$$\frac{\ln \Omega(N)}{N} = -\phi(\mathcal{L}_s, 1, 1) = \frac{1}{\pi} \int_0^1 \frac{\arctan(\eta) d\eta}{\eta}. \quad (13.2.28)$$

Expanding the integrand in a series, valid for $\eta \leq 1$, and integrating gives

$$\frac{\ln \Omega(N)}{N} = \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} = \frac{G}{\pi}, \quad (13.2.29)$$

where $G := 0.915966$ is *Catalan's constant*.

The partition function for the finite $2N_1 \times N_2$ lattice was obtained by Fisher (1961) and Kasteleyn (1961). For even N_2

$$Z(\mathcal{N}_s, \mathfrak{z}_1, \mathfrak{z}_2) = 2^{\frac{1}{2}N} \prod_{k_1=1}^{N_1} \prod_{k_2=1}^{\frac{1}{2}N_2} \left[\mathfrak{z}_1^2 \cos^2 \left(\frac{k_1 \pi}{2N_1 + 1} \right) + \mathfrak{z}_2^2 \cos^2 \left(\frac{k_2 \pi}{N_2 + 1} \right) \right] \quad (13.2.30)$$

in the present notation. Fisher used an alternative combinatorial method to evaluate the Pfaffian, while Kasteleyn, by some ingenious algebra, was able to dispense with the periodic condition (13.2.15). The formula (13.2.19) for the free energy per site follows from (13.2.30) in the thermodynamic limit, confirming that in this limit intensive quantities are independent of the original boundary conditions. By putting $\mathfrak{z}_1 = \mathfrak{z}_2 = 1$ in (13.2.30), values of $\Omega(N)$ for finite lattices can be derived. The number of ways of arranging 32 symmetric dominos on a chess board corresponds to taking $2N_1 = N_2 = 8$ and gives $\Omega(64) = 12,988,816$ which is surprisingly large. This gives $\ln \Omega(64)/64 = 0.25593$, to be compared with the value 0.291561 obtained from (13.2.29) in the thermodynamic limit.

13.2.2 The Honeycomb-Lattice Case

An arrow assignment for the honeycomb lattice $\mathcal{N}_H$ is shown in Fig. 13.3. There are clockwise arrows on three of the six edges forming an elementary cycle so that the clockwise odd condition is satisfied. The broken lines denote the boundary of a unit cell containing a first-neighbour pair of sites. The notation for the representation of L and R sites is the same as in Fig. 13.2b. To present the cell structure as clearly as possible the usual geometry of the lattice has been distorted, although parallel edges remain parallel. This does not affect the evaluation of the partition function which depends only on the topology of the vertex connections and the fugacities assigned to the lattice directions. These are shown on the bottom right of the figure.

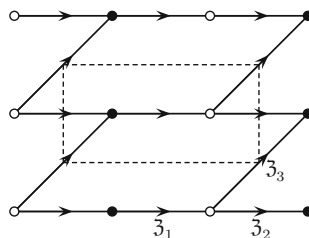


Fig. 13.3 An arrow assignment for the honeycomb lattice $\mathcal{N}_H$. The unit cell boundaries are marked by broken lines and L and R sites are represented by ● and ○ respectively

The lattice topology determines an important conservation property of the honeycomb dimer configurations. It can be seen from Fig. 3.11 that the occupation of any member of a row of horizontal edges by a dimer implies that exactly one of the rows of diagonal edges below is unoccupied. So if there are m_1 and m_2 dimers of species 1 and 2 respectively in a given row of horizontal edges then there will be $m_3 = N_1 - m_1 - m_2$ dimers in the row of diagonal edges above and the same number m_3 in the row of diagonal edges below. The number m_3 is thus the same for all rows of diagonal edges. Since the three sets of parallel edges indexed by the numbers 1, 2 and 3 are topologically equivalent a similar conservation rule applies to dimers of species 1 and 2. In the geometry of Fig. 13.3 there is thus the same number of dimers in each column of type 1 edges and in each column of type 2 edges.⁷

The analysis of the honeycomb lattice is similar to that for the square lattice, although $z(\pm\hat{y}) = 0$, since, from Fig. 13.3, there are no first-neighbour links between the sites of a given cell and those of the cells immediately above and below it. However, there are diagonal edges, in the direction $\hat{d} := (\hat{x} + \hat{y})/\sqrt{2}$, connecting the R site of a given cell with the L site of the cell above and to the right and the L site to the R site of cell below and to the left. The non-zero blocks are, therefore,

$$\begin{aligned} z(\mathbf{0}) &= \begin{pmatrix} 0 & \mathfrak{z}_1 \\ -\mathfrak{z}_1 & 0 \end{pmatrix}, \\ z(\hat{x}) &= \begin{pmatrix} 0 & 0 \\ \mathfrak{z}_2 & 0 \end{pmatrix}, \quad z(-\hat{x}) = \begin{pmatrix} 0 & -\mathfrak{z}_2 \\ 0 & 0 \end{pmatrix}, \\ z(\hat{d}) &= \begin{pmatrix} 0 & 0 \\ \mathfrak{z}_3 & 0 \end{pmatrix}, \quad z(-\hat{d}) = \begin{pmatrix} 0 & -\mathfrak{z}_3 \\ 0 & 0 \end{pmatrix}. \end{aligned} \quad (13.2.31)$$

From (17.7.10),

$$\begin{aligned} \zeta(\theta_1, \theta_2) &= z(\mathbf{0}) + z(\hat{x}) \exp(i\theta_1) + z(-\hat{x}) \exp(-i\theta_1) \\ &\quad + z(\hat{d}) \exp(i[\theta_1 + \theta_2]) + z(-\hat{d}) \exp(-i[\theta_1 + \theta_2]), \end{aligned} \quad (13.2.32)$$

and thus

$$\begin{aligned} \text{Det}\{\zeta(\theta_1, \theta_2)\} &= \mathfrak{z}_1^2 + \mathfrak{z}_2^2 + \mathfrak{z}_3^2 - 2\mathfrak{z}_1\mathfrak{z}_2 \cos(\theta_1) \\ &\quad + 2\mathfrak{z}_2\mathfrak{z}_3 \cos(\theta_2) - 2\mathfrak{z}_1\mathfrak{z}_3 \cos(\theta_1 + \theta_2). \end{aligned} \quad (13.2.33)$$

As in the case of the square lattice this yields the dimensionless free-energy density

$$\begin{aligned} \phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) &= -\frac{1}{(4\pi)^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[\mathfrak{z}_1^2 + \mathfrak{z}_2^2 + \mathfrak{z}_3^2 - 2\mathfrak{z}_1\mathfrak{z}_2 \cos(\theta_1) \\ &\quad + 2\mathfrak{z}_2\mathfrak{z}_3 \cos(\theta_2) - 2\mathfrak{z}_1\mathfrak{z}_3 \cos(\theta_1 + \theta_2)]. \end{aligned} \quad (13.2.34)$$

⁷ It can be seen from Fig. 3.10 that similar conservation rules do not apply to the square lattice. The reason is that there the number of vacant edges in a row of vertical edges depends not only on the number of dimers in the row of horizontal edges immediately below, but also on the number of dimers in the row of vertical edges below that.

With the change of variable $\theta_1 \rightarrow \pi - \theta_1$ this becomes

$$\begin{aligned} \phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = & -\frac{1}{(4\pi)^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[\mathfrak{z}_1^2 + \mathfrak{z}_2^2 + \mathfrak{z}_3^2 + 2\mathfrak{z}_1\mathfrak{z}_2 \cos(\theta_1) \\ & + 2\mathfrak{z}_2\mathfrak{z}_3 \cos(\theta_2) + 2\mathfrak{z}_1\mathfrak{z}_3 \cos(\theta_2 - \theta_1)], \end{aligned} \quad (13.2.35)$$

a result given, without derivation, by Wu (1968). As would be expected, this expression is invariant with respect permutations of the fugacities.

So far the analysis has been very similar to that for the square lattice. However, there are important differences in the thermodynamics of the two systems, as appears when (13.2.35) is transformed to a single integral expression. This takes the form

$$\begin{aligned} \phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = & -\frac{1}{(4\pi)} \int_0^\pi \ln \{ Y^{(+)}(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3; \cos(\theta)) \\ & + |Y^{(-)}(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3; \cos(\theta))| \} d\theta, \end{aligned} \quad (13.2.36)$$

where

$$Y^{(\pm)}(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3; x) := \frac{1}{2}[\mathfrak{z}_1^2 + \mathfrak{z}_2^2 - 2x \mathfrak{z}_1\mathfrak{z}_2 \pm \mathfrak{z}_3^2]. \quad (13.2.37)$$

The fugacities $\mathfrak{z}_1, \mathfrak{z}_2$ and $\mathfrak{z}_3$ are non-negative and, in order to understand the behaviour of this model, we need to determine the analytic nature of $\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3)$ in this octant of the fugacity space, which can be divided into four regions:

- I: $\mathfrak{z}_1 > \mathfrak{z}_2 + \mathfrak{z}_3, \mathfrak{z}_2 < \mathfrak{z}_3 + \mathfrak{z}_1, \mathfrak{z}_3 < \mathfrak{z}_1 + \mathfrak{z}_2,$
- II: $\mathfrak{z}_2 > \mathfrak{z}_3 + \mathfrak{z}_1, \mathfrak{z}_3 < \mathfrak{z}_1 + \mathfrak{z}_2, \mathfrak{z}_1 < \mathfrak{z}_2 + \mathfrak{z}_3,$
- III: $\mathfrak{z}_3 > \mathfrak{z}_1 + \mathfrak{z}_2, \mathfrak{z}_1 < \mathfrak{z}_2 + \mathfrak{z}_3, \mathfrak{z}_2 < \mathfrak{z}_3 + \mathfrak{z}_1,$
- IV: $\mathfrak{z}_1 < \mathfrak{z}_2 + \mathfrak{z}_3, \mathfrak{z}_2 < \mathfrak{z}_3 + \mathfrak{z}_1, \mathfrak{z}_3 < \mathfrak{z}_1 + \mathfrak{z}_2.$

The boundaries between regions I, II and III respectively and region IV are the surfaces:

$$\mathcal{C}(\text{I}) : \quad \mathfrak{z}_1 = \mathfrak{z}_2 + \mathfrak{z}_3, \quad (13.2.38)$$

$$\mathcal{C}(\text{II}) : \quad \mathfrak{z}_2 = \mathfrak{z}_3 + \mathfrak{z}_1, \quad (13.2.39)$$

$$\mathcal{C}(\text{III}) : \quad \mathfrak{z}_3 = \mathfrak{z}_1 + \mathfrak{z}_2. \quad (13.2.40)$$

For all $\theta \in [0, \pi]$, $Y^{(-)}(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3; \cos(\theta)) < 0$ in region III and, from (13.2.21) and (13.2.36),

$$\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = -\frac{1}{2} \ln(\mathfrak{z}_3), \quad \rho_1 = \rho_2 = 0, \quad \rho_3 = \frac{1}{2}. \quad (13.2.41)$$

So in region III the system is in a ‘frozen’ state of perfect order, with all dimers parallel in the direction associated with the fugacity $\mathfrak{z}_3$. By symmetry it follows that states of perfect order also exist with the dimers aligned in the other lattice directions. In region I,

$$\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = -\frac{1}{2} \ln(\mathfrak{z}_1), \quad \rho_1 = \frac{1}{2}, \quad \rho_2 = \rho_3 = 0, \quad (13.2.42)$$

and, in region II,

$$\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = -\frac{1}{2} \ln(\mathfrak{z}_2), \quad \rho_1 = \rho_3 = 0, \quad \rho_2 = \frac{1}{2}. \quad (13.2.43)$$

These frozen configurations do not admit small (localized) perturbations. Consider, for example, the case of the system in region III, where, in the geometry of Fig. 13.3, all horizontal edges are unoccupied, while all edges in each row of diagonal sites are occupied giving $m_3 = N_1$ for all such rows. Now suppose that one dimer of any row of diagonal dimers is displaced to a horizontal edge. By the conservation rule discussed above, there must be a similar displacement from all other rows of diagonal dimers, so that the disturbance is propagated throughout the lattice. The situation is similar to that in a ferroelectric ground state of the six-vertex model (Sect. 3.8.2). The honeycomb lattice situation contrasts with that on the square lattice, where there can be small disturbances in the basic configuration or in the corresponding configuration with all dimers vertical. Figure 3.10 can, for instance, be regarded as showing a perturbation in a ‘corner’ of a larger lattice, with all other dimers on this lattice in their basic configuration positions.

Consider the case where the system is in region IV. There is now a real value θ_0 in the range $(0, \pi)$ such that

$$Y^{(-)}(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3; \cos(\theta_0)) = 0. \quad (13.2.44)$$

The range of integration in (13.2.36) can be split into two parts and, using (13.2.37),

$$\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) = -\frac{\theta_0}{2\pi} \ln(\mathfrak{z}_3) - \frac{1}{4\pi} \int_{\theta_0}^{\pi} \ln[\mathfrak{z}_1^2 + \mathfrak{z}_2^2 - 2\mathfrak{z}_1\mathfrak{z}_2 \cos(\theta)] d\theta. \quad (13.2.45)$$

From this result it follows that

$$\rho_3 = \frac{\theta_0}{2\pi} = \frac{1}{2\pi} \arccos\left(\frac{\mathfrak{z}_1^2 + \mathfrak{z}_2^2 - \mathfrak{z}_3^2}{2\mathfrak{z}_1\mathfrak{z}_2}\right), \quad (13.2.46)$$

and by symmetry

$$\rho_1 = \frac{1}{2\pi} \arccos\left(\frac{\mathfrak{z}_2^2 + \mathfrak{z}_3^2 - \mathfrak{z}_1^2}{2\mathfrak{z}_2\mathfrak{z}_3}\right), \quad (13.2.47)$$

$$\rho_2 = \frac{1}{2\pi} \arccos\left(\frac{\mathfrak{z}_3^2 + \mathfrak{z}_1^2 - \mathfrak{z}_2^2}{2\mathfrak{z}_3\mathfrak{z}_1}\right). \quad (13.2.48)$$

It can easily be verified that

$$\rho_1 + \rho_2 + \rho_3 = \frac{1}{2}. \quad (13.2.49)$$

From (13.2.37) and (13.2.44), as the phase point approaches the surface $\mathcal{C}(\text{III})$, $\theta_0 \rightarrow \pi$. It then follows from (13.2.46) that $\rho_3 \rightarrow \frac{1}{2}$ and it must be the case that $\rho_2 \rightarrow 0$ and $\rho_1 \rightarrow 0$. We can also see that $\phi(\mathcal{L}_H, \mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3)$, given by (13.2.45), tends to the value $-\frac{1}{2} \ln \mathfrak{z}_3$. The dimer densities and the free-energy density are continuous over the boundary $\mathcal{C}(\text{III})$ between the regions III and IV. They are also continuous over the boundaries $\mathcal{C}(\text{I})$ and $\mathcal{C}(\text{II})$. From (13.2.41)–(13.2.43), in the ordered regions I–III,

$$\frac{\partial \rho_i}{\partial \mathfrak{z}_j} = 0, \quad i, j = 1, 2, 3, \quad (13.2.50)$$

and, from (13.2.46)–(13.2.48), in the disordered region IV,

$$\frac{\partial \rho_i}{\partial \mathfrak{z}_j} = \begin{cases} \frac{\mathfrak{z}_i}{\tau(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3)}, & \text{if } i = j, \\ \frac{\mathfrak{z}_k^2 - \mathfrak{z}_i^2 - \mathfrak{z}_j^2}{2\mathfrak{z}_j \tau(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3)}, & \text{if } i \neq j \text{ and } k \neq i, j, \end{cases} \quad (13.2.51)$$

where

$$\begin{aligned} & \tau(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3) \\ & := \pi \sqrt{(\mathfrak{z}_1 + \mathfrak{z}_2 - \mathfrak{z}_3)(\mathfrak{z}_2 + \mathfrak{z}_3 - \mathfrak{z}_1)(\mathfrak{z}_3 + \mathfrak{z}_1 - \mathfrak{z}_2)(\mathfrak{z}_1 + \mathfrak{z}_2 + \mathfrak{z}_3)}. \end{aligned} \quad (13.2.52)$$

In view of condition (13.2.49) only six of these nine partial derivatives are independent. From (13.2.38)–(13.2.40) and (13.2.52), it can be seen that, as any one of the boundaries of the disordered region is approached from within that region, the partial derivatives given by (13.2.51) exhibit singular behaviour. Specifically, as surface $\mathcal{C}(\text{I})$ is approached within region IV,

$$\frac{\partial \rho_i}{\partial \mathfrak{z}_j} \sim \frac{1}{\sqrt{\mathfrak{z}_2 + \mathfrak{z}_3 - \mathfrak{z}_1}}, \quad i, j = 1, 2, 3. \quad (13.2.53)$$

$\mathcal{C}(\text{I})$, $\mathcal{C}(\text{II})$ and $\mathcal{C}(\text{III})$ are surfaces of second-order phase transitions. This result was stated, without derivation, by Kasteleyn (1963).

Consider now the case where the fugacities $\mathfrak{z}_i$ are smooth functions of couplings $K_i := \varepsilon_i/T$, for $i = 1, 2, 3$. The system is controlled by changing the temperature, so we adopt the procedure used in Sect. 2.3.1, setting $K := K_1$ with $\kappa_i := \varepsilon_i/\varepsilon_1$, so that $K_i = \kappa_i K$, $i = 1, 2, 3$. The critical temperature T_c at which the system passes from the disordered state IV to the ordered state I is then given as the solution of

$$\mathfrak{z}_2(T_c) + \mathfrak{z}_3(T_c) = \mathfrak{z}_1(T_c), \quad (13.2.54)$$

and, inside IV and close to the surface $\mathcal{C}(\text{I})$,

$$\mathfrak{Z}_2(T) + \mathfrak{Z}_3(T) - \mathfrak{Z}_1(T) \simeq [\mathfrak{Z}'_2(T_c) + \mathfrak{Z}'_3(T_c) - \mathfrak{Z}'_1(T_c)](T - T_c), \quad (13.2.55)$$

unless

$$\mathfrak{Z}'_2(T_c) + \mathfrak{Z}'_3(T_c) = \mathfrak{Z}'_1(T_c). \quad (13.2.56)$$

From (2.3.18), (2.3.19) and (13.2.21), the internal-energy density

$$u = \sum_{i=1}^3 \kappa_i \frac{\partial \phi}{\partial K_i} = \sum_{i=1}^3 \kappa_i \frac{\partial \phi}{\partial \mathfrak{Z}_i} \frac{d\mathfrak{Z}_i}{dK_i} = - \sum_{i=1}^3 \kappa_i \rho_i \frac{d \ln \mathfrak{Z}_i}{dK_i}. \quad (13.2.57)$$

Since the dimer densities ρ_i are continuous across the critical surfaces this must also be the case for u and, from (13.2.41)–(13.2.43),

$$u = -\frac{1}{2} \kappa_i \frac{d \ln \mathfrak{Z}_i}{dK_i}, \quad i = 1, 2, 3, \quad (13.2.58)$$

in the ordered regions I, II and III respectively. The heat capacity c per lattice site, which is given from (2.3.29) by

$$c = k_B \varepsilon_1 \frac{\partial u}{\partial T}, \quad (13.2.59)$$

will involve the partial derivatives of the dimer densities with respect to the fugacities. It follows from (13.2.53) and (13.2.55) that

$$c \sim \frac{1}{\sqrt{T - T_c}}, \quad (13.2.60)$$

as $\mathcal{C}(I)$ is approached from inside region IV, unless (13.2.56) is satisfied. This gives a critical exponent $\alpha = \frac{1}{2}$.

We now consider a simple interpretation of the fugacities. Suppose that each dimer carries a small electric dipole aligned with its axis and that an electric field $\mathcal{E}$ acts on the system parallel to the direction labelled 3. Given that $\mathcal{E}$ includes the magnitude of the dipole moment, thereby being measured in units of energy, and that dipole-dipole interactions are negligible in comparison with the dipole-field interaction, this model corresponds to the case $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \mathcal{E}$ and

$$\mathfrak{Z}_3 = \exp(K) + \exp(-K) = 2 \cosh(K). \quad (13.2.61)$$

In a similar way

$$\mathfrak{Z}_1 = \mathfrak{Z}_2 = 2 \cosh\left(\frac{1}{2}K\right). \quad (13.2.62)$$

Since $\mathfrak{Z}_3 > \mathfrak{Z}_2 = \mathfrak{Z}_1$, a transition at a fixed value of $\mathcal{E}$ will occur on the surface $\mathcal{C}(\text{III})$ at a critical coupling $K_c = \mathcal{E}/T_c$ given by

$$\cosh(K_c) = 2 \cosh\left(\frac{1}{2}K_c\right), \quad (13.2.63)$$

which can be transformed to

$$\frac{\mathcal{E}}{T_c} = 2 \operatorname{arccosh}\left(\frac{1 + \sqrt{3}}{2}\right). \quad (13.2.64)$$

An equivalent relationship (again without derivation) is given by Kasteleyn (1963), who regards the dimers as polar diatomic molecules adsorbed on to a solid whose surface layer of atoms has the honeycomb form. For this model the internal energy in region III is given, from (13.2.58), by

$$u = -\frac{1}{2} \tanh(K). \quad (13.2.65)$$

A second-order transition, where, as in this model, the heat capacity has a square-root singularity for $T > T_c$ and is finite or zero for $T < T_c$ has been given the name *K-type* (K standing for Kasteleyn) by Nagle et al. (1989). These authors also use the name *O-type* (O standing for Onsager) for the kind of transition, which occurs in the two-dimensional Ising model, where the heat capacity has a logarithmic singularity on each side of the critical point (Eq. (12.2.181)). These two types of transition are further discussed in Sect. 13.4.

13.3 Vertex and Ising Model Equivalences

In Sects. 3.9.1 and 3.9.2 equivalences between special cases of dimer models and vertex and Ising models were established. The formulae for the thermodynamic functions of dimer models obtained in this chapter can now be used to investigate this further.

13.3.1 The Five-Vertex Model

Equation (3.9.11) identifies the partition function of a dimer model on the honeycomb lattice with a special case of the six-vertex model on the square lattice which violates the energy equalities (3.8.4) in two respects: (i) ε_1 is finite but $\varepsilon_2 = \infty$, so that the model is effectively a five-vertex model; (ii) $\varepsilon_3 \neq \varepsilon_4$ unless the $\mathfrak{Z}_2 = \mathfrak{Z}_3$. The relabelling of the vertex energies and Boltzmann weights given in (3.9.7) avoids a clash between vertex and dimer labelling and the relationships between these latter

and the dimer fugacities are given by (3.9.8) and (3.9.9). The energy u per lattice site for the five-vertex model on the square lattice of $\frac{1}{2}N$ sites is given, from (13.2.57), by

$$u = 2(\rho_1 \varepsilon_a + \rho_2 \varepsilon_b + \rho_3 \varepsilon_{b'}). \quad (13.3.1)$$

For the special case of the modified KDP model the vertex energy parameters satisfy (3.9.12) and we define $\varepsilon := -\varepsilon_a > 0$, $K := \varepsilon/T$, with

$$\mathfrak{z}_1 = \exp(K), \quad \mathfrak{z}_2 = \mathfrak{z}_3 = 1, \quad (13.3.2)$$

$$u = -2\rho_1 \varepsilon. \quad (13.3.3)$$

The only transition surface relevant to this model is $\mathcal{C}(I)$, which, from (13.2.38) and (13.3.2), gives

$$\exp(K_c) = 2, \quad T_c = \varepsilon / \ln(2), \quad (13.3.4)$$

for the critical temperature.

In region I, when $T < T_c$, $\rho_1 = \frac{1}{2}$ and $u = -\varepsilon$, so that for the five-vertex model every vertex is of type 1. Like the original KDP model, this modified model is in a frozen perfectly ordered ferroelectric ground state below the critical temperature. The argument used in Sect. 3.8.2 that the ferroelectric ground state, in which all vertices are of type 1, does not admit small perturbations still applies here. The only difference is that configuration graphs like those shown in Fig. 3.6 cannot any longer ‘touch’ at a vertex, since such a vertex is of type 2. The equivalent state in the honeycomb dimer model is one in which all the dimers are parallel with fugacity $\mathfrak{z}_1$. It follows that no small perturbations to this state, or the other two fully aligned states, can occur. This confirms the argument of Sect. 13.2.2.

In region IV, when $T > T_c$, it follows from (13.2.47) and (13.3.3) that

$$\rho_1 = \frac{1}{2\pi} \arccos \left[1 - \frac{1}{2} \exp(2K) \right]. \quad (13.3.5)$$

The zero-field heat capacity c_0 per vertex in the modified KDP model is thus given by

$$c_0 = k_B \frac{\partial u}{\partial T} = 2k_B K^2 \frac{\partial \rho_1}{\partial K} = \frac{k_B K^2 \exp(K_1)}{\pi \sqrt{1 - \frac{1}{4} \exp(2K_1)}}, \quad (13.3.6)$$

which exhibits a square-root singularity as $T \rightarrow T_c$. This same singular behaviour is exhibited by the corresponding six-vertex (KDP) model, except in that case there is a finite discontinuity in u at the critical temperature (see Lieb (1967c) and Lavis and Bell (1999a, Sect. 10.10)).

Now suppose that an electric field with horizontal and vertical components $\mathcal{E}^{(h)}$ and $\mathcal{E}^{(v)}$ in the plane of the lattice is applied to the modified KDP model. Given that these quantities include the magnitude of the horizontal or vertical component of the dipole moment of a polarized vertex, the vertex energies given in (3.9.12) are changed to

$$\begin{aligned}\varepsilon_a &= -\varepsilon - \mathcal{E}^{(v)} - \mathcal{E}^{(h)}, & \varepsilon_b &= -\mathcal{E}^{(v)} + \mathcal{E}^{(h)}, \\ \varepsilon_{b'} &= \mathcal{E}^{(v)} - \mathcal{E}^{(h)}, & \varepsilon_c &= 0.\end{aligned}\tag{13.3.7}$$

Since (3.9.10) is still satisfied, the model remains equivalent to the honeycomb dimer model. When

$$-\frac{1}{2}\varepsilon < \mathcal{E}^{(v)}, \quad -\frac{1}{2}\varepsilon < \mathcal{E}^{(h)},\tag{13.3.8}$$

$\varepsilon_a < \varepsilon_b$ and $\varepsilon_a < \varepsilon_{b'}$ and again the only relevant transition surface is $\mathcal{C}(I)$. For fixed values of the electric field components the system will go into a perfectly ordered state with all vertices of type 1 at a critical temperature given by

$$\exp\left(\frac{\varepsilon + \mathcal{E}^{(v)} + \mathcal{E}^{(h)}}{T_c}\right) = 2 \cosh\left(\frac{\mathcal{E}^{(v)} - \mathcal{E}^{(h)}}{T_c}\right).\tag{13.3.9}$$

When

$$\mathcal{E}^{(h)} < -\frac{1}{2}\varepsilon, \quad \mathcal{E}^{(h)} < \mathcal{E}^{(v)},\tag{13.3.10}$$

the transition is to a perfectly ordered state with all vertices of type 3 and when

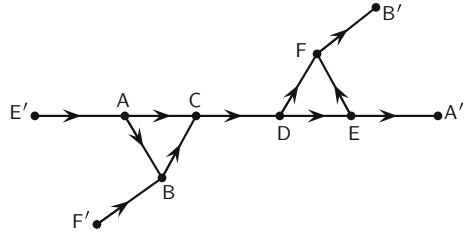
$$\mathcal{E}^{(v)} < -\frac{1}{2}\varepsilon, \quad \mathcal{E}^{(v)} < \mathcal{E}^{(h)},\tag{13.3.11}$$

the transition is to a perfectly ordered state with all vertices of type 4. In each of these cases the calculation of the transition temperature is similar to that given for the case (13.3.8).

13.3.2 The Honeycomb-Lattice Anisotropic Ising Model

Let the honeycomb lattice $\mathcal{N}_H$ consist of a $N_1 \times N_2$ array of unit cells with $N = 2N_1N_2$. From (3.9.20), it is now necessary to find $Z^{(\text{DM})}(\mathcal{N}_{\text{TT}}, v_1^{-1}, v_2^{-1}, v_3^{-1})$ for the 3-12 lattice $\mathcal{N}_{\text{TT}}$ of $3N$ sites whose unit cell is the six sites **A**, **B**, **C**, **D**, **E** and **F** shown in Fig. 13.4 with **A'**, **B'**, **E'** and **F'** being first-neighbour sites in adjacent unit cells. This lattice was derived from the honeycomb lattice and the arrow assignment for the honeycomb edges remains unchanged, while that for the triangle of new edges makes these triangles clockwise odd. The dodecagons are also clockwise odd.

Fig. 13.4 Arrow assignment for the 3-12 lattice $\mathcal{N}_{\text{TT}}$



Equation (13.2.32) can still be used for the diagonal block matrices, but these are now 6×6 , since there are six sites in the unit cell. Taking the sites of the unit cell in the literal order A, B, C, D, E and F it follows, using Fig. 13.4, that

$$\zeta(\theta_1, \theta_2) = \begin{pmatrix} 0 & 1 & 1 & 0 & -\bar{x}_2 & 0 \\ -1 & 0 & 1 & 0 & 0 & -\bar{x}_3 \\ -1 & -1 & 0 & x_1 & 0 & 0 \\ 0 & 0 & -x_1 & 0 & 1 & 1 \\ x_2 & 0 & -1 & 0 & 1 & 1 \\ 0 & x_3 & 0 & -1 & -1 & 0 \end{pmatrix}, \quad (13.3.12)$$

where $x_1 := 1/v_1$, $x_2 := \exp(i\theta_1)/v_2$, $x_3 := \exp[i(\theta_1 + \theta_2)]/v_3$, giving

$$\begin{aligned} \text{Det}\{\zeta(\theta_1, \theta_2)\} &= v_1^{-2} + v_2^{-2} + v_3^{-2} + v_1^{-2}v_2^{-2}v_3^{-2} \\ &\quad - 2v_1^{-1}v_2^{-1}(v_3^{-2} - 1)\cos(\theta_1) - 2v_2^{-1}v_3^{-1}(v_1^{-2} - 1)\cos(\theta_2) \\ &\quad - 2v_3^{-1}v_1^{-1}(v_2^{-2} - 1)\cos(\theta_1 + \theta_2). \end{aligned} \quad (13.3.13)$$

With

$$s_i := \sinh(2K_i) = \frac{2v_i}{1 - v_i^2}, \quad c_i := \cosh(2K_i) = \frac{1 + v_i^2}{1 - v_i^2}, \quad i = 1, 2, 3, \quad (13.3.14)$$

$$\begin{aligned} \text{Det}\{\zeta(\theta_1, \theta_2)\} &= \frac{1}{2}[\sinh(K_1)\sinh(K_2)\sinh(K_3)]^{-2}[c_1c_2c_3 + 1 \\ &\quad - s_1s_2\cos(\theta_1) - s_2s_3\cos(\theta_2) - s_3s_1\cos(\theta_1 + \theta_2)]. \end{aligned} \quad (13.3.15)$$

From (13.1.8), (3.9.20), (13.3.15) and (17.7.12), the dimensionless free energy per lattice site of the honeycomb lattice anisotropic Ising model in zero field is

$$\begin{aligned}
\phi^{(\text{IM})}(\mathcal{L}_\text{H}, K_1, K_2, K_3) &= -\frac{1}{2} \ln[4 \sinh(K_1) \sinh(K_2) \sinh(K_3)] \\
&\quad - \lim_{N_1 \rightarrow \infty} \lim_{N_2 \rightarrow \infty} \frac{\ln[Z(3N, v_1^{-1}, v_2^{-1}, v_3^{-1})]}{2N_1 N_2} \\
&= -\frac{3}{4} \ln(2) - \frac{1}{16\pi^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[c_1 c_2 c_3 + 1 \\
&\quad - s_1 s_2 \cos(\theta_1) - s_2 s_3 \cos(\theta_2) - s_3 s_1 \cos(\theta_1 + \theta_2)]. \tag{13.3.16}
\end{aligned}$$

The star-triangle transformation of Sect. 9.4 can be used to show that the corresponding expression for the triangular lattice is

$$\begin{aligned}
\phi^{(\text{IM})}(\mathcal{L}_\text{T}, K_1, K_2, K_3) &= -\ln(2) - \frac{1}{8\pi^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[c_1 c_2 c_3 + s_1 s_2 s_3 \\
&\quad - s_1 \cos(\theta_1) - s_2 \cos(\theta_2) - s_3 \cos(\theta_1 + \theta_2)]. \tag{13.3.17}
\end{aligned}$$

Setting $K_3 = 0, s_3 = 0, c_3 = 1$ gives (with slight changes of notation) the expression (12.2.173) for the anisotropic Ising model on the square lattice.

13.4 K-Type and O-Type Transitions

Considering for simplicity the isotropic case, we have shown in Sects. 3.9.2 and 13.3.2 that the 3-12 dimer model, with fugacity $\mathfrak{z} = v^{-1} = \coth(K) > 1$ on the original honeycomb edges and $\mathfrak{z} = 1$ on the new triangle edges, is equivalent to the honeycomb lattice Ising model ferromagnet. It follows that the transitions in this dimer model are similar to those in a two-dimensional Ising model ferromagnet, and are thus O-type according to the definition at the end of Sect. 13.2.2. With these fugacities the ground state of the 3-12 dimer model has dimers on the honeycomb edges (for example F'B or CD in Fig. 3.12) but no dimers on the triangle edges. This ground state admits localized perturbations. Dimers can, for example, be moved cyclically from the honeycomb to the triangle edges in any of the dodecagons without affecting the dimer configuration on the rest of the lattice. The dimer on F'B in Fig. 3.12 could be moved to BC, that on CD to DE and so on. We have thus encountered three types of critical behaviour in pure dimer assemblies: K-type in the honeycomb lattice model of Sect. 13.2.2, O-type on the 3-12 lattice and no transition in the square lattice model of Sect. 13.2.1. In the last two cases the ground states admit small perturbations, while for the first case the ground state is frozen with small perturbations excluded, so that the heat capacity is zero below the critical temperature.

K-type critical behaviour is not confined to dimer systems. In the six-vertex ferroelectric model with non-zero electric field there is a second-order transition at a field-dependent critical temperature (Baxter 1982b, Sect. 8.12) with a square-root singularity in the heat capacity above the critical temperature. Below the critical temperature the state is frozen, fully polarized and the heat capacity is zero. The square-root singular behaviour above the critical temperature still applies in zero

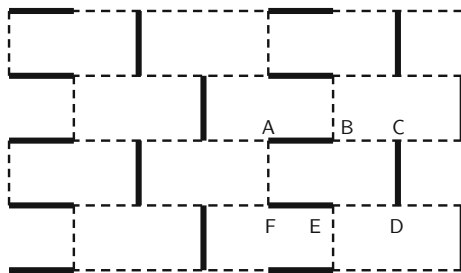


Fig. 13.5 The V_2H_2 dimer model. For the ground state, the positions of the dimers are represented by full lines and the empty edges by broken lines

field, but the transition to the completely polarized state is first-order with a latent heat (Lieb and Wu 1972). From the evidence presented so far it may seem that a square-root singularity in the heat capacity above a critical temperature is associated with a frozen state below the critical temperature. The situation is, however, not quite so simple, as we shall see.

A sequence of honeycomb lattice dimer models is discussed by Nagle et al. (1989). These models have a less simple assignment of fugacities than that for the models of Sect. 13.2.2, where the fugacities corresponded to the three edge orientations in the usual hexagonal geometry. Using the brick lattice geometry, obtained by rotating the diagonal edges into the vertical position, edge types and fugacity assignments are now made according to the arrangement of dimers in a ground state. One model of this type is called V^2H^2 . The ground state for that model is shown in Fig. 13.5, with the full line segments representing the positions of the dimers and the broken lines representing the unoccupied edges. For any arrangement of dimers a fugacity of unity is assigned to a dimer in a ground-state position and a fugacity $\beta = \exp(-\varepsilon/T)$, with $\varepsilon > 0$, to a dimer in any other position. Nagle et al. (1989) also considered dimer models on the 4-8 lattice, formed from a square lattice by replacing each vertex by a quadrangle of edges, and what they called the *SQK model*. In this model extra edges are inserted in the brick lattice to make it into a square lattice, but the dimers on these new vertical edges are given different fugacities from that on the brick lattice edges. Various applications have been made of these types of models, including that to chain conformal systems (Nagle 1975; Bell et al. 1978; Nagle 1986; Nagle et al. 1989).

The V^2H^2 model has some interesting critical properties described by Nagle et al. (1989). First, note that the ground state shown in Fig. 13.5 admits local perturbations. In the hexagon ABCDEF the dimers on AB, CD and EF, can, for example, be shifted to BC, DE and FA, without disturbing the configuration elsewhere on the lattice. Thus, as would be expected, the heat capacity below the critical temperature is not identically zero. It does, however, remain finite as the critical temperature is approached from below and has a square-root singularity as the critical temperature is approached from above. The transition is K-type, where we chose, for the reason

just seen, to provide for the possibility of non-zero subcritical values of the heat capacity.

Similar behaviour is found in six-vertex antiferroelectric models (Lieb and Wu 1972), for which there is a critical electric field $\mathcal{E}_c$ such that, for $\mathcal{E} < \mathcal{E}_c$, the ground state is sublattice ordered with zero polarization. This ground state admits small perturbations to the ordering which do not change the overall zero value of the polarization. When the temperature is reduced at a constant non-zero value of $\mathcal{E} < \mathcal{E}_c$ there is a second-order transition to the zero polarization state with a square-root heat capacity singularity. From the possibility of small perturbations we expect the heat capacity to be non-zero below the critical temperature and this is indeed the case. In the zero polarization region the free-energy density $f(\mathcal{E}, T) = f(0, T)$ so that the heat capacity $c(\mathcal{E}, T) = c(0, T)$, which is finite but non-zero. For $\mathcal{E} > \mathcal{E}_c$, antiferroelectric models, apart from the special case of the F model, undergo a transition similar to that in ferroelectric models, while at $\mathcal{E} = 0$ there is yet another kind of critical behaviour. This is a transition of infinite order in which $c(0, T)$ and all its derivatives are continuous.

As we have seen a frozen state below the transition is not a necessary condition for a K-type transition. However, lattice conservation properties may be relevant. The row conservation rule for dimers established at the beginning of Sect. 13.2.2 depends only on the topology of the brick (honeycomb) lattice and not on the assignment of fugacities. So that it applies to any brick lattice dimer system including the V^2H^2 model. Again, six-vertex models have an ice-rule arrow conservation property, which results from the permitted vertex connections and does not depend on the vertex energies. It therefore applies to both ferroelectric and antiferroelectric models.

For the eight-vertex models, the ice-rule does not apply and, as we showed in Chap. 12, there is then a new range of critical properties, including the O-type transition as a special case (Sect. 12.3.9). Both conservation rules and anisotropy play a role in determining the properties of dimer models. The six-vertex models do not have K-type second-order transitions when the field is reduced to zero, although ferroelectric models do show a square-root singularity in the heat capacity as the first-order transition is approached from above. A non-zero field obviously introduces an element of anisotropy. As shown in Sect. 13.3.1, the highly anisotropic five-vertex model has a K-type transition to the completely polarized state, even in zero field. However, the introduction of a degree of spatial anisotropy does not necessarily change the character of transitions in two-dimensional lattice systems. As may be deduced from the expressions obtained in Sect. 13.3.2 the Ising ferromagnets still have O-type transitions even when there are different couplings in different lattice directions. Obviously a similar statement can be made for the equivalent dimer systems.

Part IV

Series and Renormalization Group Methods

Introduction

Although it is possible to identify an early example of the use of a high-temperature series expansion by Opechowski (1937),¹ series expansions became powerful tools in the analysis of phase transitions only in the 1960s. Then, much of the early work was concentrated on the spin- $\frac{1}{2}$ Ising model and undertaken by a group in the Physics Department at King's College London, led by the late Prof. Cyril Domb. At the heart of the enterprise was the development of techniques for counting lattice configurations and the computation of lattice constants for weak-graph embeddings, in the case of high-temperature series, and strong-graph embeddings, in the case of low-temperature series.² An early paper by Sykes, Essam and Gaunt (1965)³ deriving low-temperature expansions on the standard two- and three-dimensional lattices was followed by a systematic presentation by Sykes, Essam, Heap and Hiley (1966) of a procedure (variously called the *partial-sum*, *shadow* or *code* method) for determining lattice constant for both weakly and strongly embedded graphs. There then followed a long sequence of papers⁴ determining low-temperature susceptibility series for the standard two- and three-dimensional lattices. A series for the square lattice was obtained to degree 11 and for the triangular and honeycomb lattices to degree 21. The expansions for the simple cubic, diamond, body-centred cubic and face-centred cubic lattices were to degrees 20, 15, 28 and 40, respectively. The procedure was also applied to the Ising model with spin greater than a half by Fox

¹ Who studied the Heisenberg model by expanding the partition function in inverse powers of temperature.

² These elements of graph theory are discussed in Sects. 16.1 and 16.4.

³ Since much of this work was published by different groups of workers with Dr. Martin Sykes as the lead author, we quote references with the list of authors given *in extenso*.

⁴ Sykes, Gaunt, Essam and Hunter (1973); Sykes, Gaunt, Mattingly, Essam and Elliott (1973); Sykes, Gaunt, Martin, Mattingly and Essam (1973); Sykes, Gaunt, Essam, Heap, Elliott and Mattingly (1973); Sykes, Gaunt, Essam and Elliott (1973); Sykes, Watts and Gaunt (1975a, b); Sykes, McKenzie, Watts and Gaunt (1975).

and Guttman (1973) and to the Ising model on the triangular lattice with a pure triplet interaction by Sykes and Watts (1975).

As will be seen from the presentation of the finite-lattice method in Sect. 14.5 it is, in general, more difficult to obtain a large number of terms of high-temperature series, except when they are provided by duality from low-temperature series. Since duality applies only when the order parameter is zero, this means that, while heat capacities can be derived from duality, this does not apply to field-related response functions (compressibility and susceptibility). Early work by Sykes, Gaunt, Roberts and Wyles (1972a) obtained high-temperature expansions for the susceptibility of the spin- $\frac{1}{2}$ Ising model on the square, triangle and honeycomb lattices to degrees 21, 16 and 32, respectively, and in a second paper Sykes, Gaunt, Roberts and Wyles (1972b) undertook the same calculations for the simple cubic, body-centred cubic and face-centred cubic lattices resulting in series to degrees 17, 15 and 12, respectively. Kim and Joseph (1975) used the finite-cluster method (Sect. 14.4) to derive high-temperature series to degree 8 for the Potts model with $\nu = 3, 4, 5$ on square, triangular and face-centred cubic lattices.

The finite-lattice method was introduced by de Neef (1975) in his doctoral thesis⁵ and, according to de Neef and Enting (1977), it is “primarily a technique for obtaining high-temperature expansions for the zero-field free energy of the Ising model and related models on a square lattice”. However, de Neef (op. cit.) had already ventured beyond this specification by applying the method to the one-dimensional Heisenberg model and de Neef and Enting (op. cit.) began what has subsequently become the major preoccupation of finite-lattice calculations, which is work on the Potts model for various lattices and values of ν . Another interesting twist to this development is that, whilst de Neef and Enting envisaged the method as mainly a high-temperature procedure, its subsequent development has been largely for the production of low-temperature series. In the list of applications to the Potts model given by Enting (1996), only five are high temperature compared to 20 which are low temperature. As he remarks, in finite-lattice calculations “high- T expansions can be calculated using a low- T style of summation working with high- T variables”. The finite-lattice method (with various extensions: Jensen and Guttman 1996; Arisue and Tabata 1997) is the most widely used method in recent work on series expansions. We shall, therefore, in Chap. 14 mainly concentrate on that method. However, it is useful to set the scene with a preliminary discussion of simple graph-counting in Sect. 14.2 and a treatment of the related finite-cluster method in Sect. 14.4.

The first widely used series-analysis technique was the *ratio method* introduced by Domb and Sykes (1956, 1957a, b) and described in Sect. 14.6.2. Although a number of variants of this method have been developed to improve the approximations to critical parameters and exponents (Guttman 1989), unless the singularity of interest lies on the circle of convergence ($w_c = r$), the ratios still do not give good estimates for critical parameters and exponents. Smooth monotonic

⁵ Where he calls it the *finite-lattice cumulant method* (ibid Sect. 5.5).

behaviour of the coefficients is required and, although oscillations arising from another known singularity on the circle of convergence can sometimes be catered for, applications to cases where the critical singularity is not at least as close as any other singularity to the origin do not yield good results.

The use of Padé approximants (Sect. 14.6.3) is due to Baker (1961) (see also, Guttmann 1989). This method has the advantage of being applicable even when $w_c > r$, since Padé approximants effectively provide an analytic continuation of the series to the region $|w| > r$. However, the only singularities it can predict are poles and it is known that many systems have singularities with non-integer exponents. This led to the use of Dlog Padé approximants, which consist of the prior application of the Dlog transformation (Sect. 14.6.4.1) before the calculation of the Padé approximant. This has the advantage of turning singularities with non-integer exponents into poles.

The equation for the Dlog Padé approximant can also be viewed as the lowest level (first-order homogeneous) *differential approximant*. This class of approximants (Sect. 14.6.4.1) is due to Guttmann and Joyce (1972) and is now the most widely used technique of series analysis. The relationship between Dlog Padé and differential approximants is paralleled by a relationship between ‘ordinary’ Padé approximants and a class of algebraic approximants (Sect. 14.6.4.2).

A Padé approximant can be viewed as a first-degree algebraic equation and Brak and Guttmann (1990) proposed the use of higher-degree algebraic equations. These (for degree greater than one) have the advantage over Padé approximants of being able to predict a singular point with rational non-integer exponents, and the disadvantage compared to differential approximants of being restricted to rational exponents. It is also the case, as is shown in Sect. 17.6, that the solution of an algebraic equation is a solution of a differential equation (but not conversely). However, from a conceptual point of view, the idea that, as for the hard-hexagon model (Example Sect. 18.5.2), a thermodynamic quantity exactly satisfies an algebraic equation is attractive.

In Sect. 14.6 the ratio and various approximant methods are introduced and applied solely to the methods for finding second-order transitions, given by singularities, and formulae for calculating their exponents. Section 14.7 discusses the use of the same approximants for the exploration of situations where the transition may be first order. In the absence of data from other sources, this tends, since it relies on matching of high- and low-temperature series, to be a rather less accurate procedure.

As we have seen, phase transitions can occur only in infinite systems and these are by their nature difficult to handle. Wilson (1971) commented that what “one would like to do is to transform the problem of calculating [the free energy] into a form where it is natural for [it] to have singularities at [the critical point] and where one may hope the nature of the singularity will be more easily seen than from [the partition function]”. He then suggested that the appropriate way to solve this

problem is in terms of the renormalization group. Although such an idea had already been mooted, in this paper Wilson made the first attempt to lay down the assumptions needed for its implementation.⁶

The approach to the renormalization group in Chap. 15 is usually referred to as the use of the *real-space renormalization group (RSRG)*. This is in contrast to renormalization group methods for continuous spin distributions, using, for example the Ginzburg–Landau local free-energy density of Sect. 5.6.1. These latter methods, in which renormalization is performed in wave-vector space, were initiated by Wilson (1975) and owe much to the influence of quantum field theory. Real-space methods, which are conceptually rather simpler, are based, in many cases, on the block-spin method of Kadanoff (1966) described in Sect. 4.2.⁷

The essential feature of all renormalization group methods is the realization that critical phenomena are associated with long wave length (small wave vector) fluctuations in wave-vector space or long-range correlations in real space. This means that an integration out of large wave vectors or a block averaging of short-range interactions in real space will leave the essential critical features of the system unchanged. The selection of renormalization group transformations contained in Chap. 15 have the common property that they lead to trajectories in the space of coupling which divide the space into regions controlled by fixed points. This parallels the picture of scaling theory presented in Chap. 4, but now we have actual numerical predictions for critical exponents rather than simply scaling-law relations between them. It should, however, be emphasized that, with the exception of the decimation of the one-dimensional Ising model in Sect. 15.5.1, the renormalization group transformation involves some form of approximation or truncation. In consequence, the scaling exponents are also approximations.

⁶ For a reflective account of the development of the renormalization group see Kadanoff (2006, 2013).

⁷ A comprehensive collection of articles on both types of renormalization group methods is contained in PTCP (6) (see the preamble to the References and Author Index for an explanation of this designation) and on real-space methods in the volume edited by Burkhardt and van Leeuwen (1982). For a description of renormalization methods in wave-vector space the reader is referred to Ma (1976a) and Amit (1978). Accounts of both approaches are given by Goldenfeld (1992), Binney et al. (1993) and Cardy (1996).

Chapter 14

Series Expansions

14.1 The Task and the Methods

In this chapter we discuss methods for obtaining coefficients in the expansions of partition functions and free energies, or their derivatives, in a series of powers of temperature-dependent variables. These series are called *low temperature*, or *high temperature*, respectively according to whether the expansion variable $w(T)$ tends to zero as $T \rightarrow 0$ or $T \rightarrow \infty$. If an expression for the coefficient of the general term were known then a series would constitute an exact solution. However, the phrase ‘series methods’ refers primarily to the term-by-term calculation of as many coefficients as possible and the deduction of results for the model from this limited information. Low- and high-temperature series expansions provide approximations to the thermodynamic functions of a model in their respective ranges of temperature and this is important for the detection of first-order transitions as we shall show in Sect. 14.7. However, series expansions are more often used to study critical properties for transitions at which the thermodynamic function is singular; in which case the interest is in obtaining critical points and exponents. Suppose that

$$F(w) := \sum_{k=0}^{\infty} f_k w^k, \quad (14.1.1)$$

is some thermodynamic function of the model of interest, with $w(T)$ appropriate to a high-temperature or low-temperature series. If

$$\lim_{k \rightarrow \infty} |f_k|^{1/k} \quad (14.1.2)$$

exists and is equal to $1/r$, the series converges in the disc $|w| < r$ in the complex w -plane, $\mathbb{C}_w$. There is always a singularity of $F(w)$ on the circle $|w| = r$, which is the *radius of convergence* of the series. If this singularity is on the positive real axis

it can be identified with $w(T_c)$,¹ where T_c is the critical temperature. When this is not the case the analysis is rather more complicated (Sect. 14.6).

In any event, as has been indicated, the methods to be described in this chapter are those for which, for a high- or low-temperature series and some integer ℓ , an ℓ -degree polynomial part

$$F_\ell(w) := \sum_{k=0}^{\ell} f_k w^k, \quad (14.1.3)$$

of the expansion (14.1.1) of $F(w)$ is known. In such circumstances, it may seem surprising that any information at all can be obtained about the function $F(w)$, given that the use of (14.1.2) to obtain a good approximation to the radius of convergence needs in general many more coefficients than are available.² More sophisticated techniques are needed and these are described in Sect. 14.6. As we shall see, the key to all these methods is the important point made by Guttman (1989, p. 8) that, in series analysis, it is a question of “fitting series coefficients to an assumed functional form—an assumed form may be implicit or explicit, but is invariably present”. Even then, of course, it is possible to obtain only approximate values for the assumed critical parameters. Nevertheless, these methods provide accurate values for critical parameters and exponents for many systems where no exact solution is known. Of course, where exact solutions are known, comparison with deductions from series methods provides a useful test for the latter.

In Sects. 14.2–14.5 we describe methods for obtaining series expansions, and for simplicity we shall concentrate mainly on the case of the ferromagnetic ($R > 0$) ν -state Potts model with Hamiltonian (3.6.53)³ and partition function $Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z})$, where $\mathfrak{X} := \exp(K)$ and $\mathfrak{Z} := \exp(L)$, with $K := R/T$ and $L := \mu/T$. In Sect. 14.2 we discuss the determination of suitable variables for low- and high-temperature expansions. Low-temperature expansions are possible only for models (like the Potts model) where there is a well-defined ground state and the successive terms in the expansion represent deviations from that state. The ‘obvious’ expansion variable for high temperatures is the inverse temperature (equivalently the coupling K) and that is the only choice for n -vector models, with $n > 1$, and for the quantum Heisenberg model. However, as we shall see in Sect. 14.2, there is a more convenient choice for the Potts model.

¹ As is discussed in Sect. 14.6, in the case of high-temperature series, a singularity on the negative real axis also corresponds to a critical temperature.

² Thus, for example, for $F(w) := (0.5 - w)^{-0.25}$, although the convergence is for a monotonically increasing sequence, the estimate for w_c from $F_{1000}(w)$ still differs from $w_c = 0.5$ by 0.0031577523. Making the same point Guttman (1989) considers the susceptibility of the spin- $\frac{1}{2}$ Ising model for which Nickel has derived 54 terms of the high-temperature series. In this case the estimate from $F_{54}(w)$ is higher than the exact $w_c = 1 + \sqrt{2}$ by about 5.4 %.

³ Knowing of course that the spin- $\frac{1}{2}$ Ising model can be recovered as the special case $\nu = 2$ (Sect. 3.6.4).

In Sects. 14.2.1 and 14.2.2 we obtain, by simple graphical reasoning, the first few terms of low- and high-temperature series for $Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z})$. As we shall see, the main concern of this chapter is to obtain the series for the free energy per lattice site, when $\mathfrak{Z} = 1$, that is for the ‘pure’ Potts model with Hamiltonian (3.6.49). Contrary to the convention used elsewhere in the text (see the footnote on p. 21), the free energy per lattice site is taken to be the quantity defined for a finite lattice $\mathcal{N}$. That is⁴

$$[Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z})]^{1/N} =: \exp[-\phi(\mathcal{N}, \mathfrak{X}, \mathfrak{Z})]. \quad (14.1.4)$$

We shall also need expansions for the order parameter⁵

$$\tilde{\rho} := 1 - \frac{\nu}{\nu - 1} \rho = 1 + \frac{\nu}{\nu - 1} \frac{\partial \phi}{\partial L} = 1 + \mathfrak{Z} \frac{\nu}{\nu - 1} \frac{\partial \phi}{\partial \mathfrak{Z}}, \quad (14.1.5)$$

defined in terms of the density ρ so as to be zero at zero temperature, when $\rho = (\nu - 1)/\nu$, and the response functions⁶

$$\frac{c_\rho}{k_B K^2} := -\frac{\partial^2 \phi}{\partial K^2} = -\mathfrak{X} \left\{ \frac{\partial \phi}{\partial \mathfrak{X}} + \mathfrak{X} \frac{\partial^2 \phi}{\partial \mathfrak{X}^2} \right\}, \quad (14.1.6)$$

$$\varphi_T := -\frac{\partial^2 \phi}{\partial L^2} = -\mathfrak{Z} \left\{ \frac{\partial \phi}{\partial \mathfrak{Z}} + \mathfrak{Z} \frac{\partial^2 \phi}{\partial \mathfrak{Z}^2} \right\}, \quad (14.1.7)$$

given respectively by (2.3.29) and (2.3.31). Although our interest is in the case $\mathfrak{Z} = 1$, it is clear, from (14.1.5)–(14.1.7), that we shall need to obtain the free-energy series with general $\mathfrak{Z}$. However, in terms of suitably chosen field variables, it will be sufficient to retain terms only to quadratic degree.

Series for the partition function are called *moment expansions* and, having defined suitable variables, we obtain the first few terms of these, at low and high temperatures for the Potts model in Sect. 14.2. The terms in the expansion of the logarithm of the partition function are called *cumulants*. These can, of course, be obtained by taking the logarithm of the moment expansion and we illustrate this for the low-temperature series in Sect. 14.2.1. However, a more direct approach to the series for the free-energy density (equivalent to the cumulant expansion) is discussed in Sect. 14.3.

For both moment and cumulant series the terms in the expansions can be represented by graphs and, although the mode of their construction at low and high

⁴ Rather confusingly, this quantity is often simply called ‘the partition function’ (see, for example Briggs et al. 1994).

⁵ The order parameter $\tilde{\rho}$, which Briggs et al. (1994) call the magnetization, differs from that used (with $\nu = 3$) in the Landau expansion in Sect. 5.4.2 by a factor of $2/\sqrt{3}$.

⁶ The response function φ_T , which Briggs et al. call the susceptibility, differs from the compressibility defined by (1.4.12) by a factor of $1/(k_B T \rho^2)$. We shall for the sake of maintaining neutrality between lattice-fluid and magnetic interpretations refer to it as the ‘field-response function’. c_ρ is, of course, the heat capacity at constant density. However, since we shall be concerned solely with calculating it for zero μ ($\mathfrak{Z} = 1$) it will be the same as the constant μ heat capacity.

temperatures differ, their edges can be understood to represent sites which are linked by an interaction. However, as we shall see there are two important differences in this respect between the moment and cumulant expansions. These are:

- (i) Moments include contributions from both one-component and multi-component graphs, whereas cumulants involve only connected graphs.
- (ii) Moments (with the variables we employ) are represented by graphs with at most one edge between a pair of sites whereas cumulants involve *multi-bond graphs*.

A consequence of the occurrence of multi-bond graphs has been the development of *finite-cluster methods* (Sect. 14.4), where the cluster is taken to be the set of all graphs occupying the same set of lattice sites irrespective of the multiplicity of the bonds.⁷ A natural development of finite-cluster methods leads to *finite-lattice methods* (Sect. 14.5).

14.2 Moment Expansions

Let the lattice $\mathcal{N}$, with $N := |\mathcal{N}|$ and coordination number $z > 2$, be one of the two- or three-dimensional lattices described in Sect. 16.2.1, with periodic boundary conditions.⁸ From (3.6.53) the Potts model partition function takes the form

$$Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z}) = \sum_{\{\sigma\}} \left\{ \prod_{\{r, r'\}}^{(1)} \mathfrak{X}^{\delta^{\mathbf{K}r}(\sigma(r) - \sigma(r'))} \right\} \left\{ \prod_{\{r\}} \mathfrak{Z}^{1 - \delta^{\mathbf{K}r}(\sigma(r))} \right\}, \quad (14.2.1)$$

of a polynomial in $\mathfrak{X}$ and $\mathfrak{Z}$. It is, of course, the case that expansions of $Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z})$ in terms of both high and low-temperature variables (as defined below) will also be polynomials. Since phase transitions occur only in the thermodynamic limit $N \rightarrow \infty$ the analysis of finite systems is of use only in so far as it yields a ‘sufficient’ number of terms which agree with the series which would arise from the corresponding infinite system.

Although this chapter will be concerned exclusively with high- and low-temperature expansions, another possible area of investigations is that of a high(positive)-chemical potential expansion at constant temperature. In the magnetic formulation in which investigations are usually set, this is the *high-field expansion*.⁹ For our

⁷ Item (ii) needs qualification if high-temperature moment series in terms of the inverse temperature are derived. These give rise to multi-bonded graphs (see Lavis and Bell 1999b, Sect. 7.6) and have led to the proposal for a finite-cluster moment method by de Neef (1975, Sect. 2.3).

⁸ We exclude the case of a linear lattice ($z = 2$), for which no phase transition occurs for non-zero temperature.

⁹ A case where such an expansion is useful is in the investigation of the zero-field first-order transition in the Ising model (see Sect. 14.7 and Enting and Baxter 1980).

purposes the inclusion of the chemical potential is to provide a means of obtaining density and compressibility expansions (in the magnetic formulation magnetization and susceptibility expansions, respectively).

14.2.1 At Low Temperatures

The variable $\mathfrak{X}$ becomes large at low temperatures and the polynomial in (14.2.1) is dominated by the ground-state term. Assuming, as we shall throughout this section, that $\mu < 0$, the ground state is $\sigma(\mathbf{r}) = 0$, for all $\mathbf{r}$. Let

$$Z_{\text{LT}}(\mathcal{N}, u, p) := Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z}) / (\mathfrak{X}^{N_E}) \quad (14.2.2)$$

$$= \sum_{\{\sigma\}} \left\{ \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} u^{1-\delta^{\text{Kr}}(\sigma(\mathbf{r})-\sigma(\mathbf{r}'))} \right\} \left\{ \prod_{\{\mathbf{r}\}} (1-p)^{1-\delta^{\text{Kr}}(\sigma(\mathbf{r}))} \right\} \quad (14.2.3)$$

$$:= 1 + N \sum_{m=1}^{\infty} a_m(N, p) u^m, \quad (14.2.4)$$

where

$$u := 1/\mathfrak{X}, \quad p = 1 - \mathfrak{Z}. \quad (14.2.5)$$

Then, from (16.2.8), (14.2.2) and (14.2.5),

$$\phi(\mathcal{N}, \mathfrak{X}, \mathfrak{Z}) = \phi_{\text{LT}}(\mathcal{N}, u, p) + \frac{1}{2} z \ln(u). \quad (14.2.6)$$

The variable u has the required property of a low-temperature variable, tending to zero as temperature tends to zero, and the coefficients in (14.2.4) are polynomials in p .¹⁰ However, in order to obtain series, for the free energy, order parameter and response function with $\mu = 0$ we need retain terms only to quadratic degree in p . Thus we require

$$Z_{\text{LT}}(\mathcal{N}, u, p) \simeq Z_{\text{LT}}^{(0)}(\mathcal{N}, u) + Z_{\text{LT}}^{(1)}(\mathcal{N}, u) p + Z_{\text{LT}}^{(2)}(\mathcal{N}, u) p^2, \quad (14.2.7)$$

which will then yield

$$\phi_{\text{LT}}(\mathcal{N}, u, p) \simeq \phi_{\text{LT}}^{(0)}(\mathcal{N}, u) + \phi_{\text{LT}}^{(1)}(\mathcal{N}, u) p + \phi_{\text{LT}}^{(2)}(\mathcal{N}, u) p^2. \quad (14.2.8)$$

¹⁰ In view of the following discussion it is convenient to include explicit linear N -dependence of the coefficients in (14.2.4). However, that does not exclude further non-linear dependence.

The series for the free energy per lattice site with $\mu = 0$ is, of course, given by $\phi_{\text{LT}}^{(0)}(\mathcal{N}, u)$. From (14.1.5)–(14.1.7) the corresponding series for the order parameter and response functions are given as expansions of

$$\tilde{\rho} = 1 - \frac{v}{v-1} \phi_{\text{LT}}^{(1)}, \quad (14.2.9)$$

$$\frac{c_\rho}{k_B K^2} = -u \left\{ \frac{\partial \phi_{\text{LT}}^{(0)}}{\partial u} + u \frac{\partial^2 \phi_{\text{LT}}^{(0)}}{\partial u^2} \right\}, \quad (14.2.10)$$

$$\varphi_T = \phi_{\text{LT}}^{(1)} - 2\phi_{\text{LT}}^{(2)}. \quad (14.2.11)$$

The series in (14.2.4) results from perturbations of the ground state, so the power of $(1-p)$ in any term is just the number of perturbed sites and the power of u is the number of pairs in unlike states. In the case $v = 2$ (equivalent to the Ising model) such pairs will consist of one ground-state site and one perturbed $\sigma = 1$ site. When $v > 2$ there will also be the possibility of pairs of unlike perturbed sites. Thus a term in the series corresponds to a section graph on the lattice formed by drawing edges between all pairs of perturbed sites together with (in the case $v > 2$) a particular distribution of perturbed states on the graph. So a particular section graph $[\mathfrak{n}, \mathfrak{k}]$ will, in general, contribute a number of terms in the series each containing a factor $(1-p)^{\mathfrak{n}}$. The number of lattice edges $l[\mathfrak{n}, \mathfrak{k}]$, which can be drawn from sites of $[\mathfrak{n}, \mathfrak{k}]$ to ground-state sites, is given by (16.4.9) and we also need $w[\mathfrak{i} | \mathfrak{n}, \mathfrak{k}]$, the number of arrangements of perturbed states on $[\mathfrak{n}, \mathfrak{k}]$ so that there are exactly $\mathfrak{i}$ edges in different perturbed states.¹¹ From (16.4.10), a graph $[\mathfrak{n}, \mathfrak{k}]$ containing $\mathfrak{i}$ unlike perturbed pairs of sites makes a contribution to the series in a term $a_m(N, p)u^m$ with

$$m = l[\mathfrak{n}, \mathfrak{k}] + \mathfrak{i} = \mathfrak{n}(z-2) + 2c[\mathfrak{n}, \mathfrak{k}] - 2\mathfrak{C}[\mathfrak{n}, \mathfrak{k}] + \mathfrak{i} \quad (14.2.12)$$

and the coefficient is

$$a_m(N, p) = \sum w[\mathfrak{i} | \mathfrak{n}, \mathfrak{k}] \overline{w}[\mathfrak{n}, \mathfrak{k}] (1-p)^{\mathfrak{n}}, \quad (14.2.13)$$

where the summation is over all graphs and configurations which contribute to this term. We now construct the first few terms of the low-temperature series for the v -state Potts model. This is done by forming a sequence of graphs with increasing values of $l[\mathfrak{n}, \mathfrak{k}]$ and then, for each graph taking the possible values of $\mathfrak{i}$. We use the square-lattice graphs in Tables 16.1 and 16.2 as an illustration, but we also consider the triangular and honeycomb lattices; for all of these the cyclomatic number $\mathfrak{C}[\mathfrak{n}, \mathfrak{k}]$ of a graph $[\mathfrak{n}, \mathfrak{k}]$ is the number of its faces. As will be seen, all the section (low-temperature) graphs in Table 16.1 are trees or have components which are trees.¹² So one strategy is to begin with these and then to make modifications to the list which arise when graphs with faces, like those in Table 16.2, make a contribution. It is clear

¹¹ Of course, for all lattices, $w[0 | \mathfrak{n}, \mathfrak{k}] = v - 1$.

¹² For simplicity we regard an isolated site as a tree (and also a chain) with no edges.

that, for $v > 2$ and a tree with n sites, i can take all values from zero to $n - 1$. The notation $\{n, f, i\}$ is used to denote all the one-component graphs with n sites, f faces and the given value of i . Since we are, at this point, considering only trees, $f = 0$. Thus, for example, $\{4, 0, i\}$ for the square lattice represents all the graphs $[4, k]$, $k = 4, \dots, 7$ in Table 16.1, with $e[4, k] = 10$. A plus sign is used to indicate a separated graph with particular components.

The first few terms for the sequence for the square lattice are:

- $m = 4$: $\{1, 0, 0\}$.
- $m = 6$: $\{2, 0, 0\}$.
- $m = 7$: $\{2, 0, 1\}$.
- $m = 8$: $\{3, 0, 0\}$; $\{1, 0, 0\} + \{1, 0, 0\}$.
- $m = 9$: $\{3, 0, 1\}$.
- $m = 10$: $\{4, 0, 0\}$; $\{3, 0, 2\}$; $\{1, 0, 0\} + \{2, 0, 0\}$.
- $m = 11$: $\{4, 0, 1\}$; $\{1, 0, 0\} + \{2, 0, 1\}$.

for the triangular lattice are:

- $m = 6$: $\{1, 0, 0\}$.
- $m = 10$: $\{2, 0, 0\}$.
- $m = 11$: $\{2, 0, 1\}$.
- $m = 12$: $\{1, 0, 0\} + \{1, 0, 0\}$.
- $m = 14$: $\{3, 0, 0\}$.
- $m = 15$: $\{3, 0, 1\}$.
- $m = 16$: $\{3, 0, 2\}$; $\{1, 0, 0\} + \{2, 0, 0\}$.

and for the honeycomb lattice are:

- $m = 3$: $\{1, 0, 0\}$.
- $m = 4$: $\{2, 0, 0\}$.
- $m = 5$: $\{3, 0, 0\}$; $\{2, 0, 1\}$.
- $m = 6$: $\{4, 0, 0\}$; $\{3, 0, 1\}$; $\{1, 0, 0\} + \{1, 0, 0\}$.
- $m = 7$: $\{5, 0, 0\}$; $\{4, 0, 1\}$; $\{3, 0, 2\}$; $\{1, 0, 0\} + \{2, 0, 0\}$.
- $m = 8$: $\{6, 0, 0\}$; $\{5, 0, 1\}$; $\{4, 0, 2\}$; $\{1, 0, 0\} + \{2, 0, 1\}$; $\{1, 0, 0\} + \{3, 0, 0\}$; $\{2, 0, 0\} + \{2, 0, 0\}$.
- $m = 9$: $\{7, 0, 0\}$; $\{6, 0, 1\}$; $\{5, 0, 2\}$; $\{4, 3\}$; $\{1, 0, 0\} + \{1, 0, 0\} + \{1, 0, 0\}$; $\{1, 0, 0\} + \{4, 0, 0\}$; $\{1, 0, 0\} + \{3, 0, 1\}$; $\{2, 0, 0\} + \{3, 0, 0\}$; $\{2, 0, 0\} + \{2, 0, 1\}$.

Consider now the role played by the first few graphs with faces ($f \neq 0$). For each lattice the lowest value of m contributed by a graph with a face is a simple cycle around one face (with $i = 0$). For the square lattice the graphs which make contributions with $m \leq 11$ are $[4, 8]$, $[5, 1]$ and $[6, 1]$ in Table 16.2. Thus to the list for the square lattice, given above, we add:

- $m = 8$: $\{4, 1, 0\}$.
- $m = 10$: $\{5, 1, 0\}$; $\{6, 2, 0\}$; $\{4, 1, 2\}$.
- $m = 11$: $\{5, 1, 1\}$; $\{4, 1, 3\}$.

It will be seen from these lists of graphical contributions to the series that graph-specific factors soon play a role in the development. However, it is of interest, particularly in connection with the high-temperature series discussion in Sect. 14.2.2, to develop the lowest-degree terms in a way which applies to all the regular lattices given in Sect. 16.2.1. We have successively the terms with

- $m = z$: single-site graphs, designated by $\{1, 0, 0\}$, with lattice constant of unity. There are $v - 1$ ways of choosing the perturbed state so $a_z(N, p) = (v - 1)(1 - p)$.
- $m = 2(z - 1)$: two-site chains, designated by $\{2, 0, 0\}$, with lattice constant $\frac{1}{2}z$ and the sites in the same perturbed state, giving $a_{2(z-1)}(N, p) = \frac{1}{2}z(v - 1)(1 - p)^2$.
- $m = 2z - 1$: two-site chains, designated by $\{2, 0, 1\}$, with the sites in different perturbed states. There are $(v - 1)(v - 2)$ ways of choosing the perturbed states so $a_{2z-1}(N, p) = \frac{1}{2}z(v - 1)(v - 2)(1 - p)^2$.

Thus we have

$$Z_{\text{LT}}(\mathcal{N}, u, p) \simeq 1 + N(v - 1) \left\{ (1 - p)u^z + \frac{1}{2}z(1 - p)^2u^{2(z-1)} + \frac{1}{2}z(v - 2)(1 - p)^2u^{2z-1} \right\}, \quad (14.2.14)$$

giving¹³

$$\phi_{\text{LT}}^{(0)}(\mathcal{N}, u) = -(v - 1) \left\{ u^z + \frac{1}{2}zu^{2(z-1)} \right\} + O(u^{2z-1}), \quad (14.2.15)$$

$$\phi_{\text{LT}}^{(1)}(\mathcal{N}, u) = (v - 1) \left\{ u^z + zu^{2(z-1)} \right\} + O(u^{2z-1}), \quad (14.2.16)$$

$$\phi_{\text{LT}}^{(2)}(\mathcal{N}, u) = -\frac{1}{2}(v - 1)zu^{2(z-1)} + O(u^{2z-1}) \quad (14.2.17)$$

and, from (14.2.9)–(14.2.11), the leading terms of low-temperature expansions of the order parameter and response functions can be derived. However, it is clear that in the term with $m = 2z - 1$ there is already an omission in the case of the honeycomb lattice, for which the chain of three sites $\{3, 0, 0\}$ also gives this value for m with, using (16.4.15), an additional contribution $a'_3(N, p) = 3(v - 1)(1 - p)^3$. This illustrates the impossibility of presenting, beyond the first few terms, a general-lattice form for the low-temperature series expansion. We thus restrict our attention to the square lattice and, using (14.2.13) and the lattice constants in Tables 16.1 and 16.2, we have

$$\begin{aligned} Z_{\text{LT}}^{(0)}(\mathcal{N}_s, u) = & 1 + N(v - 1) \left\{ u^4 + 2u^6 + 2(v - 2)u^7 \right. \\ & + \frac{1}{2} [N(v - 1) - 5v + 19] u^8 \\ & + 12(v - 2)u^9 + 2 \left[N(v - 1) + 3v^2 - 17v + 28 \right] u^{10} \\ & \left. + 2(v - 2) [N(v - 1) - 6v + 33] u^{11} \right\} + O(u^{12}), \end{aligned} \quad (14.2.18)$$

¹³ Notation: $g(u) = O(u^k)$ if $g(u)/u^k$ tends to a finite limit as $u \rightarrow 0$. If $g(u) = O(u^k)$ then $g(u) = O(u^{k'})$ for $k' \leq k$.

$$\begin{aligned}
Z_{\text{LT}}^{(1)}(\mathcal{N}_s, u) = & -N(v-1) \left\{ u^4 + 4u^6 + 4(v-2)u^7 + [N(v-1) - 5v + 27]u^8 \right. \\
& + 36(v-2)u^9 + 2[3N(v-1) + 9v^2 - 48v + 98]u^{10} \\
& \left. + 2[3N(v-1)(v-2) - 16v^2 + 160v - 256]u^{11} \right\} + O(u^{12}),
\end{aligned} \tag{14.2.19}$$

$$\begin{aligned}
Z_{\text{LT}}^{(2)}(\mathcal{N}_s, u) = & N(v-1) \left\{ 2u^6 + 2(v-2)u^7 + \frac{1}{2}[N(v-1) - 5v + 53]u^8 \right. \\
& + 36(v-2)u^9 + 2[3N(v-1) + 9v^2 - 42v + 133]u^{10} \\
& \left. + 2[3N(v-1)(v-2) - 12v^2 + 214v - 380]u^{11} \right\} + O(u^{12}).
\end{aligned} \tag{14.2.20}$$

Using the same procedure, series for the triangular and honeycomb lattices can be derived. It will be noted that the two-component graphs [2, 1] and [3, 2] make contributions to $a_8(N, p)$ and to $a_{10}(N, p)$ and $a_{11}(N, p)$ respectively which are dependent on N . This is a case of a result, which is true for all lattices, for both low- and high-temperature series and for both $\mu = 0$ and $\mu \neq 0$; that graphs with c components will make contributions to coefficients of the partition function expansion which are polynomials in N of degree $c - 1$. However, our interest is not in the partition function itself but in the free energy per lattice site and the order parameter and response functions, for which the series coefficients are N -independence (Sect. 14.3). If it were not the case the extensivity of the free energy would be violated, vitiating the relationship with thermodynamics. In particular, from (14.2.7) and (14.2.18)–(14.2.20), for the square lattice

$$\begin{aligned}
\phi_{\text{LT}}^{(0)}(\mathcal{N}_s, u) = & -(v-1) \left\{ u^4 + 2u^6 + 2(v-2)u^7 - \frac{1}{2}(5v-19)u^8 \right. \\
& + 12(v-2)u^9 + 2(3v^2 - 17v + 28)u^{10} \\
& \left. - 6(v-2)(2v-11)u^{11} \right\} + O(u^{12}),
\end{aligned} \tag{14.2.21}$$

$$\begin{aligned}
\phi_{\text{LT}}^{(1)}(\mathcal{N}_s, u) = & (v-1) \left\{ u^4 + 4u^6 + 4(v-2)u^7 - (5v-27)u^8 + 36(v-2)u^9 \right. \\
& \left. + 2(9v^2 - 48v + 98)u^{10} - 32(v-2)(v-8)u^{11} \right\} + O(u^{12}),
\end{aligned} \tag{14.2.22}$$

$$\begin{aligned}
\phi_{\text{LT}}^{(2)}(\mathcal{N}_s, u) = & -(v-1) \left\{ 2u^6 + 2(v-2)u^7 - \frac{1}{2}(5v-53)u^8 + 36(v-2)u^9 \right. \\
& \left. + 2(9v^2 - 42v + 133)u^{10} - 4(v-2)(6v-95)u^{11} \right\} + O(u^{12}).
\end{aligned} \tag{14.2.23}$$

The reasoning used here for the Potts model, whereby the low-temperature series is derived by considering perturbations from a ground state, can be used for the general-spin ferromagnetic Ising model. However, the spatial isotropy of the exchange interaction terms in classical and quantum Heisenberg Hamiltonians lead to important differences in these cases. As we have seen (Sect. 3.6.1) spontaneous

magnetization in the two-dimensional classical Heisenberg model is excluded by the Mermin–Wagner theorem. There is, therefore, no Curie temperature.¹⁴ For three-dimensional lattices there is a Curie temperature, but the low-temperature behaviour is radically different from that of the Ising model (see, for example, Ashcroft and Mermin 1976). There are no small localized perturbations from the ground state and, for the quantum Heisenberg model, the excitations are spin waves. It follows that there can be no low-temperature series of the type discussed here for these models. In fact, the changes in the spontaneous magnetization and heat capacity from their zero-temperature values are proportional to $T^{\frac{3}{2}}$. In both the quantum and classical Heisenberg models for $T < T_c$, there is a divergence in the magnetic susceptibility as $\mathcal{H} \rightarrow 0$. In the classical Heisenberg model the low-temperature heat capacity has the form $Nk_B + aT$, where a is a constant.

14.2.2 At High Temperatures

The key to the choice of a high-temperature variable for the Potts model is the duality relationship (9.2.21) between $\mathfrak{X}$ and $\mathfrak{X}^*$ expressible in the form

$$\mathfrak{X} = 1 + \frac{\nu}{\mathfrak{X}^* - 1} = \frac{\mathfrak{X} + \nu - 1}{\nu} \left\{ 1 + \frac{\nu - 1}{\mathfrak{X}^*} \right\}. \quad (14.2.24)$$

As $T \rightarrow \infty$, $\mathfrak{X} \rightarrow 1$, $\mathfrak{X}^* \rightarrow \infty$, and

$$\nu := 1/\mathfrak{X}^* = \frac{\mathfrak{X} - 1}{\mathfrak{X} + \nu - 1} \rightarrow 0, \quad (14.2.25)$$

with, from (9.2.22),

$$\mathfrak{X}^{\delta_{Kr}(\sigma)} = \frac{\mathfrak{X} + \nu - 1}{\nu} [1 + \Delta(\nu; \sigma)\nu], \quad (14.2.26)$$

where $\Delta(\nu; \sigma)$ is the *Potts delta function* defined and discussed in Sect. 18.4. In a similar way we introduce the high-temperature variable

$$q := \frac{1 - \mathfrak{Z}}{1 + \mathfrak{Z}(\nu - 1)}, \quad (14.2.27)$$

¹⁴ Although there is a Kosterlitz–Thouless transition which can be analysed using a high-temperature series expansion (Oitmaa et al. 2006, Chap. 3).

associated with the chemical potential, with

$$\mathfrak{Z}^{1-\delta^{\text{Kr}}(\sigma)} = \frac{1 + \mathfrak{Z}(\nu - 1)}{\nu} [1 + \Delta(\nu; \sigma)q]. \quad (14.2.28)$$

Substituting into (14.2.1) gives

$$Z_{\text{HT}}(\mathcal{N}, \nu, q) = \nu^{-N} Z(\mathcal{N}, \mathfrak{X}, \mathfrak{Z}) \left\{ \frac{\nu}{\mathfrak{X} + \nu - 1} \right\}^{N_{\text{E}}} \left\{ \frac{\nu}{1 + \mathfrak{Z}(\nu - 1)} \right\}^N \quad (14.2.29)$$

$$= \nu^{-N} \sum_{\{\sigma\}} \left\{ \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} [1 + \Delta(\nu; \sigma(\mathbf{r}) - \sigma(\mathbf{r}'))\nu] \right\} \\ \times \left\{ \prod_{\{\mathbf{r}\}} [1 + \Delta(\nu; \sigma(\mathbf{r}))q] \right\} \quad (14.2.30)$$

$$:= 1 + N_{\text{x}} \sum_{e=1}^{\infty} \sum_{n=0}^{\infty} b_{e,n}(N_{\text{x}}) \nu^e q^n, \quad (14.2.31)$$

where N_{x} is an extensive variable to be determined below and the coefficients in the series in (14.2.31) are obtained by expanding out the products in (14.2.30) and summing over the values of the state variables. From (16.2.8), (14.2.25), (14.2.27) and (14.2.29), for the free-energy density,

$$\phi(\mathcal{N}, \mathfrak{X}, \mathfrak{Z}) = \phi_{\text{HT}}(\mathcal{N}, u, q) - \ln(\nu) + \ln[1 + (\nu - 1)q] + \frac{1}{2}z \ln(1 - \nu). \quad (14.2.32)$$

As in the case of the low-temperature series, the aim is to obtain

$$\phi_{\text{HT}}(\mathcal{N}, \nu, q) \simeq \phi_{\text{HT}}^{(0)}(\mathcal{N}, \nu) + \phi_{\text{HT}}^{(1)}(\mathcal{N}, \nu)q + \phi_{\text{HT}}^{(2)}(\mathcal{N}, \nu)q^2, \quad (14.2.33)$$

with approximations to $\phi_{\text{HT}}^{(0)}(\mathcal{N}, \nu)$, $\phi_{\text{HT}}^{(1)}(\mathcal{N}, \nu)$ and $\phi_{\text{HT}}^{(2)}(\mathcal{N}, \nu)$ obtained to as high a power of ν as possible. However, as we see below, $\phi_{\text{HT}}^{(1)}(\mathcal{N}, \nu) = 0$ and it follows from (14.1.5)–(14.1.7), (14.2.25) and (14.2.27) that the order parameter $\tilde{\rho}$ is identically zero, with the response functions given by

$$\frac{c_{\rho}}{k_{\text{B}}K^2} = \nu^{-2} f(\nu, v) \left\{ \frac{1}{2}zg_3(\nu, v) - \frac{\partial \phi_{\text{HT}}^{(0)}}{\partial \nu} g_2(\nu, v) - \frac{\partial^2 \phi_{\text{HT}}^{(0)}}{\partial \nu^2} f(\nu, v) \right\}, \quad (14.2.34)$$

$$\varphi_T = -\nu^{-2} \{2\phi_{\text{HT}}^{(2)} + 1 - \nu\}, \quad (14.2.35)$$

where

$$f(\nu, v) := (\nu - 1)[v(\nu - 1) + 1], \quad g_n(\nu, v) := nv(\nu - 1) + n - \nu. \quad (14.2.36)$$

We have used section graphs in the development of some terms of the low-temperature series. For high-temperature series the criteria for what graphs are relevant are rather different.

A graph is said to be *zero-field* or *magnetic* according to whether it has zero or a non-zero number of odd vertices.¹⁵ This distinction is important for their use in the development of high-temperature series for the $\nu = 2$ (Ising model) case. There contributions to the zero- μ series are made only by zero-field graphs and contributions to the $\mu \neq 0$ terms are made only by magnetic graphs. This distinction no longer holds for general ν . However, an important result proved by Domb (1974a, b) is still valid, namely that the only graphs which contribute to the zero- μ high-temperature series are stars (Sect. 18.4).¹⁶

14.2.2.1 The Case $\mu = 0$

We first consider the expansion of the first product in (14.2.30) which would yield the series for $q = \mu = 0$, or equivalently the coefficients $b_{e,0}(N_x)$, $e = 1, 2, \dots$. A term of the expansion which could make a contribution to $b_{e,0}(N_x)$ would be a production of e edge Δ -functions

$$\begin{aligned} \Pi(e, k) := & \nu^{-n} \Delta(\nu; \sigma(\mathbf{r}_1) - \sigma(\mathbf{r}'_1)) \Delta(\nu; \sigma(\mathbf{r}_2) - \sigma(\mathbf{r}'_2)) \\ & \times \Delta(\nu; \sigma(\mathbf{r}_3) - \sigma(\mathbf{r}'_3)) \cdots \Delta(\nu; \sigma(\mathbf{r}_e) - \sigma(\mathbf{r}'_e)), \quad (14.2.37) \end{aligned}$$

on the e edges $(\mathbf{r}_1, \mathbf{r}'_1), (\mathbf{r}_2, \mathbf{r}'_2), \dots, (\mathbf{r}_e, \mathbf{r}'_e)$ of a graph (e, k) (Sect. 16.4.3), where $n := n(e, k)$ and, as we have shown in Sect. 18.4, the only graphs which need be considered are stars.¹⁷

At this point we should observe that, although duality between lattices has an unambiguous definition only for two-dimensional lattices,¹⁸ nothing said so far in this section implies two-dimensionality or indeed the existence of a dual lattice. The duality relationship between couplings has been used merely as an ansatz for defining a high-temperature expansion variable. However, in the remaining part of this section our analysis is restricted to a two-dimensional lattice $\mathcal{N}$ with dual lattice $\mathcal{N}^*$. The number of such sites (and edges) around a face of $\mathcal{N}$ is z^* , the coordination number of the dual lattice $\mathcal{N}^*$ (Sect. 16.2.2.1), and the number of faces of $\mathcal{N}$ is $N^* := |\mathcal{N}^*|$. So the lattice constants which feature as factors in the coefficients $b_{e,0}(N_x)$ are per face of $\mathcal{N}$ (or equivalently per site of $\mathcal{N}^*$), rather than per site of $\mathcal{N}$, and we take

¹⁵ The distinction is sometimes made between ‘magnetic graphs’ with two odd vertices and ‘hyper-magnetic graphs’ with more than two odd vertices; for simplicity we shall refer to all graphs with odd vertices simply as magnetic.

¹⁶ The question of what graphs contribute to the $\mu \neq 0$ series is more complicated, as is illustrated by the examples in Sect. 18.4.

¹⁷ Graphs with type B and C articulation points give a zero contribution and those with a type A articulation point can be considered as cases of disconnected components.

¹⁸ See the remarks at the end of Sect. 16.2.2.1.

$N_x := N^*$. Thus the leading term in the zero- μ part of the expansion (14.2.31) involves one face and configuration number one and the summation over a term like (14.2.37), where (e, k) is the cycle around a single face. From Example 18.4.1,

$$b_{z^*,0}(N^*) = (v - 1). \quad (14.2.38)$$

The next two terms, each with a lattice constant $\frac{1}{2}z^*$, will correspond to the sites around two faces which share an edge, without and with, respectively a Δ -function for the common edge. In the first instance the two faces of $\mathcal{N}$ give a single face of (e, k) and, in the second, two faces sharing a Δ -function-carrying pair. Thus, from Examples 18.4.1 and 18.4.2,

$$b_{2(z^*-1),0}(N^*) = \frac{1}{2}z^*(v - 1), \quad (14.2.39)$$

$$b_{(2z^*-1),0}(N^*) = \frac{1}{2}z^*(v - 1)(v - 2) \quad (14.2.40)$$

and, using $N^*/N = z/z^*$,

$$\phi_{\text{HT}}^{(0)}(\mathcal{N}, v) = -(z/z^*)(v - 1) \left\{ v^{z^*} + \frac{1}{2}z^*v^{2(z^*-1)} \right\} + O(v^{2z^*-1}). \quad (14.2.41)$$

14.2.2.2 The Case $\mu \neq 0$

A term of the expansion which could make a contribution to $b_{e,n'}$ would be a product of $\Pi(e, k)$ given by (14.2.37) and a single-vertex term

$$\Lambda\{n', k'\} := \Delta(v; \sigma(\mathbf{r}_1))\Delta(v; \sigma(\mathbf{r}_2))\Delta(v; \sigma(\mathbf{r}_3)) \cdots \Delta(v; \sigma(\mathbf{r}_{n'})), \quad (14.2.42)$$

which is a product of Δ -functions at n' vertices $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n'}$ of $\mathcal{N}$ in a configuration denoted by $\{n', k'\}$. In Sect. 18.4 it was shown that, for a non-zero contribution to the series, the sites of $\{n', k'\}$ must be contained within (e, k) with $2 \leq n' \leq n(e, k)$.

As we have seen, setting $\mu \neq 0$ for the low-temperature case does not add significantly to the calculation of the series expansion. In contrast $\mu \neq 0$, or equivalently $q \neq 0$, makes the high-temperature case much more complicated. However, as we have seen in the low-temperature case, such a computation is necessary, up to quadratic terms in q , to obtain the response function series. As would be anticipated the order parameter $\tilde{\rho} = 0$; this follows from (14.1.5), the definition (14.2.27) of q and the fact that none of the series coefficients contains a term linear in q . Thus, as we have already indicated above, $\phi_{\text{HT}}^{(1)}(\mathcal{N}, v) = 0$.

We now consider a number of cases of graphs which contribute to coefficients $b_{e,2}(N^*)$. The simplest of these is the graph $(1, 1)$, the single pair of vertices connected by an edge shown in Table 16.1, with a single-vertex term at each end. This will provide the sole contribution to $b_{1,2}(N^*)$. If the lattice is not restricted to the square, from (16.2.8), $\Omega(1, 1) = \frac{1}{2}Nz = \frac{1}{2}N^*z^*$. Thus, from Example 18.4.5,

$$b_{1,2}(N^*) = \frac{1}{2}z^*(\nu - 1). \quad (14.2.43)$$

As was shown in (18.4), intermediate terms of a chain can be effectively removed by summation so that the factor from $\Pi(e, k) \times \Lambda\{2, k'\}$ will still be $\nu - 1$. So for $e = 2$, from (16.4.14),

$$b_{2,2}(N^*) = \frac{1}{2}z^*(z - 1)(\nu - 1). \quad (14.2.44)$$

After this the lattice constants become more complicated because they have different contributions for different lattices. The lattice with the smallest value of z^* is the triangular lattice with $z^* = 3$. Then, according to (16.4.14), we must exclude the chain which closes around a face and calculate the lattice constant around the face separately. In general for any lattice a contribution to $b_{z^*,2}(N^*)$ arises from a simple cycle with two sites of the cycle containing single-vertex terms. As was shown in Example 18.4.7 the factor from $\Pi(e, k) \times \Lambda\{2, k'\}$ will be $(\nu - 1)(\nu - 2)$ with $\Omega(z^*, 1) = N^*$. The number of ways the two single-vertex terms can be distributed on the cycle is $\frac{1}{2}z^*(z^* - 1)$. So the contribution to $b_{z^*,2}(N^*)$ from the simple cycle round a lattice face is $\frac{1}{2}z^*(z^* - 1)(\nu - 1)(\nu - 2)$. Without extending the series into the range in which the lattices need to be distinguished

$$\phi_{\text{HT}}^{(2)}(\mathcal{N}, \nu) = -\frac{1}{2}z(\nu - 1)\{v + (z - 1)v^2\} + O(v^3). \quad (14.2.45)$$

and, from (14.2.34)–(14.2.36), (14.2.41) and (14.2.45) the expansions of the response functions can be derived.

It is clear from the examples given here and others which could be presented, using more complicated arrangements of chains and cycles (Example 18.4.9), that the calculation of high-temperature series coefficients, especially when $\mu \neq 0$, rapidly become very difficult using the kind of rather ad hoc analysis so far described. However, many systematic procedures exist for facilitating the calculations and, as we have already indicated, we shall mainly concentrate on one, the finite-lattice method described in Sect. 14.5.

14.2.3 Duality for Graphs

Comparison of (14.2.15) and (14.2.41) suggests the relationship

$$z^* \phi_{\text{LT}}^{(0)}(\mathcal{N}, \nu) = z \phi_{\text{HT}}^{(0)}(\mathcal{N}^*, \nu) \quad (14.2.46)$$

between low- and high-temperature series for the case $\mu = 0$ ($p = 0$ and $q = 0$ respectively), that is that they are related by lattice duality (the low-temperature series for $\mathcal{N}$ is the high-temperature series for $\mathcal{N}^*$, and vice-versa). This could be inferred from the Euler's relationship (16.2.7), the duality relationship (9.2.24) and

the defining relationships (14.2.2) and (14.2.29) for the low- and high-temperature partition functions respectively. However, it is of interest to see how this is realized in terms of the relationship between the graphs for the two series.

The general idea is to begin with a low-temperature (section) graph g on $\mathcal{N}$ and to add to it the $l[g]$ edges and their terminating vertices (lattice sites) which connect g to the rest of $\mathcal{N}$; $g^{(+)}$ denotes the resulting graph (which is not in general a section graph). The duality formulation of Sect. 9.2.1 can now be applied with the graph $g^{(+)}$ replacing the lattice $\mathcal{N}^{(+)}$. The dual high-temperature graph g^* is formed by drawing all the edges of $\mathcal{N}^*$ which cut the edges of $g^{(+)}$. It is clear that $\Omega(g^{(+)} | \mathcal{N}) = \Omega(g^* | \mathcal{N}^*)$.

For the purpose of establishing a duality relationship between the terms of a low- and high-temperature series, this account is not quite sufficient. We need also to be able to incorporate the feature represented by the index i in Sect. 14.2.1; that is we need to be able to fix the number of like and unlike pairs of perturbed states on $g^{(+)}$. We do this by designating the locations of all the like pairs. (To compute the whole series coefficients it will be necessary to count the number of different locations for these like pairs.) The partition function on $g^{(+)}$ is now computed with the restriction that like pairs with zero energy of interaction occur at the prescribed locations. The dual graph g^* is now formed by drawing edges only through the edges of $g^{(+)}$ not occupied by fixed like pairs. With this modification, (9.2.21) relates the couplings on $g^{(+)}$ and g^* and (9.2.23) gives

$$Z(g^{(+)}, \mathfrak{X}) = v^{-n(g^*)} (\mathfrak{X} - 1)^{e(g^*)} Z(g^*, \mathfrak{X}^*). \quad (14.2.47)$$

The next step is to relate $Z(g^*, \mathfrak{X}^*)$ to $\Pi(g^*)$, given by (14.2.37), with $(e, k) := g^*$. The key to this is to note, from (18.4.1), that

$$\Delta(v; \sigma) = -(1 - v)^{\delta_{Kr}(\sigma)}. \quad (14.2.48)$$

Thus

$$\Pi(g^*) = v^{-n(g^*)} (-1)^{e(g^*)} Z(g^*, 1 - v). \quad (14.2.49)$$

Substituting $\mathfrak{X}^* = 1 - v$ into (9.2.21) gives $\mathfrak{X} = 0$ or equivalently $K = -\infty$. This limiting value for the coupling, when substituted into $Z(g^{(+)}, \mathfrak{X})$ yields the number of arrangements of states on $g^{(+)}$ so that there are no like pairs (apart from those previously designated). This gives precisely that part of the contribution to a low-temperature series coefficient from $g^{(+)}$.¹⁹ From (14.2.47) and (14.2.48),

$$Z(g^{(+)}, 0) = \Pi(g^*), \quad (14.2.50)$$

¹⁹ Of course, it will then need to be multiplied by the appropriate lattice constant and the number of equivalent locations for the like pairs.

which establishes the relationship between low- and high-temperature series coefficients. It is of interest to check this result by calculating the partition functions for the graph [2, 1] of Table 16.1 in the two cases where the pair is constrained to be in the same state and where this is not the case and the partition functions of the two dual lattices (6, 1) and (7, 1) respectively of Table 16.2.

14.3 Cumulant Expansions

In Sect. 14.2 the first few terms of the low- and high-temperature moment series for the Potts model were derived in terms of an appropriate variable related to the first-neighbour interaction, with a contribution to a moment represented by a graph g on the lattice. In the case of low-temperature series g is a section graph corresponding to a set of edges drawn between a group of sites in states different from the ground state and terms in the moment arise from unlike pairs between g and the ambient ground state and unlike pairs within g . In the case of high-temperature series the graph g is not necessarily a section graph, but each factor in the expansion variable is related directly to an edge of g . The common feature of both low- and high-temperature graphs is that a contributing graph-edge makes only a linear contribution to the moment; all these graphs are ‘single-bonded’.

In Sect. 14.2.1 we used the partition function (moment) series to derive the free energy (cumulant) series and observed that the polynomial N -dependence of the moments arising from graphs of more than one component disappeared. The assertion made there that, in general, for all lattices and both low- and high-temperature expansions, disconnected graphs do not contribute to the cumulants needs now to be examined in more detail. In doing so we shall uncover another feature of the cumulant series, which is the role played by *multi-bonded graphs*. This can already be seen in a limited way when the process by means of which the series (14.2.8) was derived from the series (14.2.7) is considered; certain terms in (14.2.8) are partly produced by multiples of the terms in (14.2.7). This means that a factor of u arising from a particular pair of sites on a graph may be multiplied by itself and thus feature in the series term with a multiplicity greater than one. To follow this process more clearly we shall, in this section adopt a generalized approach in which the interaction variables between different pairs of sites are distinguished.

For the sake of simplicity all reference to the μ -dependence of the system is suppressed and

$$Z_{\square}(\mathcal{N}, w) := 1 + N_{\square}\{f_1 w + f_2 w^2 + \cdots\}, \quad (14.3.1)$$

is taken to represent in generic form either the low-temperature moment expansion (14.2.4) or the high-temperature moment expansion (14.2.31), with the subscript ‘ $\square$ ’ standing for ‘LT’ with $N_{\square}$ being the number of sites or ‘ $\square$ ’ standing for ‘HT’ with

$N_{\square}$ being the number of faces of the lattice.²⁰ Then

$$\Phi_{\square}(\mathcal{N}, w) = -\ln\{Z_{\square}(\mathcal{N}, w)\} := N_{\square}\{\tilde{f}_1 w + \tilde{f}_2 w^2 + \tilde{f}_3 w^3 + \dots\} \quad (14.3.2)$$

is the cumulant expansion for the (dimensionless) free energy.

Suppose that the low- or high-temperature variables u and v are replaced in (14.2.3) and (14.2.30) with different variables, $u(\mathbf{r}, \mathbf{r}')$ and $v(\mathbf{r}, \mathbf{r}')$ respectively, for each first-neighbour pair. This generalization does not affect the graphs which contribute to the low- and high-temperature series. It just involves the appearance of different sets of pair variables for each isomorphic graph. Denoting these variables generically by $w(\mathbf{r}, \mathbf{r}')$, $Z_{\square}(\mathcal{N}, w)$ in (14.3.1) is replaced by $Z_{\square}(\mathcal{N}, \{w(\mathbf{r}, \mathbf{r}')\})$ with the general term $f_e w^e$ replaced by

$$\sum_{\{g\}} f(g) \prod_g w(\mathbf{r}, \mathbf{r}'), \quad (14.3.3)$$

where the sum is over all graphs which, when the pair variables are all set to w , make a non-zero contribution to $f_e w^e$.

To obtain the cumulant expansion the Maclaurin method is used.²¹ This gives

$$\Phi_{\square}(\mathcal{N}, w) = -\left[\exp\left(w \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \mathcal{D}(\mathbf{r}, \mathbf{r}')\right) \ln[Z_{\square}(\mathcal{N}, \{w(\mathbf{r}, \mathbf{r}')\})] \right]_0, \quad (14.3.4)$$

where

$$\mathcal{D}(\mathbf{r}, \mathbf{r}') := \frac{\partial}{\partial w(\mathbf{r}, \mathbf{r}')} \quad (14.3.5)$$

and the subscript '0' indicates that all $\{w(\mathbf{r}, \mathbf{r}')\}$ are set equal to zero. Using the expansion

$$\exp\left(w \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \mathcal{D}(\mathbf{r}, \mathbf{r}')\right) = 1 + \sum_{\hat{e}=1}^{\infty} \frac{w^{\hat{e}}}{\hat{e}!} \left[\sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \mathcal{D}(\mathbf{r}, \mathbf{r}') \right]^{\hat{e}} \quad (14.3.6)$$

in (14.3.4) gives a power-series expansion in w .²²

²⁰ The variables $\mathfrak{z}$ and q are reintroduced when the analysis is particularized, respectively, to the low- and high-temperature cases.

²¹ The function $f(\mathbf{x})$ of the vector $\mathbf{x}$ can be expressed in the form

$$f(\delta\mathbf{x}) = [\exp(\delta\mathbf{x} \cdot \nabla) f(\mathbf{x})]_{\mathbf{x}=\mathbf{0}}$$

at a point $\delta\mathbf{x}$ near the origin. The expansion in powers of the components of $\delta\mathbf{x}$ is then obtained by using the power-series expansion of the exponential.

²² Equations (14.3.4) and (14.3.6) provide the starting point for the derivation of *moment-cumulant relations* between the moments $f_1, f_2, \dots$ and the cumulants $\tilde{f}_1, \tilde{f}_2, \dots$ (see e.g. Lavis and Bell 1999b, Sect. 7.5.3).

The expression in square brackets in (14.3.6) can be expanded multinomially into a sum of terms each containing a product of $\mathring{e}$ first-neighbour pair factors $\mathcal{D}(\mathbf{r}, \mathbf{r}')$. A subgraph corresponding to each product is obtained by placing a line between the first-neighbour sites $\mathbf{r}$ and $\mathbf{r}'$ for each factor $\mathcal{D}(\mathbf{r}, \mathbf{r}')$ present. However, the class of graphs obtained differs from those for the partition function since a factor $\mathcal{D}(\mathbf{r}, \mathbf{r}')$ may occur more than once, giving several edges linking a pair of vertices $\mathbf{r}$ and $\mathbf{r}'$. Thus, as indicated above, we have multi-bonded graphs. Consider the product

$$D(\mathring{e}, \mathring{k} | \mathbf{R}) := [\mathcal{D}(\mathbf{r}_1, \mathbf{r}'_1)]^{\alpha_1} [\mathcal{D}(\mathbf{r}_2, \mathbf{r}'_2)]^{\alpha_2} \cdots [\mathcal{D}(\mathbf{r}_e, \mathbf{r}'_e)]^{\alpha_e}. \quad (14.3.7)$$

The notation we have used here designates $(\mathring{e}, \mathring{k})$ as a multi-bonded graph with $\mathring{e}$ edges. Underlying this graph is the *silhouette* graph (e, k) obtained by replacing all the multiple edges by single edges. With $\alpha_1, \alpha_2, \dots, \alpha_e$ denoting the multiplicities of the edges in $(\mathring{e}, \mathring{k})$,

$$\mathring{e} = \alpha_1 + \alpha_2 + \cdots + \alpha_e. \quad (14.3.8)$$

Of course (14.3.7) contains information about a specific location for $(\mathring{e}, \mathring{k})$ on the lattice and this is formally represented by $\mathbf{R}$. The number of different values for $\mathbf{R}$ is the lattice constant $\Omega(e, k)$ of the silhouette graph. However, we also need to include in the lattice constant for $(\mathring{e}, \mathring{k})$ the factor $S(e, k | \mathring{e}, \mathring{k})$ which is the number of ways of arranging the extra $\mathring{e} - e$ edges on (e, k) to produced the topology $\mathring{k}$. Each graph $(\mathring{e}, \mathring{k})$ with location $\mathbf{R}$ will occur $\mathring{e}! A(\mathring{e}, \mathring{k})$ times, where

$$A(\mathring{e}, \mathring{k}) := \left(\prod_{i=1}^e \alpha_i! \right)^{-1}, \quad (14.3.9)$$

corresponds to all permutations of the differential operators.²³ So, from (14.3.4) and (14.3.6),

$$\Phi_{\square}(\mathcal{N}, w) = - \sum_{\mathring{e}=1}^{\infty} w^{\mathring{e}} \sum_{\{\mathring{k}\}} \Omega(e, k) S(e, k | \mathring{e}, \mathring{k}) C(\mathring{e}, \mathring{k}) A(\mathring{e}, \mathring{k}), \quad (14.3.10)$$

where

$$C(\mathring{e}, \mathring{k}) := \left[D(\mathring{e}, \mathring{k} | \mathbf{R}) \ln \{ Z_{\square}(\mathcal{N}, \{w(\mathbf{r}, \mathbf{r}')\}) \} \right]_0 \quad (14.3.11)$$

is independent of the location $\mathbf{R}$ and the order of the operators.

²³ In quantum systems these operators will not, in general, commute, but here we are concerned only with classical systems.

Before $D(\hat{e}, \hat{k} | \mathbf{R})$ acts on $\ln\{Z_{\square}(\mathcal{N}, \{w(\mathbf{r}, \mathbf{r}')\})\}$ in (14.3.11) all the variables $w(\mathbf{r}, \mathbf{r}')$ not corresponding to the edges of (e, k) can be set to zero. Let $Z_{\square}^{(e, k)}(\mathcal{N}, \{w(\mathbf{r}, \mathbf{r}')\})$ be the partition function for which this has been done. We now consider the implications of this in the low- and high-temperature cases.

14.3.1 The Low-Temperature Case

It follows, from (14.2.3), that all terms in the sum over σ are zero except those for which the two sites at the ends of a pair not included in $D(\hat{e}, \hat{k} | \mathbf{R})$ have the same value for σ , which we take to be the ground state $\sigma = 0$. This means that all sites of (e, k) connected by a edge to an exterior site must be in the ground state and the subset of those not in the ground state must form a section graph. It follows that (e, k) must be of the type for which, in Sect. 14.2.3 and at the beginning of this section, we have used the designation $\mathbf{g}^{(+)}$, that is a section graph $\mathbf{g}$ to which has been added all the edges connecting it to exterior sites, together with the terminating (ground state) sites on these edges, which we call the *boundary sites* of $\mathbf{g}^{(+)}$. The partition function $Z_{\text{LT}}(\mathcal{N}, \{u(\mathbf{r}, \mathbf{r}')\}, p)$ is reduced to

$$Z_{\text{LT}}((e, k), \{u(\mathbf{r}, \mathbf{r}')\}, p) = \sum_{\{\sigma\}} \left\{ \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(e, k)} u(\mathbf{r}, \mathbf{r}')^{1 - \delta^{\text{Kr}}(\sigma(\mathbf{r}) - \sigma(\mathbf{r}'))} \right\} \\ \times \left\{ \prod_{\{\mathbf{r}\}}^{(e, k)} (1 - p)^{1 - \delta^{\text{Kr}}(\sigma(\mathbf{r}))} \right\}, \quad (14.3.12)$$

where the summation is now restricted to the sites of the graph $\mathbf{g} := [\mathbf{n}, k]$ which is obtained from $\mathbf{g}^{(+)} := (e, k)$ by excluding the boundary sites and a non-zero contribution to $C(\hat{e}, \hat{k})$ will occur only from terms of the summation for which each edge is occupied by unlike states.

14.3.2 The High-Temperature Case

Here the partition function summations are performed over the exterior sites and, from (14.2.30), $Z_{\text{HT}}(\mathcal{N}, \{v(\mathbf{r}, \mathbf{r}')\}, q)$ is reduced to

$$Z_{\text{HT}}((e, k), \{v(\mathbf{r}, \mathbf{r}')\}, q) = v^{-n(e, k)} \\ \times \sum_{\{\sigma\}} \left\{ \prod_{\{\mathbf{r}, \mathbf{r}'\}}^{(e, k)} [1 + \Delta(v; \sigma(\mathbf{r}) - \sigma(\mathbf{r}'))v(\mathbf{r}, \mathbf{r}')] \prod_{\{\mathbf{r}\}}^{(e, k)} [1 + \Delta(v; \sigma(\mathbf{r}))q] \right\}, \quad (14.3.13)$$

where the summation is over all the sites of (e, k) , with a non-zero contribution to $C(\mathring{e}, \mathring{k})$, occurring if the corresponding product of Potts Δ -functions is non-zero.

It follows in both the low- and high-temperature cases that, if (e, k) consists of a number of components, the partition function is just a product of those for each component. The logarithm of the partition function is just the corresponding sum of partition functions and operation with $\mathcal{D}(\mathring{e}, \mathring{k} | \mathbf{R})$ yields the result of zero. Thus $C(\mathring{e}, \mathring{k}) = 0$ if $(\mathring{e}, \mathring{k})$ is a disconnected graph.

14.4 The Finite-Cluster Method

We now re-express (14.3.10) by denoting a multi-bonded graph by $\mathring{g}$ and generalizing $S(g | \mathring{g})$ so that it has the same meaning as before if g is the silhouette of $\mathring{g}$ and is zero otherwise. Then

$$\Phi_{\square}(\mathcal{N}, w) = - \sum_{\{g\}} \Omega(g) \Psi_{\square}(g, w), \quad (14.4.1)$$

where

$$\Psi_{\square}(g, w) := \sum_{\{\mathring{g}\}} w^{e(\mathring{g})} S(g | \mathring{g}) C(\mathring{g}) A(\mathring{g}). \quad (14.4.2)$$

Formula (14.4.1) is the *cluster expansion* of the free energy and the cluster function $\Psi_{\square}(g, w)$ is an infinite series in w , starting with the term $w^{e(g)} C(g)$, and summing over the cluster of graphs which have the same silhouette g . If the aim is to obtain a low- or high-temperatures series expansion of $\Phi_{\square}(\mathcal{N}, w)$ correct up to and including the terms of degree $e_{\max}$ in w then graphs can be excluded from (14.4.1) for which $e(g) > e_{\max}$. Thus we can replace (14.4.1) by

$$\overline{\Phi_{\square}(\mathcal{N}, w)}^{e_{\max}} := - \sum_{\{g\}}^{e(g) \leq e_{\max}} \Omega(g) \Psi_{\square}(g, w) \quad (14.4.3)$$

and the task is to obtain the set of cluster functions which appear in this equation. Of course, it is now possible, in principle, to proceed by calculating the terms in (14.4.2) up to degree $e_{\max}$ in w for each of the cluster functions in (14.4.3).

However, the finite-cluster method of Domb (1960, p. 322; 1974b, pp. 73–76) takes a different approach. The starting point is the observation that (14.4.1) can be generalized by replacing the lattice $\mathcal{N}$ by any graph g to give

$$\Phi_{\square}(g, w) = - \sum_{\{g'\}} \Omega(g' | g) \Psi_{\square}(g', w), \quad (14.4.4)$$

which is an explicit realization of Theorem 16.4.3. Consider now the set $\mathcal{S}$ of all graphs $\mathbf{g}$ such that $e(\mathbf{g}) \leq e_{\max}$ and suppose that $\mathcal{S}$ contains n members partially ordered, as described in Sect. 16.4.5, in the sequence $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$; that is to say, if $\mathbf{g}_i \subset \mathbf{g}_j$ then $i < j$. If $\mathbf{g}$ in (14.4.4) is a member $\mathbf{g}_i \in \mathcal{S}$ then the set of graphs in the summation on the right-hand side will consist only of members of $\mathcal{S}$ with $j \leq i$. Thus, by varying the graph $\mathbf{g}$ in (14.4.4) over the members of $\mathcal{S}$, a set of n linear equations in the n cluster functions can be obtained. This set can be expressed in the form

$$[\Phi_{\square}^{(n)}]^T = -[\Psi_{\square}^{(n)}]^T \Omega^{(n)}, \quad (14.4.5)$$

with

$$\Phi_{\square}^{(n)} := \begin{pmatrix} \Phi_{\square}(\mathbf{g}_1, w) \\ \Phi_{\square}(\mathbf{g}_2, w) \\ \vdots \\ \Phi_{\square}(\mathbf{g}_n, w) \end{pmatrix}, \quad \Psi_{\square}^{(n)} := \begin{pmatrix} \Psi_{\square}(\mathbf{g}_1, w) \\ \Psi_{\square}(\mathbf{g}_2, w) \\ \vdots \\ \Psi_{\square}(\mathbf{g}_n, w) \end{pmatrix}, \quad (14.4.6)$$

where

$$\Omega^{(n)} := \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1n-1} & \Omega_{1n} \\ 0 & \Omega_{22} & \cdots & \Omega_{2n-1} & \Omega_{2n} \\ 0 & 0 & \cdots & \Omega_{3n-1} & \Omega_{3n} \\ 0 & 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & 0 & \Omega_{nn} \end{pmatrix}, \quad \text{with } \Omega_{ji} := \Omega(\mathbf{g}_j | \mathbf{g}_i), \quad (14.4.7)$$

is the T matrix (Sect. 16.4.5). From (14.4.3),

$$\overline{\phi_{\square}(\mathcal{N}, w)}^{e_{\max}} = -[\Psi_{\square}^{(n)}]^T \boldsymbol{\omega}^{(n)}, \quad (14.4.8)$$

$\boldsymbol{\omega}^{(n)}$ is the vector of lattice constants for $\mathcal{S}$ and, from (14.4.5),

$$\overline{\phi_{\square}(\mathcal{N}, w)}^{e_{\max}} = [\Phi_{\square}^{(n)}]^T \boldsymbol{\chi}^{(n)} = \sum_{i=1}^n \chi_j^{(n)} \Phi_{\square}(\mathbf{g}_j, w), \quad (14.4.9)$$

where, as in Sects. 7.2 and 16.4.5, the weights $\chi_j^{(n)}$, $j = 1, 2, \dots, n$, of the sequence $\mathcal{S}$, are defined by

$$\boldsymbol{\chi}^{(n)} := [\Omega^{(n)}]^{-1} \boldsymbol{\omega}^{(n)}. \quad (14.4.10)$$

Example 14.4.1 To calculate the high-temperature series on the square lattice up to v^4 the sequence $\mathcal{S}$ has five members of which the first four $\mathfrak{g}_1, \mathfrak{g}_2, \mathfrak{g}_3, \mathfrak{g}_4$ are simple chains of lengths of one to four edges respectively, shown in Fig. 16.1, and $\mathfrak{g}_5$ is the simple cycle (4, 2) shown in Table 16.2.²⁴ From Table 16.1 and (16.4.14)

$$\mathbf{\Omega}^{(5)} := \begin{pmatrix} 1 & 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{\omega}^{(5)} := \begin{pmatrix} 2 \\ 6 \\ 18 \\ 50 \\ 1 \end{pmatrix}. \quad (14.4.11)$$

Thus, from (14.4.9) and (14.4.10),

$$\frac{4}{\phi_{\text{HT}}(\mathcal{N}, v, q)} = 8 \Phi_{\text{HT}}(\mathfrak{g}_1, v, q) + 24 \Phi_{\text{HT}}(\mathfrak{g}_2, v, q) - 86 \Phi_{\text{HT}}(\mathfrak{g}_3, v, q) \\ + 50 \Phi_{\text{HT}}(\mathfrak{g}_4, v, q) + \Phi_{\text{HT}}(\mathfrak{g}_5, v, q). \quad (14.4.12)$$

The high-temperature partition functions for the members of $\mathcal{S}$ are given by (14.3.13) with all the parameters $v(\mathbf{r}, \mathbf{r}')$ replaced by v . The easiest way to obtain them in explicit form is to expand the brackets and use the properties of the Potts Δ -function given in Sect. 18.4 and in particular in (18.4.18) and (18.4.20). Then

$$\Phi_{\text{HT}}(\mathfrak{g}_1, v, q) = -\ln[1 + vq^2(v-1)], \quad (14.4.13)$$

$$\Phi_{\text{HT}}(\mathfrak{g}_2, v, q) = -\ln[1 + v(2+v)q^2(v-1) \\ + v^2q^3(v-1)(v-2)], \quad (14.4.14)$$

$$\Phi_{\text{HT}}(\mathfrak{g}_3, v, q) = -\ln[1 + (3+2v+v^2)vq^2(v-1) \\ + 2(1+v)v^2q^3(v-1)(v-2) \\ + v^3q^4(v-1)(v-2)^2 + v^2q^4(v-1)^2], \quad (14.4.15)$$

$$\Phi_{\text{HT}}(\mathfrak{g}_4, v, q) = -\ln[1 + (4+3v+2v^2+v^3)vq^2(v-1) \\ + (3+4v+3v^2)v^2q^3(v-1)(v-2) \\ + (2+3v)v^3q^4(v-1)(v-2)^2 + q^5v^4(v-1)(v-2)^3 \\ + 3v^2q^4(v-1)^2 + 2v^3q^5(v-1)^2(v-2)], \quad (14.4.16)$$

²⁴ All chains of the same length but different topologies can be grouped together (e.g. graphs (3, 1)–(3, 4) in Table 16.1), since they have the same free energies and cluster functions.

$$\begin{aligned}
\Phi_{\text{HT}}(\mathfrak{g}_5, v, q) = & -\ln[1 + 4(1 + v + v^2)vq^2(v - 1) \\
& + 4(1 + 2v)v^2q^3(v - 1)(v - 2) + 4v^3q^4(v - 1)(v - 2)^2 \\
& + q^4v^4(v - 2)\{(v - 2)^3 + 1\} + 4q^3v^4(v - 2)\{(v - 2)^2 - 1\} \\
& + 6q^2v^4(v - 2)(v - 1) + v^4(v - 1) + 2v^2q^4(v - 1)^2].
\end{aligned} \tag{14.4.17}$$

Expanding these free energies to $O(v^4)$ and substituting in (14.4.12) gives, for the quantities defined in (14.2.33),

$$\phi_{\text{HT}}^{(0)}(\mathcal{N}, v) = -(v - 1)v^3 + O(v^5), \tag{14.4.18}$$

$$\phi_{\text{HT}}^{(2)}(\mathcal{N}, v) = -2(v - 1)[v + 3v^2 + 9v^3 + (3v + 19)v^4] + O(v^5). \tag{14.4.19}$$

As expected, from duality, (14.4.18) agrees with (14.2.21).²⁵ There is, of course, no agreement (and none to be expected) between (14.4.19) and (14.2.23).

A more general version of the model in Example 14.4.1, for a range of lattices and with different ‘chemical potentials’ for each Potts state, was investigated by Kim and Joseph (1975) and coefficients for the square lattice were obtained up to v^8 .²⁶ The finite-cluster method has also been used to obtain high-temperature expansions in terms of the coupling K . A method for deriving partition functions of finite clusters of spins in the classical Heisenberg model was developed by Joyce (1967) and used by Joyce and Bowers (1966) to calculate high-temperature K -expansions on various lattices. Joyce’s method was extended by Domb (1976) to give a systematic procedure for calculating the partition functions of star topologies in the n -vector model. This work was used by English et al. (1979) to obtain high-temperature series to K^{12} for general n and to K^{13} for the ($n = 3$) classical Heisenberg model.

14.5 The Finite-Lattice Method

Since the finite-lattice method is very similar to the finite-cluster method it is useful to make a comparison between them:

- (i) A *finite cluster* is just a graph. A *finite lattice* is “a cluster that displays certain characteristics of the infinite lattice” (de Neef 1975, p. 25). In particular this means that a finite lattice is a section graph. In the case where $\mathcal{N}$ is square or cubic, the finite lattice is a square or cubic block and for other lattices this result is also achieved by moving the positions of the sites while retaining the first-neighbour connectivity.
- (ii) Equation (14.4.9) is modified slightly to

²⁵ The comparative lengths of the series, illustrates the greater difficulty in obtaining high-temperature series, something which we shall see again with the finite-lattice method.

²⁶ A different expansion variable, equivalent to $v/(1 - v)$, was used in their calculations.

$$\frac{e_{\max}}{\phi_{\square}(\mathcal{N}, w)} = \frac{1}{n_c} \sum_{i=1}^n \chi_j^{(n)} \Phi_{\square}(\mathbf{g}_j, w), \quad (14.5.1)$$

where the partially-ordered sequence $\mathcal{S}$ of graphs $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ is now generated by the overlap procedure described in Sect. 16.4.6 and the factor n_c is included to make provision for the modification of the method to different lattices. In particular, as we shall see, for the honeycomb lattice the basic unit from which the graphs are constructed is a cell of two sites. Thus to obtain the correct free energy *per lattice site* $n_c = 2$. In Theorem 16.4.4 we showed that the sequence is complete, meaning that up to $\mathbf{g}_n$ there are no missing contributions to the sum in (14.5.1).

- (iii) In the *finite-cluster method* the mode of truncation of the series is determined by the cluster functions. In the *finite-lattice method* we need to develop an independent criterion for the selection of the generating set $\mathcal{S}$ for $\mathcal{S}$ so that the expansion is correct up to a certain specified degree $e_{\max}$ in w .

14.5.1 Block-Formation and Accuracy

The crucial element in determining the accuracy of the series expansion arising from (14.5.1) is the choice of the generating set $\mathcal{S}$ for $\mathcal{S}$. As indicated above the primary application of the finite-lattice method is to models on the square lattice²⁷ and we shall first consider block-formation and series accuracy in this case before discussing the modifications needed for the triangle and honeycomb lattices.

14.5.1.1 The Square Lattice

The set $\mathcal{S}$ consists of rectangular blocks of sites ($n_c = 1$) and the members of $\mathcal{S}$ generated by overlap are also rectangular blocks. The sequence is complete. So if every graph on the lattice with $e \leq e_{\max}$ is a subset of a member of $\mathcal{S}$ then the expansion of the expression on the right-hand side of (14.5.1) will have the required accuracy specified by the left-hand side. The type of blocks and the computation of accuracy depend on the nature of the graphs which feature in the series expansion and since these differ between the low- and high temperature cases they must be considered separately.

The low-temperature case:

Here a relevant graph $\mathbf{g}$ is a section graph, to which has been attached all the edges joining it to its environment together with their terminating sites, which are in the

²⁷ For which the argument can easily be extended to the simple cubic lattice.

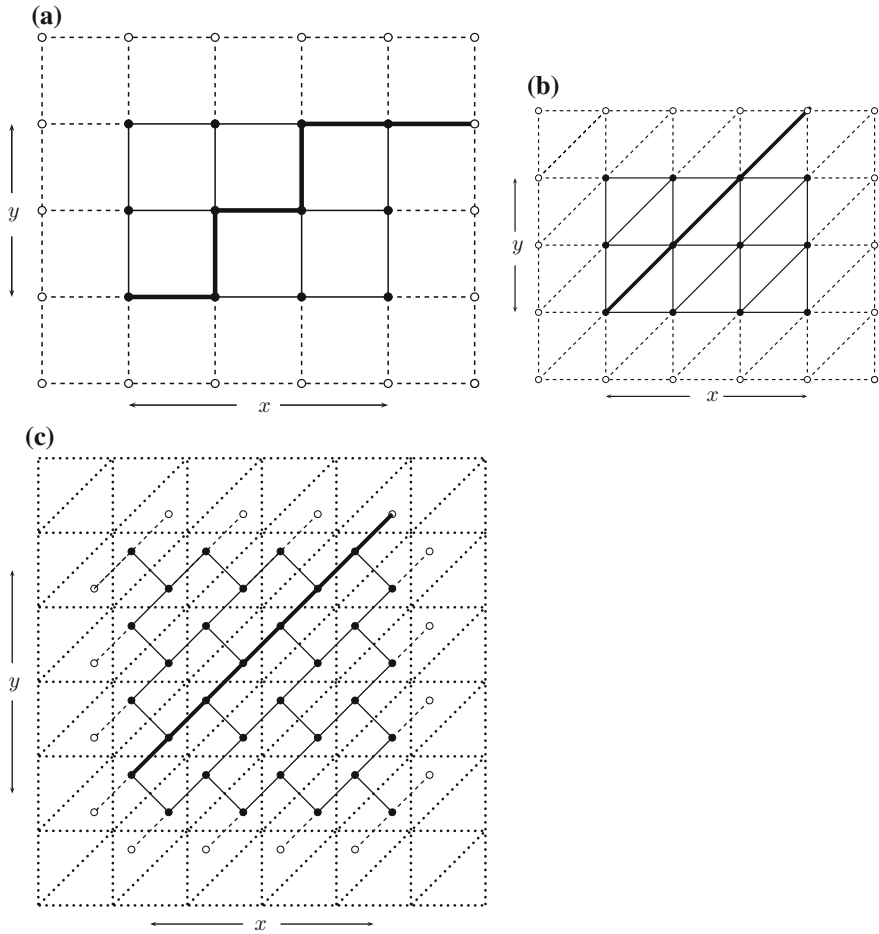


Fig. 14.1 Fully-extended chains on a low-temperature block for **a** the square lattice, **b** the triangular lattice and **c** the honeycomb lattice

ground state, to form the graph $\mathbf{g}^{(+)}$. In order to allow correct computation of the partition function for all graphs $\mathbf{g}$ on the sites of a rectangular block it must be surrounded by a line of sites in the ground state (Fig. 14.1a). In the notation of Sect. 16.4.6, a block is denoted by $[x \times y]$, where x and y are respectively the number of horizontal and vertical sites *excluding the surrounding ground-state sites* and the partition function for this block is $Z_{\text{LT}}([x \times y], u, 3)$. The power of u in $Z_{\text{LT}}([x \times y], u, 3)$ contributed by the graph $\mathbf{g}$ is given by (14.2.12) (with $z = 4$). We need to calculate the lowest power for the most fully-extended graph which cannot be accommodated in the block. This means that the graph must have only one component ($c = 1$) have all its sites in the same state ($i = 0$) and be a chain ($\mathcal{C} = 0$). The chain-configurations to be considered are then those which start at the

bottom left site and are extended on the block in a most extreme form, meaning that they cross any vertical (or horizontal) line of the block exactly once. If such a chain has n vertices then the sum of its horizontal and vertical extensions is $n - 1$. So, for all possible chain-configurations, blocks must be included for all x and y such that $x + y - 2 = n - 1$. The accuracy of the series is thus related to the sum of x and y and $\mathcal{S}$ must include all blocks with a particular sum. We conclude that the generating set $\mathfrak{S}$ must consist of all blocks with $x + y = \ell$, for some ℓ , and then $\mathcal{S}$ will consist of all blocks with $x + y \leq \ell$; ℓ is used to parameterize sequences (see, for example, Table 16.3). The first chain graph, with a configuration and extension as described here, which cannot be embedded in any member of $\mathcal{S}$ will have ℓ vertices. (In Fig. 14.1a where $\ell = 7$ an example of such a chain is shown by heavy lines.) Thus, from (14.2.12), the series will be correct to degree $e_{\max} = 2\ell + 2 - 1 = 2\ell + 1$.²⁸

The high-temperature case:

Here we must distinguish two cases, the series in v with $\mu = q = 0$ and the series of terms in v which contain a factor q^2 . The latter are exactly those needed for the response function series and the series will be accurate up to the number of edges in the most extended chain with a single-site term at each end; that is $e_{\max} = \ell - 2$. For the zero- μ series the limiting graphs are maximally extended (simple) contours. Suppose that the block is extended by one line of sites in any one direction, that is ℓ is replaced by $\ell + 1$. The most extended contour will consist of a most extended chain followed by a second chain completing the contour along another path without any intersections. This contour will have $2(\ell + 1) - 4$ edges and so $e_{\max} = 2\ell - 3$.²⁹

14.5.1.2 The Triangular Lattice

Here the argument is very similar, again $\mathcal{S}$ consists of all rectangular blocks with $x + y \leq \ell$ and $n_c = 1$, the topology of the triangular lattice being produced by adding diagonal bottom-left to top-right edges to each square (Fig. 14.1b). The remaining task is to calculate the accuracy achieved by $\mathcal{S}$ for a particular ℓ .

The low-temperature case:

For any block $[x \times y]$ the smallest chain in fully-extended form which cannot be embedded lies along the diagonal edges starting at the bottom left. It cannot be embedded in the block if it has more than the lesser of x and y sites. Applying this to the whole generating set $\mathfrak{S}$, for ℓ even the smallest non-embeddable chain of this type has $\ell/2 + 1$ vertices, and for ℓ odd it has $(\ell + 1)/2$ vertices. (In Fig. 14.1b

²⁸ For a d -dimensional hypercubic lattice the formula is $e_{\max} = 2(d - 1)(\ell + 2 - d) + 1$, with the meaning of ℓ generalized in the obvious way.

²⁹ For a d -dimensional hypercubic lattice the respective formulae are $e_{\max} = \ell - d$ and $e_{\max} = 2\ell + 1 - 2d$.

where $\ell = 7$ an example of such a chain is shown by heavy lines.) Thus, from (14.2.12) (with $z = 6$), $e_{\max} = 2\ell + 5$ if ℓ is even and $2\ell + 3$ if ℓ is odd.

The high-temperature case:

As for the square lattice we must distinguish two cases, the series in v with $\mu = q = 0$ and the series of terms in v which contain a factor q^2 . For the latter the series will be accurate up to the most extended chain with a single-site term at each end; that is $e_{\max} = (\ell - 2)/2$ if ℓ is even and $(\ell - 3)/2$ if ℓ is odd. For the zero- μ we must consider the maximally extended (simple) contours. Suppose that the block is extended by one line of sites in any one direction, that is ℓ is replaced by $\ell + 1$. If ℓ is odd the most extended chain is along the diagonal which has $(\ell - 1)/2$ edges followed by a return along the subdiagonal which has $(\ell - 3)/2$ edges plus the two connecting edges; thus $e_{\max} = \ell - 1$. If ℓ is even the block of size $\ell + 1$ has two diagonals with $\ell/2 - 1$ edges and connecting these with two edges gives again $e_{\max} = \ell - 1$.

14.5.1.3 The Honeycomb Lattice

The way that the honeycomb lattice is ‘forced’ into rectangular form is motivated by the triangle-honeycomb duality. Taking the rectangularized form of the triangular lattice, used in Fig. 14.1b and reproduced as dotted lines in Fig. 14.1c, a honeycomb lattice site is placed in each face. Thus the division of the lattice is into cells each containing two sites and x and y now represent respectively the number of horizontal and vertical cells with $\ell = x + y$, as before, but with $n_c = 2$.

The low-temperature case:

For any block $[x \times y]$ the smallest chain in fully-extended form which cannot be embedded lies along one of the two diagonal edges starting in the bottom-left cell. Either of these diagonals can be chosen if, as in Fig. 14.1c $x = y$. If $x < y$ the lower diagonal must be chosen and the smallest non-embeddable chain has $2x$ vertices; if $y < x$ the upper diagonal must be chosen and the smallest non-embeddable chain has $2y$ vertices. Applying this to the whole set $\mathfrak{S}$, the smallest non-embeddable chain of this type has ℓ vertices for ℓ even and $\ell - 1$ vertices if ℓ is odd. (In Fig. 14.1c where $\ell = 8$ an example of such a chain is shown by heavy lines.) Thus, from (14.2.12) (with $z = 3$), $e_{\max} = \ell + 1$ if ℓ is even and ℓ if ℓ is odd.

The high-temperature case:

The series terms in v which contain a factor q^2 will be accurate up to the most extended chain with a single-site term at each end; that is $e_{\max} = \ell - 2$ if ℓ is even

and $\ell - 3$ if ℓ is odd. For zero- μ the maximally extended (simple) contour will contain the two diagonal edges of length $\ell - 2$ if ℓ is even and $\ell - 3$ if ℓ is odd, starting in the bottom-left cell, and the two edges connecting them. Thus $e_{\max} = 2\ell - 2$ if ℓ is even and $2\ell - 4$ if ℓ is odd.

14.5.2 Constructing Block Partition Functions

As we have seen in Sect. 14.5.1 a graph in the sequence $\mathcal{S}$ has sites in an $x \times y$ rectangular block of cells. In the case of the square and triangular lattices each cell contains one site but in the case of the honeycomb lattice it contains two sites, an upper-left site and a lower-right site (Fig. 14.1).

Let $\tilde{Z}_{\square}(\ell, w)$ be the product of all partition functions $Z_{\square}([x \times y], w)$ for graphs in the sequence $\mathcal{S}$ for which $x + y = \ell$. Then, from (14.5.1) and (16.4.47),

$$\frac{e_{\max}}{\phi_{\square}(\mathcal{N}, w)} = -\ln[Z_{\square}(\ell, w)], \quad (14.5.2)$$

where

$$Z_{\square}(\ell, w) := \left[\frac{\tilde{Z}_{\square}(\ell - 2, w)}{\tilde{Z}_{\square}(\ell - 1, w)} \right]^3 \frac{\tilde{Z}_{\square}(\ell, w)}{\tilde{Z}_{\square}(\ell - 3, w)} \quad (14.5.3)$$

is an approximation to the N th root of the partition function.

The great advantage of the finite-lattice method based on rectangular blocks of sites is that, for a fixed y , the partition functions with increasing x can be generated using the transfer matrix methods described in Chap. 11. Here there are two differences from the discussion there and in Chap. 12. Now the slices of the lattice are columns of y sites with the lattice growing by increasing in the horizontal direction the number of columns x . Also there are no periodic boundary conditions in the vertical direction. In the case of the low-temperature series the lattice is of width $x + 2$ with the first and last columns fixed in the ground state $\sigma = 0$. In the same way each column has $y + 2$ sites with the top and bottom cells fixed in the ground state. In the case of the high-temperature series there are x columns of height y with free boundary conditions. The number of states of a column, for both the low- and high-temperature cases, is ν^y for the square and triangular lattices and ν^{2y} for the honeycomb lattice. From (11.1.9), with N_1 and N_2 replaced respectively by x and y ,

$$Z_{\square}([x \times y], w) = \tau[\mathbf{z}_{x-1}]^T \mathbf{b}^{(1)}, \quad (14.5.4)$$

where,

$$[\mathbf{z}_0]^T := [\mathbf{b}^{(F)}]^T, \quad [\mathbf{z}_x]^T = [\mathbf{z}_{x-1}]^T V \quad (14.5.5)$$

and, in (14.5.4), for the low-temperature case, $\tau = 1$. For the high-temperature case, following (14.2.30), $\tau = 1/\nu^{xy}$ for the square and triangular lattices and $\tau = 1/\nu^{2xy}$ for the honeycomb lattice.

As in Chap. 11, the transfer matrix contains the interaction between two columns of sites and half the interactions within a column. Explicitly,

$$V = V_w^{\frac{1}{2}} V_b V_w^{\frac{1}{2}}, \quad (14.5.6)$$

where V_w is a diagonal matrix. In a similar way the boundary vectors $\mathbf{b}^{(F)}$ and $\mathbf{b}^{(L)}$ contain half the interactions within the first and last columns respectively and, for the low-temperature case, the extra two columns are included in these vectors. Thus

$$\mathbf{b}^{(F)} = \mathbf{b}^{(L)} = V_w^{\frac{1}{2}} \mathbf{b}_{\square}, \quad (14.5.7)$$

where

$$\mathbf{b}_{LT} = V_b V_w \mathbf{i}_0, \quad \mathbf{b}_{HT} = \mathbf{j}, \quad (14.5.8)$$

$\mathbf{i}_0$ is the unit vector with one in the first element, fixing the first and last columns in the ground state, and $\mathbf{j}$ is the vector with every element of unity, enabling free boundary conditions in first and last columns.

For the square and triangular lattices we denote by σ the state of a column, where $[\sigma]^T := (\sigma_1, \sigma_2, \dots, \sigma_y)$, with the labelling proceeding from the top to the bottom (excluding in the low-temperature case the boundary sites with $\sigma = 0$). For the honeycomb lattices we again denote by σ the state of a column, but now $[\sigma]^T := (\sigma_1, \tilde{\sigma}_1, \sigma_2, \tilde{\sigma}_2, \dots, \sigma_y, \tilde{\sigma}_y)$, where σ_k and $\tilde{\sigma}_k$ are the states on the top-left and bottom-right sites in the k th cell. We now list the formulae for the elements of V_w and V_b for the low- and high-temperature cases and for the square, triangular and honeycomb lattices.

14.5.2.1 The Square Lattice

From (14.2.3) and (14.2.30) the diagonal elements of V_w are

$$(V_w)_{\sigma\sigma} = u^{1-\delta^{Kr}(\sigma_1)} \left\{ \prod_{k=1}^{y-1} u^{1-\delta^{Kr}(\sigma_k - \sigma_{k+1})} \right\} \left\{ \prod_{k=1}^y (1-p)^{1-\delta^{Kr}(\sigma_k)} \right\} u^{1-\delta^{Kr}(\sigma_y)} \quad (14.5.9)$$

for the low-temperature case, and

$$(V_w)_{\sigma\sigma} = \left\{ \prod_{k=1}^{y-1} [1 + \Delta(v; \sigma_k - \sigma_{k+1})v] \right\} \left\{ \prod_{k=1}^y [1 + \Delta(v; \sigma_k)q] \right\} \quad (14.5.10)$$

for the high-temperature case. The elements of V_b are

$$(V_b)_{\sigma\sigma'} = \prod_{k=1}^y u^{1-\delta^{Kr}(\sigma_k-\sigma'_k)} \quad (14.5.11)$$

for the low-temperature case, and

$$(V_b)_{\sigma\sigma'} = \prod_{k=1}^y [1 + \Delta(v; \sigma_k - \sigma'_k)v] \quad (14.5.12)$$

for the high-temperature case.

14.5.2.2 The Triangular Lattice

Here, as can be seen from Fig. 14.1a and b, the elements of V_w are the same as those for the square lattice, but the elements of V_b contain an additional diagonal interaction giving

$$(V_b)_{\sigma\sigma'} = u^{1-\delta^{Kr}(\sigma_1)} \left\{ \prod_{k=1}^y u^{1-\delta^{Kr}(\sigma_k-\sigma'_k)} \right\} \left\{ \prod_{k=2}^y u^{1-\delta^{Kr}(\sigma_k-\sigma'_{k-1})} \right\} u^{1-\delta^{Kr}(\sigma'_y)} \quad (14.5.13)$$

for the low-temperature case, and

$$(V_b)_{\sigma\sigma'} = \left\{ \prod_{k=1}^y [1 + \Delta(v; \sigma_k - \sigma'_k)v] \right\} \left\{ \prod_{k=2}^y [1 + \Delta(v; \sigma_k - \sigma'_{k-1})v] \right\} \quad (14.5.14)$$

for the high-temperature case.

14.5.2.3 The Honeycomb Lattice

In this case the elements of both V_w and V_b are more complicated because of the two sites in each cell. However, by looking at Fig. 14.1c, it is not difficult to see that

$$(V_w)_{\sigma\sigma} = u^{1-\delta^{Kr}(\sigma_1)} \left\{ \prod_{k=1}^{y-1} u^{2-\delta^{Kr}(\sigma_k-\tilde{\sigma}_k)-\delta^{Kr}(\tilde{\sigma}_k-\sigma_{k+1})} \right\} \\ \times \left\{ \prod_{k=1}^y (1-p)^{2-\delta^{Kr}(\sigma_k)-\delta^{Kr}(\tilde{\sigma}_k)} \right\} u^{2-\delta^{Kr}(\sigma_y-\tilde{\sigma}_y)-\delta^{Kr}(\tilde{\sigma}_y)} \quad (14.5.15)$$

for the low-temperature case,

$$(V_w)_{\sigma\sigma'} = \prod_{k=1}^{y-1} [1 + \Delta(v; \sigma_k - \tilde{\sigma}_k)v][1 + \Delta(v; \tilde{\sigma}_k - \sigma_{k+1})v] \\ \times [1 + \Delta(v; \sigma_y - \tilde{\sigma}_y)v] \left\{ \prod_{k=1}^y [1 + \Delta(v; \sigma_k)q][1 + \Delta(v; \tilde{\sigma}_k)q] \right\} \quad (14.5.16)$$

for the high-temperature case,

$$(V_b)_{\sigma\sigma'} = \prod_{k=1}^y u^{1-\delta^{\text{Kr}}(\tilde{\sigma}_k - \sigma'_k)} \quad (14.5.17)$$

for the low-temperature case, and

$$(V_b)_{\sigma\sigma'} = \prod_{k=1}^y [1 + \Delta(v; \tilde{\sigma}_k - \sigma'_k)v] \quad (14.5.18)$$

for the high-temperature case.

14.5.3 Calculating the Series

Using the results displayed above it is now possible to obtain series for the terms of the free-energy density to the specified level of accuracy. Techniques for doing this were developed by Enting and coworkers (Enting 1978b, 1980a; Adler et al. 1983) and although there have been some improvements in the programs, most advances shown in the papers cited below are probably due to greatly enhanced computing power. However, for the present heuristic purposes, and for small values of ℓ the reader is recommended to use one of the available computer algebra packages like MAPLE or MATHEMATICA for which no further details are needed here, except to note that, since $Z_{\square}([x \times y], w) = Z_{\square}([y \times x], w)$, it is necessary simply to obtain the more easily calculated of the two partition functions, which is usually that for which $x > y$.

For simplicity we confine our initial discussion to the ν -state Potts model on the (self-dual) square lattice and the triangle-honeycomb dual lattice pair. Then, in the case of the square and triangular lattices, for a particular value of ℓ , the order of the largest transfer matrix is $\nu^{\bar{y}}$, where $\bar{y}$ is the integer part of $\ell/2$. In the case of the honeycomb lattice, since there are two sites per cell, the transfer matrices given in the form shown in Sect. 14.5.2.3 have largest order $\nu^{2\bar{y}}$. From this it is clear that computation time increases with both ν and ℓ , with that for the honeycomb

lattice being larger than that for the square and triangular lattices.³⁰ In the following we restrict our explicit results to the 3-state model with $\ell = 6$ for the square and triangular lattices and $\ell = 5$ for the more time-consuming honeycomb case.

14.5.3.1 The Square Lattice

Using the formulae for the transfer matrices in Sect. 14.5.2.1 it is straightforward to obtain series for the low- and high-temperature free-energy densities and then for the order parameter and response functions.

For the 3-state Potts model and $\ell = 6$ the largest matrices V_w and V_b needed are of dimension 27, used to compute $Z_{LT}([3 \times 3], u)$ or $Z_{HT}([3 \times 3], v)$. The results for the free-energy densities are

$$\begin{aligned} \phi_{LT}^{(0)}(\mathcal{N}_s, u) = & -2u^4 - 4u^6 - 4u^7 - 4u^8 - 24u^9 - 16u^{10} - 60u^{11} \\ & - \frac{518}{3}u^{12} - 128u^{13} + 216u^{14} + O(u^{15}), \end{aligned} \quad (14.5.19)$$

$$\begin{aligned} \phi_{LT}^{(1)}(\mathcal{N}_s, u) = & 2u^4 + 8u^6 + 8u^7 + 24u^8 + 72u^9 + 140u^{10} + 320u^{11} \\ & + 1164u^{12} + 1560u^{13} - 892u^{14} + O(u^{15}), \end{aligned} \quad (14.5.20)$$

$$\begin{aligned} \phi_{LT}^{(2)}(\mathcal{N}_s, u) = & -4u^6 - 4u^7 - 38u^8 - 72u^9 - 352u^{10} - 616u^{11} \\ & - 3340u^{12} - 5276u^{13} + 100u^{14} + O(u^{15}) \end{aligned} \quad (14.5.21)$$

and

$$\phi_{HT}^{(0)}(\mathcal{N}_s, v) = -2v^4 - 4v^6 - 4v^7 - 4v^8 - 24v^9 + 40v^{10} + O(u^{11}), \quad (14.5.22)$$

$$\begin{aligned} \phi_{HT}^{(2)}(\mathcal{N}_s, v) = & -4v - 12v^2 - 36v^3 - 112v^4 - 192v^5 - 364v^6 - 332v^7 \\ & + 36v^8 + 1736v^9 + 5625v^{10} + O(u^{11}). \end{aligned} \quad (14.5.23)$$

Using (14.2.9)–(14.2.11) and (14.5.19)–(14.5.21), the low-temperature series for the order parameter and response functions are

$$\begin{aligned} \tilde{\rho} = & 1 - 3u^4 - 12u^6 - 12u^7 - 36u^8 - 108u^9 - 210u^{10} - 480u^{11} \\ & - 1746u^{12} - 2340u^{13} + 1338u^{14} + O(u^{15}), \end{aligned} \quad (14.5.24)$$

³⁰ A modification of the honeycomb case can be made by reformulating the transfer matrices into ones between vertical rows of sites. This reduces the order but increases the number of matrix multiplications, still leaving the honeycomb case more expensive in computing time than the square and triangular cases.

$$\begin{aligned} \frac{c_\rho}{k_B K^2} = & 32u^4 + 144u^6 + 196u^7 + 256u^8 + 1944u^9 + 1600u^{10} + 7260u^{11} \\ & + 24864u^{12} + 21632u^{13} - 42336u^{14} + O(u^{15}), \end{aligned} \quad (14.5.25)$$

$$\begin{aligned} \varphi_T = & 2u^4 + 16u^6 + 16u^7 + 100u^8 + 216u^9 + 844u^{10} + 1552u^{11} \\ & + 7844u^{12} + 12112u^{13} - 1092u^{14} + O(u^{15}) \end{aligned} \quad (14.5.26)$$

and, using (14.2.34)–(14.2.36) and (14.5.22)–(14.5.23), the corresponding high-temperature series are

$$\begin{aligned} \frac{c_\rho}{k_B K^2} = & \frac{8}{3}v + 16v^2 + \frac{368}{9}v^3 + \frac{896}{9}v^4 \\ & + \frac{2672}{9}v^5 + 616v^6 + \frac{13616}{9}v^7 + \frac{16976}{9}v^8 + O(v^9), \end{aligned} \quad (14.5.27)$$

$$\begin{aligned} \frac{\varphi_T}{v-1} = & 1 + 4v + 12v^2 + 36v^3 + 112v^4 + 192v^5 + 364v^6 \\ & + 332v^7 - 36v^8 - 1736v^9 - 5628v^{10} + O(v^{11}). \end{aligned} \quad (14.5.28)$$

De Neef and Enting (1977) obtained the series expansion for the 3-state square-lattice model for the zero- μ free-energy density up to u^{22} and this was extended by Enting (1980a), with the addition of a low-temperature series for the order parameter, to u^{31} . Further extensions, with inclusion of low-temperature series for the field-response function, were made to degree u^{35} by Adler et al. (1983). Briggs et al. (1994) presented results for Potts models with $v = 2, 3, \dots, 10$ which, for $v = 3$, extended the series for the order parameter and field response function to u^{37} . These results confirm the degree of accuracy of the low-temperature series given here as $13 = 2\ell + 1$, confirming the prediction in Sect. 14.5.1.1. Arisue and Tabata (1997) used an improvement of the finite-lattice method to extend low-temperature Potts model series on the square lattice to u^{41} , but only for values of v in the range $[5, 50]$.

Because of self-duality the accuracy of the high-temperature series (14.5.22) can be inferred from (14.5.19) to be up to $9 = 2\ell - 3$, as predicted in Sect. 14.5.1.1. There are no $\mu \neq 0$ high-temperature results for this model using the finite-lattice method.³¹ However, the work of Kim and Joseph (1975) using the finite-cluster method shows that the series (14.5.28) is accurate to degree $4 = \ell - 2$, which is again that predicted in Sect. 14.5.1.1.

14.5.3.2 The Triangular and Honeycomb Lattices

For the triangular lattice with $v = 3$ and $\ell = 6$

$$\begin{aligned} \phi_{\text{LT}}^{(0)}(\mathcal{N}_T, u) = & -2u^6 - 6u^{10} - 6u^{11} + 10u^{12} - 36u^{14} - 36u^{15} + 54u^{16} + 72u^{17} \\ & - \frac{794}{3}u^{18} + O(u^{19}), \end{aligned} \quad (14.5.29)$$

³¹ See the list of papers in Enting (1996).

$$\phi_{\text{LT}}^{(1)}(\mathcal{N}_T, u) = 2u^6 + 12u^{10} + 12u^{11} - 16u^{12} + 114u^{14} + 108u^{15} - 102u^{16} - 168u^{17} + 1016u^{18} + O(u^{19}), \quad (14.5.30)$$

$$\phi_{\text{LT}}^{(2)}(\mathcal{N}_T, u) = -6u^{10} - 6u^{11} + 2u^{12} - 126u^{14} - 108u^{15} - 30u^{16} + 72u^{17} - 1602u^{18} + O(u^{19}) \quad (14.5.31)$$

and

$$\phi_{\text{HT}}^{(0)}(\mathcal{N}_T, u) = -4v^3 - 6u^4 - 18v^5 - 34v^6 - 52v^7 + 38v^8 + \frac{1808}{3}v^9 + 2028v^{10} + O(v^{11}), \quad (14.5.32)$$

$$\phi_{\text{HT}}^{(2)}(\mathcal{N}_T, u) = -6v - 30v^2 - 136v^3 - 524v^4 - 1450v^5 - 2782vu^6 - 44v^7 + 22062v^8 + 92264v^9 + 202888v^{10} + O(u^{11}). \quad (14.5.33)$$

For the honeycomb lattice with $\nu = 3$ and $\ell = 5$

$$\phi_{\text{LT}}^{(0)}(\mathcal{N}_{\text{HC}}, u) = -2u^3 - 3u^4 - 9u^5 - 19u^6 - 52u^7 + O(u^8), \quad (14.5.34)$$

$$\phi_{\text{LT}}^{(1)}(\mathcal{N}_{\text{HC}}, u) = 2u^3 + 6u^4 + 24u^5 + 82u^6 + 290u^7 + O(u^8), \quad (14.5.35)$$

$$\phi_{\text{LT}}^{(2)}(\mathcal{N}_{\text{HC}}, u) = -3u^4 - 21u^5 - 127u^6 - 628u^7 + O(u^8) \quad (14.5.36)$$

and

$$\phi_{\text{HT}}^{(0)}(\mathcal{N}_{\text{HC}}, v) = -v^6 - 2v^{10} - 2v^{11} + O(v^{12}), \quad (14.5.37)$$

$$\phi_{\text{HT}}^{(2)}(\mathcal{N}_{\text{HC}}, v) = -3v - 6v^2 - 12v^3 - 22v^4 - 36v^5 - 67v^6 - 64v^7 - 57v^8 - 38v^9 - 46v^{10} + 142v^{11} + O(v^{12}). \quad (14.5.38)$$

From these results low- and high-temperature series can be obtained for the order parameter and response functions.

Jensen et al. (1997) used the finite-lattice method augmented by the extension procedure of Jensen and Guttmann to obtain low-temperature series for $\nu = 3$ (up to u^{34}) on the honeycomb lattice³² and Feldmann et al. (1998) used the same extension procedure to obtain the corresponding series for $\nu = 3$ (up to u^{68}) and $\nu = 4$ (up to u^{59}) on the triangular lattice and $\nu = 4$ (up to u^{30}) on the honeycomb lattice. Their results confirm the predicted level of accuracy of (14.5.29)–(14.5.31), (14.5.34)–(14.5.36).

As in the case of the square lattice there is a paucity of high-temperature results. However, the duality relationship (14.2.46) shows that $\phi_{\text{HT}}^{(0)}(\mathcal{N}_T, v)$ is accurate to v^5 as predicted in Sect. 14.5.1.2 ($5 = \ell - 1$) and $\phi_{\text{HT}}^{(0)}(\mathcal{N}_{\text{HC}}, v)$ is accurate to v^6 (for non-zero coefficients) as predicted in Sect. 14.5.1.3 ($6 = 2\ell - 4$).

³² And also, although we have not discussed this, on the kagomé lattice for $\nu = 3$ (up to u^{66}) and $\nu = 4$ (up to u^{56}).

As we have already indicated the finite-lattice method can also be used for three-dimensional lattices. In particular, on the simple cubic lattice, Guttman and Enting (1993) extended the low-temperature spin- $\frac{1}{2}$ Ising model free-energy series to degree 26 and Guttman and Enting (1994) obtain corresponding results for the 3-state Potts model to degree 43. Spin- s Ising models for $s > \frac{1}{2}$, do not, even on the square lattice, have duality properties, and it is of particular interest to obtain series approximations for such models. Adler and Enting (1984) obtained 45-term expansions for the magnetization, heat capacity and susceptibility of the spin-1 Ising model on the square lattice. The results were extended to the degree 79 by Enting et al. (1994) and higher-order spin values were investigated by Jensen et al. (1996). The finite-lattice method has also been applied to models with multi-site interactions. In particular Enting and Wu (1982) and Enting (1980b) studied the 3-state Potts model on the triangular lattice with 3-site interactions.

Guttman and Enting (1993) obtained high-temperature series for the magnetization and susceptibility for the spin- $\frac{1}{2}$ Ising model on the simple cubic lattice up to degree 26 and Guttman and Enting (1994) obtained corresponding results for the 3-state Potts model up to degree 21.

14.6 The Analysis of Series: Second-Order Transitions

Most analyses of series expansions is directed towards the extraction of critical parameters and exponents associated with singular behaviour of the thermodynamic function under examination at a phase transition³³ and here we give an account of the main methods that have been used for doing this.³⁴ If the value of either a critical parameter or exponent is known (or assumed to be known) and this value is used to estimate the value of the other quantities, this is known as a *biased* estimate; otherwise estimates are said to be *unbiased*.

Given that $F(w)$ is a function of a variable $w(T)$ which for the low- and high-temperature cases is such that $w(0) = 0$ and $w(\infty) = 0$ respectively, the interest is in the singularities of $F(w)$ in some disc in the complex w plane $\mathbb{C}_w$, centred at the origin. In the case of a low-temperature series, when $w := u$ given by (14.2.5), the physical condition $u > 0$ means that any critical point must be on the positive real axis in $\mathbb{C}_u$. In the case of a high-temperature series, when $w := v$ given by (14.2.25), v can be positive or negative corresponding respectively to $R > 0$ and $R < 0$ and, for the Ising model on loose-packed lattices, the ferromagnetic/antiferromagnetic symmetry $+R \longleftrightarrow -R$ (Sect. 3.5.1.1) guarantees that critical parameter values occur in pairs $\pm v_c$. In the rest of this discussion we shall assume that $F(w)$ has

³³ Rather less attention has been given to the way that series analysis can be used to detect first-order transitions and we shall turn to that topic in Sect. 14.7.

³⁴ For more detailed accounts the reader is referred to the review articles by Gaunt and Guttman (1965) and Guttman (1989).

exactly one singular point $w = w_c$ on the positive real axis of $\mathbb{C}_w$ ³⁵ and then, of course, the radius of convergence r of the series (14.1.1) must satisfy the condition $r \leq w_c$.

As indicated above the starting point for the series analysis of (14.1.1) is an hypothesized form for $F(w)$ and the truncated expansion (14.1.3) for some ℓ . One way to test the effectiveness of a method is to use a known analytic result, like the magnetization (12.3.283) of the square-lattice Ising model. Another is to use a range of plausible test functions. Such a programme was initiated by Hunter and Baker (1973) in relation to the ratio and Padé methods, developed further by Baker and Hunter (1973) and used to investigate the efficacy of the differential approximant method by Guttmann (1975).

14.6.1 Late-Term Analysis

It can be shown (Hille 1959) that, if $f_k \geq 0$, for all k greater than some $k_0 \geq 0$, then there exists a real positive r with $w = r$ a singular point of $F(w)$ and no singularities with $|w| < r$, giving $w_c = r$.³⁶ Since one can always choose to analyze $-F(w)$ rather than $F(w)$, the crucial point here is whether after some k_0 all the coefficients have the same sign, when, as indicated in Sect. 14.1, the critical point $w = w_c$ will be the nearest singularity to the origin and $1/w_c$ is given by the limit (14.1.2).³⁷ These remarks underline the fact that, in inferring the properties of $F(w)$, attention must be paid to what may be called (Dingle 1973, Chap. 7) the ‘late terms’ of the series (14.1.1). Important properties like sign-periodicity, and the inference just mentioned which can be drawn from all late coefficients being positive, will not, in general, be evident from the first few terms. Hence the need, given a general form for $F(w)$, for an asymptotic expression for the coefficient f_k . This is provided by the work of Darboux (1878, pp. 17–20). In general terms he showed that the asymptotic form for f_k is given by the singularities of $F(w)$ closest to the origin; that is to say on the circle of convergence. Specifically, suppose that, $F(w)$ which is analytic in some disc $|w| < r + \varepsilon$ ($\varepsilon > 0$) in $\mathbb{C}_w$, apart from algebraic singularities at $w_j = r \exp(i\theta_j)$, $j = 0, 1, \dots, s-1$, has the form

$$F(w) := \sum_{j=0}^{s-1} \left[A_j(w) \left(1 - \frac{w}{w_j} \right)^{-\lambda_j} + W_j(w) \right], \quad (14.6.1)$$

³⁵ Excluding models like those discussed in Sect. 10.4 which have upper- and lower-critical temperatures.

³⁶ There may, of course, be other singularities on the circle $|w| = r$.

³⁷ This provides a ‘simple’ method for obtaining w_c but, as we have shown by examples in Sect. 14.1, convergence is often very slow, leading to the need for the more sophisticated methods described below.

where $A_j(w)$ and $W_j(w)$ are analytic in a neighbourhood of $w = w_j$. Let

$$f_k = \sum_{j=0}^{s-1} f_{kj}, \quad (14.6.2)$$

where f_{kj} is the contribution to f_k from the singularity at $w = w_j$. Then it can be shown that, in calculating f_{kj} , $W_j(w)$ can be neglected. Replacing $A_j(w)$ by its Taylor expansion truncated at degree n gives

$$A_j(w) = \sum_{m=0}^n \left(1 - \frac{w}{w_j}\right)^m \frac{(-w_j)^m A_j^{(m)}(w_j)}{\Gamma(m+1)} \quad (14.6.3)$$

and expanding about $w = 0$ yields

$$f_{kj} = \sum_{m=0}^n \frac{(-w_j)^{m-k} A_j^{(m)}(w_j) \Gamma(m - \lambda_j + 1)}{\Gamma(k+1) \Gamma(m+1) \Gamma(m - \lambda_j - k + 1)} + O\left(\frac{1}{k^{n+2-\lambda_j}}\right). \quad (14.6.4)$$

The approximation in halting the series (14.6.3) and (14.6.4) at n is called the *n*th *Darboux approximation*. That the error in f_{kj} is $O(1/k^{n+2-\lambda_j})$ follows from the gamma function's asymptotic property $\Gamma(k+1) \Gamma(1+\alpha-k) \sim k^{\alpha+1}$.

This result was generalized to the case where algebraic singularities are confluent by Fields (1967) and Wong and Zhao (2005). Let $w_j = w_c$ with $\lambda_0 = \lambda - \Delta_j$, where $0 = \Delta_0 < \Delta_1 \leq \Delta_2 \leq \dots \leq \Delta_{s-1} < \lambda$. Then, from (14.6.4),

$$f_{kj} = \sum_{m=0}^n \frac{(-w_c)^{m-k} A_j^{(m)}(w_c) \Gamma(m + \Delta_j - \lambda + 1)}{\Gamma(k+1) \Gamma(m+1) \Gamma(m + \Delta_j - \lambda - k + 1)} + O\left(\frac{1}{k^{n+2+\Delta_j-\lambda}}\right). \quad (14.6.5)$$

The accuracy of these approximations in the case of known functions can be checked. In particular for the case of the Ising model magnetization where the only singularity on the circle of convergence is the critical point.

14.6.2 The Ratio Method

In the special case $s = 1$ of (14.6.1), where

$$F(w) := A(w) \left(1 - \frac{w}{w_c}\right)^{-\lambda} + W(w), \quad (14.6.6)$$

and w_c is real and the singularity of $F(w)$ nearest to the origin, the zeroth Darboux approximation gives

$$f_k = \frac{1}{(-w_c)^k} \frac{A^{(0)}(w_c) \Gamma(1-\lambda)}{\Gamma(k+1) \Gamma(1-\lambda-k)} \left\{ 1 + O\left(\frac{1}{k^{2-\lambda}}\right) \right\} \quad (14.6.7)$$

and

$$\frac{f_k}{f_{k-1}} = \frac{1}{w_c} \left\{ 1 + \frac{\lambda-1}{k} \right\} + O\left(\frac{1}{k^{2-\lambda}}\right). \quad (14.6.8)$$

Excluding the correction term, the ratio f_k/f_{k-1} is a linear function of $1/k$ with gradient $(\lambda-1)/w_c$ and intercept $1/w_c$ on the vertical axis. Thus, for sufficiently large k , graphical interpolation of the data can be used to determine approximations to w_c and λ . However, the case of confluent singularities is more problematic. In the zeroth Darboux approximation with one confluent singularity, from (14.6.5),

$$f_k = \frac{1}{(-w_c)^k} \frac{A^{(0)}(w_c) \Gamma(1-\lambda)}{\Gamma(k+1) \Gamma(1-\lambda-k)} \left\{ 1 + B(\lambda, \Delta; k) + O\left(k^{\lambda-2}\right) + O\left(k^{\lambda-2-\Delta}\right) \right\}, \quad (14.6.9)$$

where

$$B(\lambda, \Delta; k) := \frac{A_1^{(0)}(w_c) \Gamma(1+\Delta-\lambda) \Gamma(1-\lambda-k)}{A_0^{(0)}(w_c) \Gamma(1-\lambda) \Gamma(1+\Delta-\lambda-k)}. \quad (14.6.10)$$

From the asymptotic properties of the gamma function $B(\lambda, \Delta; k) \sim 1/k^\Delta$. So the effect of the confluent singularity is to introduce a term $O(1/k^\Delta)$ into (14.6.8). Since typically $0 < \Delta < 1$ the effect is to slow down the converge of the ratio to a range of values for which the linear interpolation is effective.

To obtain improved accuracy in the parameter estimations a number of methods have been employed to accelerate the convergence of the ratios f_k/f_{k-1} . The most widely used of these, when there are no confluent singularities, is the Neville-Aitkin polynomial interpolation scheme (Guttmann 1989, pp. 21–22; see also, Brezinski 2000).

The assumption in (14.6.6), which underpins the ratio method, is that the circle of convergence of $F(w)$ is given by the physical singularity, that is $r = w_c$. The method fails when, as in the case of the low-temperature series for the Ising model (Guttmann 1969; Domb and Guttmann 1970) there are complex singularities w_j with $|w_j| < w_c$, one or more of which fix the circle of convergence. A widely used technique in such cases is the application of a homographic transformation

$$w \longrightarrow \tilde{w} := \frac{bw}{1+aw}, \quad (14.6.11)$$

where a and b are real. This leaves the origin invariant and in the special case of *Euler's transformation*, where $b = a + 1$, it also leaves invariant the point $w = 1$. If there is one complex singularity, a and b in (14.6.11) can be chosen to transform it to

infinity. In the case where there is more than one such singularity it is still possible to choose the parameters to project them into the region of $\mathbb{C}_{w'}$ with $|w'| > \tilde{w}_c$.

14.6.3 Padé Approximants

Apart from the cases where homographic transformations can be used to extend the applicability of the ratio method, an approach is needed which provides an analytic continuation into the region $|w| > r$. This was provided by the introduction of the use of Padé approximants in statistical mechanics by Baker (1961) which initiated a development in series analysis which ultimately led to the invention of the differential approximant method by Guttmann and Joyce (1972) treated in Sect. 14.6.4.

A Padé approximant $[n/m]$, as given in Definition 18.5.1, is a rational function whose power series agrees with that of $F(w)$ up to and including the term in w^{n+m} . Clearly, rational functions, whose only singularities are poles, can be exactly represented by Padé approximants of sufficiently high order. They can also be expected to give good representations of meromorphic functions which differ from rational functions only at infinity. If the Padé approximant method were restricted to this class of functions it would be of only limited applicability. However, an intrinsic element of their utility is the preparation of the series by a prior transformation so that the expected form of the transformed function is meromorphic.

14.6.3.1 Transformations

The most widely used of these is the *Dlog transformation*³⁸

$$F(w) \longrightarrow \mathcal{DL}\{F\}(w) := \frac{d}{dw} \ln[F(w)] = \frac{F'(w)}{F(w)}, \quad (14.6.12)$$

which transforms branch-points into poles. Thus

$$F(w) = A \prod_{j=0}^{s-1} \left(1 - \frac{w}{w_j}\right)^{-\lambda_j}, \quad (14.6.13)$$

gives

$$\mathcal{DL}\{F\}(w) = \sum_{j=0}^{s-1} \frac{\lambda_j}{w_j - w}. \quad (14.6.14)$$

³⁸ The combination of this transformation with a Padé approximant is the *Dlog Padé approximant*.

A modification of this transformation can be used to obtain a biased estimate of λ if w_j is known for some j , since

$$F(w) \longrightarrow (w_j - w)\mathcal{DL}\{F\}(w) \simeq \lambda_j \quad w \rightarrow w_j - 0. \quad (14.6.15)$$

Repeated application of the Dlog transformation yielding the *double Dlog transformation* was used by Guttman (1978) to treat a Kosterlitz–Thouless singularity (Kosterlitz and Thouless 1973) where

$$F(w) \simeq A \exp\left(B \left[1 - \frac{w}{w_c}\right]^{-\lambda}\right) \quad w \rightarrow w_c - 0, \quad (14.6.16)$$

and

$$\mathcal{DL}\{\mathcal{DL}\{F\}\}(w) \simeq \frac{\lambda + 1}{w_c - w} \quad w \rightarrow w_c - 0. \quad (14.6.17)$$

In the two-dimensional XY model (Sect. 3.6.1), for which the high-temperature susceptibility series has been extensively studied (Betts 1974; Camp and Van Dyke 1975; Guttman 1978; Ferer and Velgakis 1983; Butera et al. 1989, 1993; Butera and Comi 1994) the data clearly favours methods designed to fit this kind of singularity. Transformations for the special case

$$F(w) \simeq \sum_{j=0}^{s-1} A_j \left(1 - \frac{w}{w_j}\right)^{-\lambda_j} \quad (14.6.18)$$

of (14.6.1) were considered by Baker and Hunter (1973). Here the analysis is rather complicated, but two cases are of interest:

- All $\lambda_j = \lambda$, which is known. For this case these authors applied a procedure which they call *exponent renormalization*.³⁹ They showed that the function

$$G(w) := \sum_{k=0}^{\infty} g_k w^k, \quad \text{where} \quad g_k := f_k \binom{-\lambda}{k}, \quad (14.6.19)$$

has leading power-law behaviour

$$G(w) \simeq \sum_{j=0}^{s-1} A_j \left(1 + \frac{w}{w_j}\right)^{-1}. \quad (14.6.20)$$

³⁹ This terminology is not related to the exponent renormalization described in Sect. 10.2.1.

- *The singularities are confluent (coalesce) at $w_j = w_c$.* This is the case for the heat capacity (Joyce 1975a) and magnetization (Joyce 1975b) of the Ising model with a pure three-spin interaction on a triangular lattice (*the Baxter–Wu Model* Baxter and Wu 1973). The analysis applied here is similar to exponent renormalization.

14.6.3.2 Choosing Approximants

For ease of discussion we shall suppose that our information consists of the coefficients $f_0, \dots, f_\ell$ of the power series of a function $F(w)$, which have either been obtained by the methods described in the first part of this chapter *or are the result of the prior application of a transformation to such a sequence of coefficients*.

The method for determining Padé approximants is described in Sect. 18.5.1. The task is most straightforward when the power series for $F(w)$ is *normal*, that is $X_0(n/m) \neq 0$ for all n and m . Then it is usually found that the most accurate representations of $F(w)$ are the diagonal approximants $[n/n]$ and near-diagonal approximants $[m - 1/m]$, $[n/n - 1]$. In general one might suppose that the best choice is to use all the available information, thus taking $n = m = \frac{1}{2}\ell$ if ℓ is even or $n = \frac{1}{2}(\ell \pm 1)$, $m = \frac{1}{2}(\ell \mp 1)$ if ℓ is odd, although there is no guarantee of monotonic improvement in the predictions of critical parameters with increasing n . There has been much discussion of the convergence properties of Padé approximants, particularly of $[n/n]$ for increasing n , in relation to the status of the Baker–Gammel–Wills conjecture for meromorphic functions,⁴⁰ (see, for example, Stahl 1997; Baker 2000, 2005, and references therein). However, since in practice it is possible to compute only a finite number of approximants, this is not a question to which we shall give further consideration.

In some cases $f_0, \dots, f_\ell$, can be obtained as rational numbers and the coefficients of the required approximant obtained without numerical approximation. When the calculations are with real numbers it is, of course, desirable to retain as many significant figures as possible and a rule of thumb proposed by Baker is that the number retained should be at least the number of known series coefficients. A numerically difficult situation is when the required approximant has a zero and a pole very close together and closer to the origin than the physical singularity.⁴¹ This leads to large numerical inaccuracies and corresponds to the situation where the linear equations (18.5.9)–(18.5.11) are ill-conditioned. Such approximants are called *defective* and should be avoided.

As we saw in Sect. 18.5.1, points in the X_0 -table for which $X_0(n/m) = 0$ occur in squares, within which the Padé approximants are given by Theorem 18.5.4. There they either do not exist or are equal to those on the boundary. Given a particular

⁴⁰ This conjectures that, for a function $F(w)$ meromorphic in the unit disc, there is a subsequence of the diagonal approximants which converges to $F(w)$ locally uniformly for all poles in the unit disc.

⁴¹ They cannot, of course, exactly coincide since $P_n^{(m)}(w)$ and $Q_m^{(n)}(w)$ are coprime.

value of ℓ it is an informative task to compute, using (18.5.13), all $X_0(n/m)$, for $n + m \leq \ell$, before completing the calculation of a chosen (non-defective) $[n/m]$ using (18.5.21) and (18.5.23).

14.6.4 Differential and Algebraic Approximants

The common features of Padé, differential and algebraic approximants are:

- (i) The known coefficients of the power series of the function of interest $F(w)$ are used to determine the coefficients in a number of polynomials.
- (ii) The recurrence relationships used to determine the polynomial coefficients are taken to apply to all the coefficients in the power series of a function $H(w)$ which is taken as an approximant to $F(w)$.

14.6.4.1 Differential Approximants

As we have indicated in Sect. 18.5.3 the procedure for determining the Dlog Padé approximant can be viewed as a bridge between Padé and differential approximants. On the one hand the polynomial quotient $[n/m]_{\text{D.L.}}$ is useful in its own right since it is an approximant to the function $G(w)$ which is designed to have a simpler structure than $F(w)$ (branch-points replaced by poles). On the other hand the function $H(w)$, which satisfies (18.5.36), is a first-order differential approximant to $F(w)$.⁴² In the case of the Dlog Padé approximants the recurrence equations are linear relationships between the known coefficients of $F(w)$.⁴³ Such relationships were found by M.F. Sykes to exist between the known coefficients of the spontaneous magnetization of the two-dimensional Ising model and used by him as a stable and efficient method of generating further coefficients.⁴⁴ Of course, caution is needed in using this method to generate series coefficients and there are well-known examples where it fails. For the F model the coefficients of the zero-field low-temperature polarization series fit the simple recurrence relationship $f_{j+2} = 3f_j + 4$ for $j = 4, 5, \dots, 12$; however $f_{16} = 3f_{14} - 4$ (Nagle 1969). The recurrence relations may also be non-linear (Joyce and Guttman 1973; Guttman and Enting 1988). In such cases the related polynomials are the coefficients of a differential equation, which is of order at least that of the order of the recurrence relations. An illustration of this relationship is given by Example 18.5.1 in Sect. 18.5.3.

⁴² With a small caveat. Comparing (18.5.36) with (18.5.40), with $k = 1$ and $B(w) = 0$, we see that $\tilde{P}_n^{(m)}(w) = -A_0(w)$ and $\tilde{Q}_n^{(m)}(w) = w A_1(w)$, implying, contrary to our assumption for Padé approximants that $\tilde{q}_0 = 0$ and $\tilde{q}_1 = 1$.

⁴³ Linear in the sense that the index variable appear linearly.

⁴⁴ This work by Dr. Sykes is reported by Guttman (1989) and elsewhere, but does not seem to have been published.

The most general form of a one-variable differential approximant is given by Definition 18.5.3, which includes the possibility of an inhomogeneous term $B(w)$. Of course, as we have already indicated, the type of approximant chosen (the order and the presence or absence of an inhomogeneous term) is influenced by the expected type of singular behaviour. If $m < n$ and the zeros $w_0, w_1, \dots, w_{m-1}$ of $\hat{Q}_m^{(n)}(w)$ are all distinct then the general solution of (18.5.36) is of the form (14.6.13) with $s := m$. If it is supposed that $F(w)$ is of this form then polynomials $\hat{P}_n^{(m)}(w)$ and $\hat{Q}_m^{(n)}(w)$ are obtained from the known coefficients of the series and the approximant $H(w)$ obtained by solving (18.5.36). This corresponds to the determination of a homogeneous first-order ($k = 1$) differential determinant except that (18.5.40) assumes that the expected form for $H(w)$ has a pole at the origin. Of course, this can be removed, by a little reformulation but the investigations of Rehr et al. (1980) indicate that this does not yield any advantages.

If the supposed form for $F(w)$ differs from (14.6.13) by the addition of an analytic term, then the approximant will not be a solution of (18.5.36) but can be shown to satisfy an inhomogeneous ($B(w) \neq 0$) case of (18.5.40). The use of a first-order inhomogeneous approximant was proposed by Gammel (1973) and has been found to be appropriate for low-temperature expansions of the Ising model susceptibility on three-dimensional lattices (Hunter and Baker 1979) for high-temperature expansions of the Ising model susceptibility and heat capacity on three-dimensional lattices (Fisher and Au-Yang 1979) and for low-temperature expansions for the energy, susceptibility and magnetization of the 3-state Potts model on two- and three-dimensional lattices (Bhanot et al. 1993).

In general it is found that an approximant of order greater than one is appropriate when there are confluent singularities (Rehr et al. 1980) as is the case for the Ising model susceptibility on three-dimensional lattices (Guttman and Enting 1993). However, apart from this consideration there will still be advantages in using differential approximants of some higher order. Exactly what that order should be for any given series is a question discussed by Boukraa et al. (2008). They conclude “*that there is no single unique ODE that describes any given series*. There is a minimum order linear ODE that is unique but this typically contains a very large number of apparent singularities and can only be determined from a corresponding larger number of series coefficients.”

14.6.4.2 Algebraic Approximants

The Padé approximant, $H(w)$ is just the solution of the first-order algebraic equation (18.5.6) and just as the generalization of the Dlog Padé approximant led from a first-order ordinary differential equations to one of higher order, it is natural to consider the generalization of the ordinary Padé approximant to algebraic approximants of higher degree. These are described in Sect. 18.5.4 and comparing (18.5.63) and (18.5.65) with the corresponding formulae (18.5.40) and (18.5.43) for the differential approximant it can be seen that they are related, in a formal way, by

$$\begin{aligned} \{H(w)\}^i &\longleftrightarrow \mathcal{D}^{i-1}\{H(w)\}, & s &\longleftrightarrow k+1, \\ n_i &\longleftrightarrow m_{i-1}, & n_0 &\longleftrightarrow n. \end{aligned} \quad (14.6.21)$$

In fact the relationship is more than formal since it can be shown (Theorem 17.6.1) that the solution $H(w)$ of (18.5.63) satisfies a linear differential equation. Given that this is the case Brak and Guttmann (1990) remark that:

At first sight ... it might appear that [the method of algebraic approximants] offers no more than the existing differential approximant method. As a mathematical statement this view is correct. As a method of series analysis, however this view is quite misleading ...

This they explain in terms of the need, given a particular number of known coefficients for $F(w)$, for computation of polynomial coefficients. In general fewer are needed for the algebraic approximant. This point can be illustrated by reference to Example 17.6.1. For both the algebraic equation with coefficients (17.6.8), the homogeneous differential equation with coefficients given by (17.6.9) and (17.6.10) and the inhomogeneous differential equation with coefficients given by (17.6.9) and (17.6.12), the total number of coefficients is (taking into account that one coefficient is set to one) the sum of the degrees plus two. For the algebraic equation that number is 11, and for the homogeneous and inhomogeneous differential equations, respectively 50 and 20.

This advantage for algebraic approximants must, however, be balanced against the fact that, for differential approximants of any order, one may, in principle, either derive successive coefficients of the approximant *or* solve the differential equation. In the case of algebraic approximants this is the case only at degrees two, three and four. Beyond this there are, of course, no analytic formulae for solutions of algebraic equations. Successive series coefficients can be obtained using the formulae given in Sect. 18.5.4 and singular points and exponents derived using the method described by Guttmann (1990) and illustrated in Example 18.5.2.

A subject of much interest both for differential and algebraic approximants is the determination of a differential or algebraic equation which gives $F(w)$ exactly. Algebraic approximants yield only rational critical exponents. However, most exactly solved models have exponents of this type⁴⁵ and some have been shown to have thermodynamic functions satisfying algebraic equations. An example, which we have used in Sect. 18.5.4, is provided by the work of Joyce (1988b). He has shown (Eq. (10.1)) that the reciprocal activity satisfies a quadratic equation with coefficients which are quartic polynomials in the inverse of the density function. We have used this case as an illustration of the use of algebraic approximants in Example 18.5.2. Beginning with the series expansion to degree 13 the algebraic equation has been obtained, together with its solution, its critical point and exponent. Of course, in this case the complete solution is known so that further series coefficients will simply satisfy the equation. The method advocated by Brak and Guttmann (1990) when

⁴⁵ See (12.3.280) for the Ising model, (12.3.322) for the 3-state Potts model and (13.2.2) for the honeycomb dimer model. The case of the eight-vertex model is exceptional since the exponents α , β , γ and ν vary through irrational values (Sect. 12.3.8). However, see the results of the application of the proposal of *weak universality* given in (12.3.326).

the series coefficients represent all the available information, is to successively use an increasing number of these while increasing the degree of the equation and the degrees of the polynomial coefficients. If further series coefficients need no change in the equation one may speculate that it is exact, although there is, of course, always the possibility that further coefficients may show this assumption to be false.

14.7 The Analysis of Series: First-Order Transitions

As described in detail in Sect. 4.1, a second-order phase transition is evidenced by continuous first-order free-energy derivatives (densities) and singular second-order derivatives (response functions), and a first-order transition corresponds to discontinuities in one or all of the densities. In the latter case the response functions may be finite or singular, as in the case of a first-order critical point (Sect. 4.5.6), at the transition point.

When the location and the order of the transition are known the interest is in whether the series for response functions can be analytically continued through the transition point into the metastable region, or whether it is a singularity.⁴⁶ In the first case it is to be anticipated that the analytic continuation terminates at a singularity which can be regarded (by analogy with mean-field theory, Sect. 5.1.1) as a point on a *pseudo-spinodal* line. The most well-known system with a known line of first-order transitions is the square-lattice spin- $\frac{1}{2}$ Ising model with the transition line on the zero-field axis terminating at the critical point and there has been considerable interest in the question of whether this transition line is one of singularities. This has been investigated by Enting and Baxter (1980) using a high-field series at fixed temperature. Analysing the series using forms of the ratio method they suggested that the transition line consisted of infinitely differentiable singularities. This conclusion is supported by Isakov (1984), who proved the absence of an analytic continuation for the free energy through the transition line for all $d \geq 2$.

For system for which the transition point is unknown a single series expansion (low or high temperature) is not sufficient for its location. The technique here to use both high- and low-temperature series. A first-order transition approached from above and below will then exhibit a change in slope of the free energy with a corresponding discontinuity in the density.

The ν -state Potts model provides the ideal test-bed for these methods. For the square lattice the transition point, order of the transition and the discontinuity (or otherwise) of the internal-energy density are known (Sects. 9.2.1 and 12.3.10.6) and for the simple cubic lattice they are still a matter of discussion.

⁴⁶ Although it is noted by Briggs et al. (1994) that even “with an essential singularity [at the transition point] it may be possible to define the thermodynamic functions in the metastable region by analytic continuation in the complex plane passing *around* the singularity”.

Briggs, Enting and Guttman (1994) used the finite-lattice method to extend the low-temperature series for the square-lattice Potts model for $\nu = 2, 3, \dots, 10$ (Sect. 14.5.3.1). They then used Dlog Padé approximants to determine analytic continuation through the transition point and to compare the temperature T_c^* , where a singularity occurred, with the transition temperature T_c . For the second-order transition for $\nu = 3$ their calculations gave $T_c^* = 0.99990 T_c$ and for the marginal case $\nu = 4$, where the transition is second-order (but with logarithmic corrections to the exponents, Sect. 12.3.10.6) $T_c^* = 0.99975 T_c$. For $\nu = 5$, $T_c^* = 1.00064 T_c$ and, for $\nu = 10$, $T_c^* = 1.0104 T_c$. These latter cases, implying analytic continuation through the critical point, are consonant with the known first-order transition. In the square-lattice Potts model high- and low-temperature series are related by duality. Using these, differential approximant equations were found for the free-energy and internal-energy densities and the order parameter and integrated numerically. The results clearly indicated the order of the phase transition with results for the location of the transition and, for $\nu > 4$, the latent heat and order-parameter discontinuities in good agreement with the known results. With respect to the latter, as might be expected, the agreement is better at $\nu = 10$ than at $\nu = 5$.

The investigations of Briggs et al. (1994) represented a preliminary study for the work of two of the authors (Guttman and Enting 1994) on the 3-state Potts model on the simple cubic lattice. They again used the finite-lattice method to extend the low- and high-temperature series for the free-energy density. Employing the differential approximant method an estimate for the transition temperature was obtained, in reasonable agreement with Monte-Carlo results, located at a discontinuity of slope, clearly indicating a first-order transition.

Chapter 15

Real-Space Renormalization Group Theory

15.1 The Basic Elements of the Renormalization Group

As in Sect. 16.2.1 we consider a d -dimensional lattice $\mathcal{N}$, with a lattice site $\mathbf{r}$ given in terms of unit lattice vectors by (16.2.1). With periodic boundary conditions applied, $n^{(\ell)}$ is an integer modulo $N^{(\ell)}$ and the total number of lattice sites $N := |\mathcal{N}|$ is given by (16.2.5). At each site $\mathbf{r}$ there is a microsystem with state specified by the variable $\sigma(\mathbf{r})$. For ease of presentation we shall suppose that the state variable at each lattice site is the same and that it can take ν discrete values, although this can be modified in particular applications. As in Sect. 4.3 it is assumed that the system has n independent (non-trivial) couplings. Here we need not distinguish between internal and external couplings and they will be denoted by $\zeta_1, \zeta_2, \dots, \zeta_n$, with the so-called ‘trivial coupling’ denoted by ζ_0 . The dimensionless Hamiltonian is given by (2.5.2) and the partition function by

$$Z(N, \zeta_i) = \sum_{\{\sigma\}} \exp[-\hat{H}(N, \zeta_i; \sigma)], \quad (15.1.1)$$

where including ζ_0 in the Hamiltonian (2.5.2) effectively sets the zero-point of the Hamiltonian and consequently of the dimensionless free energy Φ , which is defined for the Hamiltonian with ζ_0 set equal to zero; thus

$$\Phi(N, \zeta_i) := -\ln[\exp(N\zeta_0)Z(N, \zeta_i)], \quad (15.1.2)$$

or equivalently,

$$\Phi(N, \zeta_i) := -N\zeta_0 - \ln[Z(N, \zeta_i)]. \quad (15.1.3)$$

An important point about renormalization group theory is that it is, when linearized as in Sect. 15.3, an implementation of the scaling procedure described in Chap. 4. However, while that is able to predict only relations between critical exponents,

renormalization group methods are able to give values for these exponents. This procedure can be carried out in a variety of ways. Even within the context of real-space renormalization group (RSRG) methods it would be very difficult to present a formulation which applies to all possible schemes for all possible systems. We have not attempted this. Instead a block-spin procedure is described, which includes most simple applications of the theory. Although it will need modification in some cases, these changes will not radically affect the development.

The following steps in the development can, in part, be viewed as a formalization and generalization of the ideas introduced in Sect. 4.2.2:

- (i) The lattice $\mathcal{N}$ is divided into blocks of the form of d -dimensional parallelepipeds whose edges are given by the vectors $\lambda a^{(\ell)} \hat{\mathbf{r}}^{(\ell)}$, for some integer $\lambda > 1$, and such that in each block there are λ^d sites. A lattice $\tilde{\mathcal{N}}$ is formed of the same structure as $\mathcal{N}$ by associating a lattice site $\tilde{\mathbf{r}} \in \tilde{\mathcal{N}}$ with every block of $\mathcal{N}$. The block of sites in $\mathcal{N}$ associated with $\tilde{\mathbf{r}}$ is denoted by $\mathcal{B}(\tilde{\mathbf{r}})$ and the number $\tilde{N}$ of sites in $\tilde{\mathcal{N}}$ is

$$\tilde{N} := \lambda^{-d} N, \quad (15.1.4)$$

with lattice spacings

$$\tilde{a}^{(\ell)} := \lambda a^{(\ell)}, \quad \ell = 1, 2, \dots, d. \quad (15.1.5)$$

- (ii) The size of the lattice $\tilde{\mathcal{N}}$ is reduced by a length scaling

$$|\tilde{\mathbf{r}}| = \lambda^{-1} |\mathbf{r}|. \quad (15.1.6)$$

This means that $\tilde{\mathcal{N}}$ now differs from $\mathcal{N}$ only in the number of lattice sites.

- (iii) The conditional probability

$$\rho(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) \geq 0, \quad (15.1.7)$$

that the microstate at $\tilde{\mathbf{r}}$ is $\sigma(\tilde{\mathbf{r}})$, given a particular configuration of microstates in the block $\mathcal{B}(\tilde{\mathbf{r}})$, is called the *weight function* and it satisfies the usual formula

$$\sum_{\{\sigma(\tilde{\mathbf{r}})\}} \rho(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) = 1, \quad (15.1.8)$$

where the summation is over all the microstates of $\tilde{\mathbf{r}}$. The probability $\rho(\zeta_i; \sigma(\mathbf{r}))$ that there is a particular configuration of microstates on $\mathcal{N}$ is, from (2.1.2),

$$\rho(\zeta_i; \sigma(\mathbf{r})) = \frac{\exp[-\hat{H}(N, \zeta_i; \sigma(\mathbf{r}))]}{Z(N, \zeta_i)}. \quad (15.1.9)$$

Since the mapping is between two lattices which differ only in the number of lattice sites, the Hamiltonian function $\hat{H}$ and partition function Z will be the same in each case and the probability $p(\tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}}))$ that there is a particular configuration of microstates on $\tilde{N}$, with coupling constants $\tilde{\zeta}_i$, is

$$p(\tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}})) = \frac{\exp[-\hat{H}(\tilde{N}, \tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}}))]}{Z(\tilde{N}, \tilde{\zeta}_i)}. \quad (15.1.10)$$

- (iv) The probabilities given by (15.1.9) and (15.1.10) can be related by a conditional probability of the form (15.1.7) using the standard formula of probability theory. In general this means that the conditional probability will be a function of both sets of couplings ζ_i and $\tilde{\zeta}_i$ with no specified relationship between the two. In renormalization group theory, however, the probability formula is used to *impose* the relationship

$$p(\tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}})) := \sum_{\{\sigma(\mathbf{r})\}} \left[\prod_{\{\tilde{\mathbf{r}}\}} p(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) \right] p(\zeta_i; \sigma(\mathbf{r})) \quad (15.1.11)$$

between the two sets of couplings, where the conditional probability is independent of the values of the members of both sets.

- (v) Although the probabilities $p(\zeta_i; \sigma(\mathbf{r}))$ and $p(\tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}}))$, as they are defined in (15.1.9) and (15.1.10), involve the trivial couplings ζ_0 and $\tilde{\zeta}_0$ respectively, this dependence is purely formal. The factor $\exp(-N\zeta_0)$ can be cancelled from the numerator and denominator of (15.1.9) and similarly for $\exp(-\tilde{N}\tilde{\zeta}_0)$ in (15.1.10). We, therefore, have two arbitrary parameters in (15.1.11) and imposing the condition

$$Z(\tilde{N}, \tilde{\zeta}_i) = Z(N, \zeta_i) \quad (15.1.12)$$

fixes the difference $\tilde{N}\tilde{\zeta}_0 - N\zeta_0$.

From (15.1.9)–(15.1.12),

$$\exp[-\hat{H}(\tilde{N}, \tilde{\zeta}_i; \sigma(\tilde{\mathbf{r}}))] = Z(N, \zeta_i; \sigma(\tilde{\mathbf{r}})), \quad (15.1.13)$$

where

$$Z(N, \zeta_i; \sigma(\tilde{\mathbf{r}})) := \sum_{\{\sigma(\mathbf{r})\}} \left[\prod_{\{\tilde{\mathbf{r}}\}} p(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) \right] \exp[-\hat{H}(N, \zeta_i; \sigma(\mathbf{r}))]. \quad (15.1.14)$$

15.2 RG Transformations and Weight Functions

Equation (15.1.13) is the starting point for the development of a renormalization transformation. Ideally, by allowing the microstates of $\tilde{\mathcal{N}}$ to range over all their possible values and performing the summation on the right-hand side, we hope to obtain *recurrence relationships* of the form

$$\tilde{\zeta}_0 = \lambda^d [\zeta_0 + \mathcal{K}_0(\zeta_j)], \quad (15.2.1)$$

$$\tilde{\zeta}_i = \mathcal{K}_i(\zeta_j), \quad i = 1, 2, \dots, n, \quad (15.2.2)$$

where (15.2.1) can be neglected from present consideration as it serves only to obtain thermodynamic properties (Sect. 15.4). If (15.2.2) can be obtained they are used to perform a sequence of iterations indexed by $s = 0, 1, 2, \dots$ in the space $\mathcal{S}$ of non-trivial couplings. This produces a trajectory $\mathbf{P}_0 \rightarrow \mathbf{P}_1 \rightarrow \mathbf{P}_2 \rightarrow \dots$ in $\mathcal{S}$. It satisfies the property

$$\zeta_i^{(s+s')} = \mathcal{K}_i^{s'}(\zeta_j^{(s)}) \quad (15.2.3)$$

and is therefore a *semi-group*.¹ As in the case of general dynamic systems (see, for example, Devaney 1989) the behaviour of renormalization group trajectories in $\mathcal{S}$ could be very complicated, exhibiting fixed points, periodic points and even chaos. However, in most simple examples, it might be anticipated that the behaviour is governed by a relatively small set of fixed points. A point $\mathbf{P}^* \in \mathcal{S}$ with coupling values $\zeta_1^*, \zeta_2^*, \dots$, is a *fixed point* if

$$\zeta_i^* = \mathcal{K}_i(\zeta_j^*), \quad i = 1, 2, \dots, n. \quad (15.2.4)$$

If it exists, the set $\mathfrak{B}(\mathbf{P}^*)$ of all points $\mathbf{P}_0$ which on iteration converge to $\mathbf{P}^*$ is called the *basin of attraction* of $\mathbf{P}^*$. If an n -dimensional neighbourhood of $\mathbf{P}^*$ contains only points of $\mathfrak{B}(\mathbf{P}^*)$ then $\mathbf{P}^*$ is called a *sink*. In general fixed points can have neighbourhoods containing points for which they are *attractive*, points for which they are *repulsive* and points for which trajectories iterated neither towards nor away from the fixed point. This last type of fixed point is called *marginal* and this characteristic is often associated with the existence of a region (line or surface) of fixed points.

In Sect. 15.1 we referred to the lattice $\mathcal{N}$ having N sites with the number of sites of $\tilde{\mathcal{N}}$ being $\tilde{N} = \lambda^{-d} N < N$. However, it is clear that the development of trajectories in $\mathcal{S}$, leading to fixed points, implicitly assumes that a large (potentially infinite) number of iterations can be applied to the system. This could be the case only if the thermodynamic limit $N \rightarrow \infty$ is assumed. The method described in Sect. 15.8, whereby the form of the transformation derived for finite N is applied an indefinite number of times, is in these terms by its very structure an approximation method, even though the calculation of (15.1.13) for finite N may be exact. It is clear that

¹ Not a full group since no inverse transformation is defined.

the form of an transformation will be in part determined by the choice of weight function. Three commonly used forms are:

(a) **The decimation weight function²**

Choose a particular site $\tilde{r} \in \mathcal{B}(\tilde{r})$ so that the set $\mathcal{N}^\circ$ of all sites $\tilde{r}$ for all the blocks of $\mathcal{N}$ forms a lattice of the same type as $\mathcal{N}$. Then set

$$\rho(\sigma(\tilde{r})|\sigma(r), r \in \mathcal{B}(\tilde{r})) := \delta^{\text{Kr}}(\sigma(\tilde{r}) - \sigma(\tilde{r})). \quad (15.2.5)$$

It is clear that the effect of choosing this weight function is that the summation on the right-hand side of (15.1.13) is a partial sum over all the sites of the lattice $\mathcal{N}$ except those of $\mathcal{N}^\circ$. The effect is to ‘thin out’ the sites of $\mathcal{N}$ leaving $\mathcal{N}^\circ$ which is now identified as $\tilde{\mathcal{N}}$. For the spin- $\frac{1}{2}$ model ($\nu = 2$, $\sigma(r) = \pm 1$) equation (15.2.5) takes the simple form

$$\rho(\sigma(\tilde{r})|\sigma(r), r \in \mathcal{B}(\tilde{r})) = \frac{1}{2}[1 + \sigma(\tilde{r})\sigma(\tilde{r})]. \quad (15.2.6)$$

(b) **The majority-rule weight function**

This weight function was introduced by Niemeijer and van Leeuwen (1973, 1974). The first step in assigning $\sigma(\tilde{r})$ for the block $\mathcal{B}(\tilde{r})$ can be described in terms of the ‘winner takes all’ voting procedure used in some democracies. Given that, among the sites of $\mathcal{B}(\tilde{r})$ one of the ν microstates occurs more than any other, $\sigma(\tilde{r})$ is assigned to this value. Unless $\nu = 2$ and the number of sites λ^d in a block is odd it is clear that this is not sufficient to determine $\sigma(\tilde{r})$ for every configuration of the block. A ‘tie’ can occur in the voting procedure and a strategy must be adopted to deal with such cases. One possibility is to assign to $\sigma(\tilde{r})$ one of these predominating values on the basis of equal probabilities. In the special case of a spin- $\frac{1}{2}$ model this weight function would take the form

$$\rho(\sigma(\tilde{r})|\sigma(r), r \in \mathcal{B}(\tilde{r})) \simeq \frac{1}{2} \left[1 + \sigma(\tilde{r}) S \left(\sum_{\{r \in \mathcal{B}(\tilde{r})\}} \sigma(r) \right) \right], \quad (15.2.7)$$

where

$$S(m) = \begin{cases} -1, & \text{if } m < 0, \\ 0, & \text{if } m = 0, \\ 1, & \text{if } m > 0. \end{cases} \quad (15.2.8)$$

In some cases this may not, however, be the most appropriate choice. In their work on the Ising model using square first-neighbour blocks Nauenberg and Nienhuis (1974a, b) divided the configurations with equal numbers of up and

² The term ‘decimation’ (originating in a form of discipline used in the Roman army) is more properly used when one tenth of the sites is removed; but here it is employed more loosely to denote any proportion.

down spins between block spins up and down with probability one. The rule (one of four) which they chose ensured that the reversal of all the spins in the block reversed the block spin.

(c) **The Kadanoff–Houghton weight function**

For the spin- $\frac{1}{2}$ model the weight function

$$p(\sigma(\tilde{\mathbf{r}})|\sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) := \frac{\exp\left[\tau \sigma(\tilde{\mathbf{r}}) \sum_{\mathbf{r} \in \mathcal{X}(\tilde{\mathbf{r}})} \sigma(\mathbf{r})\right]}{2 \cosh\left[\tau \sum_{\mathbf{r} \in \mathcal{X}(\tilde{\mathbf{r}})} \sigma(\mathbf{r})\right]} \quad (15.2.9)$$

was introduced by Kadanoff and Houghton (1975). In this formula τ is a real number and $\mathcal{X}(\tilde{\mathbf{r}})$ is some subset of $\mathcal{B}(\tilde{\mathbf{r}})$. It can be shown that, in the limit $\tau \rightarrow \infty$, we recover the majority-rule weight function (15.2.7), if $\mathcal{X}(\tilde{\mathbf{r}}) = \mathcal{B}(\tilde{\mathbf{r}})$, and the decimation weight function (15.2.6), if $\mathcal{X}(\tilde{\mathbf{r}})$ is the single site $\hat{\mathbf{r}} \in \mathcal{B}(\tilde{\mathbf{r}})$.

In itself the weight function chosen does not lead to any approximation as long as the summation on the left-hand side of (15.1.13) can be evaluated exactly. This means that the calculation of thermodynamic functions (Sect. 15.4) would be exact. What can be lost are fixed points associated with particular parts of phase space. This leads to the inability to map the part of the phase space associated with these fixed points. The usual rule employed is to choose a weight function, which, when applied to the ground state in $\mathcal{N}$ associated with a phase of interest, leads to the same ground state in the lattice $\tilde{\mathcal{N}}$. This idea will be illustrated by reference to the one-dimensional Ising model in Sect. 15.5.1.

Beginning in the early 1970s the general pattern of development described here was used to construct approximate renormalization group transformations (Niemeijer and van Leeuwen 1976) and to gain a new perspective on the known solutions of some exactly-solvable models (Nelson and Fisher 1975). One problem, recognized at an early stage, was the difficulty of obtaining a consistent set of equations, of the form of (15.2.1) and (15.2.2), from (15.1.13). The number of equations generated by choosing all possible states of the microsystems on $\tilde{\mathcal{N}}$ is $v^{\tilde{N}}$. Because of the symmetries of the system many of these equations will be identical. However, within this set, there must be exactly $n + 1$ independent equations which express the new set of couplings as functions of old set. It is not guaranteed that this can be achieved. In fact, as we shall see in Sect. 15.5.2, for the two-dimensional Ising model on a square lattice with first-neighbour couplings alone, using a decimation weight function, this is impossible. Consistent equations can be obtained by adding second-neighbour and four-site interactions in $\tilde{\mathcal{N}}$. These, however, would in their turn demand the inclusion of further couplings at the next iteration. This *proliferation* of couplings led to the development of a number of approximation methods designed to ‘cut-off’ the proliferation in a more or less ad hoc manner.

Equations (15.2.1) and (15.2.2) must not only be consistent, but $\mathcal{K}_i$, $i = 0, 1, \dots, n$, must be smooth functions of the couplings ζ_j , $j = 1, 2, \dots, n$. For explicit approximation methods for finite systems this is not normally a problem. However, as was first recognized by Griffiths and Pearce (1978, 1979) (see also Griffiths 1981), this condition is not obviously true in the thermodynamic limit. The fundamental point is that, whereas we need, for the existence of phase transitions, non-smooth behaviour in the partition function (15.1.1), we require smooth behaviour from the summation on the left-hand side of (15.1.13), which differs from $Z(N, \zeta_i)$ only by the presence of the weight function. It is by no means assured that the weight function is able to suppress all non-smooth behaviour. As an example Griffiths and Pearce (1979) considered the case, for a spin- $\frac{1}{2}$ model, of the Kadanoff–Houghton weight function, where $\mathcal{X}(\tilde{\mathbf{r}}) = \hat{\mathbf{r}} \in \mathcal{B}(\tilde{\mathbf{r}})$, yielding

$$p(\sigma(\tilde{\mathbf{r}})|\sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) = \frac{\exp[\tau \sigma(\tilde{\mathbf{r}})\sigma(\hat{\mathbf{r}})]}{2 \cosh(\tau)}. \quad (15.2.10)$$

This gives, apart from a multiplicative constant, for the right-hand side of (15.1.13) the partition function of the model with the addition of an external field-type coupling $\tau \sigma(\tilde{\mathbf{r}})$ on the subset of sites $\hat{\mathcal{N}}$. Given that the model exhibits phase transitions it is not necessarily the case that the extra external field is sufficient to lead to their suppression. In particular, if the decimation limit $\tau \rightarrow \infty$ is chosen, it is possible by a judicious choice of the signs of the states on $\hat{\mathcal{N}}$ (Griffiths 1981) to cancel out the additional field, leading to zero-field models known to have phase transitions. Transformations which give unwelcome phase transitions in the modified partition function on the right-hand side of (15.1.13) are said by Griffiths and Pearce (1979) to exhibit *peculiarities*. Examples are given by these authors of schemes with and without such behaviour. One simple observation with respect to the decimation weight function can be made. This is that if, as in Sects. 15.5.1 and 15.5.2, the sites of $\mathcal{N} - \hat{\mathcal{N}}$ form disconnected clusters then no peculiarities will occur. This is clear from the fact that each cluster constitutes an effective finite-lattice system, which cannot have a phase transition.

A number of authors (Wilson 1975; Kadanoff and Houghton 1975; van Leeuwen 1975) have drawn attention to another problem with the decimation weight function. Consider any subset $\{\mathbf{r}_1, \mathbf{r}_2, \dots\}$ of $\hat{\mathcal{N}} \subset \mathcal{N}$. These sites will remain as sites of $\tilde{\mathcal{N}}$. It is clear that the expectation value $\langle \sigma(\mathbf{r}_1)\sigma(\mathbf{r}_2)\dots \rangle$ will have the same value in both $\mathcal{N}$ and $\tilde{\mathcal{N}}$. In particular the pair correlation function, defined as in (4.2.3), for the states at $\mathbf{r}_1$ and $\mathbf{r}_2$ satisfies the scaling form

$$\Gamma_2(\theta_1, \dots, \theta_n; \sigma(\mathbf{r}_2 - \mathbf{r}_1)) = \Gamma_2(\lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n; \sigma(\lambda^{-1}[\mathbf{r}_2 - \mathbf{r}_1])). \quad (15.2.11)$$

It follows from (4.5.48) that the scaling exponent associated with the coupling conjugate to the spin variable is equal to d , the physical dimension of the system. This means that everywhere on a critical surface the pair correlation function is equal to its value at the fixed point controlling the surface and, from (4.5.40) and (4.5.56), $\beta = 0$, $\delta = \infty$ and $\eta = 2 - d$. While these results are true for the one-dimensional

Ising model, where the phase transition occurs at zero temperature with a discontinuity in field (Sect. 15.5.1), they are certainly untrue for the two-dimensional Ising model (Eq. (12.3.280)) and conflict with series estimates for the three-dimensional Ising model according to which $\beta \simeq 0.33$, $\delta \simeq 4.8$ and $\eta \simeq 0.04$. Although it may be argued that this failure of an *exact* decimation transformation is hardly relevant to real situations, where in most cases decimation is used in conjunction with further approximation, it is of some interest to understand why it occurs. The basis of scaling and of the use of the renormalization group is, as has already been remarked, the perception that critical behaviour arises from long-range correlations. The application of the renormalization group in most cases averages over the short-range structure of the system leaving the long-range behaviour and thus the critical properties unaffected. As has been observed by Sneddon and Barber (1977), this is not the case with the decimation weight function which sums over some short and some long-range effects. This may be understood as a loss of information about critical behaviour leading to the absence of asymptotic decay of correlations.

15.3 Fixed Points and the Linear Renormalization Group

To analyze the nature of the fixed point given by (15.2.4) we linearize the recurrence relationships (15.2.2) about $\mathbf{P}^*$. With

$$[\Delta\zeta^{(s)}]^T := (\zeta_1^{(s)} - \zeta_1^*, \zeta_2^{(s)} - \zeta_2^*, \dots, \zeta_n^{(s)} - \zeta_n^*), \quad (15.3.1)$$

we have

$$\Delta\zeta^{(s+1)} \simeq \mathbf{L}^* \Delta\zeta^{(s)}, \quad (15.3.2)$$

where $\mathbf{L}^*$ is the fixed-point value of

$$\mathbf{L} := \begin{pmatrix} \frac{\partial \mathcal{K}_1}{\partial \zeta_1} & \frac{\partial \mathcal{K}_1}{\partial \zeta_2} & \cdots & \frac{\partial \mathcal{K}_1}{\partial \zeta_n} \\ \frac{\partial \mathcal{K}_2}{\partial \zeta_1} & \frac{\partial \mathcal{K}_2}{\partial \zeta_2} & \cdots & \frac{\partial \mathcal{K}_2}{\partial \zeta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{K}_n}{\partial \zeta_1} & \frac{\partial \mathcal{K}_n}{\partial \zeta_2} & \cdots & \frac{\partial \mathcal{K}_n}{\partial \zeta_n} \end{pmatrix}. \quad (15.3.3)$$

In general $\mathbf{L}^*$ is not symmetric and thus we have the left and right eigenvalue equations,

$$\mathbf{w}_j^T \mathbf{L}^\star = \Lambda_j \mathbf{w}_j^T, \quad (15.3.4)$$

$$\mathbf{L}^\star \mathbf{x}_j = \Lambda_j \mathbf{x}_j, \quad (15.3.5)$$

respectively, where $\mathbf{w}_j$ and $\mathbf{x}_j$ denote the left and right eigenvectors, for the eigenvalue Λ_j . Assuming that $\mathbf{L}^\star$ is a simple matrix (Sect. 17.2)

$$\mathbf{w}_j \cdot \mathbf{x}_{j'} = \delta^{\text{Kr}}(j - j'). \quad (15.3.6)$$

From (15.3.2) and (15.3.4),

$$\mathbf{w}_j \cdot \Delta \boldsymbol{\zeta}^{(s+1)} \simeq \Lambda_j \mathbf{w}_j \cdot \Delta \boldsymbol{\zeta}^{(s)}. \quad (15.3.7)$$

Suppose, as in Sect. 4.3, that there exist scaling fields

$$\theta_j = \theta_j(\Delta \zeta_i), \quad (15.3.8)$$

which are smooth functions of the couplings (and thus of $\Delta \zeta_j, j = 1, 2, \dots, n$) and such that

$$\theta_j(0) = 0, \quad (15.3.9)$$

$$\theta_j^{(s+1)} = \Lambda_j \theta_j^{(s)}. \quad (15.3.10)$$

From (15.3.6), (15.3.7) and (15.3.10) it follows that

$$\theta_j \simeq \mathbf{w}_j \cdot \Delta \boldsymbol{\zeta}, \quad (15.3.11)$$

$$\Delta \boldsymbol{\zeta} \simeq \sum_{j=1}^n \mathbf{x}_j \theta_j. \quad (15.3.12)$$

Iterating (15.3.10)

$$\theta_j^{(s)} = \Lambda_j^s \theta_j^{(0)} \quad (15.3.13)$$

and thus, from (15.3.12),

$$\Delta \boldsymbol{\zeta}^{(s)} \simeq \sum_{j=1}^n \Lambda_j^s \mathbf{x}_j \theta_j^{(0)}. \quad (15.3.14)$$

Given that the scaling fields exist, they are a set of curvilinear coordinates in a region surrounding $\mathbf{P}^\star$ with the following properties:

- (i) $\theta_j = 0$ is an $(n - 1)$ -dimensional subspace in $\mathcal{S}$, which, from (15.3.10), is invariant under the transformation. From (15.3.11) the left eigenvector $\mathbf{w}_j$ is normal to this subspace at $\mathbf{P}^*$.
- (ii) $\{\theta_i = 0, \forall i \neq j\}$ is an invariant line in $\mathcal{S}$, which, from (15.3.12), is tangential to $\mathbf{x}_j$ at $\mathbf{P}^*$.
- (iii) Supposing that all the eigenvalues are real:
 - (a) If $\Lambda_j > 1$ and $\theta_j^{(0)} \neq 0$ then $\theta_j^{(s)}$ increases as s increases. The trajectory iterates away from $\mathbf{P}^*$ with a component in the direction of $\mathbf{x}_j$ and Λ_j is *relevant*.
 - (b) If $\Lambda_j < 1$ and $\theta_j^{(0)} \neq 0$ then $\theta_j^{(s)}$ decreases as s increases. The trajectory approaches the subspace $\theta_j = 0$ and Λ_j is *irrelevant*. It will, however, approach $\mathbf{P}^*$ only if there is no relevant scaling field θ_i with $\theta_i^{(0)} \neq 0$.
 - (c) If $\Lambda_j = 1$ it is *marginal* and θ_j remains unchanged under iteration.

The exponents y_j for $\mathbf{P}^*$ are defined by the formulae

$$\Lambda_j := \lambda^{y_j}, \quad j = 1, 2, \dots, n \quad (15.3.15)$$

and we recover from (15.3.10) the scaling transformation (4.3.3). Exponents are named in the same way as their respective eigenvalues. Since $\lambda > 1$, an exponent is relevant if it is positive, marginal if it is zero and irrelevant if it is negative (Sect. 4.3). The region around $\mathbf{P}^*$ with tangent space spanned by the set of right eigenvectors corresponding to irrelevant eigenvalues is the *basin of attraction* $\mathfrak{B}(\mathbf{P}^*)$. A *sink* has only irrelevant eigenvalues. A fixed point which has a neighbourhood containing no other fixed point is called *isolated*. As indicated above this usually means that it has no marginal eigenvalues.

This analysis is now illustrated by two examples for the case $n = 3$.

- (1) $\Lambda_1 > 1, \Lambda_2 < 1, \Lambda_3 < 1$. The region around $\mathbf{P}^*$ for this case is shown in Fig. 15.1. The local part of $\mathfrak{B}(\mathbf{P}^*)$ is the surface $\theta_1 = 0$. Trajectories on this surface will converge on $\mathbf{P}^*$. Any trajectory near to the fixed point, but not on $\theta_1 = 0$ will iterate away from the fixed point, tending towards the curve $\theta_2 = \theta_3 = 0$.
- (2) $\Lambda_2 > \Lambda_1 > 1, \Lambda_3 < 1$. Λ_2 and Λ_1 are called respectively the *strong* and *weak* relevant eigenvalues. This is the case shown in Fig. 15.2. The local part of $\mathfrak{B}(\mathbf{P}^*)$ is the curve $\theta_1 = \theta_2 = 0$. Any trajectory beginning on this curve will converge on $\mathbf{P}^*$. Any other trajectory will be repelled by $\mathbf{P}^*$. If it begins at a point where $\theta_2 \neq 0$ then it will tend towards the strong direction $\theta_1 = \theta_3 = 0$. Otherwise it will tend towards the weak direction $\theta_2 = \theta_3 = 0$.

Fig. 15.1 The region around a fixed point P^* for $n = 3$, with one relevant and two irrelevant eigenvalues

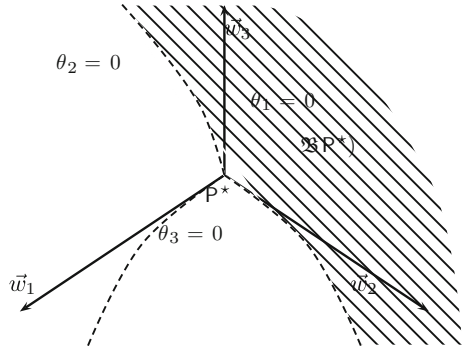
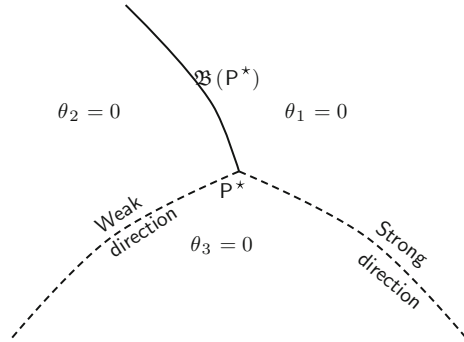


Fig. 15.2 The region around a fixed point P^* for $n = 3$, with one irrelevant and two relevant eigenvalues



15.4 Free Energy and Densities

From (15.1.3) and (15.1.12)

$$\Phi(\tilde{N}, \tilde{\zeta}_i) + \tilde{N}\tilde{\zeta}_0 = \Phi(N, \zeta_i) + N\zeta_0. \quad (15.4.1)$$

Suppose that free energy per lattice site

$$\phi(\zeta_i) := \Phi(N, \zeta_i)/N, \quad (15.4.2)$$

is, at least for large N , independent of N (Sect. 1.3), then

$$\phi(\tilde{\zeta}_i) + \tilde{\zeta}_0 = \lambda^d \phi(\zeta_i) + \lambda^d \zeta_0, \quad (15.4.3)$$

which, from (15.2.1), becomes

$$\phi(\tilde{\zeta}_i) = \lambda^d [\phi(\zeta_i) - \mathcal{K}_0(\zeta_i)]. \quad (15.4.4)$$

Linearizing the recurrence relations in a neighbourhood of a fixed point in terms of the scaling fields gives

$$\lambda^{-1} \phi(\lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) = \phi(\theta_1, \dots, \theta_n) - \mathcal{K}_0(\theta_1, \dots, \theta_n). \quad (15.4.5)$$

This is of the form of (4.3.79) derived from the Nightingale–’t Hooft scaling hypothesis (4.3.47). A difference is that here $\lambda > 1$ is an integer, whereas in scaling theory it is assumed to have any real value in this range. This can be resolved if it is assumed that (15.4.5) can be analytically continued into non-integer values of λ .

Iterating equation (15.4.4) along a trajectory from a point with couplings $\zeta_i^{(0)}$, $i = 1, 2, \dots, n$ gives

$$\phi(\zeta_i^{(0)}) = \frac{1}{\lambda^{kd}} \phi(\zeta_i^{(k)}) + \sum_{s=0}^{k-1} \frac{1}{\lambda^{sd}} \mathcal{K}_0(\zeta_i^{(s)}). \quad (15.4.6)$$

If we assume that

$$\lim_{k \rightarrow \infty} \frac{1}{\lambda^{kd}} \phi(\zeta_i^{(k)}) = 0, \quad (15.4.7)$$

it follows from (15.4.6) that the free energy per lattice site, at the initial point $\zeta_i^{(0)}$, is given by

$$\phi(\zeta_i^{(0)}) = \sum_{s=0}^{\infty} \frac{1}{\lambda^{sd}} \mathcal{K}_0(\zeta_i^{(s)}). \quad (15.4.8)$$

In practice, it is usually found that this series converges after a very few iterations (see, for example, Southern and Lavis 1980). From (4.3.117) the densities at the initial point on the trajectory are given by

$$\rho_i^{(0)} = - \left(\frac{\partial \phi}{\partial \zeta_i} \right)^{(0)}. \quad (15.4.9)$$

To obtain these functions a chain differentiation must be performed along the trajectory. Let

$$\kappa_{\ell m} := \frac{\partial \mathcal{K}_m}{\partial \zeta_\ell}, \quad (15.4.10)$$

$$\tau_{im}^{(0)} := \kappa_{im}(\zeta_j^{(0)}), \quad (15.4.11)$$

$$\tau_{im}^{(k)} := \sum_{\ell=1}^n \tau_{i\ell}^{(k-1)} \kappa_{\ell m}(\zeta_j^{(k)}), \quad k = 1, 2, \dots, s. \quad (15.4.12)$$

Then

$$\rho_i^{(0)} = \sum_{s=0}^{\infty} \frac{1}{\lambda^{sd}} \tau_{i0}^{(s)}. \quad (15.4.13)$$

The response functions, which correspond to the second-order partial derivatives of the free-energy density with respect to the couplings, can be obtained in a similar way.

15.5 Decimation for the Ising Model

15.5.1 In One Dimension

Although the one-dimensional Ising model can be solved quite simply using the transfer matrix methods of Chap. 11 (see, for example, Domb 1960, Sect. 3.2.1), it is illuminating to consider an RSRG treatment of the same problem using a decimation weight function. The only change made to the one-dimensional version of the Hamiltonian (3.5.1) is to add a constant term $-C$ at each lattice site so that the transfer matrix is

$$V := \mathfrak{z}_0 \begin{pmatrix} \mathfrak{z}_1 \mathfrak{z}_2 & \mathfrak{z}_2^{-1} \\ \mathfrak{z}_2^{-1} & \mathfrak{z}_1^{-1} \mathfrak{z}_2 \end{pmatrix}, \quad (15.5.1)$$

where $\mathfrak{z}_0 := \exp(C)$ and, for convenience, $\mathfrak{z}_1 := \mathfrak{X}$ and $\mathfrak{z}_2 := \mathfrak{Z}$. Then, with periodic boundary conditions applied so $\mathcal{N}$ forms a ring of sites, from (11.1.6),

$$Z(N, C, K, L) = \text{Trace}\{V^N\}. \quad (15.5.2)$$

$\mathcal{N}$ is divided into blocks of λ neighbouring sites and the application of the decimation weight function corresponds to fixing the spin state on one site of each block (the first, say) and performing a partial sum over the remaining spin states. The cases $\lambda = 2$ and $\lambda = 3$ have been considered in detail by Nelson and Fisher (1975). The general case can be most conveniently formulated in terms of the transfer matrix when the partition function for $\tilde{\mathcal{N}}$ is

$$Z(\tilde{N}, \tilde{C}, \tilde{K}, \tilde{L}) = \text{Trace}\{\tilde{V}^{\tilde{N}}\}, \quad (15.5.3)$$

where

$$\tilde{V} := V^\lambda. \quad (15.5.4)$$

Computing the elements of this formula for any λ yields the recurrence relationships.

We first investigate in detail the case $\lambda = 2$. It is not difficult to see that (15.5.4) yields the three relationships

$$\tilde{z}_0 \tilde{z}_1 \tilde{z}_2 = z_0^2 (z_1^2 z_2^2 + z_2^{-2}), \quad (15.5.5)$$

$$\tilde{z}_0 \tilde{z}_1^{-1} \tilde{z}_2 = z_0^2 (z_1^{-2} z_2^2 + z_2^{-2}), \quad (15.5.6)$$

$$\tilde{z}_0 \tilde{z}_2^{-1} = z_0^2 (z_1 + z_1^{-1}), \quad (15.5.7)$$

which give

$$\tilde{z}_0^4 = z_0^8 (z_1^2 z_2^2 + z_2^{-2}) (z_1^{-2} z_2^2 + z_2^{-2}) (z_1 + z_1^{-1})^2, \quad (15.5.8)$$

$$\tilde{z}_1^4 = \frac{z_1^2 z_2^2 + z_2^{-2}}{z_1^{-2} z_2^2 + z_2^{-2}}, \quad (15.5.9)$$

$$\tilde{z}_2^4 = \frac{(z_1^2 z_2^2 + z_2^{-2}) (z_1^{-2} z_2^2 + z_2^{-2})}{(z_1 + z_1^{-1})^2}. \quad (15.5.10)$$

The formulae (15.5.8)–(15.5.10) are invariant under the transformation $\tilde{z}_1 \rightarrow 1/\tilde{z}_1$, which corresponds to reversing the direction of the magnetic field. In order to obtain recurrence relationships in a form appropriate to the one-dimensional Ising model we must bear in mind that, in this case, the Curie point lies at zero temperature, implying a fixed point on the zero-field axis at $T = 0$. The appropriate variables to choose are ones which are finite at zero temperature. If we concentrate on the region $\mathcal{H} \geq 0$ and $J \geq 0$ ($L \geq 0$, $K \geq 0$) then the variables

$$u := \tilde{z}_1^{-2} = \exp(-2L), \quad (15.5.11)$$

$$t := \frac{\tilde{z}_2^2 \tilde{z}_1 - 1}{\tilde{z}_2^2 \tilde{z}_1 + 1} = \tanh\left(\frac{1}{2}L + K\right) \quad (15.5.12)$$

lie in the interval $[0, 1]$ and, from (15.5.9) and (15.5.10),

$$\tilde{u} = \frac{u^2(1+t)^2 + (1-t)^2}{2(1+t^2)}, \quad (15.5.13)$$

$$\tilde{t} = \frac{4t^2 - (1-u)(t^2 - 1)}{4 + (1-u)(t^2 - 1)}. \quad (15.5.14)$$

The first step is to determine the invariant lines. There are three of these:

- (a) $u = 1$, corresponding to $\mathcal{H} = 0$. On this line there is the one recurrence relationship

$$\tilde{t} = t^2 \quad (15.5.15)$$

with *fixed points* $t = 1$, $u = 1$ and $t = 0$, $u = 1$.

- (b) $t = 1$, corresponding to $T = 0$. On this line there is the one recurrence relationship

$$\tilde{u} = u^2 \quad (15.5.16)$$

with the *fixed points* $t = 1, u = 1$ and $t = 1, u = 0$.

- (c) Every point on the line

$$u = \left(\frac{1-t}{1+t} \right)^2 \quad (15.5.17)$$

satisfies (15.5.13) and (15.5.14). This is a *line of fixed points* corresponding to $J = 0$.

The invariant lines, fixed points and the fixed line are shown in Fig. 15.3, together with examples of trajectory flows. We now consider the linearized recurrence relationships in the neighbourhood of each fixed point.

- $t = 1, u = 1$ is the *ferromagnetic fixed point* at $T = 0, \mathcal{H} = 0$. For this

$$\begin{pmatrix} \Delta \tilde{u} \\ \Delta \tilde{t} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta t \end{pmatrix}. \quad (15.5.18)$$

The eigenvalues, eigenvectors and exponents are given in Table 15.1(a). That the magnetic field exponent $y_1 = y_{\mathcal{H}} = d$ has already been seen in Sect. 15.2 to be a result of the use of the decimation weight function. In this particular case it can also be interpreted as a consequence of the first-order transition at $T = 0$ (Sect. 4.3) when the magnetization reverses direction as the magnetic field passes through zero.

Fig. 15.3 The trajectory flows for the renormalization group transformation of the one-dimensional Ising model. The origin corresponds to $u = t = 1$ and *fixed points* are denoted by •

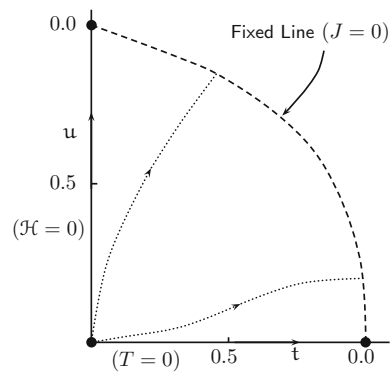


Table 15.1 The eigenvalues Λ_j , exponents y_j , and left and right eigenvectors w_j and x_j for (a) the ferromagnetic fixed point $t = 1, u = 1$, (b) the infinite temperature fixed point $t = 0, u = 1$, (c) the infinite field fixed point $t = 1, u = 0$, (d) the line of fixed points (15.5.17)

Λ_j	y_j	w_j	x_j	
2	1	(1,0)	(1,0)	} (a)
2	1	(0,1)	(0,1)	
1	0	(1,0)	$(1, -\frac{1}{4})$	} (b)
0	$-\infty$	$(\frac{1}{4}, 1)$	(0,1)	
0	$-\infty$	(1,0)	(1,0)	} (c)
1	0	(0,1)	(0,1)	
0	$-\infty$	$h(t)(1, x(t))$	$h(t)(1, y(t))$	} (d)
1	0	$h(t)(-y(t), 1)$	$h(t)(-x(t), 1)$	

- $t = 0, u = 1$ is the infinite temperature fixed point on $\mathcal{H} = 0$. For this we have

$$\begin{pmatrix} \Delta \tilde{u} \\ \Delta \tilde{t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{4} & 2 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta t \end{pmatrix}. \quad (15.5.19)$$

The eigenvalues, eigenvectors and exponents are given in Table 15.1(b). The fixed point is infinitely attractive along the zero-field axis. The marginal exponent $y_1 = 0$ corresponds to the fact that this fixed point is on the line of fixed points (15.5.17). The right eigenvector x_1 is tangential to the line of fixed points at this fixed point, where $du/dt = -4$.

- $t = 1, u = 0$ is the infinite field fixed point on $T = 0$. For this we have

$$\begin{pmatrix} \Delta \tilde{u} \\ \Delta \tilde{t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta t \end{pmatrix}. \quad (15.5.20)$$

The eigenvalues, eigenvectors and are given in Table 15.1(c). The fixed point is infinitely attractive along the zero-temperature axis. The marginal exponent $y_2 = 0$ again corresponds to the fact that this fixed point is on the line of fixed points (15.5.17). The right eigenvector x_2 is tangential to the fixed line at the fixed point where $du/dt = 0$.

- The line of fixed points (15.5.17) at some $0 \leq t \leq 1$ has linearized recurrence relationships

$$\begin{pmatrix} \Delta \tilde{u} \\ \Delta \tilde{t} \end{pmatrix} = \begin{pmatrix} \frac{(1-t)^2}{1+t^2} & \frac{-8t(1-t)}{(1+t^2)(1+t)^3} \\ \frac{(t^2-1)(1+t)^2}{4(1+t^2)} & \frac{2t}{1+t^2} \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta t \end{pmatrix}. \quad (15.5.21)$$

The eigenvalues, eigenvectors and exponents, as functions of t , are given in Table 15.1(d), where

$$h(t) := \sqrt{\frac{2t}{1+t^2}}, \quad x(t) := \frac{4(1-t)}{(1+t)^3}, \quad y(t) := \frac{(1-t)(1+t)^3}{8t}. \quad (15.5.22)$$

The line of fixed points is infinitely attractive to trajectories which approach it. The marginal exponent $y_2 = 0$ has a corresponding right eigenvector x_2 which is tangential to the fixed line, where $du/dt = -4(1-t)/(1+t)^3$.

We now calculate the free energy per lattice site. From (15.2.1), (15.5.8), (15.5.11) and (15.5.12)

$$\mathcal{K}_0(t, u) = -\frac{1}{8} \ln \left\{ \frac{2(1+u^{-1})^2(1+t^2)[u^2(1+t)^2 + (1-t)^2]}{(1-t^2)^2} \right\}. \quad (15.5.23)$$

The free energy per lattice site at some initial point $(u^{(0)}, t^{(0)})$ is now obtained by inserting this expression into the series (15.4.8), with $\lambda = 2$, $d = 1$. An interesting problem is to show that this series converges to the standard closed-form expression obtained by transfer matrix methods. Translated in the present notation this formula (Domb 1960, Sect. 3.2.1, Eq. 6) becomes

$$\phi_1(t, u) = -\ln \left\{ \frac{1}{2u^{3/4}} \left(\frac{1+t}{1-t} \right)^{1/2} \left[1 + u + \sqrt{(1-u)^2 + 4 \left(\frac{1-t}{1+t} \right)^2} \right] \right\}. \quad (15.5.24)$$

It is quite simple to write a computer program, using the recurrence relationships (15.5.13) and (15.5.14) to compute $\phi(u^{(0)}, t^{(0)})$, from (15.5.23) and (15.4.8). This will give convincing evidence that the series converges to the value given by (15.5.24). In general it would appear to be rather difficult to sum the series analytically. In the case $\mathcal{H} = 0$, ($u = 1$), however, the problem has been solved by Nauenberg (1975). His argument, translated into our notation, is as follows. From (15.5.23),

$$\mathcal{K}_0(t, 1) = -\frac{1}{4} \ln \left[4 \left(\frac{1+t^2}{1-t^2} \right) \right] \quad (15.5.25)$$

and hence, from (15.5.14) and (15.4.8),

$$\phi_1(t, 1) = - \sum_{s=1}^{\infty} \frac{1}{2^s} \left[\ln(2) + \frac{1}{2} \ln \left(\frac{1+t^{2^s}}{1-t^{2^s}} \right) \right] = - \ln \left[2 \prod_{s=1}^{\infty} \left(\frac{1+t^{2^s}}{1-t^{2^s}} \right)^{1/2^{s+1}} \right]. \quad (15.5.26)$$

Using the identity

$$\frac{1}{\sqrt{1-q^2}} = \prod_{s=1}^{\infty} \left(\frac{1+q^{2^s}}{1-q^{2^s}} \right)^{1/2^{s+1}}, \quad (15.5.27)$$

gives

$$\phi_1(t, 1) = -\ln(2) + \frac{1}{2} \ln(1-t^2), \quad (15.5.28)$$

which also follows from (15.5.24).

For the ferromagnetic fixed point the exponents given in Table 15.1 can be used together with formulae (4.5.40) and (4.5.39), to give the critical exponents

$$\alpha = 1, \quad \beta = 0, \quad \gamma = 1, \quad \delta = \infty. \quad (15.5.29)$$

The values of β and δ both arise from the first-order phase transition condition $y_1 = d$. Because the matrix in (15.5.18) is diagonal the scaling fields can be identified as

$$\theta_1 = -\Delta u \simeq 2L, \quad (15.5.30)$$

$$\theta_2 = -\Delta t \simeq 2 \exp(-2K). \quad (15.5.31)$$

From the general analysis of scaling at a critical point described in Sect. 4.5, $\theta_1 \sim \Delta L$ and $\theta_2 \sim \Delta K$. It follows from (15.5.30) that the first of these two formulae still holds in this case since $L_c = 0$. However, the fact that the critical temperature is $T_c = 0$ implies that $K_c = \infty$, leading to the different choice of scaling field θ_2 given by (15.5.31). The formula

$$\frac{\partial m}{\partial \mathcal{H}} = \frac{1}{T} \exp \left(\frac{2J}{T} \right), \quad (15.5.32)$$

(Domb 1960, Sect. 3.2.1, Eqs. 9 and 10) for the response of the magnetization to change of magnetic field, is now consistent with that derived from scaling using the value of γ in (15.5.29) and (15.5.31). A similar analysis is applicable to other one-dimensional models.

For the case $\lambda = 2$ we have considered only the ferromagnetic situation $J > 0$. This is because application of this weight function to the lattice, by removing alternate sites, maps the antiferromagnetic ground state of alternating spins into the

ferromagnetic ground state of aligned spins. In general, from (15.5.10), a point in phase space with $\mathfrak{Z}_2 < 1$ is mapped in one step into a point with $\mathfrak{Z}_2 > 1$. Although this does not affect the evaluation of the free energy which must be correct for either sign of J , (since the method is exact), it does mean that this method does not give fixed points associated with the antiferromagnetic model. In particular it is not possible to obtain the fixed point on the zero-temperature axis associated with the critical field (Sect. 3.5.1.1). One way of achieving this (Nelson and Fisher 1975) is to take $\lambda = 3$. In fact it is obvious that any odd value of λ would suffice. In the special case of zero magnetic field, it is not difficult to show that (15.5.15) generalizes to

$$\tilde{t} = t^\lambda, \quad (15.5.33)$$

which has the ferromagnetic fixed point $t = 1$ and the antiferromagnetic fixed point $t = -1$.

15.5.2 In Two Dimensions

We now explore the possibility of using the decimation procedure, described in the previous section, for the first-neighbour Ising model on a square lattice with zero magnetic field. The lattice is divided into black and white sites as shown in Fig. 15.4 and the summation in (15.1.13) is over the states on all the white sites, leaving the spins on the black sites fixed. With the spin states denoted appropriately, summing over the spin state σ_G on site G at the centre of the square $ABCD$ gives

$$2 \exp(2C) \cosh[K_1(\sigma_A + \sigma_B + \sigma_C + \sigma_D)], \quad (15.5.34)$$

where C and $K_1 := J/T$ are the trivial and first-neighbour couplings respectively. The left-hand side of (15.1.13) consists of a product of $N/2$ terms of this form, one

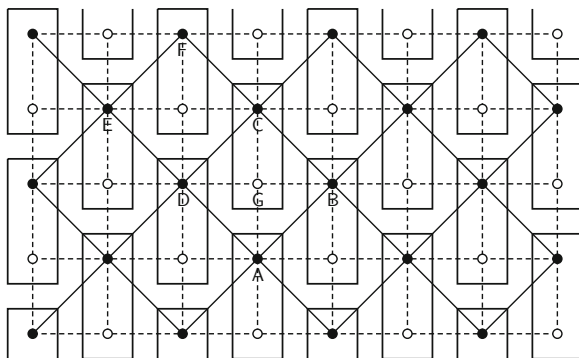


Fig. 15.4 Two site blocks for the first-neighbour Ising model on a square lattice

for each of the squares of type **ABCD** as shown in Fig. 15.4. It was pointed out by Wilson (1975) that this expression can be rewritten in the form

$$\begin{aligned} & \exp \left[2C_0 + \mathcal{K}_0 + \frac{1}{2} \mathcal{K}_1 (\sigma_A \sigma_B + \sigma_B \sigma_C + \sigma_C \sigma_D + \sigma_D \sigma_A) \right. \\ & \left. + \mathcal{K}_2 (\sigma_A \sigma_C + \sigma_B \sigma_D) + \mathcal{K}_3 \sigma_A \sigma_B \sigma_C \sigma_D \right], \end{aligned} \quad (15.5.35)$$

where

$$\mathcal{K}_0 := \ln(2) + \frac{1}{8} \ln \cosh(4K_1) + \frac{1}{2} \ln \cosh(2K_1), \quad (15.5.36)$$

$$\mathcal{K}_1 := \frac{1}{4} \ln \cosh(4K_1), \quad (15.5.37)$$

$$\mathcal{K}_2 := \frac{1}{8} \ln \cosh(4K_1), \quad (15.5.38)$$

$$\mathcal{K}_3 := \frac{1}{8} \ln \cosh(4K_1) - \frac{1}{2} \ln \cosh(2K_1). \quad (15.5.39)$$

By setting

$$\tilde{C} := 2C + \mathcal{K}_0(K_1), \quad (15.5.40)$$

$$\tilde{K}_i := \mathcal{K}_i(K_1), \quad i = 1, 2, 3, \quad (15.5.41)$$

we have recurrence relationships of the form of (15.2.1) and (15.2.2), with $d = 2$ and $\lambda = \sqrt{2}$, where $\tilde{K}_1$ is the first-neighbour coupling on the lattice of black sites and $\tilde{K}_2$ and $\tilde{K}_3$ are respectively a second-neighbour coupling and a four-site coupling on this new lattice.

The effect of decimation has been to generate new couplings. We can consider the strategy of going back to the original lattice and including second-neighbour and four-site couplings at this stage. These, however, have the effect of coupling white sites together and it is not difficult to see that this will lead to further couplings. This is the beginning of the proliferation of couplings referred to in Sect. 15.2. One possible approximation is to include a second-neighbour interaction only between black sites in the original lattice and to neglect the four-site coupling. This leads to the modification

$$\tilde{K}_1 = K_2 + \frac{1}{4} \ln \cosh(4K_1), \quad (15.5.42)$$

$$\tilde{K}_2 = \frac{1}{8} \ln \cosh(4K_1), \quad (15.5.43)$$

of (15.5.36)–(15.5.39). Following Wilson (1975) and approximating (15.5.42) and (15.5.43) by expanding to quadratic terms gives

$$\tilde{K}_1 = K_2 + 2K_1^2, \quad (15.5.44)$$

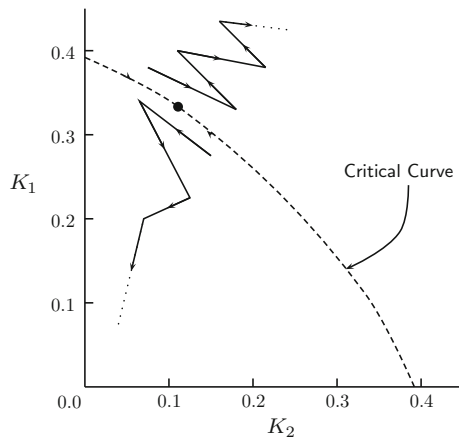
$$\tilde{K}_2 = K_1^2. \quad (15.5.45)$$

These have the fixed points $K_1 = K_2 = 0$ and $K_1 = K_2 = \infty$, which are both sinks and the fixed point $K_1^* = \frac{1}{3}$, $K_2^* = \frac{1}{9}$. Linearizing (15.5.44) and (15.5.45) about this fixed point

$$\begin{pmatrix} \Delta \tilde{K}_1 \\ \Delta \tilde{K}_2 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} & 1 \\ \frac{2}{3} & 0 \end{pmatrix} \begin{pmatrix} \Delta K_1 \\ \Delta K_2 \end{pmatrix}. \quad (15.5.46)$$

The eigenvalues of the matrix are $\Lambda_1 = \frac{1}{3}(2 + \sqrt{10}) = 1.721$ and $\Lambda_2 = \frac{1}{3}(2 - \sqrt{10}) = -0.387$. With $\lambda = \sqrt{2}$ it follows from (15.3.15) that Λ_1 gives the relevant (thermal) exponent $y_1 = y_T = 1.566$. The fixed point (K_1^*, K_2^*) lies on a critical curve, which cuts the axes in the plane of couplings at $(0, 0.392)$ and $(0.392, 0)$ (Fig. 15.5). We expected these two values to be the same, since an Ising model with only second-neighbour couplings is equivalent to two disconnected Ising models with first-neighbour couplings. Trajectories which begin at points enclosed within the region bounded by the axes and the critical curve iterate ultimately to the sink at the origin. Trajectories not beginning on the critical curve and outside this region iterate to the sink at infinity. The fixed point (K_1^*, K_2^*) on the critical curve is attractive to trajectories which begin on the curve. The irrelevant eigenvalue Λ_2 controls the attraction to the fixed point for such trajectories. That it is negative, giving an imaginary exponent y_2 , conflicts with the assumption we have made in Sect. 15.3. However, this is not a serious problem. It simply means that the iteration process will jump back forth from one side of the fixed point to the other as shown in Fig. 15.5. The crossing points of the critical curve with the axes give an approximation $T_c/J = 2.550$ to the critical temperature of the two-dimensional Ising model, to be compared with the exact value 2.269 given by (9.2.36). Also, from

Fig. 15.5 The phase diagram for the approximate treatment of the two-dimensional Ising model using the recurrence relations (15.5.44) and (15.5.45)



(4.5.53) and the value for the exponent given above, we have the approximate value for $\nu = 1/y_T = 0.639$, which should be compared with the exact value of unity (Sect. 12.3.9). Given the crudeness of the approximation the estimate of the critical temperature is surprisingly good although the value of ν is poor. It is interesting to see that if (15.5.42) and (15.5.43) are used rather than (15.5.44) and (15.5.45) the estimates for the critical temperature and ν are worse. The phase diagram again has the same structure but now $T_c/J = 1.582$ and $\nu = 0.556$. This example of ad hoc approximation, applied to a model for which exact results are known, serves to illustrate the rather arbitrary nature of the improvements and deteriorations in results which can occur. It stands as a warning when such methods are applied to models for which exact results are not available.

The problem with proliferating interactions, arises also for the other standard two and three-dimensional lattices, although it can be avoided in hierarchical lattices, where decimation amounts to a reversal of the procedure by which the hierarchical lattice was created (Kaufman and Griffiths 1981). A type of exact decimation procedure for the Ising model on a triangular lattice has been developed by Hilhorst et al. (1978, 1979). This is based on the star-triangle transformation. It avoids the problem of proliferation of interactions, but leads to spatially dependent couplings. Nevertheless the fixed point can be located analytically and the thermal exponent determined exactly.

15.6 The Kosterlitz–Thouless Transition

As we have seen the standard procedure of the renormalization group is to sum or integrate out short-distance details. The scheme of Wilson (1975) works with the Hamiltonian in wave-vector space and integrates out the high-frequency short-wavelength contributions. In the RSRG procedures discussed in Sect. 15.5 short-range detail is eliminated by removing a proportion of the lattice sites leading to a increase in the lattice spacing.

The Kosterlitz–Thouless transition in the two-dimensional XY model was discussed in Sect. 3.6.1.1. Its RSRG analysis by Kosterlitz (1974) (see also, Pierson 1997) has something of the flavour of the Wilson approach, but is performed in real-space and involves integrations which increase the lattice distance from a to $a + \Delta a$. Restoring the partition function to its original form will then involve a rescaling of the couplings.

Attention will be restricted to the case $\omega_j = \pm 1$, with ω denoting the set of winding numbers of all the vortices. In terms of³

$$\zeta_1 := \exp(K\mu), \quad \zeta_2 := \pi K, \quad (15.6.1)$$

³ This mixed notation where ζ_2 is a coupling and ζ_1 is a Boltzmann factor turns out to be the most convenient in this case.

and with the inclusion of a trivial coupling ζ_0 , the Hamiltonian (3.6.41) takes the form

$$\hat{H}(\boldsymbol{\zeta}, \boldsymbol{\omega}, a; \{\mathbf{r}_j^*\}) = V\zeta_0/a^2 - \zeta_2 \sum_{\{j \neq i\}} \omega_j \omega_i \ln \left| \frac{\mathbf{r}_j^* - \mathbf{r}_i^*}{a} \right| - N_v \ln(\zeta_1). \quad (15.6.2)$$

The number of lattice sites is $N^* = N$, which we suppose to be even and, given that the quasi-ordered state consists of bound pairs of vortices, we suppose that there can be $N_v =: 2n$ of these, each of winding number ± 1 . At this point we drop the ‘ $\star$ ’ reference to the dual lattice in the Hamiltonian and allow integrations in the partition function to range over all positions $\{\mathbf{r}_j\}$ for the vortices in the plane $\mathcal{V}$ subject only to the condition that, for every pair of v_j and v_i , $|\mathbf{r}_j - \mathbf{r}_i| \geq a$. With the order of the variable $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{2n}$, fixed, the partition function will involve integration over these variables beginning with $\mathbf{r}_1$. At the stage of integration with respect to $\mathbf{r}_j$ the only variables remaining are $\mathbf{r}_j, \mathbf{r}_{j+1}, \dots, \mathbf{r}_{2n}$, so the region of integration is

$$\mathcal{R}_j(a) := \mathcal{V} - \sum_{i=j+1}^{2n} \mathcal{D}(a; \mathbf{r}_j - \mathbf{r}_i), \quad (15.6.3)$$

where

$$\mathcal{D}(a; \mathbf{r}) := \{\mathbf{r}' : |\mathbf{r}'| < a\}. \quad (15.6.4)$$

The partition function is

$$\begin{aligned} Z(\boldsymbol{\zeta}, \boldsymbol{\omega}, a) &= \exp(-V\zeta_0/a^2) \sum_{n=1}^{N/2} \frac{\zeta_1^{2n}}{(n!)^2} \int_{\mathcal{R}_{2n}(a)} d\mathcal{V}_{2n} \cdots \int_{\mathcal{R}_1(a)} d\mathcal{V}_1 \\ &\quad \times \exp\left(2\zeta_2 \sum_{\{i>j\}} \omega_j \omega_i \ln \left| \frac{\mathbf{r}_i - \mathbf{r}_j}{a} \right| \right), \end{aligned} \quad (15.6.5)$$

where the factors $n!$ count the rearrangements of the identical vortices and $d\mathcal{V}_j$ denotes integration with respect to the variable $\mathbf{r}_j$. The aim now is to obtain renormalized parameters $\tilde{\zeta}_0$, $\tilde{\zeta}_1$ and $\tilde{\zeta}_2$ so that the invariance condition (15.1.12) holds in the form

$$Z(\tilde{\boldsymbol{\zeta}}, \boldsymbol{\omega}, a + \Delta a) = Z(\boldsymbol{\zeta}, \boldsymbol{\omega}, a) \quad (15.6.6)$$

for this system. In this case the two-dimensional volume V of the system is fixed and the number of lattice sites is reduced by the increase in lattice spacing so that $\tilde{N} = N[a/(a + \Delta a)]^2$. The regions of integration in the two partition functions in (15.6.6) are related by

$$\mathcal{R}_j(a) = \mathcal{R}_j(a + \Delta a) + \sum_{i=j+1}^{2n} \Delta \mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j), \quad (15.6.7)$$

where

$$\Delta \mathcal{D}(a; \mathbf{r}) := \mathcal{D}(a + \Delta a; \mathbf{r}) - \mathcal{D}(a; \mathbf{r}), \quad (15.6.8)$$

giving

$$\int_{\mathcal{R}_j(a)} d\mathcal{V}_j = \int_{\mathcal{R}_j(a+\Delta a)} d\mathcal{V}_j + \sum_{i=j+1}^{2n} \int_{\Delta \mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j)} d\mathcal{V}_j. \quad (15.6.9)$$

The contribution to the partition function arising from a term in the sum in (15.6.9) is $O(\Delta a)$. So when applying it to (15.6.5) the result to this order will contain only single integrals of this type. Thus there will be a contribution arising from the first term on the right of (15.6.9) plus a set of terms where the integration over $\mathcal{R}_j(a + \Delta a)$ is replaced by integration over $\Delta \mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j)$ for every $i > j$. Fixing the pair $[i, j]$ with $i > j$, the integrals over $\mathbf{r}_j$ and then $\mathbf{r}_i$ can be moved rightwards to the beginning of the order of integration when $\mathcal{R}_j(a + \Delta a)$ is replaced by

$$\mathcal{O}_i^{(j)}(a) := \mathcal{V} - \sum_{\{k \neq j\}} \mathcal{D}(a; \mathbf{r}_k - \mathbf{r}_i). \quad (15.6.10)$$

For the terms in the integrand containing $\mathbf{r}_i$ or $\mathbf{r}_j$, the factor of interest is

$$\begin{aligned} \mathcal{Z}_{ij}(\xi_2, \boldsymbol{\omega}, a) &= \int_{\mathcal{O}_i^{(j)}(a)} d\mathcal{V}_i \int_{\mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j)} d\mathcal{V}_j \\ &\times \exp \left(2 \xi_2 \left[\omega_i \sum_{\{k \neq i\}} \omega_k \ln \left| \frac{\mathbf{r}_k - \mathbf{r}_i}{a} \right| \right. \right. \\ &\quad \left. \left. + \omega_j \sum_{\{k \neq j\}} \omega_k \ln \left| \frac{\mathbf{r}_k - \mathbf{r}_j}{a} \right| - \omega_i \omega_j \ln \left| \frac{\mathbf{r}_j - \mathbf{r}_i}{a} \right| \right] \right). \end{aligned} \quad (15.6.11)$$

The integrand for the whole partition function is dominated by terms with $\omega_j = -\omega_i$ and the inner integration in (15.6.11) imposes the restriction that $\mathbf{r}_j = \mathbf{r}_i + \mathbf{q}$, where $a < |\mathbf{q}| < a + \Delta a$, giving

$$\begin{aligned} \mathcal{Z}_{ij}(\xi_2, \boldsymbol{\omega}, a) &= \int_{\mathcal{O}_i^{(j)}(a)} d\mathcal{V}_i \int_{\mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j)} d\mathcal{V}_j \\ &\times \exp \left(-\xi_2 \omega_i \sum_{\{k \neq i\}} \omega_k \ln \left| 1 + \frac{2q \cos(\alpha_k)}{|\mathbf{r}_k - \mathbf{r}_i|} + \frac{q^2}{|\mathbf{r}_k - \mathbf{r}_i|^2} \right| \right), \end{aligned} \quad (15.6.12)$$

where α_k is the angle between $\mathbf{r}_k - \mathbf{r}_i$ and $\mathbf{q}$. Expanding in powers of $q/|\mathbf{r}_k - \mathbf{r}_i|$ and then expanding the exponential and integrating with respect to q over the range

($a, a + \Delta a$) gives

$$\begin{aligned}
 \mathcal{Z}_{ij}(\zeta_2, \boldsymbol{\omega}, a) &= \int_{\mathcal{O}_i^{(j)}(a)} d\mathcal{V}_i \int_{\mathcal{D}(a; \mathbf{r}_i - \mathbf{r}_j)} d\mathcal{V}_j \\
 &\times \left(1 - \zeta_2 \omega_i \sum_{\{k \neq i\}} \omega_k \left[\frac{2q \cos(\alpha_k)}{|\mathbf{r}_k - \mathbf{r}_i|} + \frac{q^2 [1 - 2 \cos^2(\alpha_k)]}{|\mathbf{r}_k - \mathbf{r}_i|^2} \right] \right. \\
 &\quad \left. + 2\zeta_2^2 \sum_{\{k \neq i\}} \sum_{\{\ell \neq i\}} \omega_k \omega_\ell \frac{q^2 \cos(\alpha_k) \cos(\alpha_\ell)}{|\mathbf{r}_k - \mathbf{r}_i| |\mathbf{r}_\ell - \mathbf{r}_i|} \right) \\
 &= 2\pi a \Delta a \int_{\mathcal{O}_i^{(j)}(a)} d\mathcal{V}_i \left\{ 1 + \frac{\zeta_2^2 a^2}{\pi} \int_0^{2\pi} d\alpha_k \sum_{\{k \neq i\}} \sum_{\{\ell \neq i\}} \omega_k \omega_\ell \frac{\cos(\alpha_k) \cos(\beta_{k\ell} - \alpha_k)}{|\mathbf{r}_k - \mathbf{r}_i| |\mathbf{r}_\ell - \mathbf{r}_i|} \right\},
 \end{aligned} \tag{15.6.13}$$

where $\beta_{k\ell}$ is the angle between $\mathbf{r}_k - \mathbf{r}_i$ and $\mathbf{r}_\ell - \mathbf{r}_i$. Integrating with respect to α_k gives

$$\begin{aligned}
 \mathcal{Z}_{ij}(\zeta_2, \boldsymbol{\omega}, a) &= 2\pi a \Delta a \int_{\mathcal{O}_i^{(j)}(a)} d\mathcal{V}_i \left\{ 1 + \zeta_2^2 a^2 \sum_{\{k \neq i\}} \sum_{\substack{k \neq \ell \\ \{\ell \neq i\}}} \omega_k \omega_\ell \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot (\mathbf{r}_\ell - \mathbf{r}_i)}{|\mathbf{r}_k - \mathbf{r}_i|^2 |\mathbf{r}_\ell - \mathbf{r}_i|^2} \right. \\
 &\quad \left. + \zeta_2^2 a^2 \sum_{\{k \neq i\}} \frac{1}{|\mathbf{r}_k - \mathbf{r}_i|^2} \right\}
 \end{aligned} \tag{15.6.14}$$

and a final integration over $\mathcal{O}_i^{(j)}(a)$ gives

$$\mathcal{Z}_{ij}(\zeta_2, \boldsymbol{\omega}, a) = 2\pi a \Delta a \left[V - 2\pi \zeta_2^2 a^2 \sum_{\{k \neq \ell\}} \omega_k \omega_\ell \ln \left| \frac{\mathbf{r}_k - \mathbf{r}_\ell}{a} \right| \right], \tag{15.6.15}$$

(Kosterlitz 1974). When substituting into (15.6.9) to obtain the partition function we must note that the contribution from the second term has ‘lost’ two vortices with winding numbers ± 1 . But the factor in (15.6.15) has lost its dependence on i and j so the sum in (15.6.9) and the choice of j compensate for this, the residual effect being just a factor ζ_1^2 . The term to insert into the partition function is therefore

$$\begin{aligned}
 \zeta_1^2 \mathcal{Z}_{ij}(\zeta_2, \boldsymbol{\omega}, a) &= \exp \left(2\zeta_1^2 \pi a \Delta a \left[V - 4\pi \zeta_2^2 a^2 \sum_{\{k > \ell\}} \omega_k \omega_\ell \ln \left| \frac{\mathbf{r}_k - \mathbf{r}_\ell}{a} \right| \right] \right) - 1 \\
 &\quad + O(\Delta a).
 \end{aligned} \tag{15.6.16}$$

In applying (15.6.9) to (15.6.5) the a in the denominator of the logarithmic factors must be replaced by $a + \Delta a$ so that

$$\exp \left(2\zeta_2 \sum_{\{i > j\}} \omega_j \omega_i \ln \left| \frac{\mathbf{r}_i - \mathbf{r}_j}{a} \right| \right) = \exp \left(2\zeta_2 \sum_{\{i > j\}} \omega_j \omega_i \ln \left| \frac{\mathbf{r}_i - \mathbf{r}_j}{a + \Delta a} \right| - \frac{2n \Delta a \zeta_2}{a} \right). \tag{15.6.17}$$

Applying a similar change to (15.6.16) yields a neglectable factor $O([\Delta a]^2)$. So from (15.6.5), (15.6.9), (15.6.16) and (15.6.17) the transformation in couplings is given by

$$\frac{V\tilde{\zeta}_0}{(a + \Delta a)^2} = \frac{V\zeta_0}{a^2} - 2V\zeta_1^2 \pi a \Delta a, \quad (15.6.18)$$

$$\tilde{\zeta}_1(a + \Delta a)^2 = \zeta_1 a^2 \left(1 + \frac{(2 - \zeta_2)\Delta a}{a} \right), \quad (15.6.19)$$

$$\tilde{\zeta}_2 = \zeta_2 \left(1 - \frac{(2\pi\zeta_1 a^2)^2 \zeta_2 \Delta a}{a} \right). \quad (15.6.20)$$

In the limit $\Delta a \rightarrow 0$ and in terms of the variables

$$\ell := \ln(a), \quad y := 4\pi\zeta_1 a^2, \quad x := 2 - \zeta_2, \quad (15.6.21)$$

(15.6.19) and (15.6.20) can be put into the differential form

$$\frac{dx}{d\ell} = y^2 \left(1 - \frac{1}{2}x \right)^2, \quad (15.6.22)$$

$$\frac{dy}{d\ell} = xy, \quad (15.6.23)$$

for which $y = 0$ is a fixed line. In the region close to $x = 0$, where $T \simeq T_{\text{KT}}$, the *Kosterlitz–Thouless temperature* given by (3.6.35), these equations can be linearized in x to give

$$\frac{dx}{d\ell} = y^2, \quad (15.6.24)$$

$$\frac{dy}{d\ell} = xy, \quad (15.6.25)$$

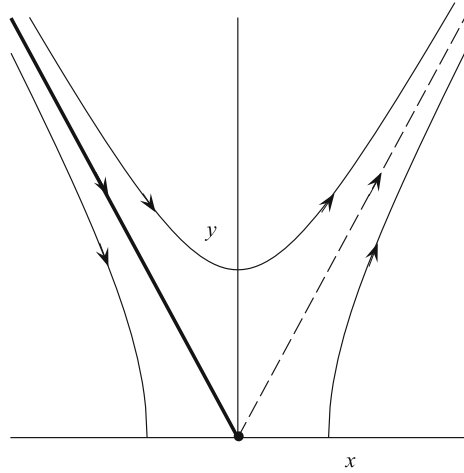
for which the solutions are two families of rectangular hyperbolae

$$x^2 - y^2 = \pm c^2 \quad \text{together with the lines } y = \pm x. \quad (15.6.26)$$

On physical grounds we are interested only in the region $y \geq 0$ and $x \geq, \leq 0$ correspond respectively to $T \geq, \leq T_{\text{KT}} := \pi J/2$. The scaling parameter ℓ governing the flow increases with a yielding increasing x everywhere and y decreasing or increasing according as $x < 0$ and $x > 0$. The renormalization flow is shown in Fig. 15.6 and the line $x = -y$ is the critical line controlled by the Kosterlitz–Thouless point $x = y = 0$. For the XY model μ is the energy of an isolated vortex leading to a single critical point on the line. For the Coulomb gas the critical line is parameterised by the fugacity μ . For a fixed value of this quantity the physical curve of the system is

$$x(K) = \pi(K_{\text{KT}} - K), \quad y(K) = 4\pi \exp(\mu K), \quad (15.6.27)$$

Fig. 15.6 The Kosterlitz–Thouless renormalization group flows. The heavy line $x = -y$ is the critical line terminating at the Kosterlitz–Thouless point at the origin



passing through the critical point $K = K_c$ given by the transcendental equation

$$K_c - K_{KT} = 4 \exp(\mu K_c). \quad (15.6.28)$$

In the low-temperature region to the left of the critical line the flow is towards the fixed line $y = 0$. Beginning at a particular (K, μ) the point on the fixed line corresponds to $\mu \rightarrow -\infty$ where $K = \bar{K}(K)$. The pair correlation function given by (3.6.25) is algebraic where, from (3.6.14) the exponent is $\eta(2; K, \mu) = K_{KT}/4\bar{K}(K) < 1$, with $\bar{K}(K) > K$. Let $\bar{K}(K) := \rho(K)K/J$ giving $\eta(K) = JK_{KT}/4\rho(K)K$. Since $\bar{K}(K) \geq K$, $\rho(K) \geq J$ with $\rho(K_{KT}) = J$, where $\eta = \frac{1}{4}$. The quantity $\rho(K)$, which vanishes discontinuously at $K = K_{KT}$ is the *spin-wave stiffness* of the system, which is related to the choice of boundary conditions (Cardy 1984, Sect. 6.4; Goldenfeld 1992, Sect. 11.3; Chaikin and Lubensky 1995, Sect. 6.1). Using the continuum form of (4.2.4) it follows from the algebraic form of the pair correlation function that the correlation length $\tau \sim L_0$ the size of the system.

Consider now the high-temperature region to the right of the critical line. Close to the critical point, where $K = K_c - \Delta K$, the physical curve intersects the rectangular hyperbola

$$y^2 - x^2 = \Delta K B, \quad (15.6.29)$$

$$B := 2A(\pi - \mu A), \quad A := 4\pi \exp(\mu K_c) = \pi(K_c - K_{KT}).$$

From (15.6.24) the scaling parameter is given by

$$\ell = \int \frac{dx}{x^2 + B\Delta K} \simeq \frac{\pi}{\sqrt{B\Delta K}}, \quad (15.6.30)$$

where, since for small ΔK the integrand is steeply peaked around $x = 0$, the limits have been extended to infinity in each direction. From (15.6.21) it is clear that the correlation length scales according to

$$\tau = \tau_0 \exp(\ell) = \tau_0 \exp\left(\frac{\pi}{\sqrt{B\Delta K}}\right). \quad (15.6.31)$$

Comparing with the scaling form (4.5.52) it is evident that there is no exponent ν . In fact the correlation length diverges at the critical curve faster than any power of $|\Delta K|$. However, it can be shown (Kosterlitz 1974) that the weak scaling laws (12.3.325) are satisfied with $\delta = 15$, $\eta = \frac{1}{4}$.

15.7 Upper-Bound and Lower-Bound Approximations

The programme for RSRG methods is to use the formula (15.1.13) to obtain the recurrence relationships (15.2.1) and (15.2.2) and then to use the methods described in Sects. 15.3 and 15.4 to obtain the critical exponents and free energy, respectively. As we have seen in Sect. 15.5.1, this scheme can be carried out exactly for some one-dimensional models, but Sect. 15.5.2 describes the problems encountered when a similar approach is attempted in two dimensions. The approximation methods which will be described in the remainder of this chapter are of two broad types. In the methods of Sects. 15.8 and 15.10 the lattice for the system is replaced by a finite subset of sites or a lattice of lower dimensionality. In the method described in Sect. 15.9 an approximation is made to the right-hand side of (15.1.13) so that the summation can be performed. In that case the modified form, when summed over the microstates on $\tilde{\mathcal{N}}$, will not necessarily reproduce the invariance condition (15.1.12) for the partition function.

There are, however, a number of methods for which this equality can be replaced by an inequality which remains valid throughout the iteration process. Suppose that the approximate recurrence relationships replacing (15.2.1) and (15.2.2) are

$$\tilde{\xi}'_0 = \lambda^d [\zeta_0 + \mathcal{U}_0(\zeta_j)], \quad (15.7.1)$$

$$\tilde{\xi}'_i = \mathcal{U}_i(\zeta_j), \quad i = 1, 2, \dots, n, \quad (15.7.2)$$

with (15.1.12) replaced by

$$Z(\tilde{\mathcal{N}}, \tilde{\xi}'_i) \leq Z(N, \zeta_i). \quad (15.7.3)$$

Then, in place of (15.4.1) and (15.4.4), we have, respectively,

$$\Phi(\tilde{\mathcal{N}}, \tilde{\xi}'_i) + \tilde{\mathcal{N}} \tilde{\xi}'_0 \geq \Phi(N, \zeta_i) + N \zeta_0, \quad (15.7.4)$$

$$\phi(\tilde{\zeta}_i') \geq \lambda^d [\phi(\zeta_i) - \mathcal{U}_0(\zeta_i)]. \quad (15.7.5)$$

Now suppose that (15.7.1) and (15.7.2) are used to generate a trajectory $\zeta_i^{(0)} \rightarrow \zeta_i^{(1)'} \rightarrow \zeta_i^{(2)'} \dots$ in phase space. By an procedure similar to that of Sect. 15.4 it follows that

$$\phi(\zeta_i^{(0)}) \leq \phi^{(\mathcal{U})}(\zeta_i^{(0)}) = \sum_{s=0}^{\infty} \frac{1}{\lambda^{sd}} \mathcal{U}_0(\zeta_i^{(s)'}). \quad (15.7.6)$$

The transformation generates an upper-bound to the free energy. By a similar argument a set of recurrence relationships

$$\tilde{\zeta}_0' = \lambda^d [\zeta_0 + \mathcal{L}_0(\zeta_j)], \quad (15.7.7)$$

$$\tilde{\zeta}_i' = \mathcal{L}_i(\zeta_j), \quad i = 1, 2, \dots, n, \quad (15.7.8)$$

with

$$Z(\tilde{N}, \tilde{\zeta}_i') \geq Z(N, \zeta_i), \quad (15.7.9)$$

will generate a lower bound

$$\phi^{(\mathcal{L})}(\zeta_i^{(0)}) = \sum_{s=0}^{\infty} \frac{1}{\lambda^{sd}} \mathcal{L}_0(\zeta_i^{(s)'}) \quad (15.7.10)$$

to the free energy.

15.7.1 An Upper-Bound Method

An approach which leads to an upper bound to the free energy was introduced by Niemeijer and van Leeuwen (1974) (see, also, Niemeijer and van Leeuwen 1976). It is based on a splitting of the Hamiltonian on the right-hand side of (15.1.13) into two parts

$$\hat{H}(N, \zeta_i, \Delta\zeta; \sigma(\mathbf{r})) = \hat{H}_0(N, \zeta_i; \sigma(\mathbf{r})) + \Delta\zeta \hat{C}(N; \sigma(\mathbf{r})), \quad (15.7.11)$$

where $\hat{H}_0$ is referred to as the *zeroth-order part*, in comparison to which the second term is small as a consequence of the small value $\Delta\zeta$ of its single coupling.⁴ The zeroth-order average $\langle A \rangle_0$ for any function $A(\sigma(\mathbf{r}))$ of the spin states is given by

⁴ The argument ζ_i now represents the set of couplings in $\hat{H}_0$ and any further couplings in the second term in (15.7.11) are absorbed as ratios in $\hat{C}$.

$$\langle A \rangle_0 := \frac{\sum_{\{\sigma(\mathbf{r})\}} \left[\prod_{\{\tilde{\mathbf{r}}\}} p(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) \right] A(\sigma(\mathbf{r})) \exp[-\hat{H}_0(N, \zeta_i; \sigma(\mathbf{r}))]}{\mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}}))}, \quad (15.7.12)$$

where $\mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}}))$ is the quantity given by (15.1.14), but using only the zeroth-order part of the Hamiltonian. From this it follows that

$$\langle \exp[-\Delta\zeta \hat{C}(N; \sigma(\mathbf{r}))] \rangle_0 = \frac{\mathcal{Z}(N, \zeta_i, \Delta\zeta; \sigma(\tilde{\mathbf{r}}))}{\mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}}))} \quad (15.7.13)$$

and (15.1.13) can be expressed in the form

$$\exp[-\hat{H}(\tilde{N}, \tilde{\zeta}_i, \tilde{\Delta}\zeta; \sigma(\tilde{\mathbf{r}}))] = \langle \exp[-\Delta\zeta \hat{C}(N; \sigma(\mathbf{r}))] \rangle_0 \mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}})). \quad (15.7.14)$$

Summing over all the spin states $\sigma(\tilde{\mathbf{r}})$ gives

$$Z(\tilde{N}, \tilde{\zeta}_i, \tilde{\Delta}\zeta) = Z(N, \zeta_i, \Delta\zeta) = \mathcal{Z}_0(N, \zeta_i) \langle \exp[-\Delta\zeta \hat{C}(N; \sigma(\mathbf{r}))] \rangle_0. \quad (15.7.15)$$

These expressions are, of course, exact. The approximation arises when the final factor on the left is replaced by a finite number of terms of its *cumulant expansion*. The simplest way to generate this cluster expansion is to consider

$$\begin{aligned} \ln[\mathcal{Z}(N, \zeta_i, \Delta\zeta; \sigma(\tilde{\mathbf{r}}))] &= \left[\exp\left(\Delta\zeta \frac{\partial}{\partial q}\right) \ln[\mathcal{Z}(N, \zeta_i, q; \sigma(\tilde{\mathbf{r}}))] \right]_{q=0} \\ &= \ln[\mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}}))] + \left[\sum_{m=1}^{\infty} (\Delta\zeta)^m \frac{\partial^m}{\partial q^m} \ln[\mathcal{Z}(N, \zeta_i, q; \sigma(\tilde{\mathbf{r}}))] \right]_{q=0}, \end{aligned} \quad (15.7.16)$$

(cf. (14.3.4)–(14.3.6) and the footnote on p. 537), which gives, from (15.7.13)

$$\begin{aligned} \langle \exp[-\Delta\zeta \hat{C}(N; \sigma(\mathbf{r}))] \rangle_0 &= \exp \left\{ -\Delta\zeta \langle \hat{C}(N; \sigma(\mathbf{r})) \rangle_0 \right. \\ &\quad + \frac{1}{2!} (\Delta\zeta)^2 \langle [\hat{C}(N; \sigma(\mathbf{r})) - \langle \hat{C}(N; \sigma(\mathbf{r})) \rangle_0]^2 \rangle_0 \\ &\quad \left. - \frac{1}{3!} (\Delta\zeta)^3 \langle [\hat{C}(N; \sigma(\mathbf{r})) - \langle \hat{C}(N; \sigma(\mathbf{r})) \rangle_0]^3 \rangle_0 + \cdots \right\}. \end{aligned} \quad (15.7.17)$$

The *first-order cumulant approximation* consists in retaining only the first term in the exponential on the right-hand side of Eq. (15.7.17) for which⁵

⁵ For any $A(\sigma(\mathbf{r}))$

$$\begin{aligned} \langle \exp(A) \rangle_0 &= \exp(A)_0 \langle \exp(A - \langle A \rangle_0) \rangle_0 \\ &= \exp(A)_0 \langle 1 + (A - \langle A \rangle_0) + \frac{1}{2} (A - \langle A \rangle_0)^2 + \cdots \rangle_0 \geq \exp(A)_0. \end{aligned}$$

$$\langle \exp[-\Delta\zeta \widehat{C}(N; \sigma(\mathbf{r}))] \rangle_0 \geq \exp[-\Delta\zeta \langle \widehat{C}(N; \sigma(\mathbf{r})) \rangle_0]. \quad (15.7.18)$$

Substituting into (15.7.14) and replacing $\tilde{\zeta}_i$ and $\widetilde{\Delta\zeta}$ by $\tilde{\zeta}'_i$ and $\widetilde{\Delta\zeta}'$ gives

$$\exp[-\widehat{H}(\tilde{N}, \tilde{\zeta}'_i, \widetilde{\Delta\zeta}'; \sigma(\tilde{\mathbf{r}}))] = \mathcal{Z}_0(N, \zeta_i; \sigma(\tilde{\mathbf{r}})) \exp[-\Delta\zeta \widehat{C}(N; \sigma(\mathbf{r}))]_0. \quad (15.7.19)$$

The corresponding partition function is again obtained by summing over the spin states $\sigma(\tilde{\mathbf{r}})$ to give

$$Z(\tilde{N}, \tilde{\zeta}'_i, \widetilde{\Delta\zeta}') = Z_0(N, \zeta_i) \exp \langle -\Delta\zeta \widehat{C}(N; \sigma(\mathbf{r})) \rangle_0. \quad (15.7.20)$$

The upper-bound condition (15.7.3) then follows from (15.7.15) and (15.7.18).

Example 15.7.1 We return to the problem, discussed in Sect. 15.5.2, of using decimation for the two-dimensional Ising model, adopting a first-order cumulant approximation to construct an upper-bound transformation. The zeroth-order Hamiltonian $\widehat{H}_0$ is taken to include the trivial coupling C and the first-neighbour coupling K_1 . Terms corresponding to a second-neighbour coupling K_2 and a four-site coupling K_3 are included in $\Delta\zeta \widehat{C}$ with K_2 playing the role of $\Delta\zeta$ and $K_3 = \beta K_2$. Because the four-site term links together pairs of white sites in Fig. 15.4 the method must be based on pairs of neighbouring squares like **ABCD** and **CDEF**. If the Hamiltonian is taken to be made up of overlapping pairs of squares, each square is counted four times and thus factors arising from a single square should be reduced by a factor of four. Denoting the contribution to any factor from the pair of squares by the symbol $[..]_{\square\square}$

$$\begin{aligned} & [\mathcal{Z}_0(N, C, K_1; \sigma(\tilde{\mathbf{r}}))]_{\square\square} \\ &= \exp(C) \{4 \cosh[K_1(\sigma_A + \sigma_B + \sigma_C + \sigma_D)] \cosh[K_1(\sigma_E + \sigma_F + \sigma_C + \sigma_D)]\}^{1/4}, \end{aligned} \quad (15.7.21)$$

$$\begin{aligned} & [\exp \langle -K_2 \widehat{C}(N, \beta; \sigma(\mathbf{r})) \rangle_0]_{\square\square} = \exp \left[K_2 \left\{ \alpha^{(1)} (\sigma_A \sigma_B + \sigma_E \sigma_F) \right. \right. \\ & \quad + \alpha^{(2)} \sigma_C \sigma_D + \alpha^{(3)} (\sigma_A \sigma_D + \sigma_B \sigma_C + \sigma_E \sigma_D + \sigma_F \sigma_C) \\ & \quad \left. \left. + (1 + \beta \sigma_C \sigma_D) \tanh[K_1(\sigma_A + \sigma_B + \sigma_C + \sigma_D)] \tanh[K_1(\sigma_E + \sigma_F + \sigma_C + \sigma_D)] \right\} \right], \end{aligned} \quad (15.7.22)$$

where

$$2\alpha^{(1)} + \alpha^{(2)} + 4\alpha^{(3)} = 1. \quad (15.7.23)$$

As in the simpler situation described in Sect. 15.5.2 the product of the right-hand sides of (15.7.21) and (15.7.22) can be re-expressed in the form

$$\begin{aligned}
& \exp \left\{ C + \frac{1}{2} \mathcal{K}_0 + \mathcal{K}_1^{(1)} (\sigma_A \sigma_B + \sigma_E \sigma_F) + \mathcal{K}_1^{(2)} \sigma_C \sigma_D \right. \\
& \quad + \mathcal{K}_1^{(3)} (\sigma_A \sigma_D + \sigma_B \sigma_C + \sigma_E \sigma_D + \sigma_F \sigma_C) \\
& \quad + \frac{1}{4} \mathcal{K}_2 (\sigma_A \sigma_C + \sigma_B \sigma_D + \sigma_C \sigma_E + \sigma_D \sigma_F) \\
& \quad + \frac{1}{4} \mathcal{K}_3 \sigma_C \sigma_D (\sigma_A \sigma_B + \sigma_E \sigma_F) \\
& \quad + \mathcal{K}_4 (\sigma_A \sigma_F + \sigma_B \sigma_E) + \mathcal{K}_5 (\sigma_A \sigma_E + \sigma_B \sigma_F) + \mathcal{K}_6 \sigma_A \sigma_B \sigma_E \sigma_F \\
& \quad + \mathcal{K}_7 \sigma_C \sigma_D (\sigma_A \sigma_E + \sigma_B \sigma_F) + \mathcal{K}_8 \sigma_C \sigma_D (\sigma_A \sigma_F + \sigma_B \sigma_E) \\
& \quad + \mathcal{K}_9 [\sigma_A \sigma_B (\sigma_C \sigma_F + \sigma_D \sigma_E) + \sigma_E \sigma_F (\sigma_A \sigma_D + \sigma_B \sigma_C)] \\
& \quad + \mathcal{K}_{10} [\sigma_A \sigma_B (\sigma_D \sigma_F + \sigma_C \sigma_E) + \sigma_E \sigma_F (\sigma_A \sigma_C + \sigma_B \sigma_D)] \\
& \quad \left. + \mathcal{K}_{11} \sigma_A \sigma_B \sigma_C \sigma_D \sigma_E \sigma_F \right\}, \tag{15.7.24}
\end{aligned}$$

where, using the compact notation

$$\text{th}_\ell := \tanh(\ell K_1), \quad \text{lnc}_\ell := \ln \cosh(\ell K_1), \quad \ell = 2, 4, \tag{15.7.25}$$

and reverting to the use of the coupling $K_3 = \beta K_2$,

$$\mathcal{K}_0 := \frac{1}{16} (K_2 + K_3) [\text{th}_4 + 2 \text{th}_2]^2 + \ln(2) + \frac{1}{8} \text{lnc}_4 + \frac{1}{2} \text{lnc}_2, \tag{15.7.26}$$

$$\mathcal{K}_1^{(1)} := \frac{1}{32} (K_2 + K_3) [\text{th}_4^2 - 4 \text{th}_2^2] + K_2 \alpha^{(1)} + \frac{1}{32} \text{lnc}_4, \tag{15.7.27}$$

$$\mathcal{K}_1^{(2)} := \frac{1}{32} (K_2 + K_3) [\text{th}_4 + 2 \text{th}_2]^2 + K_2 \alpha^{(2)} + \frac{1}{16} \text{lnc}_4, \tag{15.7.28}$$

$$\mathcal{K}_1^{(3)} := \frac{1}{32} (K_2 + K_3) \text{th}_4 [\tanh_4 + 2 \text{th}_2] + K_2 \alpha^{(3)} + \frac{1}{32} \text{lnc}_4, \tag{15.7.29}$$

$$\mathcal{K}_2 := \frac{1}{8} (K_2 + K_3) \tanh(4K_1) [\text{th}_4 + 2 \text{th}_2] + \frac{1}{8} \text{lnc}_4, \tag{15.7.30}$$

$$\mathcal{K}_3 := \frac{1}{8} [\text{th}_4^2 - 4 \text{th}_2^2] + \frac{1}{8} \text{lnc}_4 - \frac{1}{2} \text{lnc}_2, \tag{15.7.31}$$

$$\mathcal{K}_4 = \mathcal{K}_5 := \frac{1}{32} [(K_2 + K_3) \text{th}_4 + 4K_2(1 - \beta) \text{th}_2], \tag{15.7.32}$$

$$\mathcal{K}_6 := \frac{1}{32} (K_2 + K_3) [\text{th}_4 - 2 \text{th}_2]^2, \tag{15.7.33}$$

$$\mathcal{K}_7 = \mathcal{K}_8 := \frac{1}{32} [(K_2 + K_3) \text{th}_4 - 4K_2(1 - \beta) \text{th}_2], \tag{15.7.34}$$

$$\mathcal{K}_9 = \mathcal{K}_{10} := \frac{1}{32} (K_2 + K_3) \text{th}_4 [\text{th}_4 - 2 \text{th}_2], \tag{15.7.35}$$

$$\mathcal{K}_{11} = \frac{1}{32} (K_2 + K_3) [\text{th}_4 - 2 \text{th}_2]^2. \tag{15.7.36}$$

With

$$\mathcal{K}_1(K_1, K_2, K_3) := 2\mathcal{K}_1^{(1)}(K_1, K_2, K_3) + \mathcal{K}_1^{(2)}(K_1, K_2, K_3) + 4\mathcal{K}_1^{(3)}(K_1, K_2, K_3), \tag{15.7.37}$$

recurrence relationships of the form (15.2.1) and (15.2.2) with $d = 2$, $\lambda = \sqrt{2}$ and $n = 11$ can now be constructed. They reduce to the recurrence relationships given by (15.5.36)–(15.5.41), when $K_2 = K_3 = 0$, $\tilde{K}_1$ and $\tilde{K}_2$ are the first- and second-neighbour couplings on a square like ABCD and $\tilde{K}_3$ is a four-site coupling on such a square. The couplings $\tilde{K}_i$, $i = 4, 5, \dots, 11$ have been generated by the transformation and link sites. Sneddon and Barber (1977) have investigated the approximation where these proliferating couplings are neglected and only first-neighbour, second-neighbour and the four-site coupling on the square are retained. They show that the only fixed point not corresponding to either zero or infinite values of the couplings is $K_1^* = 0.3683$, $K_2^* = 0.1275$, $K_3^* = -0.0303$. Linearizing about this fixed point yields the one relevant eigenvalue $\Lambda_1 = 1.4287$, giving the thermal exponent $y_1 = y_T = 1.029$ with $\nu = 1/y_T = 0.9714$. This represents a considerable improvement over the results of the simpler approximation described in Sect. 15.5.2. It is tempting to suppose that cumulant approximations to higher orders would yield further improvement. However this does not seem to be the case. Kadanoff (1976a) reports that unpublished results by Houghton and himself yielded respectively $y_T = 1.002$ and $y_T = 1.7$, for the second- and third-order extension to the method described here. This problem with the cumulant approximation does not appear to be related to the difficulties associated with the decimation weight function. Hsu and Gunton (1977) have used the Kadanoff weight function (15.2.9) to investigate second- and third-order cumulant expansions for the Ising model on the square and simple cubic lattice. They find that the best results are achieved in the decimation limit $\tau \rightarrow \infty$ and that, although reasonable results are achieved for the critical temperature and magnetic field exponent, the results for the thermal exponent are poor.

15.7.2 A Lower-Bound Method

A starting point for developing a class of lower-bound transformations can be achieved if the right-hand side of (15.1.13) is modified by replacing $\hat{H}(N, \zeta_i; \sigma(\mathbf{r}))$ by $\hat{H}(N, \zeta_i; \sigma(\mathbf{r})) + \hat{V}(N, \zeta_i; \sigma(\mathbf{r}))$. Then we have

$$\begin{aligned} \exp[-\hat{H}(\tilde{N}, \tilde{\zeta}_i'; \sigma(\tilde{\mathbf{r}}))] &= \sum_{\{\sigma(\mathbf{r})\}} \left[\prod_{\{\tilde{\mathbf{r}}\}} p(\sigma(\tilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\tilde{\mathbf{r}})) \right] \\ &\times \exp[-\hat{H}(N, \zeta_i; \sigma(\mathbf{r})) - \hat{V}(N, \zeta_i; \sigma(\mathbf{r}))], \end{aligned} \quad (15.7.38)$$

where the functional form of the Hamiltonian on the left-hand side is the same as that in (15.1.13), but the renormalized couplings take different values denoted by $\tilde{\zeta}_i'$. Summing (15.7.38) over the microstates on $\tilde{\mathcal{N}}$ gives

$$\begin{aligned} Z(\tilde{N}, \tilde{\zeta}_i') &= Z(N, \zeta_i) \langle \exp[-\hat{V}(N, \zeta_i; \sigma(\mathbf{r}))] \rangle \\ &\geq Z(N, \zeta_i) [1 - \langle \hat{V}(N, \zeta_i; \sigma(\mathbf{r})) \rangle]. \end{aligned} \quad (15.7.39)$$

If $\widehat{V}(N, \xi_i; \sigma(\mathbf{r}))$ is chosen so that

$$\langle \widehat{V}(N, \xi_i; \sigma(\mathbf{r})) \rangle = 0, \quad (15.7.40)$$

the transformation satisfies the lower-bound condition (15.7.9).

The possibility of achieving a lower-bound transformation with condition (15.7.40) by defining a $\widehat{V}(N, \xi_i; \sigma(\mathbf{r}))$ which moved the interaction between the microsystems of one pair of sites to another pair was first discussed by Kadanoff (1975) (see also, Kadanoff et al. 1976).

Consider the zero-field first-neighbour Ising model on a hypercubic lattice with Hamiltonian

$$\widehat{H}(N, C, K; \sigma(\mathbf{r})) := -NC - K \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma(\mathbf{r})\sigma(\mathbf{r}'), \quad (15.7.41)$$

where $K := J/T$. Now choose any two first-neighbour pairs of sites $[\mathbf{r}_1, \mathbf{r}'_1]$ and $[\mathbf{r}_2, \mathbf{r}'_2]$. Remove the interaction (or *bond*) from the first pair and add it to the bond on the second pair. Thus

$$\widehat{V}(N, K; \sigma(\mathbf{r})) = K[\sigma(\mathbf{r}_1)\sigma(\mathbf{r}'_1) - \sigma(\mathbf{r}_2)\sigma(\mathbf{r}'_2)] \quad (15.7.42)$$

and

$$\langle \widehat{V}(N, K; \sigma(\mathbf{r})) \rangle = K[\langle \sigma(\mathbf{r}_1)\sigma(\mathbf{r}'_1) \rangle - \langle \sigma(\mathbf{r}_2)\sigma(\mathbf{r}'_2) \rangle] = 0. \quad (15.7.43)$$

Of course, this result does not in itself lead to a tractable renormalization scheme. It is necessary to move bonds in a systematic way throughout the lattice so that the summation on the right-hand side of (15.7.38) can be performed, leading to a Hamiltonian $\widehat{H}(\tilde{N}, \tilde{C}, \tilde{K}; \sigma(\tilde{\mathbf{r}}))$ of the same form (15.7.41) as $\widehat{H}(N, C, K; \sigma(\mathbf{r}))$. This can be achieved using the *Migdal–Kadanoff transformation*, first invented by Migdal (1975) and reinterpreted using bond moving ideas by Kadanoff (1976a). In Kadanoff's formulation of this method bonds are moved so that sites can be eliminated by decimation.

Example 15.7.2 Consider the case of the zero-field first-neighbour Ising model on a square lattice. Taking a block of $\lambda \times \lambda$ squares, as shown in Fig. 15.7 (for the case $\lambda = 3$) and move the horizontal interior bonds to the upper edge and the vertical interior bonds to the left edge. The interior sites do not now interact with their neighbours, so they will each contribute a factor of two to the summation on the left-hand side of (15.7.38). The coupling on each segment of the boundary will be λK . These sites form a one-dimensional chain connecting the corners to which we can apply the decimation formula (15.5.33), which together with (15.5.12) gives

$$\tilde{K} = \operatorname{arctanh}\{[\tanh(\lambda K)]^\lambda\}. \quad (15.7.44)$$

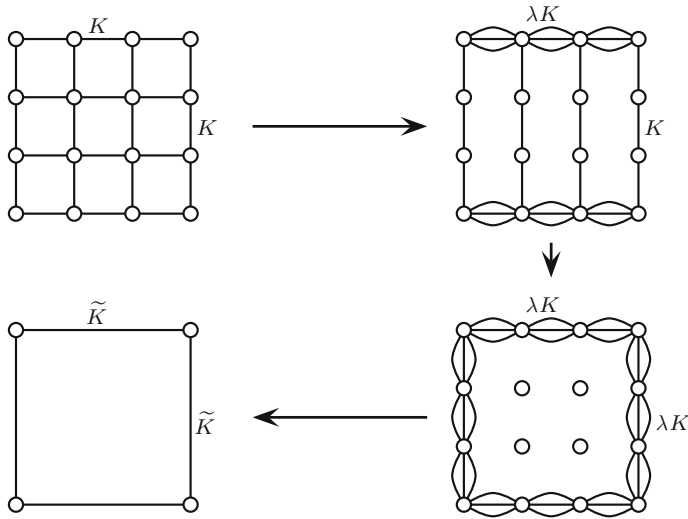


Fig. 15.7 Bond moving and decimation operations on a block of 3×3 sites for the Migdal–Kadanoff transformation applied to the zero-field first-neighbour Ising model on a square lattice

The recurrence relation

$$\tilde{K} = \operatorname{arctanh}\{[\tanh(\lambda^{d-1}K)]^\lambda\}, \quad (15.7.45)$$

obtained by Migdal (1975) for the d -dimensional hypercubic lattice is a simple generalization of (15.7.44), which, with $\lambda = 2$, can be re-expressed in the form

$$\tilde{K} = \frac{1}{2} \ln[\cosh(4K)], \quad (15.7.46)$$

giving the critical fixed point $K^* = K_c = 0.304689$ with $T_c/J = 3.282$ and $\nu = 1/y_T = \ln(2)/\ln[2 \tanh(4K_c)] = 1.3383$. Neither the critical coupling nor the exponent compare well with the exact values for the two-dimensional Ising model ($K_c = 0.4407$, $\nu = 1$, see Sect. 12.3.9) and the comparison deteriorates with increasing λ giving for K_c the values 0.2406 and 0.2018 for $\lambda = 3$ and 4 respectively.

An interesting modification of the method has been proposed by Kadanoff (1976a). He considered the anisotropic Ising model on the d -dimensional hypercubic lattice with different couplings in each axis direction. In the Migdal formulation this would simply lead to the result that each coupling satisfied (15.7.45) and the fixed points for different values of λ would have all the couplings equal. In this modified method, instead of the bond moving for all the axes preceding decimation, bond moving and decimation are performed alternately.

Example 15.7.3 Consider the case of the square lattice K_h and K_v denoting horizontal and vertical couplings. First all the internal horizontal bonds are moved to the edges of the $\lambda \times \lambda$ square as shown in the first line of Fig. 15.7. Decimation is then performed in the vertical direction after which the vertical bonds are moved to the edges of the square. Finally decimation is carried out in the horizontal direction. The recurrence relations achieved by this procedure are

$$\begin{aligned}\tilde{K}_h &= \operatorname{arctanh}\{[\tanh(\lambda K_h)]^\lambda\}, \\ \tilde{K}_v &= \lambda \operatorname{arctanh}\{[\tanh(K_v)]^\lambda\}.\end{aligned}\tag{15.7.47}$$

It is clear that, for any λ , the critical fixed point is now given by

$$K_h^* = \lambda^{-1} K_v^* = K^*,\tag{15.7.48}$$

where K^* is the fixed point value given by (15.7.44). Substituting for λ from (15.7.48) into one of the Eq. (15.7.47) gives

$$K_v^* \ln[\tanh(K_v^*)] = K_h^* \ln[\tanh(K_h^*)].\tag{15.7.49}$$

Equation (15.7.49) has branches⁶

$$K_v^* = K_h^*\tag{15.7.50}$$

and

$$\sinh(2K_v^*) \sinh(2K_h^*) = 1.\tag{15.7.51}$$

The formula (15.7.50), which is satisfied by the Migdal critical fixed point, disagrees with (15.7.48), except for $\lambda = 1$, when, in the isotropic case, (15.7.51) becomes the formula (9.2.36) which gives the exact critical temperature for the isotropic Ising model. In general (15.7.51) gives the critical curve for the anisotropic Ising model (Baxter 1982b). It is, however, a curious property of the method that the individual points on this curve are achieved by choosing different (including non-integer) values of λ . This is because, for any chosen value of λ , the couplings renormalize independently and thus can yield only a finite number of discrete points, not a continuous curve. It can be shown that both the Migdal method and the Kadanoff modification lead to the same values for the exponent ν , with the ‘best’ value being achieved in the limit $\lambda \rightarrow 1$ when $\nu = 1.326$.

⁶ Equation (15.7.50) is obvious. Equation (15.7.51), which is equivalent to $K_v^* = -\frac{1}{2} \ln[\tanh(K_h^*)]$, follows from the fact that this latter equation is invariant under interchange of K_v^* and K_h^* .

15.8 Finite-Lattice Approximations

Methods of this type apply the procedure of Sect. 15.1 to a lattice with a small number of sites with the proviso that, in order for the renormalization group transformation to be able to be applied once, it must be the case that $\tilde{N} \geq \lambda^d$ and thus $N \geq \lambda^{2d}$. Within this framework there is still the choice of weight function and of whether to apply periodic boundary conditions to the lattice. Having made these choices the recurrence relationships can be computed exactly. As has been remarked in Sect. 15.2 the approximation here consists of obtaining the fixed points and phase diagram by iterating the recurrence relationships a large number of times. Early applications of this method to the Ising model were by Nauenberg and Nienhuis (1974a, b) who used four-site blocks on a square lattice and Tjon (1974), who used four three-site blocks on a triangular lattice. The former authors used periodic boundary conditions and the latter did not. Rather than present these calculations, which are described by Niemeijer and van Leeuwen (1976), we shall use as an example the work of Schick et al. (1976, 1977).

Example 15.8.1 Here a nine-site triangular lattice $\mathcal{N}$ with periodic boundary conditions is used to study the antiferromagnetic Ising model in a magnetic field. Given that antiferromagnetism leads to sublattice ordering on the three sublattices **A**, **B** and **C** shown in Fig. 15.8, the blocks are chosen to reflect this phenomenon. As is shown in Fig. 15.8 the three sites A_1 , A_2 and A_3 of sublattice **A** form a block with block-spin site located at $\hat{A}$, with a similar scheme applying to sublattices **B** and **C**. The Hamiltonian for this model has the form

$$\begin{aligned} \hat{H}(N, C, L, K, M; \sigma) := & -NC - \sum_{\Delta} \left\{ \frac{1}{6}L(\sigma_A + \sigma_B + \sigma_C) + M\sigma_A\sigma_B\sigma_C \right. \\ & \left. + \frac{1}{2}K(\sigma_A\sigma_B + \sigma_B\sigma_C + \sigma_C\sigma_A) \right\}, \end{aligned} \quad (15.8.1)$$

where C, L and K are, as in Sect. 15.5.1, the trivial, magnetic field and first-neighbour couplings and M is a three-spin coupling. $|\mathcal{N}| = N = 9$ and the summation is over

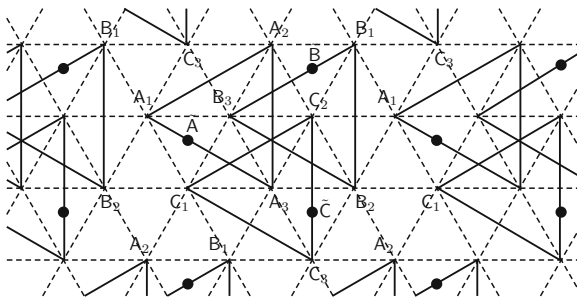


Fig. 15.8 The sublattice and block structure for the triangular lattice showing the form of the periodic boundary conditions

all twenty-four triangles of the lattice. This yields a lattice with $|\tilde{\mathcal{N}}| =: \tilde{N} = 3$ and the only triangle in the summation is $\tilde{\mathbf{A}}\tilde{\mathbf{B}}\tilde{\mathbf{C}}$, which must be counted six times because of the periodic boundary conditions. Thus, in this transformation, $\lambda = \sqrt{3}$. Defining

$$\begin{aligned} \mathfrak{Z}_0 &:= \exp(C), & \mathfrak{Z}_1 &:= \exp(L), \\ \mathfrak{Z}_2 &:= \exp(K), & \mathfrak{Z}_3 &:= \exp(M), \end{aligned} \quad (15.8.2)$$

(15.1.13) yields the four independent equations

$$[\tilde{\mathfrak{Z}}_0]^3 [\tilde{\mathfrak{Z}}_1]^3 [\tilde{\mathfrak{Z}}_2]^9 [\tilde{\mathfrak{Z}}_3]^6 = \mathfrak{Z}_0^9 \mathcal{Z}(+1, +1, +1), \quad (15.8.3)$$

$$[\tilde{\mathfrak{Z}}_0]^3 [\tilde{\mathfrak{Z}}_1]^{-3} [\tilde{\mathfrak{Z}}_2]^9 [\tilde{\mathfrak{Z}}_3]^{-6} = \mathfrak{Z}_0^9 \mathcal{Z}(-1, -1, -1), \quad (15.8.4)$$

$$[\tilde{\mathfrak{Z}}_0]^3 [\tilde{\mathfrak{Z}}_1][\tilde{\mathfrak{Z}}_2]^3 [\tilde{\mathfrak{Z}}_3]^{-6} = \mathfrak{Z}_0^9 \mathcal{Z}(+1, +1, -1), \quad (15.8.5)$$

$$[\tilde{\mathfrak{Z}}_0]^3 [\tilde{\mathfrak{Z}}_1]^{-1} [\tilde{\mathfrak{Z}}_2]^{-3} [\tilde{\mathfrak{Z}}_3]^6 = \mathfrak{Z}_0^9 \mathcal{Z}(-1, -1, +1), \quad (15.8.6)$$

where $\mathcal{Z}(\sigma_{\tilde{\mathbf{A}}}, \sigma_{\tilde{\mathbf{B}}}, \sigma_{\tilde{\mathbf{C}}})$ is the expression for the right-hand side of (15.1.13), excluding the trivial coupling, for the designated values of the block spins. These are given (Schick et al. 1977) by

$$\begin{aligned} \mathcal{Z}(+1, +1, +1) &:= \mathfrak{Z}_1^9 \mathfrak{Z}_2^{27} \mathfrak{Z}_3^{18} + 9\mathfrak{Z}_1^7 \mathfrak{Z}_2^{15} \mathfrak{Z}_3^6 + 27\mathfrak{Z}_1^5 \mathfrak{Z}_2^7 \mathfrak{Z}_3^2 \\ &\quad + 18\mathfrak{Z}_1^3 \mathfrak{Z}_2^3 \mathfrak{Z}_3^{-2} + 9\mathfrak{Z}_1^3 \mathfrak{Z}_2^3 \mathfrak{Z}_3^6, \end{aligned} \quad (15.8.7)$$

$$\begin{aligned} \mathcal{Z}(-1, -1, -1) &:= \mathfrak{Z}_1^{-9} \mathfrak{Z}_2^{27} \mathfrak{Z}_3^{-18} + 9\mathfrak{Z}_1^{-7} \mathfrak{Z}_2^{15} \mathfrak{Z}_3^{-6} + 27\mathfrak{Z}_1^{-5} \mathfrak{Z}_2^7 \mathfrak{Z}_3^{-2} \\ &\quad + 18\mathfrak{Z}_1^{-3} \mathfrak{Z}_2^3 \mathfrak{Z}_3^2 + 9\mathfrak{Z}_1^{-3} \mathfrak{Z}_2^3 \mathfrak{Z}_3^{-6}, \end{aligned} \quad (15.8.8)$$

$$\begin{aligned} \mathcal{Z}(+1, +1, -1) &:= \mathfrak{Z}_1^3 \mathfrak{Z}_2^{-9} \mathfrak{Z}_3^{-18} + 6\mathfrak{Z}_1 \mathfrak{Z}_2^{-9} \mathfrak{Z}_3^{-6} + 3\mathfrak{Z}_1^5 \mathfrak{Z}_2^3 \mathfrak{Z}_3^{-6} \\ &\quad + 9\mathfrak{Z}_1^{-1} \mathfrak{Z}_2^{-5} \mathfrak{Z}_3^{-2} + 18\mathfrak{Z}_1^3 \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^{-2} \\ &\quad + 18\mathfrak{Z}_1 \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^2 + 9\mathfrak{Z}_1 \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^{-6}, \end{aligned} \quad (15.8.9)$$

$$\begin{aligned} \mathcal{Z}(-1, -1, +1) &:= \mathfrak{Z}_1^{-3} \mathfrak{Z}_2^{-9} \mathfrak{Z}_3^{18} + 6\mathfrak{Z}_1^{-1} \mathfrak{Z}_2^{-9} \mathfrak{Z}_3^6 + 3\mathfrak{Z}_1^{-5} \mathfrak{Z}_2^3 \mathfrak{Z}_3^6 \\ &\quad + 9\mathfrak{Z}_1^1 \mathfrak{Z}_2^{-5} \mathfrak{Z}_3^2 + 18\mathfrak{Z}_1^{-3} \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^2 \\ &\quad + 18\mathfrak{Z}_1^{-1} \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^{-2} + 9\mathfrak{Z}_1^{-1} \mathfrak{Z}_2^{-1} \mathfrak{Z}_3^6. \end{aligned} \quad (15.8.10)$$

From (15.8.3)–(15.8.10),

$$\begin{aligned} \tilde{z}_0 &= \tilde{z}_0^3 [\mathcal{Z}(+1, +1, +1)\mathcal{Z}(-1, -1, -1)]^{1/24} \\ &\quad \times [\mathcal{Z}(+1, +1, -1)\mathcal{Z}(-1, -1, +1)]^{1/8}, \end{aligned} \quad (15.8.11)$$

$$\tilde{z}_1 = \left[\frac{\mathcal{Z}(+1, +1, +1)\mathcal{Z}(+1, +1, -1)}{\mathcal{Z}(-1, -1, -1)\mathcal{Z}(-1, -1, +1)} \right]^{1/8}, \quad (15.8.12)$$

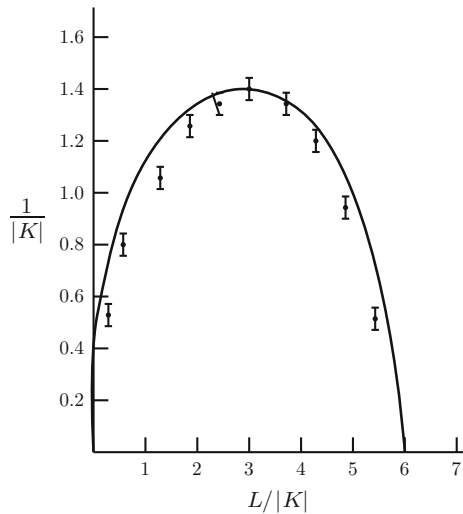
$$\tilde{z}_2 = \left[\frac{\mathcal{Z}(+1, +1, +1)\mathcal{Z}(-1, -1, -1)}{\mathcal{Z}(+1, +1, -1)\mathcal{Z}(-1, -1, +1)} \right]^{1/24}, \quad (15.8.13)$$

$$\tilde{z}_3 = \left[\frac{\mathcal{Z}(+1, +1, +1)}{\mathcal{Z}(-1, -1, -1)} \right]^{1/48} \left[\frac{\mathcal{Z}(-1, -1, +1)}{\mathcal{Z}(+1, +1, -1)} \right]^{1/16}. \quad (15.8.14)$$

These are the recurrence relationships for the model. It can be seen that there is not a consistent solution given simply by setting $\tilde{z}_3 = 1$, ($M = 0$). However, there is a solution for $\tilde{z}_2$ alone when $\tilde{z}_1 = \tilde{z}_3 = 1$. This is because of the symmetry of the terms of the Hamiltonian. For this type of transformation, once a basic set of sites (a triangle in this case) is chosen, all or none of the possible terms in the Hamiltonian corresponding to a particular symmetry must be included. The term corresponding to K is invariant under spin reversal and it is the only possible function of the spins on a triangle (apart from the trivial coupling C) with this property. The terms corresponding to L and M are the only possible combinations of spins which change sign with spin reversal and either both or neither must be included in the Hamiltonian. When they are present the recurrence relationships are invariant under the simultaneous application of $\tilde{z}_1 \leftrightarrow 1/\tilde{z}_1$, $\tilde{z}_3 \leftrightarrow 1/\tilde{z}_3$.

The results derived for this model for the antiferromagnetic case ($K < 0$) in the plane $M = 0$ are presented by Schick et al. (1976) and shown in Fig. 15.9. The transition for $L > 0$ is to the antiferromagnetic phase with ground state $(+1, +1, -1)$ and that for $L < 0$ to the antiferromagnetic phase with ground state $(-1, -1, +1)$. The results show good agreement with the Monte Carlo results of Metcalf (1973). As is well-known there is no non-zero Néel temperature on the zero-field axis and the outer termination points of the curves are at the critical field values $\mathcal{H}_c = \pm 6|J|$. The factor six arises as the coordination number of the lattice. The zero-point entropy per lattice site at $\mathcal{H} = 0$ is obtained by the method described in Sect. 15.4 and yields the value 0.324, which is very close to the exact value 0.32306 (Domb 1960, p. 213). The transitions for $L >, < 0$ are controlled respectively by the fixed points $L^* = \pm\infty$, $K^* = -\infty$, $M^* = \pm 0.113$, $|L^*| + 6K^* + 6|M^*| \sim -0.91$. The two antiferromagnetic phases are each three-fold degenerate corresponding to the three choices for the alignment of spins on the sublattices. This set of six ground states is the same as that of the 3-state antiferromagnetic Potts model (Schick and Griffiths 1977). Alexander (1975) has shown that this leads to a cubic invariant in the Landau theory

Fig. 15.9 Phase diagram for the continuous phase transition in the plane $M = 0$. The data points are the Monte Carlo results of Metcalf (1973). (Reprinted from Schick et al. (1976), by permission of the publisher Elsevier Science)



formulation of exactly the form exhibited by the 3-state Potts model (Sect. 5.4.2). This implies a first-order transition, a result which for the Potts model is confirmed by other methods (see the discussion in Sect. 5.4.2).

The physical motivation for the work of Schick et al. (1977) is to formulate a model for the order-disorder transition which occurs for gases adsorbed on monolayer films; in particular the case of helium adsorbed on graphite. While the general fit of the curves in Fig. 15.9 with experimental data for this model is good the work by Bretz (1977) for this system yielded a value $\alpha = 0.34$. The thermal exponent derived from these model calculations has the value 0.956, which from Eq. (4.5.39) gives the value $\alpha = -0.093$ for the heat capacity exponent. One might speculate that the exponent -0.26 for this fixed point is associated with a dangerous irrelevant variable (Sect. 4.8). To achieve the value $\alpha = 0.34$ it follows from (4.5.39) that the effective dimension of the system would need to be $\bar{d} = 1.587$. This in turn would imply from Eqs. (4.8.6) and (4.8.9) that the divergence associated with the correction to scaling has exponent 1.5885. This, however, is mere speculation and further calculations would be needed to settle this point. The global phase diagrams for $M \neq 0$ for both signs of K are described in detail by Schick et al. (1977). They show that a very rich phase structure is exhibited with first-order phase transition surfaces, and lines of critical, tricritical and critical end points. This renormalization scheme has been used by Schick and Griffiths (1977) for the 3-state Potts model and by Young and Lavis (1979) for a type of bonded lattice fluid which is an extension of the 3-state Potts model. This latter work was extended by Southern and Lavis (1979, 1980) to include the whole class of terms corresponding to the spin-1 model. A spin-1 model renormalized according to this method was used by Lavis et al. (1982) to model phase transitions in monolayers at air/water and oil/water interfaces and Lavis and Quinn (1983) used a spin- $\frac{1}{2}$ model to investigate ferrimagnetic ordering on a triangular lattice.

In general RSRG procedures on finite lattices give qualitatively satisfactory phase diagrams, although the values of critical exponents tend to be rather disappointing particularly for models exhibiting sublattice ordering. Although one might expect the results to improve with the choice of larger lattices, the method then becomes much more difficult to handle. In the work of Lavis and Southern (1984), which treated a spin-1 model with directional bonding on a body-centred cubic lattice of sixteen sites the recurrence relations have thirty-three thousand terms, which placed the computations near the limit of tractability at that time.

15.9 Variational Approximations

As we indicated in Sect. 15.2 a crucial factor in renormalization group calculations is the choice of weight function. Although an exact transformation would not be affected by this choice, the results of approximate transformations can change quite dramatically when the weight function is changed. In Sects. 15.5.1 and 15.5.2 and Examples 15.7.1–15.7.3 cases of the use of the decimation weight function have been explored and, in Sect. 15.8, the majority-rule weight function was employed. The possibility of using a weight function containing a parameter which can be varied to improve the approximation was considered by Kadanoff and Houghton (1975). They used the weight function (15.2.9), with parameter τ , for transformations of the spin- $\frac{1}{2}$ Ising model on the square lattice. The parameter τ was chosen by using the criterion that two different evaluations of the magnetic field exponent agreed with each other.

A form of variational approximation, designed to optimize the lower-bound condition (15.7.9) was developed by Kadanoff (1975) (see also Kadanoff et al. 1976). Known as the *one-hypercube approximation*, it is applicable to a hypercubic lattice of any dimension d . Suppose that the centres of the hypercubes of the lattice are denoted, in terms of the set of independent unit vectors $\hat{\mathbf{r}}^{(1)}, \hat{\mathbf{r}}^{(2)}, \dots, \hat{\mathbf{r}}^{(d)}$, by vectors of the form

$$\boldsymbol{\rho} := a(n^{(1)}\hat{\mathbf{r}}^{(1)} + n^{(2)}\hat{\mathbf{r}}^{(2)} + \dots + n^{(d)}\hat{\mathbf{r}}^{(d)}), \quad (15.9.1)$$

where $n^{(\ell)}$ is an integer modulo $N^{(\ell)}$, with the set of *even* integers $N^{(1)}, N^{(2)}, \dots, N^{(d)}$ satisfying (16.2.5). Let the Hamiltonian be such that all the couplings can be expressed in terms of interactions between microsystems on the sites of a single hypercube. Then

$$\hat{H}(N, \zeta_i; \sigma(\mathbf{r})) := -N\zeta_0 - \sum_{i=1}^n \zeta_i \sum_{\{\boldsymbol{\rho}\}} q_i(\boldsymbol{\rho}), \quad (15.9.2)$$

where $q_i(\boldsymbol{\rho})$ is a function of the $\sigma(\mathbf{r})$ at the sites of the cube with centre $\boldsymbol{\rho}$. We assume that the system satisfies the homogeneity condition that $\langle q_i(\boldsymbol{\rho}) \rangle$ is independent of $\boldsymbol{\rho}$.

If in (15.7.38) we choose the form

$$\widehat{V}(N, X_i(\boldsymbol{\rho}); \sigma(\mathbf{r})) := \sum_{i=1}^n \sum_{\{\boldsymbol{\rho}\}} X_i(\boldsymbol{\rho}) q_i(\boldsymbol{\rho}), \quad (15.9.3)$$

where

$$\sum_{\{\boldsymbol{\rho}\}} X_i(\boldsymbol{\rho}) = 0, \quad (15.9.4)$$

the condition (15.7.40) is satisfied, leading to the lower-bound condition (15.7.9); (15.7.42) is a particular case of (15.9.3). The quantities $X_i(\boldsymbol{\rho})$ must be chosen to achieve a tractable summation on the right-hand side of (15.7.38) with the renormalized Hamiltonian, now denoted by $\widehat{H}(\widetilde{N}, \widetilde{\xi}_i; \sigma(\widetilde{\mathbf{r}}))$, of the same form as $\widehat{H}(N, \xi_i; \sigma(\mathbf{r}))$. In the one-hypercube approximation this is achieved by dividing the hypercubes of the lattice into subsets. The ‘blue’ subset, with centres denoted by $\boldsymbol{\rho}_b$, consists of all hypercubes for which all the integers $n^{(\ell)}$ in equation (15.9.1) are odd and the ‘red’ subset with centres denoted by $\boldsymbol{\rho}_r$, consists of all hypercubes for which all the integers $n^{(\ell)}$ are even. There are $N/2^d$ hypercubes in each of the blue and red subsets and the remaining $N(1 - 2^{1-d})$ hypercubes, with centres denoted by $\boldsymbol{\rho}_g$ will be called ‘green’. It easily be checked that the choice

$$X_i(\boldsymbol{\rho}) := \begin{cases} \zeta_i - Y_i, & \text{if } \boldsymbol{\rho} = \boldsymbol{\rho}_r, \\ \zeta_i, & \text{if } \boldsymbol{\rho} = \boldsymbol{\rho}_g, \\ -(2^d - 1)\zeta_i + Y_i, & \text{if } \boldsymbol{\rho} = \boldsymbol{\rho}_b, \end{cases} \quad (15.9.5)$$

satisfies (15.9.4) for any choice of the new parameters Y_i , which are chosen to ensure that the summation on the right-hand side of (15.7.38) is tractable.

For the one-hypercube approximation applied to a spin- $\frac{1}{2}$ model, the renormalization scheme is constructed with the sites $\widetilde{\mathbf{r}} \in \widetilde{\mathcal{N}}$ coinciding with the points $\boldsymbol{\rho}_r$. The scaling parameter $\lambda = 2$, and the Kadanoff–Houghton weight function (15.2.9) is used with $\mathcal{X}(\widetilde{\mathbf{r}}) = \mathcal{B}(\widetilde{\mathbf{r}})$ taken to be the sites on the red hypercube with centre $\widetilde{\mathbf{r}} = \boldsymbol{\rho}_r$. Then

$$\prod_{\{\widetilde{\mathbf{r}}\}} \rho(\sigma(\widetilde{\mathbf{r}}) | \sigma(\mathbf{r}), \mathbf{r} \in \mathcal{B}(\widetilde{\mathbf{r}})) = \exp \left\{ \sum_{\{\widetilde{\mathbf{r}}\}} [\tau \sigma(\widetilde{\mathbf{r}}) \mathfrak{p}(\widetilde{\mathbf{r}}) - \ln(2 \cosh(\tau \mathfrak{p}(\widetilde{\mathbf{r}})))] \right\}, \quad (15.9.6)$$

where $\mathfrak{p}(\widetilde{\mathbf{r}})$ is the sum of the variables $\sigma(\mathbf{r})$ on all the sites of the red hypercube with centre $\widetilde{\mathbf{r}}$. The aim is to reduce the dependence of the argument of the summation on the right-hand side of (15.7.38) to only those terms involving the interactions between the microstates on blue hypercubes. The first term in the exponential on

the right-hand side of (15.9.6) is no problem since each site at the corner of one red hypercube is also at the corner of one blue hypercube. The second term can be ‘moved’ from red to blue hypercubes by imposing the condition

$$\sum_{i=1}^n Y_i q_i(\tilde{\mathbf{r}}) = \ln[2 \cosh(\tau \mathbf{p}(\tilde{\mathbf{r}}))], \quad (15.9.7)$$

which, when all choices are made for the states on the corners of the cube, is equivalent to 2^d linear equations for the n variables. This means, as was the case of the finite-lattice methods of Sect. 15.8, the number n of non-trivial coupling included in the Hamiltonian must be enough to ensure a consistent solution. A detailed analysis of this method for the square lattice was performed by Burkhardt (1976b) for a Hamiltonian with no magnetic field but with first-neighbour, second-neighbour and four-site couplings. The value of the parameter τ which maximized the free energy for an initial Hamiltonian equal to its critical fixed point value was 0.761. For this he obtained $T_c/J = 2.066$, $\nu = 0.983$ and $\delta = 15.36$. The recurrence relationships for this system have an invariant subspace with equal first- and second-neighbour interactions. Slightly improved results can be achieved by beginning with an initial Hamiltonian with no second-neighbour interactions and performing a preliminary exact decimation to produce a Hamiltonian within this subspace. This was the method employed by Kadanoff (1975) and Kadanoff et al. (1976). A modification of this method applicable to both square and triangular lattices was developed by Southern (1978) and for the body-centred cubic lattice by Burkhardt and Eisenriegler (1978). It has also been generalized to spin-1 models by Burkhardt (1976a) and Burkhardt and Knops (1977) and to the 3-state Potts model by Burkhardt et al. (1976).

15.10 Phenomenological Renormalization

The idea of finite-size scaling was introduced in Sect. 4.10 as an extension of the scaling methods for infinite systems treated in the rest of Chap. 4. A renormalization group justification for this has been given by Suzuki (1977) and Brézin (1982) using momentum-space methods and by Barber (1983) using real-space methods. We now discuss an RSRG method based on finite-size scaling which was developed by Nightingale (1976).

In Sect. 4.10 we described scaling for a d -dimensional system $\mathcal{L}_{\mathfrak{d}}(\aleph)$, which was infinite in $\mathfrak{d}$ dimensions and for which the size of the system for the remaining $d - \mathfrak{d}$ dimensions was represented by a parameter $\aleph$ called the thickness. The essential feature of finite-size scaling is that, for such a system with scaling fields $\theta_1, \theta_2, \dots, \theta_n$, $1/\aleph$ is treated as a scaling field θ_{n+1} . We now adopt a slight change of notation and express the correlation length $\tau(\theta_1, \dots, \theta_n, \theta_{n+1})$ for $\mathcal{L}_{\mathfrak{d}}(\aleph)$ as $\tau^{(\aleph)}(\theta_1, \dots, \theta_n)$. Equation (4.10.6) can now be used to relate the correlation lengths of two similar systems $\mathcal{L}_{\mathfrak{d}}(\aleph)$ and $\mathcal{L}_{\mathfrak{d}}(\tilde{\aleph})$ with couplings ζ_i and $\tilde{\zeta}_i$ and thicknesses $\aleph$ and $\tilde{\aleph}$ respectively.

Taking $\aleph > \tilde{\aleph}$ and making the substitution

$$\lambda = \aleph / \tilde{\aleph}, \quad (15.10.1)$$

we have

$$\tau^{(\aleph)}(\theta_1, \dots, \theta_n) = \lambda \tau^{(\tilde{\aleph})}(\tilde{\theta}_1, \dots, \tilde{\theta}_n), \quad (15.10.2)$$

where

$$\tilde{\theta}_j = \theta_j(\tilde{\zeta}_i) = \lambda^{y_j} \theta_j(\zeta_i) \quad (15.10.3)$$

relates the scaling fields θ_j and $\tilde{\theta}_j$ for $\mathcal{L}_\partial(\aleph)$ and $\mathcal{L}_\partial(\tilde{\aleph})$. Equation (15.10.2) forms the basis of Nightingale's phenomenological renormalization method. For the infinite system $\mathcal{L}$, defined by the thermodynamic limit $\mathcal{L}_\partial(\aleph) \rightarrow \mathcal{L}$, as $\aleph \rightarrow \infty$, the correlation length τ satisfies an equation similar to (15.10.2) except that the same function appears on both sides of the equation in place of the correlation functions for two systems of different thickness. If, however, we suppose that, in the limit $\aleph, \tilde{\aleph} \rightarrow \infty$ with λ fixed, the correlation lengths for the finite systems approach that of the infinite system, then (15.10.2) can be taken as an approximation to the scaling form for $\mathcal{L}$. This has the advantage, as we shall see below, that the correlation length for $\mathcal{L}_\partial(\aleph)$ can be calculated explicitly as functions of the couplings ζ_i , $i = 1, 2, \dots, n$, at least for fairly small values of $\aleph$, and we can reasonably suppose that the approximation will increase in accuracy as $\aleph$ is increased. Equation (15.10.2) can be rewritten as an expression relating the couplings for $\mathcal{L}_\partial(\aleph)$ and $\mathcal{L}_\partial(\tilde{\aleph})$. Thus

$$\tau^{(\aleph)}(\zeta_1, \dots, \zeta_n) = \lambda \tau^{(\tilde{\aleph})}(\tilde{\zeta}_1, \dots, \tilde{\zeta}_n). \quad (15.10.4)$$

If we suppose that there is a renormalization transformation of the form of (15.2.2) relating the coupling of $\mathcal{L}_\partial(\aleph)$ and $\mathcal{L}_\partial(\tilde{\aleph})$ then a fixed point will satisfy

$$\tau^{(\aleph)}(\zeta_1^*, \dots, \zeta_n^*) = \lambda \tau^{(\tilde{\aleph})}(\zeta_1^*, \dots, \zeta_n^*). \quad (15.10.5)$$

It is, however, clear that for $n > 1$ this transformation is not completely specified by (15.10.4).

A number of methods for dealing with systems with more than one coupling have been developed and will be described below. We first concentrate on the case $n = 1$, with $\zeta_1 = K := J/T$, where J is some interaction energy between microsystems. The renormalization transformation $K \rightarrow \tilde{K} = \mathcal{K}(K)$ is now defined by

$$\tau^{(\aleph)}(K) = \lambda \tau^{(\tilde{\aleph})}(\tilde{K}) \quad (15.10.6)$$

and the fixed points by

$$\mathfrak{r}^{(\aleph)}(K^\star) = \lambda \mathfrak{r}^{(\aleph)}(K^\star). \quad (15.10.7)$$

If this transformation is a satisfactory approximation to a system with a non-zero critical temperature T_c , then we expect a finite non-zero critical fixed point solution $K^\star = K_c := J/T_c$ to equation (15.10.7). The scaling field can be chosen as

$$\theta_1 = K - K^\star. \quad (15.10.8)$$

From (15.10.3) the thermal exponent $y_K = 1/\nu$ is given by

$$\lambda^{y_K} = \left(\frac{\partial \tilde{K}}{\partial K} \right)^\star \quad (15.10.9)$$

and, from (15.10.6),

$$y_K = \frac{\ln \left\{ \left(\frac{\partial \mathfrak{r}^{(\aleph)}}{\partial K} \right)^\star \left[\left(\frac{\partial \mathfrak{r}^{(\aleph)}}{\partial \tilde{K}} \right)^\star \right]^{-1} \right\}}{\ln(\lambda)} - 1. \quad (15.10.10)$$

In Chap. 11 we considered the application of the transfer matrix method to a two-dimensional system in which the thermodynamic limit is taken in one dimension. This analysis applies equally well, but with more analytic complexity, to a d -dimensional system infinite in one dimension ($\mathfrak{d} = 1$) and, subject to the assumptions described there, the correlation length for a system, for which the ground state of the ordered phase has a periodicity η , is given in terms of the ratio of the two largest eigenvalues of the transfer matrix by (11.1.36). This formula provides the starting point for most phenomenological renormalization calculations. It is clearly of importance to check the accuracy of the method for models for which the eigenvalues of the transfer method are known exactly for any width of system. This has been done for the isotropic and anisotropic square lattice spin- $\frac{1}{2}$ Ising models by Nightingale (1976) and Sneddon (1978), respectively, for the triangular lattice spin- $\frac{1}{2}$ Ising model by Kinzel and Schick (1981) and for the symmetric eight-vertex model ($K^{(\mathfrak{f})} = K^{(\mathfrak{b})}$ in Sect. 12.3) by Nightingale (1977). As an example we consider the simplest case of the isotropic square lattice spin- $\frac{1}{2}$ Ising model.

15.10.1 The Square-Lattice Ising Model

Consider a lattice of $N_1 N_2 = N$ sites with the thermodynamic limit being taken for $N_1 \rightarrow \infty$ and the thickness of the lattice $\aleph := N_2$. The complete set of eigenvalues for this model is given by Eqs. (12.2.136), (12.2.137) and (12.2.149) with $K^{(\mathfrak{h})} = K^{(\text{ch})} = K$. The largest and next largest eigenvalues correspond to taking all plus

signs with $s = +1$ and $s = -1$ respectively and in the present notation are given by⁷

$$\Lambda_1(\aleph; K) = [2 \sinh(2K)]^{\frac{1}{2}\aleph} \exp \left\{ \frac{1}{2} [\omega_1^{(\aleph)}(K) + \cdots + \omega_{2\aleph-1}^{(\aleph)}(K)] \right\} \quad (15.10.11)$$

and

$$\Lambda_2(\aleph; K) := [2 \sinh(2K)]^{\frac{1}{2}\aleph} \exp \left\{ \frac{1}{2} [\omega_0^{(\aleph)}(K) + \cdots + \omega_{2\aleph-2}^{(\aleph)}(K)] \right\}, \quad (15.10.12)$$

respectively, where

$$\omega_r^{(\aleph)}(K) := \omega(K, K; \frac{r\pi}{2\aleph}), \quad 0, 1, \dots, 2\aleph - 1, \quad (15.10.13)$$

and

$$\cosh[\omega_r^{(\aleph)}(K)] := \cosh(2K) \coth(2K) - \cos\left(\frac{r\pi}{\aleph}\right). \quad (15.10.14)$$

From (11.1.36) the correlation length is

$$\tau^{(\aleph)}(K) = -\frac{a}{\sqrt{2}} \left[\ln \left| \frac{\Lambda_2(\aleph; K)}{\Lambda_1(\aleph; K)} \right| \right]^{-1}, \quad (15.10.15)$$

where we have used the result $c(2) = \frac{1}{4}$, obtained in Sect. 5.6.3. Equations (15.10.7) and (15.10.10) can now be used, together with (15.10.15), to obtain K_c and y_K . It follows from these equations that K_c is given by the crossing point of the curves for $g^{(\aleph)}(K)$ and $g^{(\tilde{\aleph})}(K)$ plotted as functions of K , where

$$g^{(\aleph)}(K) := \left| \frac{\Lambda_2(\aleph; K)}{\Lambda_1(\aleph; K)} \right|^{\aleph}. \quad (15.10.16)$$

Nightingale (1976) considered various values of $\aleph$ and $\tilde{\aleph}$ and showed that the best results were obtained when $\tilde{\aleph} = \aleph - 1$. As may be expected the results improved with increasing $\aleph$, but even with quite small lattice thicknesses the results are remarkably good. Plots for $g^{(10)}(K)$ and $g^{(9)}(K)$ are shown in Fig. 15.10. The crossing point of the curves is at $K_c = 0.440286$ and using (15.10.10) the result $y_K = 1.0067$ is obtained for the critical exponent. The corresponding results for $\tilde{\aleph} = 49$ and $\aleph = 50$ are $K_c = 0.44068412$, $y_K = 1.000233$, the exact results being $K_c = 0.44068679$ and $y_K = 1$. A similar high level of accuracy was also achieved for the other exactly solved models in the work listed above.

⁷ We omit the factor $\ell^{(s)}(\gamma) \left(\frac{k}{2}\right)^{\frac{1}{2}N_2}$, which, as explained in Sect. 12.2.7, is an artifact of our method of derivation.

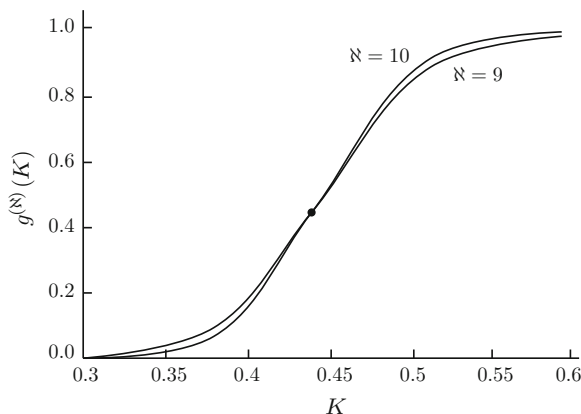


Fig. 15.10 Plots for $g^{(N)}(K)$ with $N = 9, 10$ for the spin- $\frac{1}{2}$ isotropic Ising model on a square lattice

15.10.2 Other Models

The checks on phenomenological renormalization are, of course, more limited for systems which have not been exactly solved. In some cases the critical temperature is known from the transformation methods of Chap. 9 and for the square-lattice Potts model it is given by (9.2.28). This value can also be derived from a mapping to a staggered six-vertex model (Sect. 3.8.2.1), with the value of the exponent α given by exploiting the relationship to the staggered F model. The paper of Sneddon (1978) contains, in addition to the calculations for the spin- $\frac{1}{2}$ anisotropic Ising model, results for spin values $s = 1, \frac{3}{2}, 2, \frac{5}{2}$. For quite small lattice thicknesses ($N = 5, 4, 3$, with $\tilde{N} = N - 1$) values of the critical temperature correspond quite closely with high-temperature series and lower-bound renormalization group results obtained by Burkhardt and Swendsen (1976). A series of phenomenological renormalization calculations by Nightingale and Blöte (1980), Blöte et al. (1981) and Blöte and Nightingale (1982) for the ν -state Potts model again give encouraging agreement with the exact results available. The method was applied to a triangular Ising ferrimagnetic model by Lavis and Quinn (1987). In that case the sublattice structure gives a periodicity of three rows in the ground state of the ordered phase. By taking $N = 6, \tilde{N} = 3$ they were able to obtain results for the critical temperature in various cases within 1% of the exact values obtained by Lavis and Quinn (1983), using transformation methods.

15.10.3 More Than One Coupling

We now consider the case $n > 1$. Here, as was observed above, the transformation cannot be completely specified by the one equation (15.10.4) and the fixed point equation (15.10.5) will yield an $(n-1)$ -dimensional surface in the space of couplings. The simplest method is to set $\zeta_i := \omega_i \zeta_1$ for $i = 2, 3, \dots, n$ and then carry out the renormalization procedure for ζ_1 at fixed values of $\omega_2, \omega_3, \dots, \omega_n$. This was the approach of Sneddon (1978) for the anisotropic Ising model and Kinzel and Schick (1981) for the triangular Ising antiferromagnet in a field. In the latter case (15.10.4) becomes

$$\tau^{(\mathbb{N})}(K, L) = \lambda \tau^{(\tilde{\mathbb{N}})}(\tilde{K}, \tilde{L}), \quad (15.10.17)$$

where, as for the zero-field model, $K := J/T$ and $L := \mathcal{H}/T$. If an equation of the form of (15.10.17) is used for a model with a critical temperature on the zero-field axis then the magnetic field exponent y_L can be evaluated using a procedure similar to that contained in Eqs. (15.10.9) and (15.10.10). Since however, in most situations, the correlation length is an even function of $\mathcal{H}$ (or L) it is necessary to use the second derivative, giving

$$y_L = \frac{\ln \left\{ \left(\frac{\partial^2 \tau^{(\mathbb{N})}}{\partial L^2} \right)^* \left[\left(\frac{\partial^2 \tau^{(\tilde{\mathbb{N}})}}{\partial \tilde{L}^2} \right)^* \right]^{-1} \right\}}{2 \ln(\lambda)} - \frac{1}{2}. \quad (15.10.18)$$

If, as in the case of the antiferromagnetic model of Kinzel and Schick (1981), (15.10.17) is applied by renormalizing K at fixed values of $L/K = \mathcal{H}/J$ then a fixed point corresponds to any point on the critical curve defined by

$$\tau^{(\mathbb{N})}(K, L) = \lambda \tau^{(\tilde{\mathbb{N}})}(K, L), \quad (15.10.19)$$

in the plane of couplings. The difference between this method and those based on other procedures, which allow renormalization in more than one coupling, is that here each point of the curve is a fixed point, whereas other methods produce curves containing one fixed point, which controls the whole curve. This has the consequence that the single relevant thermal exponent y_K will be a function of $\mathcal{H}/J$. There is also the problem that, if y_K is evaluated using (15.10.10), it corresponds to the exponent parallel to the L axis, which is not in general the relevant direction with respect to the critical curve. Following Kinzel and Schick (1981), (15.10.10) can be generalized to define the exponent

$$y(\hat{\mathbf{v}}) := \frac{\ln \left\{ (\nabla \tau^{(\mathbb{N})} \cdot \hat{\mathbf{v}})^* [(\nabla \tau^{(\tilde{\mathbb{N}})} \cdot \hat{\mathbf{v}})^*]^{-1} \right\}}{\ln(\lambda)} - 1 \quad (15.10.20)$$

for some arbitrary unit vector $\hat{\mathbf{v}}$ in the plane, where ‘ $\star$ ’ now simply indicates evaluation at some particularly chosen point $(K^\star, L^\star)$ on the critical curve. If $\hat{\mathbf{v}} = (1, 0)$, the unit vector parallel to the L axis, (15.10.20) reduces to (15.10.10). The vector $\mathbf{u}^\star := (\nabla \mathbf{r}^{(\aleph)} - \lambda \nabla \mathbf{r}^{(\tilde{\aleph})})^\star$ is orthogonal to the critical curve at $(K^\star, L^\star)$ and, if $\hat{\mathbf{v}}$ is tangential to the curve (and thus orthogonal to $\mathbf{u}^\star$), $y(\hat{\mathbf{v}}) = 0$. This marginal exponent is not surprising given that the whole of the critical curve consists of fixed points. For different $(K^\star, L^\star)$ the exponent $y(\hat{\mathbf{v}})$ will vary as the direction of $\hat{\mathbf{v}}$ is varied. According to the analysis of the linear renormalization group about a fixed point, described in Sect. 15.3, all scaling directions not along the critical curve should yield the same exponent. For the present method this will be the case only when $\nabla \mathbf{r}^{(\aleph)}$ and $\nabla \mathbf{r}^{(\tilde{\aleph})}$ are parallel. For this reason Kinzel and Schick (1981) chose values of $(K^\star, L^\star)$ which satisfied this condition as yielding the ‘best’ value of y_K , which is now given by

$$y_K = \frac{\ln \{ |\nabla \mathbf{r}^{(\aleph)}|^\star [|\nabla \mathbf{r}^{(\tilde{\aleph})}|^\star]^{-1} \}}{\ln(\lambda)} - 1. \quad (15.10.21)$$

With $\aleph = 9$, $\tilde{\aleph} = 6$ they obtained the value $y_K = 1.2002$ for the thermal exponent on the antiferromagnetic critical curve for the triangular Ising model. It is supposed that this transition is in the universality class of the 3-state Potts model, (Alexander 1975) for which $\alpha = \frac{1}{3}$. From (4.5.39) $y_K = 2/(2 - \alpha) = \frac{6}{5}$.

Most phenomenological renormalization calculations have concentrated on cases where the phase transitions were known, or expected, to be continuous. Exceptions to this are the ν -state Potts model calculations of Blöte et al. (1981) and Roomany and Wyld (1981). The latter authors reported that the method failed to differentiate between the second-order transition expected for $\nu \leq 4$ and the first-order transition expected for $\nu \geq 5$. Wood and Osbaldstun (1982, 1983) have argued that this inability to distinguish between continuous and first-order transitions is indeed a general characteristic of the method. They developed a general procedure based on phenomenological renormalization for determining an approximated form for phase diagrams including the location of regions of multiphase coexistence. They applied the procedure to the ν -state Potts model and it was applied to a triangular Ising ferrimagnet by Lavis and Quinn (1987).

15.11 Other Renormalization Group Methods

In this chapter we have attempted to outline the basic ideas of the RSRG and to provide a brief account of some of the main methods used. In keeping with the general approach of this book we have omitted applications to dynamic phenomena and quantum systems, for which the reader is referred respectively to the articles by Mazenko and Valls and Pfeuty et al. in Burkhardt and van Leeuwen (1982).

One approach to the statistical mechanics of lattice systems, of which limitations of space have prevented a discussion, is *Monte Carlo simulation* (see, for example, Binder 1986). The possibility of improving the results of a Monte Carlo calculations by combining them with a renormalization group procedure was first suggested by Ma (1976b). In the form developed by Swendsen (1979a, b) this *Monte Carlo renormalization method* uses a standard Monte Carlo simulation to obtain a large number of configurations near to criticality. From these correlation functions can be calculated. A renormalization group transformation can now be applied directly to each configuration as many times as is feasible. The last iteration should leave a lattice still large compared to the range of the fixed point Hamiltonian. Approximations to the matrix L^* , defined by Eq. (15.3.3), can now be calculated from correlation functions between the lattices corresponding to each stage of the renormalization procedure and a sequence of estimates of the critical exponents obtained. Swendsen (1979a) applied the method to the two-dimensional spin- $\frac{1}{2}$ Ising model and the 3-state Potts model using an real-space block-spin procedure with the majority-rule weight function. He obtained results which compared well with other renormalization group methods and were better than those obtained by Monte Carlo simulations alone. In a subsequent paper (Swendsen 1979b) he used different weight functions. This demonstrated that the accuracy and reliability of the results can be strongly affected by the choice made. This problem is one experienced by other renormalization group procedures (Sect. 15.9). The method has also been applied to a variety of other models. In particular the application to the three-dimensional Ising model by Pawley et al. (1984) gives results in good agreement with those achieved by series methods.

As indicated on p. 520, RSRG methods represent one of the two ways in which the renormalization group can be applied, the other being the use of renormalization in wave-vector space. This latter method, which involves the elimination of the short-range details of the system by integration over large wave vectors, has led to the ϵ -expansion method, (for detailed accounts see, for example, Widom 1965; Ma 1976a), where $\epsilon = 4 - d$. For the Ising model the ϵ -expansion confirms the prediction of Ginzburg–Landau theory in Sect. 5.6.3 that the upper critical dimension is at $\epsilon = 0$ ($d_u = 4$). The changeover in critical behaviour in this context can be identified with the change between relevance and irrelevance of one of the parameters at the Gaussian fixed point. Critical exponents are here obtained as expansions in powers of ϵ . For the three-dimensional Ising model the expansion to order ϵ^2 gives good agreement with other methods.

Part V
Mathematical Appendices

Chapter 16

Graphs and Lattices

16.1 Graphs

16.1.1 Introduction

A *graph* g is a set of *vertices* (points) some or all of which are connected by *edges* (lines)¹; the number of vertices of the graph g is denoted by $n(g)$ and the number of edges by $e(g)$. The number of edges incident at a vertex of g is called its *valency*. A vertex is called *odd* or *even* according to whether its valency is an odd or even number. Since each edge connects two vertices, the sum of all the vertex valencies is equal to $2e(g)$. There is, therefore, either zero or an even number of odd vertices.

A *chain* is a graph where the vertices can be ordered (possibly with repetition) in such a way that each is connected by an edge to the next vertex in the ordering.² A chain is *elementary* if each of its vertices is distinct. A *cycle* is a finite chain which begins and ends at the same vertex. A cycle is *elementary* if all its vertices are distinct,³ and its *length* is the number of its vertices or edges. An elementary cycle is a *polygon*.

A graph is *connected* if, for any pair of vertices, it contains a chain going from one to the other. A *multiply-connected* graph is one for which any two vertices are linked by at least two chains, which have no vertices in common except the end-points. A

¹ In this section we present only those elements of graph theory needed for the discussion of statistical mechanical lattice systems. For more general presentations see, for example, Berge (1962) or Wilson and Watkins (1990). In the case of graphs which are lattices (Sect. 16.2) or parts of lattices (Sect. 16.4) the vertices are usually referred to as ‘sites’ and the edges are ‘first-neighbour pairs’ or ‘second-neighbour pairs’ etc., according to whether they are represented by lines joining closest or next-closest sites. The term *bond* is also used in our discussion in Sect. 14.3 of the case where a number of bonds replace a single edge of a graph. For the moment we shall suppose that a pair of vertices is connected by at most one edge.

² Of course, if the chain is *finite* this condition does not apply to the last vertex.

³ The term ‘simple’ is sometimes used here for both chains and cycles in place of ‘elementary’. We follow the usage of Berge (1962) who uses ‘simple’ for the weaker condition which just disallows repeated edges along the ordering of vertices.

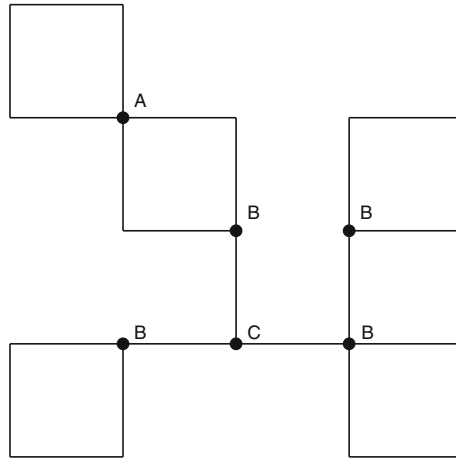


Fig. 16.1 Types of articulation point

graph which is not connected is *separated* and consists of several disjoint connected parts or *components*, where *disjoint* is defined as having no vertices in common. The number of components of the graph g is denoted by $c(g)$.

A connected graph with no cycles is called a *tree*⁴ and if all vertices have the same valency, it is a *Bethe lattice*.⁵ A Bethe lattice where each vertex is of valency two is an infinite chain.

An *articulation point* is a vertex of a connected graph such that the removal of this vertex and the edges incident there divides the graph into disjoint parts. The alternative name for an articulation point is a *cut vertex* and analogously to this there is a *cut edge* (also called a *bridge*) whose removal from the graph will also divide it into disjoint parts. A connected graph without articulation points is called a *star*. The simplest star is just a single edge (and two vertices) and it can be shown quite easily that a graph with three or more vertices is multiply connected if and only if it is a star. Every vertex of a tree is an articulation point and any graph which is neither a tree nor a star consists of a collection of stars linked by articulation points. Articulation points are of three kinds. Those at which

- A: no bridges are incident,
- B: bridges and other edges are incident.
- C: only bridges are incident.

These three types of articulation point are illustrated in Fig. 16.1.

⁴ There is unfortunately some variation in the literature in the definition of a tree. Thus Lavis and Bell (1999b, Sect. A.7.2) define a tree to be any connected graph, which is not a star. According to this definition a tree is composed of stars connected at articulation points. Then a tree with no closed cycles of edges is called a *Cayley tree*. To complicate things further the term ‘Cayley tree’ is sometimes used to signify the graph which we have called a Bethe lattice.

⁵ In the case of a finite lattice this definition is modified by excluding from the valency condition sites of valency one (sometimes called *leaves*), which terminate the lattice.

16.1.2 The Cyclomatic Number

For a finite graph g which is not a tree it is useful to be able to count the number of cycles it contains; however, this idea needs to be qualified by defining the idea of *independent cycles*. One way to do this (Berge 1962, p. 28) is to *orient* g by arbitrarily assigning, by means of an arrow, a direction to each of its edges (Sect. 16.1.4). Then for any cycle of e edges, traversed in a chosen direction an e -dimensional vector can be constructed with its i th entry equal to the number of times the cycle passes through the i th edge in the positive direction minus the number of times it passes through the edge in the negative direction. A basis set of vectors for all cycles on g represents a maximal set of independent cycles. The number of cycles in this set is $\mathcal{C}(g)$, the *cyclomatic number* of g and it can be shown (Berge 1962, p. 29)⁶ that

$$\mathcal{C}(g) = c(g) + e(g) - n(g). \quad (16.1.1)$$

Little reference is needed to the space in which these vertices and edges exist, except to say that a *planar graph* is one which can be drawn on a plane without any edges intersecting. In such a case a *face* of the graph is a region of the plane, bounded by an elementary cycle, with no vertices except those on its boundary. The subset of elementary cycles which are the boundaries of faces are called *face cycles* and the number of faces $f(g)$ of a planar graph g is equal to its cyclomatic number. For a one-component graph it follows from (16.1.1) that

$$n(g) + f(g) - e(g) = 1. \quad (16.1.2)$$

This is *Euler's relationship* (Wilson and Watkins 1990, p. 219). It can be generalized to surfaces with different topologies. Thus for a graph on the surface of a sphere the exterior of the graph is counted as an additional face giving

$$n(g) + f(g) - e(g) = 2 \quad (16.1.3)$$

and, in general,

$$n(g) + f(g) - e(g) = C, \quad (16.1.4)$$

where, for graphs with simply-connected faces covering a two-dimensional manifold of genus g , $C = 2(1 - g)$.

⁶ In fact Berge *defines* the cyclomatic number by (16.1.1) and then proves it equal to the number of independent cycles.

16.1.3 Triangulation of Graphs

Consider a one-component planar graph g bounded by an elementary cycle denoted by c . The length of c is $n(c) = e(c)$ and, denoting the number of interior (non-boundary) vertices and edges of g by $n_i(g)$ and $e_i(g)$ respectively, it follows from (16.1.2) that

$$n_i(g) + f(g) - e_i(g) = 1. \quad (16.1.5)$$

If g is *triangular* (meaning that face cycles are all of length three), every edge of c is a side of exactly one triangular face and every interior edge is a side of exactly two faces. So $3f(g) = n(c) + 2e_i(g)$ and, substituting into (16.1.5), gives

$$f(g) = n(c) + 2n_i(g) - 2. \quad (16.1.6)$$

In general g can be turned into a triangular graph g' (*triangulated*) by adding extra edges, but no extra vertices so that each face of g' is a triangle. Let c_k , $k = 1, 2, \dots, f(g)$ be the face cycles of g . It is clear that triangulation can be implemented for any plane graph and that it can, in general, be done in a number of different ways. However, from (16.1.6), it follows that the face bounded by c_k will be divided into

$$\Upsilon(c_k) := n(c_k) - 2 \quad (16.1.7)$$

faces by any process of triangulation. $\Upsilon(c_k)$ is the *triangulation number* of c_k and

$$\Upsilon(g) = \sum_{k=1}^{f(g)} [n(c_k) - 2] \quad (16.1.8)$$

is the triangulation number of g .

16.1.4 Oriented Graphs

As indicated in Sect. 16.1.2 a graph g is oriented if each of its edges is given a direction by marking it with an arrow. The face cycles c_k have a clockwise and an anticlockwise *orientation* and an arrow on a edge of c_k is said to be clockwise or anticlockwise with respect to c_k according to whether it is in the direction of the clockwise or anticlockwise orientation of c_k . Let ω_k be the number of clockwise arrows on c_k and ω the number of clockwise arrows on the boundary c of g . Any arrow which is clockwise with respect to c is also clockwise with respect to one face cycle. Any interior edge separates two of the faces of g and the arrow on this edge is clockwise with respect to one face cycle and anticlockwise with respect to another.

Thus we have, using (16.1.5),

$$\sum_{k=1}^{f(g)} \omega_k = \omega + e_i(g) = \omega + f(g) + n_i(g) - 1, \quad (16.1.9)$$

and it follows that

$$\omega = \sum_{k=1}^{f(g)} [\omega_k - 1] - n_i(g) + 1. \quad (16.1.10)$$

16.1.5 The Dual Graph

For any planar graph g a *dual graph* g^* can be formed consisting of vertices placed one in each face of g with an edge joining the vertices of g^* in adjacent faces. It is clear that the various forms (16.1.2)–(16.1.4) of Euler's relationship, with $f(g)$ replaced by $n(g^*)$, apply, with C varying according to the way that the boundary of g is handled.

Consider now a finite one-component graph g on a plane, bounded by a simple cycle c and with no dangling edges (edges terminating at one end in a vertex of valency one). This graph will have $l(g) := e(c)$ boundary edges and $e_i(g) = e(g) - e(c)$ interior edges, with a face on either side. Since a site of g^* is placed in each face of g and edges are drawn to connect sites in neighbouring faces, $n(g^*) = f(g)$, $e(g^*) = e_i(g)$ and, from (16.1.2),

$$n(g) + n(g^*) - 1 = e(g) = e(g^*) + l(g). \quad (16.1.11)$$

Now suppose g is surrounded by another simple closed curve $\bar{c}$ and $g^{(+)}$ denotes the graph formed from the union of g with edges joining it to $\bar{c}$, together with sites on $\bar{c}$ terminating these edges. Then

$$n(g^{(+)}) = n(g) + l(g), \quad e(g^{(+)}) = e(g) + l(g). \quad (16.1.12)$$

The dual graph g^* is now formed by placing a site in every face within $\bar{c}$. This increases the number of sites and edges of g^* by $l(g)$ and $2l(g)$, respectively, giving

$$n(g) + n(g^*) - 1 = e(g) + l(g) = e(g^*). \quad (16.1.13)$$

16.2 Lattices

16.2.1 Types of Regular Lattices

A regular lattice $\mathcal{L}$ in a d -dimensional space is an array of points, called the *sites* of the lattice, which has translational symmetry in all d dimensions. The number z of first-neighbours (or nearest-neighbours) to a site is called the *coordination number*.⁷ Second-neighbour and further-neighbour sites are defined in a similar way and if needed we denote the number of n th neighbours by z_n , ($z_1 = z$).

Diagrams of the one-, two- and three-dimensional regular lattices are shown in Figs. 16.2–16.10. Sites are shown by black or white circles and first-neighbour pairs are connected by line segments. The linear, square and simple cubic lattices are the $d = 1, 2, 3$ versions of a *hypercubic lattice*. The body-centred cubic lattice of Fig. 16.8 can be obtained from the simple cubic of Fig. 16.7 by placing a new site at the centre of each cube of eight sites. The face-centred cubic lattice of Fig. 16.9 can be obtained from the simple cubic by placing a new site at the centre of each square of four sites. The diamond lattice of Fig. 16.10 can be obtained from the body-centred cubic by eliminating alternate sites.

A *loose-packed (or bipartite) lattice* is one which can be divided into two equal sublattices in such a way that all the first neighbours of a site of one sublattice belong to the other sublattice; a *close-packed lattice* is one for which this is not possible. The linear, square, honeycomb, simple cubic, body-centred cubic and diamond lattices are loose-packed and the triangular, kagomé and face-centred cubic lattices are close-packed. In the diagrams of the loose-packed lattices the sites of one sublattice are denoted by $\bullet$ and the other by $\circ$. First-neighbour pairs of sites are linked by full line segments and second-neighbour $\bullet\text{--}\bullet$ pairs by broken line segments. The corresponding $\circ\text{--}\circ$ links are not shown.

The close-packed triangular and kagomé lattices are divisible into three sublattices, denoted in Figs. 16.4 and 16.6 by black, white and grey disks. Each triangular face contains a site of each sublattice. In the case of the triangular lattice the three sublattices are equivalent and any two of the three sublattices forms a honeycomb lattice. The three different sublattices of the kagomé lattice are distinguished by the orientations of the edges passing through them and the equivalence is achieved by rotation of the lattice. If, in the case of a d -dimensional lattice, one site is chosen as the origin, all the other sites are given by

$$\mathbf{r} = a^{(1)} n^{(1)} \hat{\mathbf{r}}^{(1)} + a^{(2)} n^{(2)} \hat{\mathbf{r}}^{(2)} + \dots + a^{(d)} n^{(d)} \hat{\mathbf{r}}^{(d)}, \quad (16.2.1)$$

where the $n^{(\ell)}$ range over all integer values and the $\hat{\mathbf{r}}^{(\ell)}$ are the unit lattice vectors with respective lattice spacings $a^{(\ell)}$. For a hypercubic lattice the unit lattice vectors are orthogonal and $a^{(\ell)} = a$, for $\ell = 1, 2, \dots, d$. For other regular lattices this is not the case.

⁷ For all the regular lattices, except the dice lattice (Sect. 16.2.2.1), each lattice site has the same coordination number. The dice lattice has sites of coordination numbers three and six.

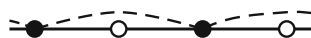


Fig. 16.2 The linear lattice; $d = 1$, $z = 2$

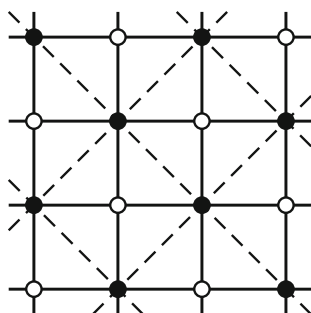


Fig. 16.3 The square lattice; $d = 2$, $z = 4$

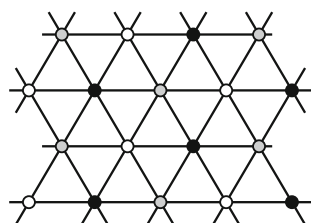


Fig. 16.4 The triangular lattice; $d = 2$, $z = 6$

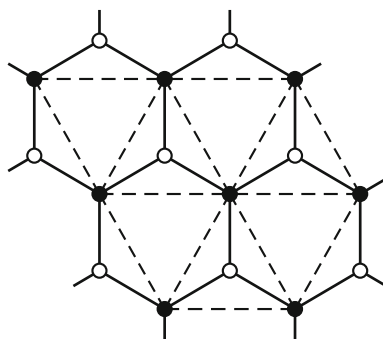


Fig. 16.5 The honeycomb; $d = 2$, $z = 3$

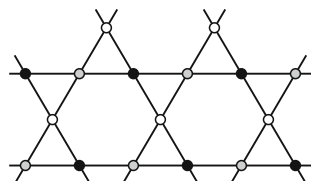


Fig. 16.6 The kagomé lattice; $d = 2$, $z = 4$

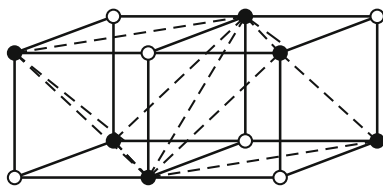


Fig. 16.7 The simple cubic lattice; $d = 3$, $z = 6$

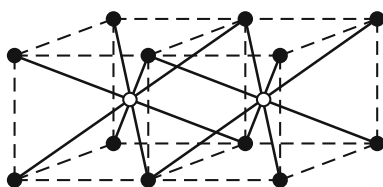


Fig. 16.8 The body-centred cubic lattice; $d = 3$, $z = 8$

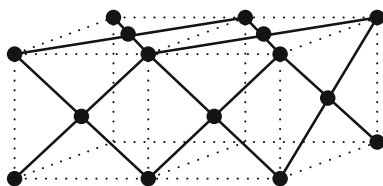


Fig. 16.9 The face-centred cubic lattice; $d = 3$, $z = 12$

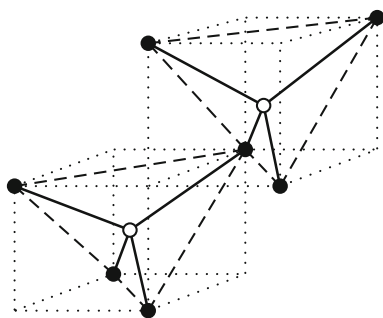


Fig. 16.10 The diamond lattice; $d = 3$, $z = 4$

Example 16.2.1 For the square lattice

$$\hat{\mathbf{r}}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{r}}^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (16.2.2)$$

and $a^{(1)} = a^{(2)} = a$.

Example 16.2.2 For the body-centred cubic lattice

$$\hat{\mathbf{r}}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{r}}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \hat{\mathbf{r}}^{(3)} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad (16.2.3)$$

and, with a the first-neighbour lattice spacing, $a^{(1)} = a^{(2)} = 2a/\sqrt{3}$, $a^{(3)} = a$.

Example 16.2.3 Although the triangular and kagomé lattices are two-dimensional the site locations are most easily expressed in terms of three lattice directions

$$\hat{\mathbf{r}}^{(1)} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \hat{\mathbf{r}}^{(2)} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \hat{\mathbf{r}}^{(3)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\sqrt{3}}{2}(\hat{\mathbf{r}}^{(1)} + \hat{\mathbf{r}}^{(2)}), \quad (16.2.4)$$

with $a^{(1)} = a^{(2)} = a^{(3)} = a$. For the triangular lattice with a site chosen as the origin, all the other sites are given by (16.2.1) with $d = 3$. For the kagomé lattice, from a given site other sites, are given in only two of the three lattice directions.

According to the discussion up to this point a lattice will have an infinite number of sites. We shall, however, want to deal with lattices with a finite number of sites. These are denoted by $\mathcal{N}$, with $|\mathcal{N}| := N$, the number of lattice sites, being an extensive variable of the system. There are in general two ways to restrict the size of a lattice. One is simply to truncate it in all dimensions to produce a finite portion of the infinite lattice. This will destroy the translational symmetry of the lattice and produce a subset $\mathcal{N}_b$ of boundary sites which no longer have the same coordination number as the bulk of the lattice. Although, under certain conditions,⁸ the effect of the boundary can be neglected in the limit of large N . Alternatively we can apply *periodic boundary conditions*. This is done by imposing the condition that $n^{(\ell)}$ in (16.2.1) is an integer modulo $N^{(\ell)}$, for $\ell = 1, 2, \dots, d$, and

$$N = \prod_{\ell=1}^d N^{(\ell)}. \quad (16.2.5)$$

The set of all vectors $\mathbf{r}_R$ which satisfy the condition

$$\exp(i\mathbf{r} \cdot \mathbf{r}_R) = 1 \quad (16.2.6)$$

form the *reciprocal lattice* $\mathcal{N}_R$ of $\mathcal{N}$ (Chaikin and Lubensky 1995, p. 45). The reciprocal lattice of a hypercubic lattice of spacing a is a hypercubic lattice of spacing $2\pi/a$. That of a triangular lattice is a triangular lattice and of a honeycomb lattice is a honeycomb lattice, but body-centred and face-centred cubic lattices form reciprocal pairs.

⁸ Essentially those is which the van Hove thermodynamic limit (Sect. 8.1) can be taken.

16.2.2 Lattice Transformations

A fundamental part of the transformation methods considered in Chaps. 9 and 10 is a transformation of the lattice $\mathcal{N} \rightarrow \mathring{\mathcal{N}}$, which can be pictured as the addition (in a translationally symmetric way) of a site on each⁹ of (i) the faces, or (ii) the edges, of $\mathcal{N}$, followed by either (a) removing the sites of $\mathcal{N}$ to leave the lattice $\mathring{\mathcal{N}}$ simply consisting of the added sites, or (b) leaving the sites of $\mathcal{N}$ in place so that $\mathring{\mathcal{N}}$ consists of all original and added sites. There are thus four pairs of choices:

- (i) + (a) when $\mathring{\mathcal{N}} := \mathcal{N}^*$ is the *dual lattice*.
- (ii) + (a) when $\mathring{\mathcal{N}} := \mathcal{N}^\oplus$ is the *medial lattice*.
- (i) + (b) when $\mathring{\mathcal{N}} := \mathcal{N} + \mathcal{N}^*$ is a *face-decorated lattice*.
- (ii) + (b) when $\mathring{\mathcal{N}} := \mathcal{N} + \mathcal{N}^\oplus$ is an *edge-decorated lattice*.

Each of these cases is now considered.

16.2.2.1 The Dual Lattice

An account of the construction of a dual graph $\mathfrak{g}^*$ for a planar graph $\mathfrak{g}$ is given in Sect. 16.1.5. In general the points of the dual of such a graph can be placed, one to each face, at any locations in the faces. However, for the regular two-dimensional lattices listed above¹⁰ the dual lattices $\mathcal{N}^*$ are also regular lattices formed by placing the sites of $\mathcal{N}^*$ at the points of coincidence of the orthogonal bisectors of the edges of $\mathcal{N}$.¹¹ It can be seen from Fig. 16.11a that the square lattice is *self-dual*, and, from Fig. 16.11b, that the dual of the honeycomb lattice is the triangular lattice and the dual of the triangular lattice is the honeycomb lattice. In each of these cases all first-neighbour pairs are connected by edges in the dual lattice. However, the dual of the kagomé lattice is the *dice lattice*. This is obtained from a triangular lattice by removing one-third of the edges to produce a mosaic of rhombi; meaning that its elementary polygons are not section graphs (Sect. 16.4.3). Although sites in adjacent faces of the kagomé lattice are first neighbours in the dice lattice, not all first neighbours in the dice lattice correspond to sites in adjacent faces of the kagomé lattice and are not connected by edges, so sites have coordination number three or six.

Let the sites of the dice lattice be separated into those originating from the centres of the hexagons of the kagomé lattice (with $z = 6$), which form a triangular lattice, and those originating from the triangles of the kagomé lattice (with $z = 3$), which form its honeycomb dual. In Sect. 16.3 we considered the graph $\mathfrak{g}_\diamond$ formed from the

⁹ Variations of this where more than one site is added to a face or edge, or a site is added to only a translationally invariant subset of the edges or faces is obviously possible. A case where sites are added to only the downward-pointing triangles of a triangular lattice is described in Sect. 16.2.2.4 and used in the star-triangle transformation of Sect. 9.4.

¹⁰ It is assumed that the lattices are finite with the cases of different boundary conditions considered below.

¹¹ The case of three-dimensional lattices is discussed below.

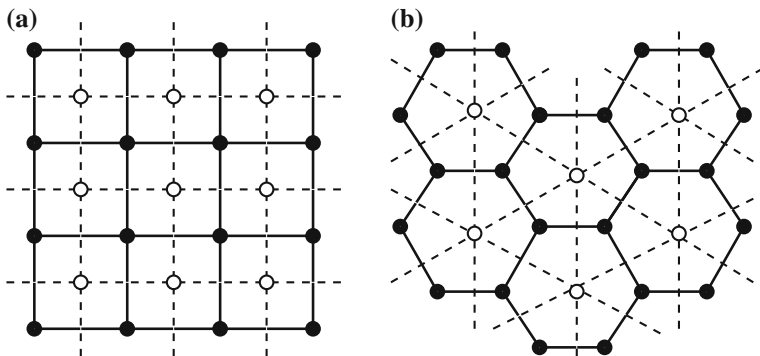


Fig. 16.11 Dual lattices: **a** square (self-dual); **b** honeycomb and triangular

vertices of a graph $\mathcal{g}$ and its dual $\mathcal{g}^*$ and the edges connecting them. In this case $\mathcal{g}_\diamond$ for the honeycomb/triangle dual pair is just the dice lattice and, since its faces are rhombi, it follows from Theorem 16.3.2 and its corollary that *rapidity lines*¹² can be drawn. It is easy to see that these simply consist of the original kagomé lattice. A similar analysis of the square lattice shows that its rapidity lattice is also a square lattice. However, the lattice $\mathcal{g}_\diamond$ formed in the same way from the kagomé/dice pair does not have a rhombic embedding so, according to Theorem 16.3.2, there is no rapidity lattice.

The number of dual lattice sites $N^* := |\mathcal{N}^*|$ is the number of faces of $\mathcal{N}$ and this is related to the number of sites N and edges N_E of $\mathcal{N}$ by Euler's relationship (16.1.4) in the form

$$N + N^* = N_E + C. \quad (16.2.7)$$

Neglecting consideration of boundary conditions, it is clear that the transformation $\mathcal{N} \rightarrow \mathcal{N}^*$ is self-inverse, in the sense that the dual of the dual of a lattice is the lattice itself; lattices form duality pairs, triangular/honeycomb, kagomé/dice, square/square. We now consider the implications of different kinds of boundary conditions. There are four possibilities:

- (I) Toroidal boundary conditions are applied and each site of $\mathcal{N}$ has coordination number z . The formula (16.2.7) is satisfied with $C = 0$ and the self-inversion of $\mathcal{N} \rightarrow \mathcal{N}^*$ is preserved. Now, apart from for the kagomé lattice, each site of $\mathcal{N}^*$ has the same coordination number, denoted by z^* , with¹³

$$N_E = \frac{1}{2}zN = \frac{1}{2}z^*N^* = N_E^*, \quad (16.2.8)$$

$$(z + z^*) = zz^*. \quad (16.2.9)$$

¹² See Sect. 16.3 for an explication of rapidity lines and lattices.

¹³ The relationship (16.2.9) between the coordination numbers of interior sites of a lattice and its dual has, of course, general validity.

For the¹⁴

- (a) self-dual square lattice $z = 4$, $N^* = N$ and $N_E = 2N$.
- (b) honeycomb/triangular dual pair $z = 3$, $z^* = 6$, $N^* = N/2$, $N_E = 3N/2$.
- (c) kagomé/dice pair $z = 4$ and the dice lattice has one-third of its sites with coordination number six and two-thirds with coordination number three, giving $N = N^*$ and $N_E = 2N$.

- (II) $\mathcal{N}$ lies in a bounded region of the plane and the exterior is counted as a face. In Fig. 16.11 this corresponds to connecting all the ‘dangling’ edges to a single ‘exterior site’ of the dual lattice. This is the case described in Sect. 16.1.5, where the topology is that of a sphere, and (16.1.3), or equivalently (16.2.7) with $C = 2$, applies. Here $N_E^* = N_E$ but, of course, the coordination number of the exterior dual-lattice site is not that of the other dual-lattice sites. The transformation $\mathcal{N} \rightarrow \mathcal{N}^*$ is still self-inverse in the sense that repeated application of the transformation produces the correct sequence of lattices (in the topological sense) and the numbers N and N^* are preserved.
- (III) $\mathcal{N}$ lies in a bounded region of the plane with the exterior not counted as a face. In Fig. 16.11 this corresponds to excluding the ‘dangling’ edges leading to the exterior from the dual lattice. The formula (16.2.7) is satisfied with $C = 1$ and

$$\frac{1}{2}zN > N_E > \frac{1}{2}z^*N^* > N_E^*. \quad (16.2.10)$$

Repeated application of $\mathcal{N} \rightarrow \mathcal{N}^*$ leads in the case of Fig. 16.11a to a succession of square lattices and in the cases of Fig. 16.11b to an alternation of honeycomb and triangular lattices, but there is a progressive reduction in the number of lattice sites.

- (IV) $\mathcal{N}$ is finite and contained within a larger lattice of the same structure. A simple cycle of boundary sites surrounds $\mathcal{N}$, with N_L edges joining $\mathcal{N}$ to the boundary sites; $\mathcal{N}^{(+)}$ denotes the lattice formed by the union of $\mathcal{N}$ with the edges joining $\mathcal{N}$ to the boundary together with their terminating boundary sites. From the discussion in Sect. 16.1.5

$$N^{(+)} = N + N_L, \quad (16.2.11)$$

$$N_E^{(+)} = N_E + N_L = \frac{1}{2}\{Nz + N_L\}. \quad (16.2.12)$$

The dual lattice $\mathcal{N}^*$ is now formed from every site lying within the boundary and, from (16.1.13),

¹⁴ Toroidal boundary conditions can be applied to the honeycomb and kagomé lattices only when N is divisible by two and three respectively and this will be assumed, where necessary, in all future discussion.

$$N + N^* - 1 = N_E^{(+)} = N_E^*. \quad (16.2.13)$$

Before leaving the topic of duality it is of interest to consider its extension to higher dimensions; in particular to the simple cubic lattice (Fig. 16.7), with $z = 6$. Two methods of designating the sites of the dual lattice suggest themselves. The first, advocated by Syozi (1972), is to place the sites of $\mathcal{N}^*$ on the faces of $\mathcal{N}$ at the locations of the sites added to $\mathcal{N}$ to transform it to a face-centred cubic lattice (Fig. 16.9). $\mathcal{N}^*$ is an octahedral lattice with $z^* = 8$ and $N^* = 3N$. Alternatively, as proposed by Savit (1980) and Chao (1982), the sites of $\mathcal{N}^*$ could be placed in the cubes of $\mathcal{N}$ at the locations of the sites added to $\mathcal{N}$ to transform it to a body-centred cubic lattice (Fig. 16.8). In this case the simple cubic lattice is self-dual. In a similar way Chao (1982) proposed a duality relationship between the face-centred cubic lattice and the diamond lattice, using these transformations to study bond percolation on the lattices, and Savit (1980) has developed a general scheme of duality for d -dimensional hypercubic lattices.

16.2.2.2 The Medial Lattice

Suppose for simplicity that toroidal boundary conditions are applied to $\mathcal{N}$.¹⁵ In Figs. 16.12 and 16.13 the sites of $\mathcal{N}$ and the medial lattice $\mathcal{N}^\circ$, for which $|\mathcal{N}^\circ| = N_E = Nz/2$, are marked as black and grey respectively. The medial lattice of a square lattice is a square lattice (Fig. 16.12) which is the rapidity lattice for the lattice and its dual, for which it is, of course, also the medial lattice (Sect. 16.2.2.1). The medial lattice for the honeycomb lattice is a kagomé lattice (Fig. 16.13), which is the rapidity lattice for the triangular/honeycomb dual pair and also the medial lattice for the triangular lattice. Forming the medial lattice of a kagomé lattice yields, if only first-neighbour lattice edges are drawn, a set of disconnected triangles; while drawing both first- and second-neighbour edges yields a lattice with triangular, hexagonal and rectangular faces. This is consonant with the fact that a rapidity lattice does not exist for the kagomé/dice dual pair (Sect. 16.2.2.1).

16.2.2.3 Edge Decoration

In the context of both edge and face decoration the sites of $\mathcal{N}$ and the decorating sites ($\mathcal{N}^\circ$ and $\mathcal{N}^*$ respectively) are referred to as *primary and secondary sites*, respectively. The decoration of the edges of a lattice can be unambiguously implemented whatever the dimension of the lattice, but $\mathcal{N} + \mathcal{N}^\circ$ may or may not be one of the regular lattice types. Both the square lattice (black sites) and its dual $\mathcal{N}^*$ are edge-decorated by the same medial (square) lattice (Fig. 16.12), but the resulting lattice in each case is not one of the regular lattice types of Sect. 16.2.1. For the primary

¹⁵ Which, as indicated above, carries the assumption for the honeycomb and kagomé lattices that N is divisible by two and three respectively.

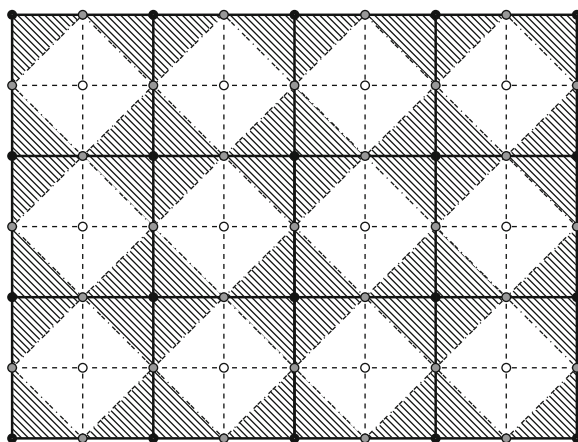


Fig. 16.12 Decoration of a square lattice, shown with *black sites and lines*. The dual square lattice is shown with *broken lines and white sites* and the square medial lattice with *chain lines and grey sites*

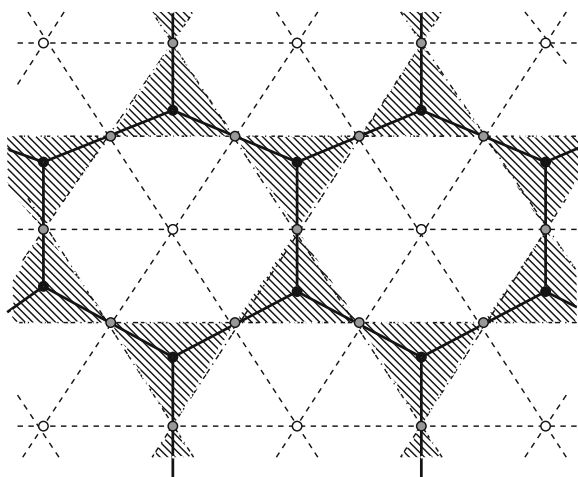
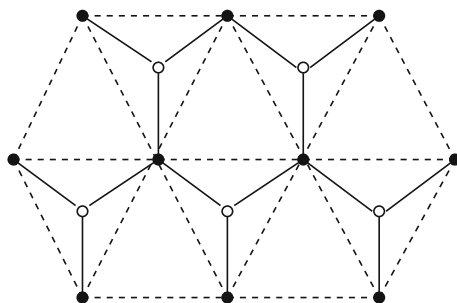


Fig. 16.13 Decoration of a honeycomb lattice, shown with *black sites and lines*. The dual triangular lattice is shown with *broken lines and white sites* and the kagomé medial lattice with *chain lines and grey sites*

honeycomb lattice (black sites) (Fig. 16.13) edge-decoration with the kagomé lattice (grey sites) does not yield a standard regular lattice. However, for the triangular lattice (white sites), edge decoration with the kagomé lattice yields a triangular lattice with four times as many sites.

Fig. 16.14 Decoration of half the faces of the triangular lattice



16.2.2.4 Face Decoration

Unlike edge-decoration, face-decoration is confined to two-dimensional lattices.¹⁶ Adding decoration sites to every face of a square lattice yields a square lattice with $2N$ sites, to every face of a triangular lattice yields a triangular lattice of $3N$ sites and to every face of a honeycomb lattice yields a triangular lattice of $3N/2$ sites.

If, as shown in Fig. 16.14, decoration sites are added to every downward-pointing triangle of a triangular lattice, and if the original edges (of broken lines) are removed and replaced by new first-neighbour edges between black and white sites, then a honeycomb lattice is produced. Also (reversing the role of primary and secondary sites in Fig. 16.13) it can be seen that face decoration of the hexagons of a kagomé lattice produces a triangle lattice.

Dedecoration involves the removal of sites and is unproblematic if it is simply the reverse of a *known* decoration procedure. However, if this is not the case then there will often be more than one way of applying dedecoration. As we have already seen dedecoration using the white sites of the triangular lattice in Fig. 16.13 will yield another triangular lattice, whereas dedecoration using the black sites will yield a kagomé lattice. It can also be seen that the decorated square lattice of Fig. 16.12 can be obtained by taking square groups of four squares of a square lattice and removing the middle site. The removal of sites from a lattice is also an important technique in real-space renormalization methods (see Chap. 15) where it is called *decimation*.

16.3 Rapidity Graphs and Lattices

Consider a plane graph $g^{\otimes}$ consisting of a finite number of lines which do not self-intersect and which intersect with each other only in pairs and at most once. Recall that, according to the definition in Sect. 16.1.1, a plane graph simply consists of

¹⁶ Although, of course, forms of decoration which are not edge-decoration can be devised and used for higher-dimensional lattices. Indeed, one might regard both the body-centered cubic lattice of Fig. 16.8 and the face-centred cubic lattice of Fig. 16.9 as decorations of the simple cubic lattice of Fig. 16.7.

vertices connected by non-crossing edges. It is clear from the construction process for $g^{\oplus}$ that it has another property beyond this basic definition, namely that a succession of edges are taken to be *parts of a single line*. For Baxter (1978) the identification of the lines is obvious since they are taken to be straight (as we have portrayed them in Fig. 16.15). However, this restriction is not necessary (Costa-Santos 2006); since each vertex has valency of four, the lines can be identified by taking each vertex as a crossing point (not a touching point) of two lines.

Definition 16.3.1 A plane graph which can be decomposed into lines which do not self-intersect or form a simple closed curve and which intersect with each other only in pairs, and at most once, is said to have the *rapidity property*, consist of *rapidity lines*, or be a *rapidity graph (or lattice)*.

Of course, it is not the case that all plane graphs with vertices of valency four are rapidity graphs. However, it must be the case that all two-dimensional regular lattices with vertices of valency (coordination number) four are rapidity lattices. As can be seen from Sect. 16.2.1 these are the square and kagomé lattices. Our interest in rapidity graphs is as hosts for eight-vertex models with arrow configurations defined as in Sect. 3.8.1. So it is clear that there is no loss of generality in representing the lines of $g^{\oplus}$ as straight; these are labelled L_i , $i = 1, 2, \dots$. No assumption has been made about the lengths of these lines and it can be assumed that the graph $g^{\oplus}$ is contained within the interior of a simple closed curve $\mathcal{C}$, which is not part of $g^{\oplus}$ and is represented in Fig. 16.15 as a circle. However we shall need to consider the graph $\mathcal{E}(g^{\oplus})$ obtained by extending, where necessary, the lines of $g^{\oplus}$ to include all possible vertices of non-parallel pairs. In the case of L_i and L_j the intersection is denoted by V_{ij} . Let *angle variables* α_{ij} and $\bar{\alpha}_{ij}$ be associated with the supplementary angles at V_{ij} , where

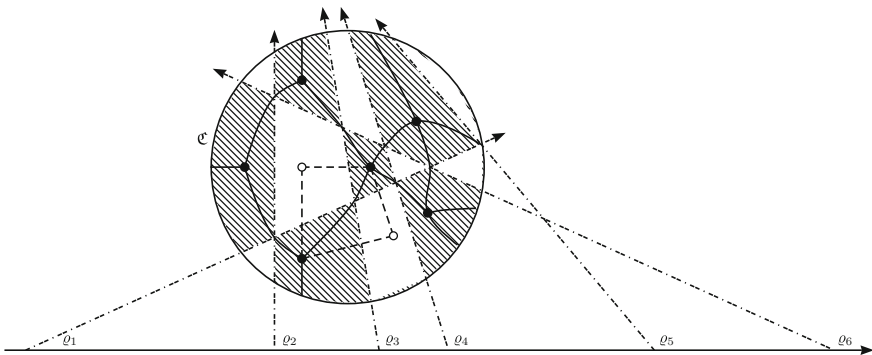


Fig. 16.15 The lines of $\mathcal{E}(g^{\oplus})$ are marked as *chain lines* and of g as continuous lines with *black vertices*. The vertices of $g^{\star}$ (only two vertices marked, *white*) are in the unshaded regions, with edges (not marked) also passing through the intersection points of the edges of $g^{\oplus}$ and g . Four edges of $g_{\diamond}$, forming a rhombus are shown (by *broken lines*)

$$\alpha_{ij} + \bar{\alpha}_{ij} = \lambda, \quad (16.3.1)$$

for some parameter λ . For any polygon c of m edges and vertices of $\mathcal{E}(\mathfrak{g}^{\otimes})$ the sum of the interior angle variables (referred to simply as the *interior sum*) is denoted by $IS(c)$. Denoting in a similar way the sum of the exterior angle variables (the *exterior sum*) by $ES(c)$, it follows, from (16.3.1), that

$$IS(c) + ES(c) = m\lambda. \quad (16.3.2)$$

A line is now needed which intersects all the lines of $\mathcal{E}(\mathfrak{g}^{\otimes})$. Such a line may already be part of $\mathcal{E}(\mathfrak{g}^{\otimes})$, but if it is not¹⁷ it must be added to produce a graph $\mathcal{EB}(\mathfrak{g}^{\otimes})$. This line will be referred to as the *base line* and will be denoted by L_0 , with the corresponding extension of the designation of angle variables to α_{i0} and $\bar{\alpha}_{i0}$, $i = 1, 2, \dots$. The triangles formed by L_0 and all other pairs of non-parallel lines are called *base triangles* and, of course, every vertex of $\mathcal{E}(\mathfrak{g}^{\otimes})$ is a vertex of a base triangle. In Fig. 16.15 a simple $\mathfrak{g}^{\otimes}$ is shown with the base line drawn horizontally at the bottom of the figure. A direction is given to the base line (to the right in Fig. 16.15), then to the lines of $\mathcal{E}(\mathfrak{g}^{\otimes})$ so that their crossings of the base line with respect to its own direction are from right to left (upwards in Fig. 16.15). The two supplementary angles at a vertex V_{ij} can now be distinguished by line direction. Refining our notation slightly we designate α_{ij} as the angle variable for the angle occupied by the heavy line in Fig. 16.16a and $\bar{\alpha}_{ij}$ as that for the angle occupied by the heavy line in Fig. 16.16b. Let $\varrho_i := \alpha_{i0}$, $i = 1, 2, \dots$. These quantities, associated each with one line of the lattice and completed by setting $\varrho_0 = 0$, are *rapidities*. Given that the interior sum of each base triangle is λ , α_{ij} for all i and j can now be expressed in terms of rapidities. If the indices of the lines (as in Fig. 16.15) increase along the base line from left to right and if the vertex V_{ij} for $i < j$ is above the base line¹⁸

$$\alpha_{ij} = \varrho_j - \varrho_i, \quad \bar{\alpha}_{ij} = \lambda - \varrho_j + \varrho_i. \quad (16.3.3)$$

If V_{ij} is below the base line i and j and the relative slopes of the two lines in Fig. 16.16a are interchanged.

The following is a pivotal result for the discussion in Chap. 12:

Theorem 16.3.1 *If the interior sum of every base triangle is λ , the interior sum of every polygon of m edges is $(m - 2)\lambda$.*

Proof For such a polygon there can be only two crossings of rapidity lines of type (a). With the graph oriented as in Fig. 16.15 with the base line horizontal, these will

¹⁷ As would be the case where every line of $\mathfrak{g}^{\otimes}$ is parallel to at least one other line.

¹⁸ It should be noted that our definitions of the variables are formulated to achieve consistency with Baxter (1978). They differ from those used in Baxter (1986) and Costa-Santos (2006). Specifically the difference involves the interchange $\alpha_{ij} \leftrightarrow \lambda - \alpha_{ij}$, corresponding to an interchange between the edge variables of the dual lattice edges passing through every site of $\mathfrak{g}^{\otimes}$. This leads to the replacement of $(m - 2)\lambda$ by 2λ in the statement of Theorem 16.3.1 (see Baxter 1986, Eq. (8.2)).

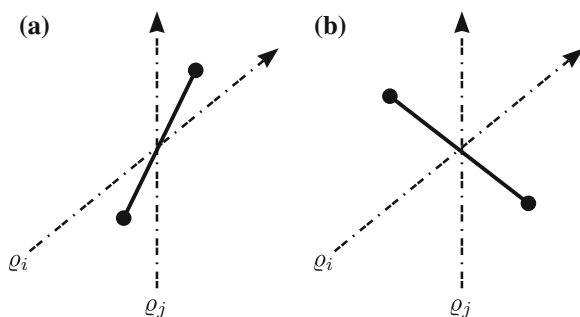


Fig. 16.16 Assignment of angle variables in terms of rapidities

be the lowest and highest vertices, representing respectively a source and a sink for the direction arrows. In the interior angle variables at the ends of a particular edge the rapidity appears with opposite signs, thus cancelling in IS(c). So when the angle variables are summed the $(m - 2)$ type (b) vertices contribute $(m - 2)\lambda$. $\square$

Rapidities are relative, in the sense that only their differences appear in the analysis. They have effectively been fixed by setting the rapidity of the base line to zero. A different choice of base line would lead to a set of rapidities differing by an additive factor.

The faces of $\mathfrak{g}^{\circledast}$ can be marked by shading or not shading so that no two shaded/unshaded faces share an edge. A point can then be placed in every shaded face, with the points joined by edges passing through the surrounding crossing points of the lines of $\mathfrak{g}^{\circledast}$. This yields a graph $\mathfrak{g}$, and repetition of the procedure for the unshaded faces yields a second graph $\mathfrak{g}^{\star}$, which is the dual of $\mathfrak{g}$. The graph $\mathfrak{g}^{\circledast}$ now comprises the set of *rapidity lines* for the dual pair $\mathfrak{g}, \mathfrak{g}^{\star}$.

The reverse of this procedure poses an interesting question, namely: given a graph $\mathfrak{g}$, is it possible to draw a set of lines which have the rapidity property and for which the intersections lie on the edges of $\mathfrak{g}$, with exactly one per edge? To answer this question the first step is to form a graph $\mathfrak{g}_{\diamond}$ whose vertices are all the vertices of $\mathfrak{g}$ and its dual $\mathfrak{g}^{\star}$ and whose edges are, for every vertex of $\mathfrak{g}^{\star}$, the lines connecting it to the vertices of $\mathfrak{g}$ on the boundary of the face in which it is situated. Clearly every face of $\mathfrak{g}_{\diamond}$ is (topologically speaking) a quadrilateral with the edges of $\mathfrak{g}$ and $\mathfrak{g}^{\star}$ as diagonals. A set of *train tracks* are now defined as sequences of faces of $\mathfrak{g}_{\diamond}$ connected along opposite edges and the following theorem can be proved (Kenyon and Schlenker 2005):

Theorem 16.3.2 *The planar graph $\mathfrak{g}_{\diamond}$ has a rhombic embedding¹⁹ in the plane if and only if*

- (i) *No train track crosses itself or forms a simple closed curve.*
- (ii) *Two distinct train tracks cross each other at most once.*

¹⁹ A *rhombic embedding* (also called a *isoradial embedding*) of quadratic graph is a continuous displacement of its vertices and edges which, while leaving all the edge/vertex connections unchanged, transforms every edge into a straight line of unit length; meaning that every face becomes a rhombus.

It is clear that to each train track there corresponds a unique curve drawn through the intersections of the diagonals of its constituent quadrilaterals and that the set of train tracks satisfy properties (i) and (ii) of Theorem 16.3.2 if and only if the set of curves are rapidity lines. Thus:

Corollary 16.3.1 *Rapidity lines, and hence a rapidity graph $\mathfrak{g}^{\otimes}$, exist for a graph $\mathfrak{g}$ if and only if the derived quadratic graph $\mathfrak{g}_{\diamond}$ has a rhombic embedding.*

Since an edge of $\mathfrak{g}$ and an edge of $\mathfrak{g}^{\star}$ passes through every vertex of $\mathfrak{g}^{\otimes}$ (as shown by the heavy lines in Fig. 16.16) they can be associated with the angle variables at the site. Given that the polygon $\mathfrak{c}$ is a face cycle, with the face containing a site of $\mathfrak{g}$ (say), the sum of the angle variables for the edges meeting at this vertex is equal to the interior sum $(m - 2)\lambda$. And, if a polygon is formed with the edges of $\mathfrak{g}^{\star}$ passing through the vertices of $\mathfrak{c}$, the sum of the angle variables for the edges of this polygon is equal to the exterior sum 2λ .

The two regular two-dimensional four-coordinated lattices with the rapidity property are the square lattice and the kagomé lattice. Each is the medial lattice (and hence denoted by $\mathcal{N}^{\otimes}$) of a dual pair of lattices $\mathcal{N}$ and $\mathcal{N}^{\star}$ with sites situated on the faces of $\mathcal{N}^{\otimes}$ (Sect. 16.2.2.2). For the square lattice, $\mathcal{N}$ and $\mathcal{N}^{\star}$ are also square lattices (tilted through 45°), with $N^{\otimes} = 2N$, $N^{\star} = N$. For the kagomé lattice, $\mathcal{N}$ and $\mathcal{N}^{\star}$ are the honeycomb and triangular members of the dual pair, with $N^{\otimes} = \frac{3}{2}N$, $N^{\star} = \frac{1}{2}N$.

16.4 Lattice Graphs

The vertices of a *lattice graph* $\mathfrak{g}$ form a subset of the sites of the lattice $\mathcal{N}$, restricted so that edges can be drawn only between first-neighbour sites.²⁰ This is said to be a *weak embedding* of $\mathfrak{g}$ in $\mathcal{N}$. It is not necessarily the case that there is an edge drawn between every pair of first-neighbour sites, but if this is so then the graph is called a *section graph*.²¹ All low-temperature graphs are section graphs (Sect. 14.2.1) and if $\mathfrak{g}$ is a section graph in $\mathcal{N}$ then this is called a *strong embedding*.²² A graph is said to be a *spanning graph of $\mathcal{N}$* if it contains every vertex of $\mathcal{N}$. Of course, a spanning section graph (strongly embedded) in $\mathcal{N}$ is just $\mathcal{N}$ itself. Examples of weakly embedded spanning graphs are the augmented graphs defined in Sect. 16.4.1 and used in Sect. 3.8.2.1 and the dimer covering of Sect. 3.9 and Chap. 13.

²⁰ Although it is convenient to allow the topological equivalence between the triangular lattice and a square lattice with edges between all second-neighbour pairs with one of the two orientations.

²¹ In some cases, like for example in the discussion of graphs on the body-centred cubic lattice in Example 16.4.7, the account given here is generalized to the case where there are two different kinds of edges, corresponding to first-neighbour and second-neighbour pairs of sites. A section graph is then one which has edges connecting all of its first-neighbour and second-neighbour pairs.

²² In this terminology, rather unfortunately, strong embeddings are subclasses of weak embeddings.

16.4.1 Augmented Graphs and the Whitney Polynomial

Suppose the lattice graph $\mathfrak{g} \subseteq \mathcal{N}$ is augmented by adding all the sites of $\mathcal{N} \setminus \mathfrak{g}$ as single site components. No additional edges are added in this process and so $e(\mathfrak{g})$ and $\mathfrak{C}(\mathfrak{g})$ are unchanged with

$$n(\mathfrak{g}) \rightarrow N, \quad c(\mathfrak{g}) \rightarrow c(\mathfrak{g}) + N - n(\mathfrak{g}). \quad (16.4.1)$$

An important quantity underlying many problems in statistical mechanics is the *Whitney polynomial* (Whitney 1932)

$$W(x, y) := \sum_{\{\mathfrak{g}\}}^{(\text{AG})} x^{c(\mathfrak{g})} y^{e(\mathfrak{g})} = \sum_{\{\mathfrak{g}\}} x^{N+c(\mathfrak{g})-n(\mathfrak{g})} y^{e(\mathfrak{g})}, \quad (16.4.2)$$

where ‘(AG)’ indicates that the summation is over all graphs on the lattice augmented in the way described here.

16.4.2 Hopping Matrices and the Canonical Flux Distribution

Let $\mathcal{N}$ be a two-dimensional lattice of N sites labelled $j = 1, 2, \dots, N$ and T be a Hermitian hopping matrix with elements satisfying (13.1.2) and (13.1.3). For convenience the length $e(\mathfrak{c}) = n(\mathfrak{c})$ of the elementary cycle $\mathfrak{c}$ on $\mathcal{N}$ (Sect. 16.1.1) is denoted by ℓ with an ordered labelling of vertices $(o_1, o_2, \dots, o_\ell)$. The *flux* $\Phi(T; \mathfrak{c})$ of T through $\mathfrak{c}$, with the orientation specified by the order of the vertices, is

$$\Phi(T; \mathfrak{c}) := \text{Arg}\{t(o_1, o_2) \cdots t(o_{\ell-1}, o_\ell) t(o_\ell, o_1)\}. \quad (16.4.3)$$

If the orientation of $\mathfrak{c}$ is reversed by specifying its vertices in the reverse order $(o_\ell, o_{\ell-1}, \dots, o_1)$ then the flux changes sign. The cycle $\mathfrak{c}$ will circumscribe $f(\mathfrak{c})$ faces circumscribed by face cycles $\mathfrak{c}_k$, $k = 1, 2, \dots, f(\mathfrak{c})$. With the orientations around the face cycles the same as that around $\mathfrak{c}$, it is obvious that

$$\Phi(T; \mathfrak{c}) = \sum_{k=1}^{f(\mathfrak{c})} \Phi(T; \mathfrak{c}_k). \quad (16.4.4)$$

Consider now a gauge transformation on the hopping matrix expressible in the equivalent forms

$$\begin{aligned} t'(i, j) &:= \exp(-i\varphi(i)) t(i, j) \exp(i\varphi(j)), \\ T' &:= U^\dagger T U, \quad \langle i | U | j \rangle := \delta^{\text{Kr}}(i - j) \exp(i\varphi(j)). \end{aligned} \quad (16.4.5)$$

It can then be shown that:

Theorem 16.4.1 *Two hopping matrices T' and T with the same hopping amplitudes have the same fluxes through all elementary cycles on $\mathcal{N}$ if and only if they are related by the gauge transformation (16.4.5).*

That the fluxes are invariant under the gauge transformation is obvious. That the identity of fluxes implies the existence of the gauge transformation was proved by Lieb and Loss (1993, p. 342) who also showed (p. 343) that:

Theorem 16.4.2 *If numbers $\Phi^{(k)}$ are assigned to the face cycles c_k of $\mathcal{N}$ then there is an assignment of real numbers $\theta(i, j)$ to the first-neighbour pairs of the lattice with $\theta(j, i) = -\theta(i, j)$ and $0 \leq |\theta(i, j)| < 2\pi$ such that the hopping matrix T with elements*

$$t(i, j) = |t(i, j)| \exp[i\theta(i, j)], \quad (16.4.6)$$

has the property

$$\Phi(T; c_k) =, \quad \forall c_k. \quad (16.4.7)$$

It follows from Theorem 16.4.1 that the numbers $\theta(i, j)$ are unique to within the gauge transformation (16.4.5).

The work of Lieb and Loss (1993) was motivated by a number of physical realizations, in each of which the ground-state energy of the system is dependent on the choice of fluxes $\{\Phi^{(k)}\}$. In particular, for what is called the *half-filled band* case, the ground-state energy is twice the sum of the negative eigenvalues of $-T$ and part of the aim of their work was to determine a flux distribution which minimised this quantity. They were able to show in some special cases (with their work completed by Lieb (1994)) that this was achieved by taking the triangulation as described in Sect. 16.1.3 and (with the same orientation) assigning a flux $\frac{1}{2}\pi$ to each triangular face. Then the flux through c_k is $\frac{1}{2}\pi \Upsilon(c_k)$, where, as defined in Sect. 16.1.3, $\Upsilon(c_k)$ is the triangulation number of c_k and is independent of the mode of triangulation. It follows from Theorem 16.4.2 that there are hopping matrices T with

$$\Phi(T; c) = \frac{1}{2}\pi \sum_{k=1}^{f(c)} \Upsilon(c_k). \quad (16.4.8)$$

Such matrices are said to satisfy the *canonical flux distribution*. Given a set of hopping amplitudes, the T satisfying (16.4.8) is unique to within the gauge transformation.

16.4.3 Embeddings and Topologies

The *topology* of a graph g comprises the number of its vertices and the way in which they are linked by edges. Two graphs with the same topology may have components

with different *shapes*. Any graph without isolated vertices can be specified uniquely by giving the location of all its edges. The denotation used for such a graph is (e, k) , where e is the number of edges and the index k is used to distinguish between shapes. Thus the one-component graph with three sites connected by two edges on the square lattice can have two shapes. These are given in Table 16.1 as $(2, 1)$, with two collinear edges, and $(2, 2)$, with two edges at right angles.²³ Graphs with the same topology and with components having the same shapes are called *isomorphic* meaning that one can be brought into coincidence with the other by some combination of translations, rotations and reflections acting independently on each component.²⁴ Graphs which are not isomorphic are said to be *distinct*. A selection of graphs on the square lattice which are useful for illustrating the discussion are given in Tables 16.1 and 16.2. The number of components and sites of (e, k) are denoted respectively by $c(e, k)$ and $n(e, k)$.

For section graphs the topology and shape is fixed once the relative location of the vertices is given. We thus are able to use an alternative notation $[n, k]$ exclusively for section graphs, where again k distinguishes distinct graphs. The number of components and edges of the section graph $[n, k]$ are denoted respectively by $c[n, k]$ and $e[n, k]$. For section graphs and low-temperature expansions a useful parameter for $[n, k]$ is

$$l[n, k] := nz - 2e[n, k], \quad (16.4.9)$$

the number of links between a graph and exterior sites on the lattice. This is included in Tables 16.1 and 16.2. From (16.1.1) and (16.4.9)

$$l[n, k] = 2c[n, k] + n(z - 2) - 2\mathcal{C}[n, k], \quad (16.4.10)$$

where, as we have noted above, the cyclomatic number $\mathcal{C}[n, k]$ of the plane lattice graph $[n, k]$ is the number of its faces.

The use of round and square brackets to specify graphs is standard notation (see, for example, Sykes et al. 1966). Of course, all section graphs which do not have isolated vertices can be given in both notations (see Tables 16.1 and 16.2).

16.4.4 Lattice Constants

The number of ways that a graph g can be placed on g' is denoted by $\Omega(g|g')$, called *the lattice constant of the graph g on the graph g'* . In general $\Omega(g|g')$ is the number of weak embeddings of g on g' . If g and g' are section graphs then the embedding is strong and we indicate this by the notation $\Omega[g|g']$.

²³ The significance of this way of labeling graphs is described below.

²⁴ The ‘shape of a graph’ is, in the case of a multi-component graph, taken to mean the shapes of all the components.

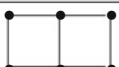
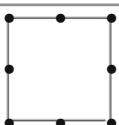
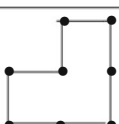
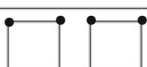
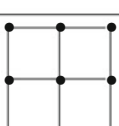
Table 16.1 Some distinct graphs on the square lattice giving both specifications (e, k) and $[n, k]$, where possible

Graph	c	e	n	(e, k)	$\overline{\omega}(e, k)$	$[n, k]$	$\overline{\omega}[n, k]$	$l[n, k]$
	1	0	1	—	—	$[1, 1]$	1	4
	2	0	2	—	—	$[2, 1]$	$\frac{1}{2}(N - 5)$	8
	1	1	2	$(1, 1)$	2	$[2, 2]$	2	6
	3	0	3	—	—	$[3, 1]$	$\frac{1}{6}(N^2 - 15N + 62)$	12
	2	1	3	—	—	$[3, 2]$	$2(N - 8)$	10
	1	2	3	$(2, 1)$	2	$[3, 3]$	2	8
	1	2	3	$(2, 2)$	4	$[3, 4]$	4	8
	1	2	4	—	—	$[4, 1]$	$2(N - 11)$	12
	1	2	4	—	—	$[4, 2]$	$4(N - 10)$	12
	2	2	4	$(2, 3)$	$2N - 7$	$[4, 3]$	$2N - 23$	12
	1	3	4	$(3, 1)$	2	$[4, 4]$	2	10
	1	3	4	$(3, 2)$	8	$[4, 5]$	8	10
	1	3	4	$(3, 3)$	4	$[4, 6]$	4	10
	1	3	4	$(3, 4)$	4	—	—	
	1	3	4	$(3, 5)$	4	$[4, 7]$	4	10
	1	4	5	$(4, 1)$	2	$[5, 1]$	2	12

The graphs are listed in a partially ordered dictionary sequence using the pair $\{e, n\}$. The designation of k is arbitrary and simply reflects the choices of displayed graphs. $l[n, k]$, which is need in Chap. 14, is the number of edges which link the section graph to external sites

When we consider the placing of a graph g on the lattice $\mathcal{N}$ itself it is clear that the lattice constant will depend not just on the number of lattice sites $N = |\mathcal{N}|$, but also on the ‘shape’ of the lattice. Different boundary conditions will lead to different lattice constants. The only useful way to resolve this problem is to assume periodic

Table 16.2 Continuation of Table 16.1

Graph	c	e	n	(e, k)	$\overline{\omega}(e, k)$	$[\mathfrak{n}, k]$	$\overline{\omega}[\mathfrak{n}, k]$	$l[\mathfrak{n}, k]$
	1	4	4	(4, 2)	1	[4, 8]	1	8
	1	5	5	(5, 2)	8	[5, 2]	8	10
	1	6	6	(6, 1)	2	—	—	
	1	7	6	(7, 1)	2	[6, 1]	2	10
	1	8	8	(8, 1)	1	—	—	
	1	8	8	(8, 2)	2	—	—	
	1	8	8	(8, 3)	4	—	—	
	2	8	8	(8, 4)	$\frac{1}{2}(N - 9)$	[8, 1]	$\frac{1}{2}(N - 21)$	16
	1	10	8	(10, 1)	2	[8, 2]	2	12
	1	12	9	(12, 1)	1	[9, 1]	1	12

boundary conditions so that each site is of valency z and the number of edges and hence the lattice constant of $[2, 2]$, a single edge connecting a first-neighbour pair of sites, is $\Omega[2, 2] = \frac{1}{2}zN$.²⁵

²⁵ We simplify the case of graphs placed on the whole lattice to $\Omega(\mathfrak{g}) := \Omega(\mathfrak{g} \mid \mathcal{N})$. In the case where $\mathfrak{g}$ is (e, k) or $[\mathfrak{n}, k]$ we denote the lattice constants (on $\mathcal{N}$) by $\Omega(e, k)$ and $\Omega[\mathfrak{n}, k]$, respectively.

We define $\bar{\omega}(e, k) := \Omega(e, k)/N$ with the *lattice constant per site* $\omega(e, k)$ being the N -independent term in $\bar{\omega}(e, k)$.²⁶ Of course, for a connected graph, $\bar{\omega}(e, k) = \omega(e, k)$. A similar notation is used for section graphs.

If a section graph without isolated vertices has alternative designations (e, k) and $[\mathfrak{n}, k']$ then $\bar{\omega}(e, k) = \bar{\omega}[\mathfrak{n}, k]$ if the graph has one component (Tables 16.1 and 16.2). However, if the graph has more than one component the strong embedding of $[\mathfrak{n}, k]$ on $\mathcal{N}$ must be such that sites of the two components cannot be joined by edges; thus $\bar{\omega}(e, k) > \bar{\omega}[\mathfrak{n}, k]$, (see graphs (2, 3) of Table 16.1 and (8, 4) of Table 16.2).

The calculation of lattice constants for small graphs can be done quite easily. As a simple example, consider all the graphs with e edges. Since the lattice has $\frac{1}{2}Nz$ edges it must be the case that

$$\sum_{\{k\}} \Omega(e, k) = \frac{(\frac{1}{2}Nz)!}{(\frac{1}{2}Nz - e)!e!} \quad (16.4.11)$$

and in a similar way, with the counting based on sites rather than edges,

$$\sum_{\{k\}} \Omega[\mathfrak{n}, k] = \frac{N!}{(N - \mathfrak{n})!\mathfrak{n}!}. \quad (16.4.12)$$

Example 16.4.1 The lattice constants for the three graphs (2, 1), (2, 2) and (2, 3), shown in Table 16.1 are easy to calculate and they provide a simple verification of (16.4.11), where the number of ways of arranging two edges on the lattice is $N(2N - 1)$. In the case of section graphs with $\mathfrak{n} = 3$ the most difficult lattice constant is for $[3, 1]$, three disconnected vertices. The easiest way to find this is from (16.4.12) which gives

$$\bar{\omega}[3, 1] = \frac{1}{6}(N - 1)(N - 2) - \bar{\omega}[3, 2] - \bar{\omega}[3, 3] - \bar{\omega}[3, 4]. \quad (16.4.13)$$

16.4.4.1 Chains

Graphs which are chains play an important part in the discussion and it is convenient to have a special notation to denote a chain of particular length irrespective of its topology on the lattice. Thus

- $(e, \mathbb{C})$ denotes a chain of e edges.
- $(e, \mathbb{E}\mathbb{C})$ denotes an elementary chain of e edges.
- $[\mathfrak{n}, \mathbb{C}]$ denotes a chain of $\mathfrak{n}$ sites which is a section graph.
- $[\mathfrak{n}, \mathbb{E}\mathbb{C}]$ denotes an elementary chain of $\mathfrak{n}$ sites which is a section graph.

²⁶ The term ‘lattice constant’ will whenever the meaning is clear be used to refer to ‘per site’ quantities.

Considering only the regular two-dimensional lattices shown in Sect. 16.2.1 and excluding the kagomé lattice (which has two types of lattice face), it is clear that all chains with $e < z^*$ are elementary. This restriction is to prevent the chain closing in on itself, which could happen whenever e is at least as great as the number of edges around a lattice face, which is the coordination number z^* of the dual lattice. In the case of chains of length z^* there is one chain that needs to be excluded from the count and this is that which passes around a lattice face (a cycle with one site designated as the two ends of the chain), which can be placed on the lattice in $N^*z^* = Nz$ ways. Thus to extend slightly the formula of Lavis and Bell (1999b, Sect. A.7.4)

$$\Omega(e, \mathbb{E}\mathbb{C}) = \begin{cases} \frac{1}{2}Nz(z-1)^{e-1}, & e < z^*, \\ \frac{1}{2}Nz(z-1)^{e-1} - Nz, & e = z^*. \end{cases} \quad (16.4.14)$$

Clearly $n(e, \mathbb{E}\mathbb{C}) = e + 1$ and the formula (16.4.14) can be modified in such a way as to apply to section graphs. The additional consideration here is to prevent a chain containing all the sites around a lattice face, which can be placed on the lattice in $N^*z^* = Nz$ ways. Such a chain is allowed with $e = z^* - 1$ in formula (16.4.14), but for section graphs must be excluded since the strong embedding condition would force the inclusion of the final edge. Thus we have

$$\Omega[n, \mathbb{E}\mathbb{C}] = \begin{cases} \frac{1}{2}Nz(z-1)^{n-2}, & n < z^*, \\ \frac{1}{2}Nz(z-1)^{n-2} - Nz, & n = z^*. \end{cases} \quad (16.4.15)$$

Example 16.4.2 The two-link chains in Table 16.1 satisfy both (16.4.14) and (16.4.15). For the three-link chains $n = z^* = 4$, so the non-section graph (3, 4) is excluded from the counting in (16.4.15).

Powerful computer programs are now available for calculating lattice constants (Oitmaa et al. 2006, Chap. 2).

16.4.5 Partially-Ordered Sequences of Graphs and the T Matrix

If g' is a subgraph of g we write $g' \subseteq g$.²⁷ The relation ' $\subseteq$ ' defines a *partial ordering* of the connected graphs of $\mathcal{N}$. Thus, for the square lattice with $n \leq 4$, a partial ordering of the connected graphs is given by the order in which they appear in

²⁷ This includes the case where g' and g are isomorphic; when this is not the case we write $g' \subset g$ and say that g' is a *proper subgraph* of g .

Table 16.1.²⁸ Let $\mathcal{S}$ be the partially-ordered sequence $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \dots, \mathbf{g}_n$, of n graphs such that if $\mathbf{g}_i \subset \mathbf{g}_j$ then $i < j$.²⁹

For any graph $\mathbf{g}$ and the sequence $\mathcal{S}$, there is a set $\Pi_n(\mathbf{g}) := \{\Omega(\mathbf{g}_i | \mathbf{g}) : i = 1, 2, \dots, n\}$ of lattice constants. It is clear that term by term the sets $\Pi_n(\mathbf{g})$ and $\Pi_n(\mathbf{g}')$, for two *non-intersecting* graphs $\mathbf{g}$ and $\mathbf{g}'$, satisfy the extensive property

$$\Omega(\mathbf{g}_i | \mathbf{g} \cup \mathbf{g}') = \Omega(\mathbf{g}_i | \mathbf{g}) + \Omega(\mathbf{g}_i | \mathbf{g}') \quad (16.4.16)$$

and this is represented by

$$\Pi_n(\mathbf{g} \cup \mathbf{g}') = \Pi_n(\mathbf{g}) + \Pi_n(\mathbf{g}'). \quad (16.4.17)$$

Now suppose that $\mathcal{E}$ is a function defined on the graphs of $\mathcal{N}$ with the extensive property

$$\mathcal{E}(\mathbf{g} \cup \mathbf{g}') = \mathcal{E}(\mathbf{g}) + \mathcal{E}(\mathbf{g}'). \quad (16.4.18)$$

If it is further supposed that the dependence of $\mathcal{E}$ on $\mathbf{g}$ is determined explicitly by the set $\Pi_n(\mathbf{g})$, that is $\mathcal{E}(\mathbf{g}) := E(\Pi_n(\mathbf{g}))$, then, from (16.4.17) and (16.4.18),

$$E(\Pi_n(\mathbf{g}) + \Pi_n(\mathbf{g}')) = E(\Pi_n(\mathbf{g})) + E(\Pi_n(\mathbf{g}')). \quad (16.4.19)$$

We now use a result proved by Sykes et al. (1966, Theorem 3):

Theorem 16.4.3 *If $\mathcal{E}(\mathbf{g}) := E(\Pi_n(\mathbf{g}))$ satisfies (16.4.19) then*

$$\mathcal{E}(\mathbf{g}) = \sum_{i=1}^n \theta_i \Omega(\mathbf{g}_i | \mathbf{g}), \quad (16.4.20)$$

for some set of parameters $\{\theta_i\}$.³⁰

Now applying (16.4.20) to each of the graphs in the sequence $\mathcal{S}$

$$\mathcal{E}(\mathbf{g}_j) = \sum_{i=1}^n \theta_i \Omega_{ij}, \quad (16.4.21)$$

where $\Omega_{ij} := \Omega(\mathbf{g}_i | \mathbf{g}_j)$. In matrix form this is

$$[\mathcal{E}^{(n)}]^\top = [\boldsymbol{\theta}^{(n)}]^\top \boldsymbol{\Omega}^{(n)}, \quad (16.4.22)$$

²⁸ Permutations of the last five lines of the table give other partial orderings.

²⁹ We discuss the details of methods of generating such a sequence below.

³⁰ In the case where $\mathcal{E}(\mathbf{g})$ is a function of other variables, for example those of a thermodynamic system, then that dependence is carried over into the set $\{\theta_i\}$.

where

$$\boldsymbol{\varepsilon}^{(n)} := \begin{pmatrix} \mathcal{E}(\mathfrak{g}_1) \\ \mathcal{E}(\mathfrak{g}_2) \\ \vdots \\ \mathcal{E}(\mathfrak{g}_n) \end{pmatrix}, \quad \boldsymbol{\theta}^{(n)} := \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix} \quad (16.4.23)$$

and the triangular array

$$\boldsymbol{\Omega}^{(n)} := \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1n-1} & \Omega_{1n} \\ 0 & \Omega_{22} & \cdots & \Omega_{2n-1} & \Omega_{2n} \\ 0 & 0 & \cdots & \Omega_{3n-1} & \Omega_{3n} \\ 0 & 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \ddots & 0 & \Omega_{nn} \end{pmatrix} \quad (16.4.24)$$

is the *T matrix* of Domb (1974b), where, of course, the diagonal elements $\Omega_{ii} = 1$.

For the lattice $\mathcal{N}$, $\Pi_n(\mathcal{N}) := \{N\omega_i : i = 1, 2, \dots, n\}$, where $\omega_i := \omega(\mathfrak{g}_i)$, and, from (16.4.21) and (16.4.22),

$$\mathcal{E}(\mathcal{N}) = N[\boldsymbol{\theta}^{(n)}]^\top \boldsymbol{\omega}^{(n)} = N[\boldsymbol{\varepsilon}^{(n)}]^\top [\boldsymbol{\Omega}^{(n)}]^{-1} \boldsymbol{\omega}^{(n)} = N[\boldsymbol{\varepsilon}^{(n)}]^\top \boldsymbol{\chi}^{(n)}, \quad (16.4.25)$$

where

$$\boldsymbol{\omega}^{(n)} = \boldsymbol{\Omega}^{(n)} \boldsymbol{\chi}^{(n)} \quad (16.4.26)$$

and

$$\boldsymbol{\omega}^{(n)} := \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{pmatrix}, \quad \boldsymbol{\chi}^{(n)} := \begin{pmatrix} \chi_1^{(n)} \\ \chi_2^{(n)} \\ \vdots \\ \chi_n^{(n)} \end{pmatrix}. \quad (16.4.27)$$

Finally from (16.4.25)

$$\mathcal{E}(\mathcal{N}) = N \sum_{i=1}^n \mathcal{E}(\mathfrak{g}_i) \chi_i^{(n)}. \quad (16.4.28)$$

Determination of the *weights* $\chi_j^{(n)}$, $j = 1, 2, \dots, n$ of the sequence $\mathcal{S}$, defined by (16.4.26) and (16.4.27), play an important part in the use of the KHDeB method of Sect. 7.2, the finite-cluster method of Sect. 14.4 and the finite-lattice method of Sect. 14.5.

Two things are obvious about the set of weights: (i) $\chi_n^{(n)} = \omega_n$, and (ii) if members of $\mathcal{S}$ for which $\chi_j^{(n)} = 0$ are removed and the remaining weights recalculated then they remain unchanged. These properties can be confirmed for Example 16.4.4 with the weights recalculated for the sequence g_3, g_5, g_6 . However, neither of these properties is of significant practical use in calculating weights. It is of greater importance to know whether a sequence $\mathcal{S}$ is *complete*. This is defined in the following way:

Definition 16.4.1 The partially-ordered sequence $\mathcal{S}$ of graphs $g_1, g_2, \dots, g_n$ is complete if there does not exist a graph g' , which is not a member of $\mathcal{S}$, is a proper subgraph of at least one member of $\mathcal{S}$, and which when added to the sequence has a non-zero weight.

Of course, it is clear that not all the weights of a complete sequence are necessarily non-zero.

16.4.6 Generating the Partially-Ordered Sequence

Partially-ordered sequences of graphs together with the formalism of Sect. 16.4.5 are used in the discussion of the finite-cluster method of Sect. 14.4, the KHDeB cluster-variation method described in Sect. 7.2 and the finite-lattice method of Sect. 14.5. In the first of these the sequence of graphs consists simply of all the subgraphs of a particular graph and there is no restriction to section graphs. In the case of the KHDeB and finite-lattice methods the graphs are section graphs and the modes of generation of the sequences are both special cases of the *overlap procedure* defined in the following way:

- (i) Choose an arbitrary subset $\mathfrak{S}$ of m connected section subgraphs of $\mathcal{N}$, none of which is a proper subgraph of any other.
- (ii) Find all distinct subgraphs obtained by overlapping the members of $\mathfrak{S}$, then the overlap graphs of these overlaps and so on until the process is exhausted. All the graphs of this sequence will be section graphs. If any are not connected then included in the sequence are all of their components.
- (iii) Let the number of members of $\mathcal{S}$ (including the members of $\mathfrak{S}$) be n and label them as specified in Sect. 16.4.5 for a partially-ordered sequence.

Then the following theorem is of importance:

Theorem 16.4.4 A partially-ordered sequence generated by the overlap procedure is complete.

Proof Let g' be a graph which is a proper subgraph of at least one member of $\mathcal{S}$. Then there exists a $g_k \in \mathcal{S}$ such that $g' \subset g_k$ but $g' \not\subset g_j$, for all $j < k$. This means that g' is not contained in the overlap of copies of g_k and³¹

³¹ The use of both round and square brackets in this equation indicates that, whereas the members of $\mathcal{S}$ are section graphs, this has not been assumed for g' .

$$\begin{aligned}\Omega(\mathfrak{g}' | \mathfrak{g}_j) &= \Omega(\mathfrak{g}' | \mathfrak{g}_k) \Omega[\mathfrak{g}_k | \mathfrak{g}_j], & \forall \mathfrak{g}_j \in \mathcal{S}, \\ \omega(\mathfrak{g}') &= \Omega(\mathfrak{g}' | \mathfrak{g}_k) \omega[\mathfrak{g}_k].\end{aligned}\tag{16.4.29}$$

The line of the weight-generating Eq. (16.4.26) for $\mathfrak{g}_k$ is

$$\chi_k^{(n)} + \Omega[\mathfrak{g}_k | \mathfrak{g}_{k+1}] \chi_{k+1}^{(n)} + \cdots + \Omega[\mathfrak{g}_k | \mathfrak{g}_n] \chi_n^{(n)} = \omega[\mathfrak{g}_k].\tag{16.4.30}$$

If $\mathfrak{g}'$ were added to $\mathcal{S}$ and had weight $\chi(\mathfrak{g}')$, the relevant line in (16.4.26), from (16.4.26), would be

$$\begin{aligned}\chi(\mathfrak{g}') + \Omega(\mathfrak{g}' | \mathfrak{g}_k) \{ \chi_k^{(n)} + \Omega[\mathfrak{g}_k | \mathfrak{g}_{k+1}] \chi_{k+1}^{(n)} + \cdots + \Omega[\mathfrak{g}_k | \mathfrak{g}_n] \chi_n^{(n)} \} \\ = \Omega(\mathfrak{g}' | \mathfrak{g}_k) \omega[\mathfrak{g}_k]\end{aligned}\tag{16.4.31}$$

and it follows, from (16.4.30), $\chi(\mathfrak{g}') = 0$; $\mathcal{S}$ is complete. $\square$

A general procedure for obtaining weights, when the sequence consists of rectangular blocks generated by some overlapping procedure on the square lattice, was given by Enting (1978a). The first step is to count the two graphs arising from the vertical and horizontal orientations of an oblong block separately. Now $\omega_j = 1$ for all $j = 1, 2, \dots, 9$. Suppose a rectangular block of x horizontal sites by y vertical sites on the square lattice is denoted by $[x \times y]$. Then it is not difficult to see that the elements of the T matrix $\mathbf{\Omega}^{(n)}$ are given by

$$\langle a \times b | \mathbf{\Omega}^{(n)} | x \times y \rangle = \begin{cases} (x - a + 1)(y - b + 1), & \text{if } a \leq x \text{ and } b \leq y, \\ 0, & \text{otherwise.} \end{cases}\tag{16.4.32}$$

It can then be shown (Enting 1978a) that

$$\langle x \times y | [\mathbf{\Omega}^{(n)}]^{-1} | a \times b \rangle = \eta(x - a) \eta(y - b),\tag{16.4.33}$$

where

$$\eta(p) = \delta^{\text{Kr}}(p) - 2\delta^{\text{Kr}}(p + 1) + \delta^{\text{Kr}}(p + 2).\tag{16.4.34}$$

From (16.4.26),

$$\langle x \times y | \mathbf{\chi}^{(n)} = \sum_{\{[a \times b]\}} \eta(x - a) \eta(y - b),\tag{16.4.35}$$

where the summation is over all the blocks $[a \times b]$ in the sequence. As we shall now see the method of evaluation of the right-hand side of (16.4.35) depends on the choice of $\mathcal{S}$.

16.4.6.1 KHDeB Sequences

The simplest application of the overlap procedure for generating a partially-ordered sequence $\mathcal{S}$ is that used in the KHDeB method and also in the finite-lattice series method of Enting and Baxter (1977). Here $m = 1$ and $\mathfrak{S}$ consists of one member.

Example 16.4.3 For the triangular lattice with $\mathfrak{S}$ taken as a single triangle the sequence $\mathcal{S}$ is

g_1 : a single site, g_2 : a single edge, g_3 : a single triangle.

Then

$$\mathbf{\Omega}^{(3)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{\omega}^{(3)} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}. \quad (16.4.36)$$

Solving (16.4.26) for (16.4.36) gives

$$\chi_1^{(3)} = 1, \quad \chi_2^{(2)} = -3, \quad \chi_3^{(3)} = 2. \quad (16.4.37)$$

Example 16.4.4 For the square lattice, with $\mathfrak{S}$ the single square [4, 8],³² $\mathcal{S}$ is

$g_1 := [1, 1]$: a single site,
 $g_2 := [2, 2]$: a first-neighbour pair,
 $g_3 := [4, 8]$: a first-neighbour square

and

$$\mathbf{\Omega}^{(3)} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{\omega}^{(3)} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}. \quad (16.4.38)$$

Solving (16.4.26) for (16.4.38) gives

$$\chi_1^{(3)} = 1, \quad \chi_2^{(3)} = -2, \quad \chi_3^{(6)} = 1. \quad (16.4.39)$$

If $\mathfrak{S}$ is the square block of four first-neighbour squares [9, 1], $\mathcal{S}$ is extended by adding

$g_4 := [3, 3]$: three co-linear first-neighbour sites,
 $g_5 := [6, 1]$: a block of two first-neighbour squares,
 $g_6 := [9, 1]$: the generating block of four first-neighbour squares.

giving

³² In all the following discussion graphs on the square lattice are referred to by their designation $[n, k]$ as given in Tables 16.1 and 16.2.

$$\Omega^{(6)} = \begin{pmatrix} 1 & 2 & 4 & 3 & 6 & 9 \\ 0 & 1 & 4 & 2 & 7 & 12 \\ 0 & 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 & 2 & 6 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \omega^{(6)} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}. \quad (16.4.40)$$

Solving (16.4.26) for (16.4.40) gives

$$\chi_1^{(6)} = 0, \quad \chi_2^{(6)} = 0, \quad \chi_3^{(6)} = 1, \quad \chi_4^{(6)} = 0, \quad \chi_5^{(6)} = -2, \quad \chi_6^{(6)} = 1. \quad (16.4.41)$$

It should be noted that the sequence generated by some $\mathbf{g}_j \in \mathcal{S}$ for $j < n$ does not necessarily contain all $\mathbf{g}_i \in \mathcal{S}$ for $i \leq j$. Here for example, $\mathbf{g}_1$ and $\mathbf{g}_2$ will be generated by $\mathbf{g}_3$ or $\mathbf{g}_4$, but $\mathbf{g}_3$ will not be generated by $\mathbf{g}_4$.

In the special case of (16.4.35) where the sequence is derived by choosing $\mathfrak{S}$ with one member $[k_h \times k_v]$ the summations can be taken independently over the ranges $a = 1, 2, \dots, k_h$ and $b = 1, 2, \dots, k_v$. From (16.4.34) and (16.4.35),

$$\sum_{m=1}^k \eta(q-m) = \delta^{\text{Kr}}(q-k) - \delta^{\text{Kr}}(q-k+1), \quad (16.4.42)$$

$$\begin{aligned} \langle x \times y | \chi^{(n)} = [\delta^{\text{Kr}}(x-k_h) - \delta^{\text{Kr}}(x-k_h+1)] \\ \times [\delta^{\text{Kr}}(y-k_v) - \delta^{\text{Kr}}(y-k_v+1)]. \end{aligned} \quad (16.4.43)$$

It follows that there are four non-zero weights. In the list given by (16.4.41) two weights are combined in $\chi_5^{(6)}$ and, from (16.4.43),

$$\chi_5^{(6)} = \langle 2 \times 3 | \chi^{(n)} + \langle 3 \times 2 | \chi^{(n)} = -2. \quad (16.4.44)$$

16.4.6.2 The De Neef–Enting Sequence

A more complicated initial set of graphs is used with the overlap method to generate the partially-ordered sequence of graphs in the finite-lattice series method described in Sect. 14.5. This is a procedure specifically intended to be used for the square lattice, although it can be adapted for the kagomé, honeycomb and triangular lattices (Sect. 14.5). The set $\mathfrak{S}$ consists of all rectangular blocks of sites with $x + y = \ell$, for some specified integer ℓ , and then $\mathcal{S}$ includes all rectangular graphs with $x + y \leq \ell$. Equations (16.4.33)–(16.4.35) still apply but the summations in (16.4.35) are no longer independent. From (16.4.34) and (16.4.35),

$$\begin{aligned}
\sum_{[a,b]} \eta(x-a)\eta(y-b) &= \sum_{a=1}^{\ell-1} \sum_{b=1}^{\ell-a} \eta(x-a)\eta(y-b) \\
&= \sum_{a=1}^{\ell-1} \eta(x-a)[\delta^{\text{Kr}}(y-\ell+a) - \delta^{\text{Kr}}(y-\ell+a+1)] \\
&= \eta(x+y-\ell) - \eta(x+y-\ell+1), \tag{16.4.45}
\end{aligned}$$

$$\begin{aligned}
(x \times y | \mathbf{x}^{(n)}) &= \delta^{\text{Kr}}(x+y-\ell) - 3\delta^{\text{Kr}}(x+y-\ell+1) \\
&\quad + 3\delta^{\text{Kr}}(x+y-\ell+2) - \delta^{\text{Kr}}(x+y-\ell+3). \tag{16.4.46}
\end{aligned}$$

All rectangles with the same value of $x + y$ have the same weight which is zero unless

$$\begin{aligned}
x + y = \ell, & \quad \text{when} \quad \chi_i^{(n)} = +1, \\
x + y = \ell - 1, & \quad \text{when} \quad \chi_i^{(n)} = -3, \\
x + y = \ell - 2, & \quad \text{when} \quad \chi_i^{(n)} = +3, \\
x + y = \ell - 3, & \quad \text{when} \quad \chi_i^{(n)} = -1.
\end{aligned} \tag{16.4.47}$$

Example 16.4.5 With $\ell = 6$ the set $\mathfrak{S}$ consists of the block $[3 \times 3]$, which is the graph $[9, 1]$ used to generate the sequence in Example 16.4.4, together with blocks $[4 \times 2]$, $[2 \times 4]$, $[5 \times 1]$ and $[1 \times 5]$. Using this set the complete sequence with designations $\mathfrak{g}_i$, $[x \times y]$ and $[n, k]$ and with the values of $\chi_i^{(n)}$ given by (16.4.45) is listed in Table 16.3.

16.4.7 Incorporating Sublattices

As indicated in Sect. 16.2.1 the loose-packed (bipartite) lattices can be divided into two equal sublattices α, β so that each first-neighbour pair consists of one α site and one β site. For models based on sublattice order, like the antiferromagnetic models described in Sect. 3.5.1.1, it is necessary to modify the above analysis in order to distinguish isomorphic graphs which differ in relation to the two sublattices. Thus, for the graphs of Example 16.4.4, when the square lattice (Fig. 16.3) is divided into two equivalent sublattices α and β , the single-site graph classification $\mathfrak{g}_1$ needs to be split into two according to whether the site is an α or a β , $\mathfrak{g}_4$ needs to be split into two according to whether the centre site is an α or a β and $\mathfrak{g}_6$ need to be split into two according to whether the centre and corner sites are α s or β s. Now with $m = 2$ and $\mathfrak{S}$ consisting of two blocks of four squares $\mathfrak{S}$ has nine members and a calculation similar to that in Example 16.4.4 gives

$$\begin{aligned}
\chi_1^{(9)} = -1, \quad \chi_2^{(9)} = -1, \quad \chi_3^{(9)} = 4, \quad \chi_4^{(9)} = 1, \quad \chi_5^{(9)} = -1, \quad \chi_6^{(9)} = -1, \\
\chi_7^{(9)} = -1, \quad \chi_8^{(9)} = \frac{1}{2}, \quad \chi_9^{(9)} = \frac{1}{2}.
\end{aligned} \tag{16.4.48}$$

However, it is not necessary to restrict attention to interpenetrating sublattices.

Table 16.3 The
de Neef–Enting sequence
of graphs for $\ell = 6$

$\mathfrak{g}_i$	$[x \times y]$	$[\mathfrak{n}, \mathfrak{k}]$	$\chi_i^{(15)}$
$\mathfrak{g}_1$	$[1 \times 1]$	$[1, 1]$	0
$\mathfrak{g}_2$	$[2 \times 1]$	$[2, 2]$	−1
$\mathfrak{g}_3$	$[1 \times 2]$	$[2, 2]$	−1
$\mathfrak{g}_4$	$[3 \times 1]$	$[3, 3]$	+3
$\mathfrak{g}_5$	$[1 \times 3]$	$[3, 3]$	+3
$\mathfrak{g}_6$	$[2 \times 2]$	$[4, 8]$	+3
$\mathfrak{g}_7$	$[4 \times 1]$	$[4, 4]$	−3
$\mathfrak{g}_8$	$[1 \times 4]$	$[4, 4]$	−3
$\mathfrak{g}_9$	$[3 \times 2]$	$[6, 1]$	−3
$\mathfrak{g}_{10}$	$[2 \times 3]$	$[6, 1]$	−3
$\mathfrak{g}_{11}$	$[5 \times 1]$	$[5, 1]$	+1
$\mathfrak{g}_{12}$	$[1 \times 5]$	$[5, 1]$	+1
$\mathfrak{g}_{13}$	$[4 \times 2]$	$[8, 2]$	+1
$\mathfrak{g}_{14}$	$[2 \times 4]$	$[8, 2]$	+1
$\mathfrak{g}_{15}$	$[3 \times 3]$	$[9, 1]$	+1

Example 16.4.6 Consider a simple cubic lattice with a layered sublattice structure as in the model for a metamagnet described in Sect. 6.2. The equivalent sublattices are labelled $\mathfrak{a}$ and $\mathfrak{b}$. Let $\mathfrak{S}$ be the one graph consisting of a cube of eight first-neighbour sites, four $\mathfrak{a}$ sites on one face and four $\mathfrak{b}$ sites on the opposite face. This generates the sequence:

- $\mathfrak{g}_1$: a single $\mathfrak{a}$ site,

$\mathfrak{g}_3$: a single $\mathfrak{a}$ – $\mathfrak{a}$ pair,

$\mathfrak{g}_5$: a single $\mathfrak{b}$ – $\mathfrak{b}$ pair,

$\mathfrak{g}_7$: a square of two $\mathfrak{a}$ and two $\mathfrak{b}$ sites,

$\mathfrak{g}_9$: the cube $\mathfrak{S}$.
- $\mathfrak{g}_2$: a single $\mathfrak{b}$ site,

$\mathfrak{g}_4$: a single $\mathfrak{a}$ – $\mathfrak{b}$ pair,

$\mathfrak{g}_6$: a square of four $\mathfrak{a}$ sites,

$\mathfrak{g}_8$: a square of four $\mathfrak{b}$ sites,

Then

$$\mathbf{\Omega}^{(9)} = \begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 4 & 2 & 0 & 4 \\ 0 & 1 & 0 & 1 & 2 & 0 & 2 & 4 & 4 \\ 0 & 0 & 1 & 0 & 0 & 4 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{\omega}^{(9)} = \begin{pmatrix} 1/2 \\ 1/2 \\ 1 \\ 1 \\ 1 \\ 1/2 \\ 1 \\ 1/2 \\ 1 \end{pmatrix}. \quad (16.4.49)$$

Solving (16.4.26) for (16.4.49) gives

$$\begin{aligned} \chi_1^{(9)} &= \frac{1}{2}, \quad \chi_2^{(9)} = \frac{1}{2}, \quad \chi_3^{(9)} = 2, \quad \chi_4^{(9)} = 0, \quad \chi_5^{(9)} = 2, \quad \chi_6^{(9)} = -\frac{1}{2}, \\ \chi_7^{(9)} &= -3, \quad \chi_8^{(9)} = -\frac{1}{2}, \quad \chi_9^{(9)} = -1. \end{aligned} \quad (16.4.50)$$

We can also consider cases where there are four equivalent sublattices.

Example 16.4.7 Consider a body-centred cubic lattice (Fig. 16.8) divided, as shown in Fig. 7.2b into four equivalent face-centred cubic sublattices (Fig. 16.9) labelled a, b, c and d. The pairs a–c, a–d, b–c, b–d combine to form diamond lattices (Fig. 16.10). Let $\mathfrak{S}$ be the one (section) graph consisting of a tetrahedron of four sites, one in each sublattice connected by first-neighbour and second-neighbour edges.³³ This generates the sequence:

g_1 : a single a site,	g_2 : a single b site,
g_3 : a single c site,	g_4 : a single d site,
g_5 : a single a–c first-neighbour pair,	g_6 : a single a–d first-neighbour pair,
g_7 : a single b–c first-neighbour pair,	g_8 : a single b–d first-neighbour pair,
g_9 : a single a–b second-neighbour pair,	g_{10} : a single c–d second-neighbour pair,
g_{11} : a a–b–c triangular face,	g_{12} : a a–b–d triangular face,
g_{13} : a a–c–d triangular face,	g_{14} : a b–c–d triangular face,
g_{15} : the tetrahedron $\mathfrak{S}$.	

³³ This involves the generalization of the notion of a graph to the case of two types of edges mentioned in a footnote on p. 637.

Then

$$\Omega^{(15)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \omega^{(15)} = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1 \\ 1 \\ 1 \\ 1 \\ 3/2 \\ 3/2 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 6 \end{pmatrix}. \quad (16.4.51)$$

Solving (16.4.26) for (16.4.51) gives

$$\begin{aligned} \chi_1^{(15)} &= -\frac{1}{4}, \quad \chi_2^{(15)} = -\frac{1}{4}, \quad \chi_3^{(15)} = -\frac{1}{4}, \quad \chi_4^{(9)} = -\frac{1}{4}, \quad \chi_5^{(15)} = 1, \\ \chi_6^{(15)} &= 1, \quad \chi_7^{(15)} = 1, \quad \chi_8^{(15)} = 1, \quad \chi_9^{(15)} = \frac{3}{2}, \quad \chi_{10}^{(15)} = \frac{3}{2}, \\ \chi_{11}^{(15)} &= -3, \quad \chi_{12}^{(15)} = -3, \quad \chi_{13}^{(15)} = -3, \quad \chi_{14}^{(15)} = -3, \quad \chi_{15}^{(15)} = 6. \end{aligned} \quad (16.4.52)$$

16.4.8 The Guggenheim–McGlashan Approach

A generalization of the quasi-chemical method of Guggenheim (1935) was developed by Guggenheim and McGlashan (1951). This method can be formulated in terms of the partially-ordered sequence $\mathcal{S}$ used in the KHDeB approximation, but, while it retains the capability of including irreducible energy contributions from more complicated graphs than first-neighbour pairs, it is much simpler to use, and in consequence rather less accurate.

Consider a partially-ordered sequence of graphs $\mathcal{S} := \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n\}$, where $\mathbf{g}_1$ is the only type of single site graph, $\mathbf{g}_2$ is the only type of first-neighbour pair graph and $n > 2$.³⁴ The crucial step in this approach is that the lattice $\mathcal{N}$ of N sites is replaced by a *pseudo-lattice* $\tilde{\mathcal{N}}$ of N sites made up of graphs $\mathbf{g}_n$ overlapping only by single sites $\mathbf{g}_1$. The number of $\mathbf{g}_n$ is taken so that the number of $\mathbf{g}_2$ in $\mathcal{N}$ and $\tilde{\mathcal{N}}$ is the same. Let $\tilde{\omega}_j$, $j = 1, 2, \dots, n$ be the lattice constants of $\tilde{\mathcal{N}}$. Then

³⁴ We are for simplicity, at this stage, excluding sublattice ordering of the type described in Sect. 16.4.7.

$$\tilde{\omega}_1 = \omega_1, \quad \tilde{\omega}_2 = \omega_2, \quad \tilde{\omega}_j = \Omega_{jn} \tilde{\omega}_n, \quad j = 2, \dots, n, \quad (16.4.53)$$

from which it follows that

$$\tilde{\omega}_j = \omega_2 \Omega_{jn} / \Omega_{2n}, \quad j = 2, 3, \dots, n. \quad (16.4.54)$$

Replacing $\boldsymbol{\omega}^{(n)}$ in (16.4.26) with the vector with elements given by (16.4.54) we have

$$\frac{\omega_2}{\Omega_{2n}} \boldsymbol{\Omega}_n^{(n)} + \left[\omega_1 - \omega_2 \frac{\Omega_{1n}}{\Omega_{2n}} \right] \mathbf{i}_1 = \boldsymbol{\Omega}^{(n)} \tilde{\boldsymbol{\chi}}^{(n)}, \quad (16.4.55)$$

where $\boldsymbol{\Omega}_n^{(n)}$ is the n th column of $\boldsymbol{\Omega}^{(n)}$ and $\mathbf{i}_1$ is the unit vector with elements $\delta^{\text{Kr}}(k, 1)$, $k = 1, 2, \dots, n$. Since $[\boldsymbol{\Omega}^{(n)}]^{-1}$ is upper triangular with diagonal elements unity,

$$\begin{aligned} \tilde{\chi}_1^{(n)} &= \omega_1 - \omega_2 \frac{\Omega_{1n}}{\Omega_{2n}}, \\ \tilde{\chi}_j^{(n)} &= 0, \quad j = 2, \dots, n-1, \quad \tilde{\chi}_n^{(n)} = \frac{\omega_2}{\Omega_{2n}}. \end{aligned} \quad (16.4.56)$$

It is easy to verify these results for Example 16.4.4. In particular, when $\mathfrak{S}$ is the single square,

$$\tilde{\boldsymbol{\omega}}^{(3)} = \begin{pmatrix} 1 \\ 2 \\ 1/2 \end{pmatrix} \quad \tilde{\boldsymbol{\chi}}^{(3)} = \begin{pmatrix} -1 \\ 0 \\ 1/2 \end{pmatrix}. \quad (16.4.57)$$

The extension of this method to the case where the lattice is divided into a number of sublattices is straightforward. Now the pseudo-lattice $\tilde{\mathcal{N}}$ is constructed so that the numbers of sites and first-neighbour pairs of all types are maintained constant. We illustrate this with the case described in Example 16.4.7, of a body-centred cubic lattice divided into four sublattices as shown in Fig. 7.2b. Then, from (16.4.51),

$$\begin{aligned} \tilde{\omega}_j &= \omega_j = \frac{1}{4}, & j &= 1, \dots, 4, \\ \tilde{\omega}_j &= \omega_j = 1, & j &= 5, \dots, 8, \\ \tilde{\omega}_j &= \Omega_{j,15} \tilde{\omega}_{15} = 1, & j &= 9, \dots, 15, \end{aligned} \quad (16.4.58)$$

giving

$$\begin{aligned}\tilde{\chi}_j^{(15)} &= -\frac{3}{4}, & j &= 1, \dots, 4, \\ \tilde{\chi}_j^{(15)} &= 0, & j &= 5, \dots, 14, \\ \tilde{\chi}_{15}^{(15)} &= 1.\end{aligned}\tag{16.4.59}$$

Finally we may for this method need to adjust some of the energy parameters. This will not be the case if the only contributions to the energy are from single sites and first-neighbour pairs since the Guggenheim method counts correctly the number of sites and first-neighbour pairs. However, suppose ε_j is an energy contribution to some configurations of $\mathfrak{g}_j$, which does not occur for any subgraphs of $\mathfrak{g}_j$. Since there are $N\omega_j$ $\mathfrak{g}_j$ of these on $\mathcal{N}$ and $N\tilde{\omega}_j$ $\mathfrak{g}_j$ on the Guggenheim pseudo-lattice $\tilde{\mathcal{N}}$ the replacement

$$\varepsilon_j \rightarrow \frac{\omega_j}{\tilde{\omega}_j} \varepsilon_j \tag{16.4.60}$$

is needed.

16.4.9 Further Results

The analysis in this section is needed for the calculation of the cluster-variation entropy in (7.2.4). Here we consider the case, of relevance to the KHDeB hierarchy, where the partially-ordered sequence of connected section graphs $\mathcal{S}$ is generated by one graph ($m = 1$, $\mathcal{S}$ having one member). Let $\mathcal{S}$ have n members with $\mathfrak{g}_n$ the generating graph and let $\mathbf{\Omega}^{(k)}$ be the matrix formed from the leading $k \times k$ submatrix of $\mathbf{\Omega}^{(n)}$, with $\mathbf{\omega}^{(k)}$ the column vector formed by taking the first k elements of $\mathbf{\omega}^{(n)}$. Then, as in (16.4.26), $\mathbf{\chi}^{(k)}$ is defined to be the solution of

$$\mathbf{\omega}^{(k)} = \mathbf{\Omega}^{(k)} \mathbf{\chi}^{(k)}.\tag{16.4.61}$$

Because $\mathbf{\Omega}^{(n)}$ is an upper triangular matrix with diagonal elements equal to one, it is not difficult to see that the inverse of $\mathbf{\Omega}^{(k)}$ is the leading $k \times k$ submatrix of $[\mathbf{\Omega}^{(n)}]^{-1}$. That is

$$[\mathbf{\Omega}^{(k)}]^{-1} = \{[\mathbf{\Omega}^{(n)}]^{-1}\}^{(k)}.\tag{16.4.62}$$

So

$$\mathbf{\chi}^{(k)} = \{[\mathbf{\Omega}^{(n)}]^{-1}\}^{(k)} \mathbf{\omega}^{(k)}.\tag{16.4.63}$$

However, it is easy to see, using Example 16.4.4, that $\mathbf{x}^{(k)}$ is *not*, in general, the same as the vector formed from the first k elements of $\mathbf{x}^{(n)}$. The latter we shall denote by $[\mathbf{x}^{(n)}]^{(k)}$.

With $\mathbf{\Omega}_i^{(k)}$ representing the vector formed from the i th row of $\mathbf{\Omega}^{(k)}$, it follows from (16.4.61), that

$$\omega_i = \mathbf{\Omega}_i^{(k-1)} \cdot \mathbf{x}^{(k-1)}, \quad \omega_i = \mathbf{\Omega}_i^{(k)} \cdot \mathbf{x}^{(k)}. \quad (16.4.64)$$

By considering the operation in (16.4.61) of the last row of $\mathbf{\Omega}^{(k)}$ on $\mathbf{x}^{(k)}$ it is clear that the last element of this vector is ω_k . So,

$$\mathbf{\Omega}_i^{(k-1)} \cdot \mathbf{x}^{(k-1)} = \mathbf{\Omega}_i^{(k-1)} \cdot [\mathbf{x}^{(k)}]^{(k-1)} + \Omega_{i\ k} \omega_k \quad (16.4.65)$$

and, from (16.4.64),

$$\mathbf{\Omega}^{(k-1)} \left(\mathbf{x}^{(k-1)} - [\mathbf{x}^{(k)}]^{(k-1)} \right) = \mathbf{v}^{(k)}, \quad (16.4.66)$$

where

$$\mathbf{v}^{(k)} := \begin{pmatrix} \Omega_{1\ k} \omega_k \\ \Omega_{2\ k} \omega_k \\ \vdots \\ \Omega_{k-1\ k} \omega_k \end{pmatrix} \quad (16.4.67)$$

is a $(k - 1)$ -dimensional vector.

Chapter 17

Algebra

17.1 Catastrophe Theory

According to Stewart (1997, p. 7), “catastrophe theory is not a single thread of ideas; it resembles more closely a web, with innumerable interconnected strands; these include physical intuition and experiment, geometry, algebra, calculus, topology, singularity theory and many others”. The name catastrophe theory was coined by Thom (1972) with particular reference to the application of this web of ideas to biological morphogenesis and the development of languages (Thom 1974) and it was soon evident that it could be applied to a much wider range of problems. In particular thermodynamics and the theory of phase transitions (Fowler 1972; Schulman and Revzen 1972; Schulman 1973; Lavis and Bell 1977; Gaite 1992; Gaite et al. 1998). Of particular relevance is the relationship between Landau theory, as described in Chap. 5, and catastrophe theory. In the former, physical arguments are used to justify the replacement of the full free-energy density with a ‘sufficient’ number of terms of its Taylor expansion about a critical (or tricritical) point. The purpose of this appendix is to use catastrophe theory to bring a degree of mathematical precision to this important question of sufficiency.¹

17.1.1 Equivalence and Determinancy

Our main interest will be in scalar-valued functions $\mathbb{R}^n \rightarrow \mathbb{R}$, however, for some purposes it will be necessary to give definitions for vector-valued functions $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

¹ It is not our aim to give a comprehensive account of catastrophe theory (there are many of these, see e.g. Saunders 1980; Gilmore 1981; Poston and Stewart 1997). We simply wish to trace a thread through catastrophe theory leading to the results we require. In doing so we quote some definitions and theorems; for further details of these together with the proofs of the theorems (or references where they can be found) the reader is referred to Poston and Stewart.

Definition 17.1.1 A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is k -fold differentiable, or of class $\mathcal{C}^k$, if all its m components have continuous partial derivatives of all orders up to k . A class $\mathcal{C}^\infty$ function is said to be smooth.

Definition 17.1.2 Given a polynomial function $p : \mathbb{R}^n \rightarrow \mathbb{R}$, we denote by $J^k p$ the truncated polynomial obtained by omitting all terms of degree greater than k . Given a $\mathcal{C}^k$ function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the k -jet $J^k f$ of f is the polynomial obtained by truncating its Taylor series about $\mathbf{x} = \mathbf{0}$ at degree k .^{2,3}

J^k is, of course, a linear operator with respect to sums of functions and for products of functions we use it in the sense that $J^k fg$ means $J^k(fg)$. It is also obvious that $J^k fg = J^k f(J^k g) = J^k(J^k f)g = J^k(J^k f)(J^k g)$.

Definition 17.1.3 Let $\mathcal{U}$ and $\mathcal{V}$ be open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively with $f : \mathcal{U} \rightarrow \mathbb{R}^m$ and $f(\mathcal{U}) := \mathcal{V}$. Then f is a diffeomorphism on $\mathcal{U}$ if it is: (i) smooth, (ii) has an inverse $g : \mathcal{V} \rightarrow \mathbb{R}^n$ which is also smooth. f is a local diffeomorphism at $\mathbf{x} = \mathbf{0}$ if it is a diffeomorphism on some $\mathcal{U}$ containing $\mathbf{x} = \mathbf{0}$.

Example 17.1.1 $\sin(x)$ is a local diffeomorphism at $x = 0$ since it is smooth with a smooth inverse in the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Definition 17.1.4 Two smooth functions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ are equivalent at $\mathbf{x} = \mathbf{0}$ if there exists a local diffeomorphism $\eta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ at $\mathbf{x} = \mathbf{0}$ and a constant γ so that, in some open set $\mathcal{U}$ containing $\mathbf{x} = \mathbf{0}$, $g(\mathbf{x}) = f(\eta(\mathbf{x})) + \gamma$. If in addition

$$\left(\frac{\partial \eta_i}{\partial x_j} \right)_0 = \delta^{\text{Kr}}(i - j), \quad i, j = 1, 2, \dots, n, \quad (17.1.1)$$

the functions are strongly equivalent.

For the application to Landau theory we shall need strong equivalence and we shall concentrate on that case.

Example 17.1.2 It can be shown that $f(x) := x^2$ and $g(x) := x^2 - x^4$ are strongly equivalent at $x = 0$ by taking

$$\eta(x) := x\sqrt{1 - x^2}, \quad \text{which has an inverse in } \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Definition 17.1.5 A smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is k -determinate at $\mathbf{x} = \mathbf{0}$ if, for all $g : \mathbb{R}^n \rightarrow \mathbb{R}$ of order $k + 1$, $J^k f + g$ is equivalent at $\mathbf{x} = \mathbf{0}$ to $J^k f$. If the equivalence is strong then f is strongly k -determinate.

² For easy of discussion we shall, in this and following definitions, choose our reference point as the origin in the space $\mathbb{R}^n$ of $\mathbf{x}$. Translation to a point $\mathbf{x}_0 \neq \mathbf{0}$ is trivial.

³ The terms of the polynomial p are monomials. The degree of a monomial is the sum of the degrees of each of its n variables. The degree of p is the largest of the degrees of its constituent monomials and the order of p is the smallest of the degrees of its constituent monomials. A homogeneous polynomial has terms all of the same degree, which is then both the degree and the order of the polynomial.

It will be useful to have a mechanism for determining whether any f is strongly k -determinate, and to explore this question the following notation will be useful:

Definition 17.1.6

- (a) $\mathcal{E}_n^k$ is the vector space of polynomials $p : \mathbb{R}^n \rightarrow \mathbb{R}$ of degree less than or equal to k .
- (b) $\mathcal{J}_n^k$ is the subspace of $\mathcal{E}_n^k$ consisting of polynomials with zero constant term, that is of order greater than or equal to one.
- (c) $\mathcal{J}_n^k$ is the subspace of $\mathcal{J}_n^k$ consisting of polynomials $p \in \mathcal{J}_n^k$ with $(\nabla p)_0 = \mathbf{0}$, that is of order greater than or equal to two.
- (d) $\mathcal{M}_n^k$ is the subspace of $\mathcal{J}_n^k$ consisting all homogeneous polynomials of degree (and order) k .

We use the notation $\mathcal{E}_1^k \stackrel{\text{B}}{=} \{1, x, x^2, \dots, x^k\}$ meaning that $\mathcal{E}_1^k$ is the vector space of polynomials with basis given by the monomials $x^i, i = 0, 1, 2, \dots, k$, with a similar notation applied to other spaces. The dimension $\dim[\mathcal{M}_n^k]$ of $\mathcal{M}_n^k$ is the number of monomials of order k . Thus

$$\dim[\mathcal{M}_n^k] = \frac{(n-1+k)!}{(n-1)!k!}, \quad \dim[\mathcal{E}_n^k] = \sum_{i=0}^k \frac{(n-1+i)!}{(n-1)!i!} = \frac{(n+k)!}{n!k!}, \quad (17.1.2)$$

$$\dim[\mathcal{J}_n^k] = \dim[\mathcal{E}_n^k] - 1, \quad \dim[\mathcal{J}_n^k] = \dim[\mathcal{E}_n^k] - n - 1.$$

Definition 17.1.7 Let $\mathcal{Z}$ be any one of the vector spaces of Definition 17.1.6 and let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces of $\mathcal{Z}$. Then $\mathcal{X}$ and $\mathcal{Y}$ are transverse if the dimension of their intersection is as small as possible. That is

$$\dim[\mathcal{X} \cap \mathcal{Y}] = \max \{0, \dim[\mathcal{X}] + \dim[\mathcal{Y}] - \dim[\mathcal{Z}]\}. \quad (17.1.3)$$

Let $\mathcal{Z} \stackrel{\text{B}}{=} \{v_1, v_2, \dots, v_\zeta\}$ and $\mathcal{X} \stackrel{\text{B}}{=} \{v_1, v_2, \dots, v_\xi\}$, where $\zeta := \dim[\mathcal{Z}]$, $\xi := \dim[\mathcal{X}]$ and $v_i, i = 1, 2, \dots, \zeta$ are monomials.

Definition 17.1.8 When $\xi < \zeta$, $\{v_{\xi+1}, \dots, v_\zeta\}$ is a cobasis of $\mathcal{X}$ with respect to $\mathcal{Z}$.

With $\overline{\mathcal{X}} \stackrel{\text{B}}{=} \{v_{\xi+1}, \dots, v_\zeta\}$, the codimension of $\mathcal{X}$ with respect to $\mathcal{Z}$ is given by

$$\text{cod}[\mathcal{X}|\mathcal{Z}] := \dim[\overline{\mathcal{X}}] = \dim[\mathcal{Z}] - \dim[\mathcal{X}]. \quad (17.1.4)$$

It follows from (17.1.3) that $\mathcal{X}$ and $\mathcal{Y}$ are transverse if and only if either $\mathcal{Y} \subseteq \overline{\mathcal{X}}$ or $\overline{\mathcal{X}} \subseteq \mathcal{Y}$. In particular, of course $\mathcal{X}$ and $\overline{\mathcal{X}}$ are transverse.

Definition 17.1.9 Let $\mathcal{X}$ and $\mathcal{Y}$ be any subspaces of the vector spaces given in Definition 17.1.6. The product space $\mathcal{X} \star \mathcal{Y}$ consists of all linear combinations of products of one member of each of $\mathcal{X}$ and $\mathcal{Y}$. That is

$$\mathcal{X} \star \mathcal{Y} := \{x_1 y_1 + \dots + x_r y_r \mid \forall r, x_i \in \mathcal{X}, y_i \in \mathcal{Y}\}.$$

We denote by $J^k(\mathcal{X} \star \mathcal{Y})$ the space of truncated members of $\mathcal{X} \star \mathcal{Y}$ and by $\Delta_k(f)$ the subspace of $\mathcal{E}_n^k$ spanned by

$$\left\{ J^k \frac{\partial f}{\partial x_1}, \dots, J^k \frac{\partial f}{\partial x_n} \right\}.$$

It is clear that $\Delta_k(f) = \Delta_k(J^{k+i}f)$, for $i \geq 1$. The following theorem allows us to ascertain whether the function f is strongly k -determinate:

Theorem 17.1.1 *A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is strongly k -determinate at $\mathbf{x} = \mathbf{0}$ if and only if*

$$\mathcal{M}_n^{k+1} \subseteq J^{k+1}(\mathcal{J}_n^{k+1} \star \Delta_{k-1}(f)) \quad (17.1.5)$$

and, since the order of the elements of $\mathcal{J}_n^{k+1}$ is at least two, $\Delta_{k-1}(f)$ in (17.1.5) can be replaced by $\Delta_{k+i}(f)$, for $i = 0, 1, 2, \dots$

From this result it is straightforward to show that:

Corollary 17.1.1 *If a function is strongly k -determinate at $\mathbf{x} = \mathbf{0}$ then it is strongly k' -determinate at $\mathbf{x} = \mathbf{0}$, for all $k' > k$.*

Definition 17.1.10 The smallest k for which f is k -determinate at $\mathbf{x} = \mathbf{0}$ is called the determinacy of f at $\mathbf{x} = \mathbf{0}$. A function whose determinacy is finite is called finitely-determined. Strong determinacy is defined in a similar way.

17.1.2 Critical Points, Codimension and Unfoldings

Definition 17.1.11 The smooth function $f : \mathbb{R} \rightarrow \mathbb{R}^n$ has an equilibrium point at $\mathbf{x} = \mathbf{0}$ if $(\nabla f)_0 = \mathbf{0}$. If in addition the Hessian matrix of second-order partial derivatives of f is singular then $\mathbf{x} = \mathbf{0}$ is a critical point.⁴

We are now concerned with the problem of classifying critical points.

Definition 17.1.12 $\mathcal{T}_k(f) := J^k(\mathcal{E}_n^k \star \Delta_k(f))$ is the tangent space⁵ to $J^k f$ at $\mathbf{x} = \mathbf{0}$. This is a subspace of $\mathcal{J}_n^k$.

⁴ In catastrophe theory what we have called an equilibrium point is designated a *critical point* with our critical point being a *degenerate* critical point. Our nomenclature is an attempt at uniformity with the rest of the book. However, even this is a compromise since a spinodal point, representing the meeting of a line of stable and unstable equilibrium points (Figs. 5.1a and 5.2), would according to the current definition be a critical point. In some ways it would be better to use the language of bifurcation theory (Iooss and Joseph 1989) in which thermodynamic critical points are singular bifurcations and a spinodal point is a regular turning-point bifurcation.

⁵ The reasoning underlying this definition is given by Poston and Stewart (1997, p. 138).

Definition 17.1.13 The codimension $\text{cod}[f]$ of f is given by

$$\text{cod}[f] := \max_{k \geq 2} \{\text{cod}[\mathcal{T}_k(f)|\mathcal{J}_n^k]\}. \quad (17.1.6)$$

Theorem 17.1.2 If f has strong determinancy s , then,

$$\text{cod}[f] = \text{cod}[\mathcal{T}_s(f)|\mathcal{J}_n^s]. \quad (17.1.7)$$

From Theorem 17.1.1, $\mathcal{M}_n^{s+i} \subseteq \mathcal{T}_{s+i}(f)$ for $i = 1, 2, \dots$ and it follows that $\text{cod}[\mathcal{T}_k(f)|\mathcal{J}_n^k]$ increases with k until $k = s$, after which it remains constant yielding (17.1.7).

Theorem 17.1.3 Given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, with an equilibrium point at $\mathbf{x} = \mathbf{0}$, any function $F(\mathbf{x}) := \varepsilon g(\mathbf{x}) + f(\mathbf{x})$, for ε sufficiently small has at most $\text{cod}[f] + 1$ equilibrium points near to $\mathbf{x} = \mathbf{0}$.

Definition 17.1.14 Consider an m -parameter family of functions $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ with values $F(\mathbf{x}; \mathbf{c})$, where the vector $\mathbf{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$ of parameters define the members of the family. F is an m -unfolding of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ if $F(\mathbf{x}; \mathbf{0}) = f(\mathbf{x})$.

Definition 17.1.15 An r -unfolding G of f is induced from the m -unfolding F of Definition 17.1.14 by:

- (a) a smooth mapping $\mathbf{w} : \mathbb{R}^r \rightarrow \mathbb{R}^m$ so that
 $\mathbf{t} := (t_1, t_2, \dots, t_r) \rightarrow \mathbf{w}(\mathbf{t}) := (w_1(\mathbf{t}), w_2(\mathbf{t}), \dots, w_m(\mathbf{t})).$
- (b) a smooth $\mathbf{t}$ -dependent map $\eta : \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}$ such that, for each (fixed) $\mathbf{t}$
 $\mathbf{x} \rightarrow \eta(\mathbf{x}; \mathbf{t})$ is local diffeomorphism at $\mathbf{x} = \mathbf{0}$.
- (c) a smooth map $\gamma : \mathbb{R}^r \rightarrow \mathbb{R}$,

if, in some open set $\mathcal{U}$ containing $\mathbf{x} = \mathbf{0}$ and in a neighbourhood of $\mathbf{t} = \mathbf{0}$,

$$G(\mathbf{x}; \mathbf{t}) = F(\eta(\mathbf{x}; \mathbf{t}); \mathbf{w}(\mathbf{t})) + \gamma(\mathbf{t}).$$

If G is induced from F with $r = m$ then the unfoldings are equivalent. If in addition

$$\left(\frac{\partial \eta_i}{\partial x_j} \right)_0 = \delta^{\text{Kr}}(i - j), \quad i, j = 1, 2, \dots, n,$$

for all $\mathbf{t}$ in a neighbourhood of $\mathbf{t} = \mathbf{0}$ the unfoldings are strongly equivalent.

Definition 17.1.16 An m -unfolding F of f is versal if all other unfoldings of f can be induced from it. It is universal if no versal unfolding of f exists of dimension less than m .

Theorem 17.1.4 Universal unfoldings of equivalent functions are equivalent.

Definition 17.1.17 The subspace $\mathcal{V}^k(F)$ of $\mathcal{J}_n^k$ is that spanned by the polynomials

$$v_j^k(\mathbf{x}) := \frac{\partial \mathbf{J}^k F(\mathbf{x}; 0, \dots, 0, c_j, 0, \dots, 0)}{\partial c_j}, \quad j = 1, 2, \dots, m. \quad (17.1.8)$$

Theorem 17.1.5 An m -unfolding F of f , where f is k -determinate, is versal if and only if $\mathcal{V}^k(F)$ and $\mathcal{T}_k(f)$ are transverse. It is universal if and only if $m = \text{cod}[f]$.

Let f have an equilibrium point at $\mathbf{x} = \mathbf{0}$ and strong determinacy s then it follows from Theorems 17.1.2 and 17.1.5 that

$$F(\mathbf{x}; \mathbf{c}) := \sum_{j=1}^{\text{cod}[f]} c_j v_j^s(\mathbf{x}) + f(\mathbf{x}) \quad (17.1.9)$$

is a universal unfolding of f if $\{v_1^s, v_2^s, \dots, v_{\text{cod}[f]}^s\}$ is a cobasis of $\mathcal{T}^s(f)$ in $\mathcal{J}_n^s$; we assume that the parameters $\mathbf{c}$ are chosen so this is the case. Now f is equivalent to $\mathbf{J}^s f$ and to $\mathbf{J}^{s+1} f$. However, $\mathcal{T}_s(f) = \mathcal{T}_s(\mathbf{J}^{s+1} f)$ but not necessarily $\mathcal{T}_s(f) = \mathcal{T}_s(\mathbf{J}^s f)$. So

$$\tilde{F}(\mathbf{x}; \mathbf{c}) := \sum_{j=1}^{\text{cod}[f]} c_j v_j^s(\mathbf{x}) + \mathbf{J}^{s+1} f(\mathbf{x}) \quad (17.1.10)$$

is a universal unfolding of $\mathbf{J}^{s+1} f$, which is equivalent to F . However

$$\tilde{F}'(\mathbf{x}; \mathbf{c}) := \sum_{j=1}^{\text{cod}[f]} c_j v_j^s(\mathbf{x}) + \mathbf{J}^s f(\mathbf{x}) \quad (17.1.11)$$

is a universal unfolding of $\mathbf{J}^s f$, and thus equivalent to F , only if $\mathcal{T}_s(\mathbf{J}^s f)$ is transverse to $\mathcal{V}^s(F)$. A universal unfolding $F(\mathbf{x}; \mathbf{c})$ of $f(\mathbf{x})$, with a critical point at $\mathbf{x} = \mathbf{0}$ and small values of the unfolding parameters, is the most general perturbation of $f(\mathbf{x})$ and it will thus, from Theorem 17.1.3, have at most $\text{cod}[f] + 1$ equilibrium points in a neighbourhood of $\mathbf{x} = \mathbf{0}$.

Example 17.1.3 Let f be a smooth function with a power series of the form

$$f(x) := a_0 + \sum_{i=s}^{\infty} a_i x^i, \quad a_s \neq 0, \quad s \geq 2. \quad (17.1.12)$$

It is clear that the presence of the constant a_0 does not affect the calculation and

$$J^{k-1} \frac{df}{dx} = \begin{cases} \sum_{i=s}^k i a_i x^{i-1}, & k \geq s, \\ 0, & k < s. \end{cases} \quad (17.1.13)$$

Then

$$x^{s+1} = J^{s+1} \frac{x^2}{s a_s} J^{s-1} \frac{df}{dx}. \quad (17.1.14)$$

So the function f has strong determinancy s and there exists a diffeomorphism $y(x)$ with $y'(0) = 1$ such that

$$f(y(x)) = a_0 + x^s. \quad (17.1.15)$$

The terms in the power series of $y(x)$ can be determined iteratively and in some particular cases an exact form is derivable. Thus, for example if $s = 2$ and $a_i = 1$, $i \geq 2$ then $y(x) = -\frac{1}{2} (x^2 - \sqrt{x^4 + 4x^2})$.

From (17.1.13),

$$\mathcal{T}_k(f) \stackrel{\text{B}}{=} \begin{cases} \{0\}, & k < (s-1), \\ \{x^{s-1}, x^s, \dots, x^k\}, & k \geq (s-1) \end{cases} \quad (17.1.16)$$

and $\text{cod}[f] = s - 2$. We note that the formulae for $\mathcal{T}_k(f)$ do not depend on $a_i \neq 0$ for $i > s$. The element x^j in the basis, for $s \leq j \leq k$, can be generated by the product of $x^{s-1} \in \Delta_s(f)$ and $x^{j-s+1} \in \mathcal{E}_1^k$. From this it follows that $\mathcal{T}_k(f) = \mathcal{T}_k(J^{s+1}f) = \mathcal{T}_k(J^s f)$. In particular consider the case $s = 4$ for which f has a critical point at $x = 0$ of codimension two. Suppose that $a_4 = 1$ and that the coefficients a_k , $k \geq 5$, are such that f has no further equilibrium points in a neighbourhood of $x = 0$. Now let $g(x) = x^j$, for some positive integer j , so that the equilibrium points for $F(x) = \varepsilon x^j + f(x)$ are given by

$$0 = j\varepsilon x^{j-1} + 4x^3 + \sum_{i=5}^{\infty} i a_i x^{i-1}. \quad (17.1.17)$$

It is clear that, for $j \geq 4$, F has a critical point at $x = 0$ and no other equilibrium points in a neighbour of the origin. The remaining cases are:

- (i) $j = 3$: where there is a critical point at $x = 0$ and an equilibrium point at $x = -\frac{3}{4}\varepsilon$.
- (ii) $j = 2$: where there is an equilibrium point at $x = 0$ and at $x = \pm\sqrt{-\varepsilon/2}$ for $\varepsilon < 0$.
- (iii) $j = 1$: where there is an equilibrium point at the real cube root of $-\varepsilon/4$.

This is in accord with Theorem 17.1.3 which asserts that the number of equilibrium points must be less than or equal to three. Consider now the $(s - 2)$ -unfolding

$$F(x; \mathbf{c}) := a_0 + \sum_{j=1}^{s-2} c_j x^j + \sum_{i=s}^{\infty} a_i x^i \quad (17.1.18)$$

of (17.1.12). Then $\mathcal{V}^s(F) \stackrel{\text{B}}{=} \{x, x^2, \dots, x^{s-2}\}$ and since this is a cobasis of $\mathcal{T}^s(f) = \mathcal{T}^s(\mathcal{J}^{s+1}f) \stackrel{\text{B}}{=} \{x^{s-1}, x^s\}$ in $\mathcal{J}_1^s$, (17.1.18) is a universal unfolding, equivalent to the universal unfolding

$$\tilde{F}(x; \mathbf{c}) := a_0 + \sum_{j=1}^{s-2} c_j x^j + a_s x^s + a_{s+1} x^{s+1} \quad (17.1.19)$$

of $\mathcal{J}^{s+1}f$ and since, as already indicated, $\mathcal{T}_s(\mathcal{J}^s f) = \mathcal{T}_s(\mathcal{J}^{s+1}f)$

$$\tilde{F}'(x; \mathbf{c}) := a_0 + \sum_{j=1}^{s-2} c_j x^j + a_s x^s \quad (17.1.20)$$

is also equivalent to F .

Example 17.1.4 For $n = 2$ the general cubic case is

$$f(x_1, x_2) := a_0 x_1^3 + a_1 x_1^2 x_2 + a_2 x_1 x_2^2 + a_3 x_2^3 + \mathcal{O}(x_1^i x_2^{4-i}). \quad (17.1.21)$$

To show that f has strong determinacy of 3 we must, according to Theorem 17.1.1, show that the five monomials $x_1^4, x_1^3 x_2, x_1^2 x_2^2, x_1 x_2^3, x_2^4$ can be expressed as linear combinations of some of the six quantities $x_1^j x_2^{2-j} f_1(x_1, x_2), x_1^j x_2^{2-j} f_2(x_1, x_2), j = 0, 1, 2$, where

$$\begin{aligned} f_1(x_1, x_2) &:= \mathcal{J}^2 \frac{\partial f}{\partial x_1} = 3a_0 x_1^2 + 2a_1 x_1 x_2 + a_2 x_2^2, \\ f_2(x_1, x_2) &:= \mathcal{J}^2 \frac{\partial f}{\partial x_2} = a_1 x_1^2 + 2a_2 x_1 x_2 + 3a_3 x_2^2, \end{aligned} \quad (17.1.22)$$

This involves finding a solution to six equations in five variables. We have one surplus equation⁶ and one needs to be eliminated. Which this will be will depend on the values of the parameters a_0, a_1, a_2, a_3 . If we were to drop the equation for $x_2^2 f_2(x_1, x_2)$ then the problem is to solve $\mathbf{A}\boldsymbol{\chi} = \boldsymbol{\phi}$, where

⁶ Unlike in the corresponding quartic case when we have exactly the right number; that is six equations and six monomials.

$$\chi := \begin{pmatrix} x_1^4 \\ x_1^3 x_2 \\ x_1^2 x_2^2 \\ x_1 x_2^3 \\ x_2^4 \end{pmatrix}, \quad \phi := \begin{pmatrix} x_1^2 f_1(x_1, x_2) \\ x_1 x_2 f_1^{(2)}(x_1, x_2) \\ x_2^2 f_1(x_1, x_2) \\ x_1^2 f_2(x_1, x_2) \\ x_1 x_2 f_2(x_1, x_2) \end{pmatrix}, \quad (17.1.23)$$

$$A := \begin{pmatrix} 3a_0 & 2a_1 & a_2 & 0 & 0 \\ 0 & 3a_0 & 2a_1 & a_2 & 0 \\ 0 & 0 & 3a_0 & 2a_1 & a_2 \\ a_1 & 2a_2 & 3a_3 & 0 & 0 \\ 0 & a_1 & 2a_2 & 3a_3 & 0 \end{pmatrix}. \quad (17.1.24)$$

Thus a sufficient condition for f to be strongly 3-determinate is that

$$\text{Det}[A] = 3a_2(27a_0^2a_3^2 - 18a_0a_1a_2a_3 + 4a_0a_2^3 + 4a_1^3a_3 - a_1^2a_2^2) \neq 0. \quad (17.1.25)$$

Of course, this is not a necessary condition since, even if this particular matrix A is singular (if, for example, $a_2 = 0$), a non-singular matrix may be achievable by a different choice of five equations. However, suppose that the coefficients are such that (17.1.25) is true. Then f has strong determinancy 3 and we need to derive a universal unfolding. To do so we must compute a basis for $\mathcal{T}^3(f)$, which will consist of some or all of the monomials in x_1 and x_2 of degrees one, two and three. The inclusion of the cubic monomials could involve the quartic terms in f , but if they appear by virtue only of the cubic terms then, as in Example 17.1.3, we have the added bonus that $\mathcal{T}^3(f) = \mathcal{T}^3(\mathcal{J}^4 f) = \mathcal{T}^3(\mathcal{J}^3 f)$. For this to be the case the four monomials x_1^3 , $x_1^2 x_2$, $x_1 x_2^2$ and x_2^3 need to be expressible as linear combinations of the four quantities $x_1 f_1(x_1, x_2)$, $x_2 f_1(x_1, x_2)$, $x_1 f_2(x_1, x_2)$ and $x_2 f_2(x_1, x_2)$. This, as in the calculation for determinancy, is equivalent to the determinant of a matrix being non-zero. In this case the condition is

$$27a_0^2a_3^2 - 18a_0a_1a_2a_3 + 4a_0a_2^3 + 4a_1^3a_3 - a_1^2a_2^2 \neq 0. \quad (17.1.26)$$

If this is satisfied then the only question concerns the quadratic contributions, since $\mathcal{T}^3(f)$ cannot contain linear terms. The two equations (17.1.22) are not sufficient to express all three monomials x_1^2 , $x_1 x_2$ and x_2^2 as linear combinations of $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$. However, it is usually possible to express two of the monomials in terms of linear combinations of $f_1(x_1, x_2)$, $f_2(x_1, x_2)$ and the third monomial showing that $\text{cod}[f] = 3$. In particular it is easy to verify that, if

$$3(a_0a_2 + a_1a_3) - a_1^2 - a_2^2 \neq 0, \quad (17.1.27)$$

$x_1 x_2$ and $x_1^2 - x_2^2$ can be expressed in terms of $x_1^2 + x_2^2$. Thus

$$\mathcal{T}^3(f) \stackrel{\text{B}}{=} \{x_1 x_2, x_1^2 - x_2^2, x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3\} \quad (17.1.28)$$

with cobasis $\{x_1, x_2, x_1^2 + x_2^2\}$ and

$$F(x_1x_2; \mathbf{c}) := c_1x_1 + c_2x_2 + c_2(x_1^2 + x_2^2) + a_0x_1^3 + a_1x_1^2x_2 + a_2x_1x_2^2 + a_3x_2^3 + O(x_1^4x_2^{-i}) \quad (17.1.29)$$

is a universal unfolding of (17.1.21). This is equivalent to J^4F and will be equivalent to J^3F if (17.1.26) is satisfied.

17.1.3 Symmetry Considerations

In the application of catastrophe theory to statistical mechanics, the function of interest, which is the free-energy density, may be restricted by certain symmetry properties. The simplest example of this is the mean-field model for ferromagnetism, where the one parameter is m the magnetization density and the free-energy density with zero field is invariant under the symmetry group $\mathcal{S}_2 : m \rightarrow -m$. The Landau expansion, apart from the linear term in the magnetic field, is thus an expansion in the invariants of $\mathcal{S}_2$, that is the even powers of m . In a similar way the free-energy density for the mean-field model for metamagnetism (Sects. 4.6, 5.5 and 6.2) is, in the absence of a staggered field, an even function of the order parameter q . A more complicated case is that of the 3-state Potts model described in Sect. 5.4.2, where there are now two parameters m and p which satisfy the symmetry group $\mathcal{S}_3$, with generators given by (5.4.4) and (5.4.7). The Landau expansion of the free-energy density now contains only powers of the invariants $v_e(p, m)$ and $v_o(p, m)$ given by (5.4.9) and (5.4.10) respectively. In general

$$f(\mathbf{x}) := \sum_{i=s}^{\infty} a_i v_i(\mathbf{x}), \quad (17.1.30)$$

where $a_s \neq 0$ and the functions $v_i(\mathbf{x})$ are homogeneous in the components of $\mathbf{x}$, of increasing order with respect to i and invariant under the operations of the symmetry group. We shall also suppose that $v_i(\mathbf{0}) = 0$, with the order of $v_s(\mathbf{x})$ being greater than or equal to two, so that f has a critical point at $\mathbf{x} = \mathbf{0}$. Of the three parts of the analysis to: (i) obtain the strong determinacy, (ii) calculate the codimension, (iii) deriving a universal unfolding, only parts (ii) and (iii) are affected by the symmetry restriction. If k_i is the order of $v_i(\mathbf{x})$ and f has strong determinacy k_s then f is strongly equivalent to $a_s v_s(\mathbf{x})$, or indeed to $a_s v_s(\mathbf{x}) + a_{s+1} v_{s+1}(\mathbf{x})$ if we have need to retain a second term. The questions of codimension and unfolding will be treated by means of examples.

Example 17.1.5 Consider the modification

$$f(x) := a_0 + \sum_{i=s}^{\infty} a_{2i} x^{2i}, \quad (17.1.31)$$

of (17.1.12) when f is an even function, which is thus invariant under the symmetry group $\mathcal{S}_2 : x \rightarrow -x$. As already seen in Example 17.1.3, f has strong determinancy $2s$. The definitions (17.1.6) and (17.1.7) for codimension must now be modified to take into account the restriction to even functions.

$$\frac{2}{\text{cod}}[f] = \dim[\frac{2}{\mathcal{J}_1^{2s}}] - \dim[\frac{2}{\mathcal{T}_{2s}(f)}], \quad (17.1.32)$$

where $\frac{2}{\mathcal{J}_1^{2s}}$ and $\frac{2}{\mathcal{T}_{2s}(f)}$ are the subspaces of $\mathcal{J}_1^{2s}$ and $\mathcal{T}_{2s}(f)$ consisting of polynomials invariant under $\mathcal{S}_2$; that is even polynomials. Thus, from (17.1.2) and (17.1.16),

$$\dim[\frac{2}{\mathcal{J}_1^{2s}}] = s, \quad \dim[\frac{2}{\mathcal{T}_{2s}(f)}] = 1, \quad (17.1.33)$$

giving $\frac{2}{\text{cod}}[f] = s - 1$. The $(s - 1)$ -unfolding

$$F(x; c) := a_0 + \sum_{j=1}^{s-1} a_{2j} x^{2j} + \sum_{i=s}^{\infty} a_{2i} x^{2i} \quad (17.1.34)$$

of (17.1.31) is universal in the sense that the space spanned by $\{x^2, x^4, \dots, x^{2(s-1)}\}$ is transverse to $\frac{2}{\mathcal{T}_{2s}(f)}$ in $\frac{2}{\mathcal{J}_1^{2s}}$ and

$$\tilde{F}(x; c) := a_0 + \sum_{j=1}^{s-2} a_{2j} x^{2j} + a_{2s} x^{2s} + a_{2s+2} x^{2s+2}, \quad (17.1.35)$$

is a universal unfolding of $\mathcal{J}^{2s+2} f$. Again the unfolding of $\mathcal{J}^{2s} f$ with the term of degree $2s + 2$ omitted from (17.1.35) is equivalent to (17.1.34).

Example 17.1.6 At the critical point in the 3-state Potts model the leading term in the Landau expansion of the free-energy density is the cubic invariant $v_o(x_1, x_2) = x_1(x_1^2 - 3x_2^2)$. Thus we are interested in

$$f(x_1, x_2) := v_o(x, x_2) + a[v_E(x_1, x_2)]^2 + \text{terms in combinations of } v_o(x, x_2) \text{ and } v_E(x, x_2) \text{ of increasing degree}, \quad (17.1.36)$$

where $v_E = x_1^2 + x_2^2$. This is the special case of Example 17.1.4 with $a_0 = 1$, $a_1 = 0$, $a_2 = -3$ and $a_3 = 0$. Condition (17.1.25) is satisfied and f has strong determinancy of 3. The inequalities (17.1.26) and (17.1.27) are also satisfied so $\text{cod}[f] = 3$. However, we are interest in

$$\frac{3}{\text{cod}}[f] = \dim[\frac{3}{\mathcal{J}_2^3}] - \dim[\frac{3}{\mathcal{T}_3(f)}], \quad (17.1.37)$$

where

$$\frac{3}{\mathcal{J}_2^3} \stackrel{\text{B}}{=} \{v_E(x_1, x_2), v_O(x_1, x_2)\}, \quad \frac{3}{\mathcal{T}_3(f)} \stackrel{\text{B}}{=} \{v_O(x_1, x_2)\}, \quad (17.1.38)$$

are the subspaces of $\mathcal{J}_2^3$ and $\mathcal{T}_3(f)$ consisting of polynomials invariant under $\mathcal{S}_3$. Thus $\frac{3}{\text{cod } [f]} = 1$ and

$$F(x_1, x_2; c) := cv_E(x_1, x_2) + v_O(x, x_2) + a[v_E(x_1, x_2)]^2 + \text{terms in combinations of } v_O(x, x_2) \text{ and } v_E(x, x_2) \text{ of increasing degree}, \quad (17.1.39)$$

is a universal of unfolding of f and this is equivalent to

$$\tilde{F}(x_1, x_2; c) := cv_E(x_1, x_2) + v_O(x, x_2) + a[v_E(x_1, x_2)]^2. \quad (17.1.40)$$

17.2 Matrix Algebra

All matrices in this section are taken to be square and of dimension M with possibly complex elements. A^T denotes the transpose of A and $\bar{A}$ the matrix obtained from A by taking the conjugate complex of every element. $A^\dagger := \bar{A}^T$ is the *Hermitian transpose* of A . If $A^\dagger = A$, then A is said to be *Hermitian* and, if $A^\dagger = -A$, to be *skew-Hermitian*. If $\text{Det}\{A\} \neq 0$ then A is *non-singular*, the rows and columns of A form sets of independent vectors, none of its eigenvalues is zero and the inverse A^{-1} of A exists. Henceforth, in this section we suppose that all the matrices are non-singular. A matrix A is *orthogonal* if it is real with $A^{-1} = A^T$ and it is *unitary* if $A^{-1} = A^\dagger$. Thus a real unitary matrix is orthogonal. For an orthogonal or unitary matrix A , $|\text{Det}\{A\}| = 1$.

The set of distinct eigenvalues of a matrix is its *spectrum* and the *algebraic multiplicity* of an eigenvalue is its multiplicity as a root of the characteristic equation. Clearly the sum of the algebraic multiplicities of all the eigenvalues is M . Eigenvectors associated with different eigenvalues are independent (Lancaster and Tismenetsky 1985, p. 150) and the *geometric multiplicity* of an eigenvalue is the number of independent eigenvectors with which it is associated. The geometric multiplicity of an eigenvalue does not exceed its algebraic multiplicity (Lancaster and Tismenetsky 1985, p. 160) and if they are equal it is called the *degeneracy* of the eigenvalue, with the eigenvalue said to be degenerate or non-degenerate according as its degeneracy is greater than or equal to one. A matrix is called *simple* if, for every eigenvalue, the geometric multiplicity is equal to the algebraic multiplicity; meaning that it has M independent eigenvectors. Non-simple matrices are called *defective*. If an eigenvalue of a simple matrix is non-degenerate then the eigenvector associated with it is (to within a multiplicative factor) uniquely given; if it has degeneracy ω then the eigenvectors associated with it span a subspace of dimension ω . Any set of vectors spanning this subspace can be chosen as the eigenvectors for this eigenvalue.

Matrices A and A' are *similar* if there exists a matrix U such that

$$A' = U^{-1}AU. \quad (17.2.1)$$

Such matrices have the same spectrum and if x is a right eigenvector of A associated with the eigenvalue Λ then $x' := U^{-1}x$ is a right eigenvector of A' also associated with Λ (Lancaster and Tismenetsky 1985, p. 152). More generally if

$$A' = \mu U^{-1}AU, \quad (17.2.2)$$

for some scalar μ , then the eigenvalues of A' are μ times the eigenvalues of A .

17.2.1 Diagonalizability

A matrix is said to be *diagonalizable* if it is similar to a diagonal matrix. The matrix A is diagonalizable if and only if it is simple (Lancaster and Tismenetsky 1985, p. 146), with

$$X^{-1}AX = \Lambda, \quad (17.2.3)$$

$$A = X\Lambda X^{-1}, \quad (17.2.4)$$

where Λ is the diagonal matrix with eigenvalues along the diagonal and the columns of X are the right eigenvectors ordered to make the correct associations with the eigenvalues. The rows of $Y := X^{-1}$ are the corresponding left eigenvectors.

A matrix A for which $[A, A^\dagger] = 0$ is said to be *normal*; unitary, Hermitian, skew-Hermitian and diagonal matrices are all normal. A matrix is normal if and only if it is diagonalizable using a unitary matrix (Lancaster and Tismenetsky 1985, p. 175); so that (17.2.3) becomes $X^\dagger AX = \Lambda$. Not all matrices are unitarily diagonalizable, but they are all unitarily similar to an upper triangular matrix (Lancaster and Tismenetsky 1985, p. 176).

It should be noted that diagonalizability is stronger than simply the condition that the matrix can be made diagonal by some product with other matrices; for a non-singular matrix this can be done in many ways.

Example 17.2.1 Let A , B and the products AB and BA all be simple and X_{AB} and X_{BA} be, respectively, the matrices containing the right eigenvectors of AB and BA as columns with diagonal matrices Λ_{AB} and Λ_{BA} containing the respective sets of eigenvalues. Then it is simple to show that

$$\Lambda_{AB} X_{AB}^{-1} A X_{BA} = X_{AB}^{-1} A X_{BA} \Lambda_{BA}, \quad \Lambda_{BA} X_{BA}^{-1} B X_{AB} = X_{BA}^{-1} B X_{AB} \Lambda_{AB}, \quad (17.2.5)$$

from which it follows that $\Lambda_{AB} = \Lambda_{BA}$ and (unless these matrices are just some multiple of the unit matrix) that $X_{BA}^{-1} \mathbf{B} X_{AB}$ and $X_{AB}^{-1} \mathbf{A} X_{BA}$ are diagonal.

The complete eigen-problem can be summarized in the equations⁷

$$\begin{aligned}
 \mathbf{x}_m &:= X|m\rangle, & \mathbf{y}_m &:= [\langle m|Y]^\top, \\
 \delta^{\text{Kr}}(m - m') \Lambda_m &:= \langle m|\Lambda|m'\rangle, \\
 \mathbf{A}\mathbf{x}_m &= \mathbf{x}_m \Lambda_m, & [\mathbf{y}_m]^\top \mathbf{A} &= \Lambda_m [\mathbf{y}_m]^\top, \\
 \mathbf{x}_m \cdot \mathbf{y}_{m'} &= \delta^{\text{Kr}}(m - m').
 \end{aligned} \tag{17.2.6}$$

For a normal matrix the eigenvectors are orthonormal, meaning that $\mathbf{y}_m = \bar{\mathbf{x}}_m$.

A matrix containing as columns a complete set of right eigenvectors of $\mathbf{A}$ is called an *eigenmatrix* of $\mathbf{A}$. A simple matrix has at least one eigenmatrix. Let $\mathcal{S}_E(\mathbf{A})$ be the set of eigenmatrices of $\mathbf{A}$. The eigenvalues of $\mathbf{A}$ are all non-degenerate if and only if the eigenmatrix is unique, that is $\mathcal{S}_E(\mathbf{A})$ has exactly one element. Otherwise there will be a set of eigenmatrices formed by taking different combinations of columns in the degenerate subspaces.

17.2.2 Commutativity

For the simple (non-identity, non-singular) matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$:

- (i) $[\mathbf{B}, \mathbf{C}] = \mathbf{0}$ if and only if $\mathbf{B}$ and $\mathbf{C}$ have a complete set of common eigenvectors; that is $\mathcal{S}_E(\mathbf{B}, \mathbf{C})$, denoting the set of common eigenmatrices of $\mathbf{B}$ and $\mathbf{C}$, has at least one member (Lancaster and Tismenetsky 1985, p. 420). If $\mathbf{B}$ (say) has no degenerate eigenvalues then $\mathcal{S}_E(\mathbf{B}, \mathbf{C})$ has exactly one member; though this does not mean that $\mathbf{C}$ has no degenerate eigenvalues.
- (ii) In general non-commuting matrices $\mathbf{B}$ and $\mathbf{C}$ can have some common eigenvectors, but if $\mathbf{BC} = \alpha \mathbf{CB}$ for some scalar $\alpha \neq 1$ (called α -related) they will have none. A particular case is anticommuting matrices for which $\alpha = -1$.
- (iii) All, none or exactly one of $[\mathbf{A}, \mathbf{B}] = \mathbf{0}$, $[\mathbf{A}, \mathbf{C}] = \mathbf{0}$ and $[\mathbf{A}, \mathbf{BC}] = \mathbf{0}$ holds.
- (iv) If $[\mathbf{A}, \mathbf{B}] = \mathbf{0}$ and $[\mathbf{A}, \mathbf{C}] = \mathbf{0}$ then:
 - (a) Every non-degenerate eigenvector of $\mathbf{A}$ is an eigenvector of $\mathbf{B}$ and $\mathbf{C}$.
 - (b) If every eigenvalue of $\mathbf{A}$ is non-degenerate then $[\mathbf{B}, \mathbf{C}] = \mathbf{0}$. But not conversely; that is, $[\mathbf{B}, \mathbf{C}] = \mathbf{0}$ does not imply that $\mathbf{A}$, or indeed $\mathbf{B}$ or $\mathbf{C}$, has no degenerate eigenvalues.
 - (c) If $\mathbf{B}$ and $\mathbf{C}$ are α -related all eigenvalues of $\mathbf{A}$ are degenerate.

⁷ Here and elsewhere in the book it is convenient to use the notation $\langle i|\mathbf{F}|j\rangle$ to denote the i, j th element of any matrix $\mathbf{F}$ with its i th row and j th column denoted respectively by $\langle i|\mathbf{F}$ and $\mathbf{F}|j\rangle$.

17.2.3 Reducibility

Let $\mathbf{e}_m$ be the unit vector with $\langle j | \mathbf{e}_m = \delta^{jr}(m - j)$. A permutation matrix $\mathbf{P}$ has columns consisting of the M unit vectors $\{\mathbf{e}_m\}$ in some order. $\mathbf{P}$ is clearly orthogonal (and unitary). Premultiplication and postmultiplication of a matrix $\mathbf{A}$ by $\mathbf{P}$ permutes the rows and columns respectively. A matrix $\mathbf{A}$ is *reducible* if there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix} \quad (17.2.7)$$

where the diagonal blocks $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square and of dimension less than M . If such a permutation matrix does not exist then $\mathbf{A}$ is said to be *irreducible*. The eigenvalues of $\mathbf{A}$, included according to their algebraic multiplicities, are the eigenvalues of $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$. If $\mathbf{A}_{11}$ or $\mathbf{A}_{22}$ are themselves reducible they can be reduced by permutation to generalize (17.2.7) to a form with diagonal blocks which are irreducible.

It can be shown (Gantmacher 1959, p. 63) that a matrix is irreducible if and only if, for any pair of indices m and m' , there exists a positive integer n such that $\langle m | \mathbf{V}^n | m' \rangle > 0$.

17.2.4 Theorems of Perron and Frobenius

Here we restrict attention to a real simple matrix $\mathbf{V}$, which is either positive or non-negative. Clearly a positive matrix is irreducible and for such, according to a theorem of Perron (1907) (Gantmacher 1959, p. 64):

Theorem 17.2.1 *The eigenvalue of a positive matrix $\mathbf{V}$ of largest magnitude is positive and non-degenerate with associated positive eigenvector.*

This is an important result for statistical mechanics since the transfer matrices of many systems, for example the Ising model, are positive matrices.

However, other models, in particular vertex models and models with some kind of first-neighbour exclusion, have transfer matrices which are merely non-negative. Then according to a theorem of Frobenius (1912) (Gantmacher 1959, p. 65):

Theorem 17.2.2 *A non-negative matrix $\mathbf{V}$, which is also irreducible, has a (real and positive) eigenvalue equal to the spectral radius*

$$\mathcal{SR}(\mathbf{V}) := \max_{\{m\}} |\Lambda_m|, \quad (17.2.8)$$

with associated positive eigenvector. The set of points in a complex plane which represent the spectrum is invariant under rotation about the origin with the smallest angle for which this is the case being $2\pi/p$, where $p > 0$ is an integer called the period of the matrix.

The matrix is said to be *periodic* or *aperiodic* according to whether $p > 1$ or $p = 1$. The non-negative matrix V is said to be *primitive* if V^n is positive for some positive integer n . Clearly a primitive matrix must be irreducible and it can be shown (Gantmacher 1959, p. 97) that V is primitive if and only if it is aperiodic; the period of a matrix is also call its *index of imprimitivity*. Theorem 17.2.1 applies with the restriction to a ‘positive matrix’ replaced by that to a ‘primitive matrix’.

For a periodic matrix, p must be a divisor of M and the eigenvalues fall into sets

$$\Lambda_{sp+1} \exp(2\pi i k/p), \quad k = 0, 1, \dots, p-1, \quad s = 0, 1, \dots, r-1, \quad (17.2.9)$$

where $M = rp$, with the set with $s = 0$ lying on the spectral circle. It follows that

$$\text{Trace}\{V^n\} = \left[\sum_{s=0}^{r-1} \Lambda_{sp+1} \right] \left[\sum_{k=0}^{p-1} \exp(2\pi i nk/p) \right] = \widetilde{\delta^{\text{Kr}}}(p; n) \sum_{s=0}^{r-1} \Lambda_{sp+1}, \quad (17.2.10)$$

where $\widetilde{\delta^{\text{Kr}}}(p; n)$ is given by (18.1.10). The matrix V^n is traceless unless n is a multiple of p .⁸

In the characteristic equation

$$\text{Det}\{\Lambda I - V\} := \sum_{m=0}^M v_m \Lambda^m = 0 \quad (17.2.11)$$

the coefficient v_m is a product of $M - m$ elements of V . Since, from Theorem 17.2.2, the spectrum is invariant under rotation through a smallest angle $2\pi/p$, it must be the case that $\exp(2\pi i [M - m]/p) = 1$, for all m with $v_m \neq 0$. So p is the greatest common divisor of the set $\{M - m\}$, for all m with $v_m \neq 0$.

There remains the case where the non-negative matrix V is reducible. We first note that if the eigenvalue $\Lambda_1 = \mathcal{SR}(V)$ has degeneracy $\omega > 1$ then V must be reducible. Suppose that it is reduced in the form which is a generalisation of (17.2.7) with irreducible diagonal blocks V_{jj} , $j = 1, 2, \dots, q$. Since the eigenvalues of V , counting according to their degeneracies coincide with the eigenvalues of the diagonal blocks, $q \geq \omega$ with ω blocks (let them be V_{jj} , $j = 1, 2, \dots, \omega$) having spectral radius $\mathcal{SR}(V)$ and a non-degenerate eigenvalue Λ_1 . The eigenvalues on the spectral circle of V are just the total of all the eigenvalues on the spectral circles of V_{jj} , $j = 1, 2, \dots, \omega$.

⁸ For $n = 1$ this conclusion is part of Frobenius’ theorem, where it shown that the elements of the matrix can be permuted to give a set of square zero matrices along the diagonal (Gantmacher 1959, p. 65).

17.2.5 Direct Products and Traces

In Chap. 12 we study matrices like $\mathbf{Q}_k$, whose rows and columns are indexed by vectors $\boldsymbol{\varsigma} := (\varsigma_1, \varsigma_2, \dots, \varsigma_N)$, where $\varsigma_\ell = \pm 1$, and for which

$$\langle \boldsymbol{\varsigma} | \mathbf{Q}_k | \boldsymbol{\varsigma}' \rangle = \text{Trace} \left\{ \prod_{\ell=1}^N \mathbf{T}_k \begin{pmatrix} \varsigma'_\ell \\ \varsigma_\ell \end{pmatrix} \right\}. \quad (17.2.12)$$

Then it is straightforward to show (Lloyd 1953) that

$$\langle \boldsymbol{\varsigma} | \mathbf{Q}_1 \mathbf{Q}_2 | \boldsymbol{\varsigma}' \rangle = \text{Trace} \left\{ \prod_{\ell=1}^N \sum_{q=\pm 1} \mathbf{T}_1 \begin{pmatrix} q \\ \varsigma_\ell \end{pmatrix} \otimes \mathbf{T}_2 \begin{pmatrix} \varsigma'_\ell \\ q \end{pmatrix} \right\}. \quad (17.2.13)$$

It should be noted that there is no necessary relationship between the dimension of the matrices $\mathbf{Q}_k$ and $\mathbf{T}_k$, with, in the latter case, a possible variation with k . Suppose $\mathbf{T}_k$ is of dimension L_k with rows and columns indexed by i_k and j_k respectively. Then the most convenient way to index the rows and columns of a direct product is using double indices.

$$\left\langle i_2 \left| \mathbf{T}_1 \begin{pmatrix} q \\ \varsigma_\ell \end{pmatrix} \otimes \mathbf{T}_2 \begin{pmatrix} \varsigma'_\ell \\ q \end{pmatrix} \right| j_2 \right\rangle = \langle i_1 | \mathbf{T}_1 \begin{pmatrix} q \\ \varsigma_\ell \end{pmatrix} | j_1 \rangle \langle i_2 | \mathbf{T}_2 \begin{pmatrix} \varsigma'_\ell \\ q \end{pmatrix} | j_2 \rangle. \quad (17.2.14)$$

17.2.6 Defective Matrices

So far in the discussion we have considered only simple matrices where the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity. Consider now the case where a particular eigenvalue Λ has algebraic multiplicity ω_Λ and geometric multiplicity $\omega_G < \omega_\Lambda$. For this particular eigenvalue the reduction to diagonal form is replaced by ω_G blocks. These may be of dimension one, and simply contain the eigenvalue, or be of dimension greater than one, in which case they have Λ along the diagonal and one along the upper diagonal. The sum of the dimensions of the blocks is ω_Λ . Such a matrix is said to be of *Jordan normal form* and it can be shown (Lancaster and Tismenetsky 1985, Sect. 6.5) that:

Theorem 17.2.3 *Any complex square matrix is similar to a matrix of Jordan normal form, with the latter matrix being unique apart from the ordering of the blocks.*

It is also straightforward to show that:

Theorem 17.2.4 *If $\mathbf{A}$ and $\mathbf{B}$ are commuting complex square matrices, then every non-degenerate eigenvector of $\mathbf{A}$ is an eigenvector of $\mathbf{B}$.*

17.2.7 Groups of Matrices

The *orthogonal group* of n -dimensional orthogonal matrices is denoted by $O(n)$. The *special orthogonal group* of n -dimensional orthogonal matrices with determinant $+1$ is denoted by $SO(n)$. The effect of $A \in O(n)$ is a distance-preserving mapping on n -dimensional Euclidian space which leave the origin invariant. If $\text{Det}\{A\} = 1$, that is $A \in SO(n)$, then A is a rotation. If $\text{Det}\{A\} = -1$ then it is a combination of a rotation and a reflection. The *unitary group* of n -dimensional unitary matrices is denoted by $U(n)$. $U(1)$ is just multiplication by complex numbers of unit modulus, which can be re-expressed in two-dimensional matrix form showing that $U(1)$ is isomorphic to $SO(2)$.

17.3 Groups and Representations

A generic designation of a group⁹ is used in the form $\mathcal{Q} := \{Q^{(0)}, Q^{(1)}, Q^{(2)}, \dots\}$ with the group operation represented multiplicatively and $Q^{(0)} := I$, the identity element. A *generating set* of a group is a subset such that every element of the group can be expressed as the combination (under the group operation) of finitely many elements of the subset and their inverses and we use the notation $\mathcal{Q} := \langle Q_1, Q_2, \dots, Q_k \rangle$ to denote the group in terms of its set of generators. Apart from the case of the n -vector model in Sect. 3.3, we are concerned simply with finite groups and the order of $\mathcal{Q}$ is denoted by $n(\mathcal{Q})$.

The use of group representation theory in statistical mechanics arises because of the presence of certain symmetries in the Hamiltonian. Specifically:

(A) *Translations, rotations or reflections of the lattice.* The rotations and reflections of a lattice form a *point group* which leaves one point P fixed.¹⁰ A subgroup of all point groups is:

- A *cyclic group* $\mathcal{C}_n := \langle C_n \rangle = \{I, C_n, C_n^2, \dots, C_n^{n-1}\}$, where C_n is a rotation through $2\pi/n$ radians about an axis A through the specified point.¹¹

For infinite three-dimensional lattices and two-dimensional lattices with A orthogonal to the plane of the lattice the only n -fold rotational symmetries consistent with translational symmetry are for $n = 1, 2, 3, 4, 6$ (Falicov 1966, pp. 79–81). However, we shall be interested in transfer matrix theory for lattices,

⁹ For the discussion in this section a working knowledge of group theory is assumed. For more extensive treatments of symmetry groups and their representations and for proofs of theorems see, for example, McWeeny (1963), Falicov (1966), Leech and Newman (1969), Serre (1977) or Lavis and Bell (1999b, Sect. A.3).

¹⁰ The choice of point is arbitrary since they are related by translation.

¹¹ Any group with a single generator is a cyclic group. In this case the cyclic group is also a rotation group, meaning that it is physically realized as a set of rotation in three-dimensional space.

with periodic boundary conditions, forming a cylinder, where $\mathbf{A}$ is the axis of the cylinder (Sect. 17.3.1). In this case n can be arbitrarily large, so the discussion is formulated in this context. The other uniaxial point groups are:

- $\mathcal{C}_{nh} := \langle C_n, \Sigma_h \rangle$, where Σ_h is a reflection in a plane Π_h through $\mathbf{P}$ and orthogonal to $\mathbf{A}$.
- $\mathcal{J}_n := \langle J_n \rangle$, where $J_n = C_n \Sigma_h$. This is the group of improper rotations.¹² Clearly $J_n^m = C_n^m$ when m is even and $J_n^m = C_n^m \Sigma_h$, when m is odd. Thus, if n is odd $\mathcal{J}_n = \mathcal{C}_{nh}$, but if n is even $\mathcal{J}_n$ is a proper subgroup of $\mathcal{C}_{nh}$.
- $\mathcal{C}_{nv} := \langle C_n, \Sigma_0 \rangle$, where Σ_0 is a reflection in a plane Π_0 containing the axis $\mathbf{P}$. This group contains reflections $\Sigma_m := C_n^m \Sigma_0 C_n^{-m}$ in the planes $\Pi_m := C_n^m \Pi_0$, for $m = 0, 1, \dots, \frac{n}{2} - 1$, when n is even and $m = 0, 1, \dots, n - 1$, when n is odd.

This completes the uniaxial point groups.¹³ Now return to the cyclic group $\mathcal{C}_n$ and add an axis $\mathbf{A}_0$ through $\mathbf{P}$ and in the plane Π_h about which there is a two-fold rotation $\mathcal{U}_0$. This gives:

- The *dihedral group* $\mathcal{D}_n := \langle C_n, \mathcal{U}_0 \rangle$. In a way similar to that for $\mathcal{C}_{nv}$, $\mathcal{D}_n$ contains two-fold rotations $\mathcal{U}_m := C_n^m \mathcal{U}_0 C_n^{-m}$ about the axes $\mathbf{A}_m := C_n^m \mathbf{A}_0$, for $m = 0, 1, \dots, \frac{n}{2} - 1$, when n is even and $m = 0, 1, \dots, n - 1$, when n is odd.

Reflections can now be added to the dihedral group:

- $\mathcal{D}_{nh} := \langle C_n, \mathcal{U}_0, \Sigma_h \rangle$.
- $\mathcal{D}_{nd} := \langle C_n, \mathcal{U}_0, \Sigma'_0 \rangle$, where Σ'_0 is a reflection through a plane Π'_0 which bisects the angle between planes Π_0 and Π_1 defined for $\mathcal{C}_{nv}$. Again it is clear that the group contains planes of reflection bisecting the angles between the planes Π_m and Π_{m+1} , for $m = 0, 1, \dots, \frac{n}{2} - 1$, when n is even and $m = 0, 1, \dots, n - 1$, when n is odd.

This completes the set of point groups which have at most one axis of n -fold rotation with $n > 2$. It also encompasses, with $n \leq 6$, twenty-seven of the thirty-two crystallographic point groups.¹⁴ The remaining point groups for three-dimensional lattices are the cubic groups with more than one n -fold axis with $n > 2$. There are five of these with $n \leq 6$, thus completing the set of thirty-two crystallographic point groups (Falicov 1966, pp. 88–92).

All the group operations defined here are *isometries*, that is, they preserve distances, and an isometry is said to be *direct* if it preserves orientation or ordering.

¹² $\mathcal{S}_n$ is usually used in place of $\mathcal{J}_n$, but this clashes with the notation for the symmetric group.

¹³ With $n = 1, 2, 3, 4, 6$, and given that $\mathcal{C}_{1v} = \mathcal{C}_{1h}$, $\mathcal{J}_1 = \mathcal{C}_{1h}$ and $\mathcal{J}_3 = \mathcal{C}_{3h}$ this gives $4 \times 5 - 3 = 17$ of the crystallographic point groups.

¹⁴ With $n = 1, 2, 3, 4, 6$, and given that $\mathcal{D}_1 = \mathcal{C}_2$, $\mathcal{D}_h = \mathcal{C}_{2v}$, $\mathcal{D}_{1d} = \mathcal{C}_{2h}$ and $\mathcal{D}_{4d}$ and $\mathcal{D}_{6d}$ are excluded because they have respectively improper rotations with $n = 8$ and $n = 12$, this gives $3 \times 5 - 5 = 10$ additional crystallographic point groups.

In this context we can think of this in terms of the ordering (clockwise or anticlockwise) of sites round a lattice face with respect to a particular (unaltered direction). An isometry not preserving this property is called *indirect*. In the case of two-dimensional lattices with the only group operations in the plane of the lattice the symmetry group consists of translations and the point group $\mathcal{C}_n$, for some n . Translations and rotations are direct isometries and reflections are indirect isometries.

- (B) *Operations on the microstates*. The largest group that can operate on a system of ν discrete microstates is the symmetric group $\mathcal{S}_\nu$ of all permutations. This applies to the zero-field standard ν -state Potts model (Sect. 3.6.4). In other cases the symmetry may be reduced, as for the zero-field ν -state planar clock model (Sect. 3.6) which is invariant under the cyclic group $\mathcal{C}_\nu$. The only case of continuous symmetry which will concern us is the n -vector model of Sect. 3.3 (and the special case $n = 2$ for the XY model in Sect. 3.6.1) for which the symmetry group is $\mathcal{O}(n)$ the group of orthogonal n -dimensional matrices (Sect. 17.2.7).
- (C) *Operations on the independent couplings*. The primary instance of this is for operations on the direction of the magnetic field. In particular its reversal of direction, which for the Ising model simply means a change of sign.

The total symmetry group $\mathcal{Q}$ will, in general, be the direct product of the groups representing the different symmetries. However, there are exceptions to this, so that if $\mathcal{F} \times \mathcal{G}$ is the direct product of two of the three symmetry types listed above, the Hamiltonian is invariant under only a proper subgroup of $\mathcal{F} \times \mathcal{G}$. The simplest example of this occurs in the spin- $\frac{1}{2}$ Ising model (Sect. 3.5.1) where the group of operations on the states and on the field consist of inversion (change of sign), but where the Hamiltonian is invariant only when the spins and the field are inverted *simultaneously*.

17.3.1 Representations

For a discussion of representations the idea of conjugacy is important. Two elements $Q^{(s)}, Q^{(s')} \in \mathcal{Q}$ are said to be *conjugate* if there exists an element $Q^{(s'')} \in \mathcal{Q}$ such that $Q^{(s)} = [Q^{(s'')}]^{-1} Q^{(s')} Q^{(s')}$. The set of all elements conjugate to a particular element is called a *conjugacy class* of $\mathcal{Q}$ and any two elements are in the same conjugacy class if and only if they are conjugate; $\mathcal{Q}$ decomposes into a set of $\kappa(\mathcal{Q})$ mutually exclusive conjugacy classes $\mathcal{Q}_k, k = 1, 2, \dots, \kappa$. Any element which commutes with all the other elements of the group is conjugate only to itself; if $\mathcal{Q}$ is commutative each element is a conjugacy class.

If, for all s , a square $M \times M$ dimensional matrix $\mathbf{R}(Q^{(s)}) := \mathbf{Q}^{(s)}$ can be assigned to $Q^{(s)}$ such that $\mathbf{Q}^{(s)} \mathbf{Q}^{(s')} = \mathbf{Q}^{(s'')}$ whenever $Q^{(s)} Q^{(s')} = Q^{(s'')}$, then the set of matrices $\mathcal{R}(\mathcal{Q}) := \{\mathbf{I}, \mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(n-1)}\}$ is an M -dimensional *representation* of $\mathcal{Q}$. It must, of course, be the case that $\mathbf{R}(\mathbf{I}) = \mathbf{I}$, the M -dimensional unit matrix, and $\mathbf{R}([Q^{(s)}]^{-1}) = [\mathbf{Q}^{(s)}]^{-1}$. For ease of notation $\mathbf{Q}^{(0)}$ is taken to denote $\mathbf{I}$, where

necessary. It is clear that, for any group, one possible assignment is $\mathbf{R}(Q^{(s)}) = (1)$, for ever s . This representation (usually denoted by A) is called the *one-dimensional symmetric representation*. If two representations, of the same dimension, $\mathcal{R}(Q)$ and $\mathcal{R}'(Q)$ are related by the similarity transformation $\mathbf{R}'(Q^{(s)}) = \mathbf{U}^{-1} \mathbf{R}(Q^{(s)}) \mathbf{U}$, for all s , then the representations are said to be *equivalent*. A representation $\mathcal{R}(Q)$ is said to be *reducible*¹⁵ if it is equivalent to another representation $\mathcal{R}'(Q)$ whose matrices have the block-diagonal form

$$\mathbf{R}'(Q^{(s)}) = \mathbf{R}_1(Q^{(s)}) \oplus \mathbf{R}_2(Q^{(s)}) \oplus \mathbf{R}_3(Q^{(s)}) \oplus \cdots, \quad (17.3.1)$$

where the blocks are of the same, non-zero, dimension for all $Q^{(s)}$. It is clear that, for any i , the set of matrices $\mathbf{R}_i(Q^{(s)})$, $s = 0, 1, \dots, n-1$ is itself a representation. Any representation which is not reducible is call *irreducible* and it follows that any reducible representation is equivalent to a representation which has matrices consisting of blocks which are irreducible representations.

Unitary matrices, yielding *unitary representations*, play a particularly important role in the theory of group representations. In particular, a representation in terms of permutation matrices is unitary. The following theorems are of importance:

Theorem 17.3.1 *Any representation of a finite group in terms of non-singular matrices is equivalent to a unitary representation.*

Theorem 17.3.2 *Any two equivalent unitary representations can be related by a unitary transformation.*

We shall henceforth assume that the matrices of all representations are unitary. In particular it follows from Theorem 17.3.1 that all the irreducible representations can be expressed in terms of unitary matrices. For one-dimensional irreducible representations this form is, of course, the unique expression of the representation. A one-dimensional matrix (η) belonging to a one-dimensional representation of a group must satisfy the unitary condition $\eta^{-1} = \bar{\eta}$, from which it follows that η must be a complex root of unity. For the representation $\mathcal{R}(Q)$,

$$\chi(Q^{(s)}) := \text{Trace}\{\mathbf{Q}^{(s)}\} \quad (17.3.2)$$

is the *character* of $Q^{(s)}$ and the set of characters $\chi(Q^{(s)})$, $s = 0, 1, \dots, n-1$ is called the *character system* of Q in $\mathcal{R}(Q)$. Since $\text{Trace}\{\mathbf{P}^{-1} \mathbf{Q}^{(s)} \mathbf{P}\} = \text{Trace}\{\mathbf{Q}^{(s)}\}$, it follows that all equivalent representations have the same character set. The same type of argument also shows that all operators of the group in the same conjugacy class have the same character. We can conveniently regard all equivalent representations as essentially the *same* representation *realized* in different forms. This representation has a character set, with a character given for each conjugacy class. An important property of a group is the collection of character sets for the irreducible representations with the useful result that:

¹⁵ ‘Reducibility’ for a representation has a slightly different meaning from that given in (17.2.7) for a single matrix.

Theorem 17.3.3 *The number of (inequivalent) irreducible representations of a group is equal to the number of conjugacy classes in the group.*

We denote the irreducible representations of $\mathcal{Q}$ by $\Theta(k)$, $k = 1, 2, \dots, \kappa$, where $\Theta(1) = A$ is the one-dimensional symmetric representation, and $e(k)$ is the dimension of $\Theta(k)$. If we denote the (unitary) matrix for $Q^{(s)}$ in the irreducible representation $\Theta(k)$ by $\mathbf{R}^{(k)}(Q^{(s)})$ then the character of $Q^{(s)}$ in the representation $\Theta(k)$ is $\chi^{(k)}(Q^{(s)}) = \text{Trace}\{\mathbf{R}^{(k)}(Q^{(s)})\}$. It is clear that

$$\chi^{(k)}(Q^{(0)}) = e(k), \quad k = 1, 2, \dots, \kappa \quad (17.3.3)$$

and, since the modulus of the trace of a unitary matrix cannot exceed its dimension,

$$|\chi^{(k)}(Q^{(s)})| \leq e(k), \quad s = 1, 2, \dots, n-1, \quad k = 1, 2, \dots, \kappa, \quad (17.3.4)$$

with

$$|\chi^{(k)}(Q^{(s)})| = 1, \quad s = 0, 1, \dots, n-1, \quad \text{if } e(k) = 1. \quad (17.3.5)$$

It can be shown that:

Theorem 17.3.4 *If the elements of $\mathbf{R}^{(k)}(Q^{(s)})$ are denoted by $R_{ij}^{(k)}(Q^{(s)})$ for $i, j = 1, 2, \dots, e(k)$ and $\omega(s)$ is the number of elements in the conjugacy class containing $Q^{(s)}$ then*

$$\sum_{s=0}^{n-1} \bar{R}_{ij}^{(k)}(Q^{(s)}) R_{i'j'}^{(k')}(Q^{(s)}) = \frac{n}{e(k)} \delta^{kr} (k - k') \delta^{kr} (i - i') \delta^{kr} (j - j'), \quad (17.3.6)$$

$$\sum_{k=1}^{\kappa} \bar{\chi}^{(k)}(Q^{(s)}) \chi^{(k)}(Q^{(s')}) = \begin{cases} n/\omega(s), & Q^{(s)} \text{ and } Q^{(s')} \text{ conjugate,} \\ 0, & Q^{(s)} \text{ and } Q^{(s')} \text{ not conjugate.} \end{cases} \quad (17.3.7)$$

Equations (17.3.6) and (17.3.7) are the *Schur orthogonality relations*. From (17.3.6),

$$\sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)} Q^{(s')}) \chi^{(k')}(Q^{(s)} Q^{(s'')}) = \frac{n}{e(k)} \delta^{kr} (k - k') \chi^{(k)}([Q^{(s')}]^{-1} Q^{(s'')}). \quad (17.3.8)$$

Since $Q^{(0)} = I$ is in a conjugacy class on its own, (17.3.7), with $s = s' = 0$, yields

$$\sum_{k=1}^{\kappa} e^2(k) = n. \quad (17.3.9)$$

Given that $\mathcal{R}(\mathcal{Q})$ is reducible, it follows from the above discussion that there exists a unitary matrix which will reduce all the matrices $\mathbf{Q}^{(s)}$ to block-diagonal form with

blocks corresponding to the irreducible representations. The number of blocks $m(k)$ corresponding to $\Theta(k)$ is called the *multiplicity* of $\Theta(k)$. Since the blocks for any $\Theta(k)$ can be collected together by the operation of a permutation matrix, which is itself unitary, there exists a unitary matrix U such that, for each $Q^{(s)}$,

$$U^\dagger Q^{(s)} U = \mathring{Q}^{(s)} := \mathring{Q}^{(s)}(1) \oplus \mathring{Q}^{(s)}(2) \oplus \cdots \oplus \mathring{Q}^{(s)}(\kappa) \quad (17.3.10)$$

where $\mathring{Q}^{(s)}(k)$ is the block corresponding to the representation $\Theta(k)$. It is of dimension $w(k) = m(k)e(k)$ and itself consists of $m(k)$ identical blocks of dimension $e(k)$. That is

$$\mathring{Q}^{(s)}(k) = \bigoplus_{m(k)} R^{(k)}(Q^{(s)}) := \overbrace{R^{(k)}(Q^{(s)}) \oplus \cdots \oplus R^{(k)}(Q^{(s)})}^{m(k) \text{ times}}. \quad (17.3.11)$$

The realization of $Q^{(s)}$ given by (17.3.10) and (17.3.11) is called its *canonical form*, with $\mathring{Q}^{(s)}$ the *Jordan normal form* (Sect. 17.2.6) for $Q^{(s)}$. For a one-dimensional representation $\Theta(k)$, $\mathring{Q}^{(s)}(k)$ is diagonal with diagonal elements $\chi^{(k)}(Q^{(s)})$, which are the eigenvalues of $\mathring{Q}^{(s)}(k)$ (and hence of $\mathring{Q}^{(s)}$ and $Q^{(s)}$). For a representation $\Theta(k)$ of dimension $e(k) > 1$, the matrices $\mathring{Q}^{(s)}(k)$ are of Jordan form, with the sum of the $e(k)$ equal eigenvalues along the diagonal of $R^{(k)}(Q^{(s)})$ giving $\chi^{(k)}(Q^{(s)})$. By taking the trace of (17.3.10) we have

$$\chi(Q^{(s)}) = \sum_{k=1}^{\kappa} m(k) \chi^{(k)}(Q^{(s)}). \quad (17.3.12)$$

It then follows from (17.3.3) and (17.3.8) that

$$m(k) = \frac{1}{n} \sum_{s=0}^{n-1} \bar{\chi}(Q^{(s)}) \chi^{(k)}(Q^{(s)}). \quad (17.3.13)$$

The characters of the irreducible representations of a group are presented in the form of a character table.¹⁶ If, therefore, $\mathcal{Q}$ is a standard group, or the direct product of standard groups, (17.3.13) can be used to determine the multiplicities of the irreducible representations in the arbitrary representation $\mathcal{R}(\mathcal{Q})$. Even if the character table of $\mathcal{Q}$ is not known we can still determine whether a particular representation is irreducible since, from (17.3.12) and (17.3.13),

$$\frac{1}{n} \sum_{s=0}^{n-1} \bar{\chi}(Q^{(s)}) \chi(Q^{(s)}) = \sum_{k=1}^{\kappa} m^2(k). \quad (17.3.14)$$

¹⁶ See the references given above or Conway et al. (1985).

The right-hand side of this equation is unity if and only if only one irreducible representation appears with multiplicity one in $\mathcal{R}(\mathcal{Q})$.

Apart from the ordering of the blocks, the canonical form for $\mathcal{Q}^{(s)}$ given by (17.3.10) and (17.3.11) is unique. Suppose that $\mathbf{W}$ is a unitary matrix such that

$$\mathbf{W}^\dagger \mathcal{Q}^{(s)} \mathbf{W} = \tilde{\mathcal{Q}}^{(s)} := \tilde{\mathcal{Q}}^{(s)}(1) \oplus \tilde{\mathcal{Q}}^{(s)}(2) \oplus \cdots \oplus \tilde{\mathcal{Q}}^{(s)}(\kappa), \quad s = 0, 1, \dots, n-1, \quad (17.3.15)$$

where

$$\frac{1}{n} \sum_{s=0}^{n-1} \text{Trace}\{\tilde{\mathcal{Q}}^{(s)}(k)\} \bar{\chi}^{(k')}(Q^{(s)}) = \delta^{kr}(k - k') m(k). \quad (17.3.16)$$

This represents a non-unique reduction of the matrices of $\mathcal{Q}$ to a set of κ reducible representations each of which involves only one irreducible representation. The final reduction to canonical form can be made if a set of unitary matrices $\mathbf{L}(k)$, $k = 1, 2, \dots, \kappa$ can be found such that

$$\mathbf{L}^\dagger(k) \tilde{\mathcal{Q}}^{(s)}(k) \mathbf{L}(k) = \mathring{\mathcal{Q}}^{(s)}(k), \quad s = 0, 1, \dots, n-1. \quad (17.3.17)$$

Then

$$\mathbf{U} = \mathbf{W} \mathbf{L}, \quad (17.3.18)$$

where

$$\mathbf{L} := \mathbf{L}(1) \oplus \mathbf{L}(2) \oplus \cdots \oplus \mathbf{L}(\kappa). \quad (17.3.19)$$

The matrix $\mathbf{W}$ can often be obtained in a straightforward way; computing the blocks of $\mathbf{L}$ is more difficult. However, the existence of the matrix $\mathbf{L}$, even without it being known is useful, as in the proof of the following result:

Theorem 17.3.5 *For the elements of the matrices $\tilde{\mathcal{Q}}^{(s)}(k)$ of (17.3.15)*

$$\sum_{s=0}^{n-1} \overline{\tilde{\mathcal{Q}}_{ij}^{(s)}(k)} \tilde{\mathcal{Q}}_{i'j'}^{(s)}(k') = 0 \quad (17.3.20)$$

for all i, j, i' and j' , whenever $k \neq k'$.

17.3.2 Permutation Representations and Equivalence Classes

As in Sect. 11.3 we now consider the group $\mathcal{Q} := \{I, Q^{(1)}, Q^{(2)}, \dots, Q^{(n-1)}\}$ in relation to a set of 'states' τ_m , $m = 1, 2, \dots, M$, such that each $Q^{(s)}$ maps a state τ_m into another state $Q^{(s)}\tau_m$, with $Q^{(s)}\tau_m \neq Q^{(s)}\tau_{m'}$ if $\tau_m \neq \tau_{m'}$.

The effect of the group $\mathcal{Q}$ on the states $\{\tau_m\}$ is given, using the representation $\mathcal{R}(\mathcal{Q}) := \{\mathbf{I}, \mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(n-1)}\}$ of M -dimensional permutation matrices, by (11.3.1). This representation of the group $\mathcal{Q}$ is called the *permutation representation* associated with $\{\tau_m\}$. If $M = n$ and if the states can be labelled by the index $s = 0, 1, \dots, n-1$ so that $\mathbf{Q}^{(s)}\tau_{s'} = \tau_{s''}$ whenever $\mathbf{Q}^{(s)}\mathbf{Q}^{(s')} = \mathbf{Q}^{(s'')}$, then the representation is called *regular*. It is not difficult to show that:

Theorem 17.3.6 *If the number of states $M = n$ and if, for some arbitrarily chosen state denoted (without loss of generality) by τ_1 , all the states $\mathbf{Q}^{(s)}\tau_1$, $s = 0, 1, \dots, n-1$, are distinct then the permutation representation associated with*

$$\tau_{s+1} = \mathbf{Q}^{(s)}\tau_1, \quad s = 0, 1, \dots, n-1, \quad (17.3.21)$$

is regular.

It is clear that, for the regular representation, $\mathbf{Q}^{(s')}\tau_s \neq \tau_s$ unless $s' = 0$ and thus

$$\chi(\mathbf{Q}^{(s)}) = \text{Trace}\{\mathbf{Q}^{(s)}\} = n \delta^{\text{Kr}}(s), \quad s = 0, 1, \dots, n-1. \quad (17.3.22)$$

Then, from (17.3.3) and (17.3.13), $m(k) = e(k)$; each irreducible representation occurs in the regular representation with a multiplicity equal to its dimension.

An *equivalence class* is a subset of the states $\tau_m, m = 1, 2, \dots, M$, which is closed under the operations of the group $\mathcal{Q}$ and is not further divisible into closed subsets. Suppose that the states are ordered in equivalence classes $\mathfrak{E}(\alpha)$, $\alpha = 1, 2, \dots, \gamma$, where $\mathfrak{E}(\alpha) := \{\tau_1(\alpha), \tau_2(\alpha), \dots, \tau_{p(\alpha)}(\alpha)\}$. The following result is useful in determining these equivalence classes:

Theorem 17.3.7 *The equivalence class $\mathfrak{E}(\alpha)$ can be generated by applying all the elements of $\mathcal{Q}$ to any one state of $\mathfrak{E}(\alpha)$. The number of states $p(\alpha)$ in $\mathfrak{E}(\alpha)$ is a divisor of n .*

It follows from (11.3.1) that each $\mathbf{Q}^{(s)}$ has the block-diagonal form

$$\mathbf{Q}^{(s)} = \mathbf{Q}^{(s)}(1) \oplus \mathbf{Q}^{(s)}(2) \oplus \dots \oplus \mathbf{Q}^{(s)}(\gamma), \quad (17.3.23)$$

with $\mathcal{R}(\alpha; \mathcal{Q}) := \{\mathbf{I}(\alpha), \mathbf{Q}^{(1)}(\alpha), \dots, \mathbf{Q}^{(n-1)}(\alpha)\}$, a $p(\alpha)$ -dimensional representation of $\mathcal{Q}$ operating on the equivalence class $\mathfrak{E}(\alpha)$, for $\alpha = 1, 2, \dots, \gamma$.

A formula of the same form as (17.3.15) now applies to the matrices of each of these representations, namely¹⁷

$$\mathbf{W}^\dagger(\alpha) \mathbf{Q}^{(s)}(\alpha) \mathbf{W}(\alpha) = \tilde{\mathbf{Q}}^{(s)}(\alpha) := \bigoplus_{k=1}^{\kappa} \tilde{\mathbf{Q}}^{(s)}(\alpha; k). \quad (17.3.24)$$

¹⁷ There is, of course, a similar specialization of (17.3.17), but we shall be concerned only with block-diagonalization with $\mathbf{W}(\alpha)$ and not the final reduction to canonical form.

The matrix $W(\alpha)$ has the effect of reducing $Q^{(s)}(\alpha)$ to block-diagonal form with the blocks corresponding to the irreducible representations in order. But to achieve the reduction given in (17.3.15), we need a further ordering between equivalence classes. Thus

$$W = [W(1) \oplus W(2) \oplus \cdots \oplus W(\gamma)]\Pi, \quad (17.3.25)$$

where Π is a permutation matrix, chosen to achieve the final ordering described above. The following theorem is useful in determining the multiplicities of the irreducible representations within equivalence classes:

Theorem 17.3.8 *For any equivalence class $\mathfrak{E}(\alpha)$ the multiplicity $m(\alpha; k)$ of the irreducible representation $\Theta(k)$ of dimension $e(k)$ satisfies the condition*

$$m(\alpha; k) \leq e(k), \quad \text{with} \quad m(\alpha; 1) = e(1) = 1. \quad (17.3.26)$$

17.3.3 Block Diagonalization Within an Equivalence Class

We now consider the problem of determining unitary matrices $W(\alpha)$ which produce the block-diagonalization given by (17.3.24). For this the following result is useful:

Theorem 17.3.9 *The set of operators*

$$P_{j\ell}^{(k)} := \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{R}_{j\ell}^{(k)}(Q^{(s)})Q^{(s)}, \quad (17.3.27)$$

for $j, \ell = 1, 2, \dots, e(k)$, $k = 1, 2, \dots, \kappa$, transform according to the formulae

$$Q^{(s)}P_{j\ell}^{(k)} = \sum_{i=1}^{e(k)} R_{ij}^{(k)}(Q^{(s)})P_{i\ell}^{(k)}, \quad P_{j\ell}^{(k)}Q^{(s)} = \sum_{i=1}^{e(k)} R_{li}^{(k)}(Q^{(s)})P_{ji}^{(k)} \quad (17.3.28)$$

and satisfies the product rule

$$P_{j'\ell'}^{(k')}P_{j\ell}^{(k)} = \delta^{k'r}(k - k')\delta^{k'r}(\ell' - j)P_{j'\ell}^{(k)}. \quad (17.3.29)$$

Using (17.3.27), let

$$T^{(k)} := \sum_{j=1}^{e(k)} P_{jj}^{(k)} = \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)})Q^{(s)}, \quad k = 1, 2, \dots, \kappa. \quad (17.3.30)$$

It follows from (17.3.28) and (17.3.29) that

$$Q^{(s)}T^{(k)} = T^{(k)}Q^{(s)}, \quad s = 0, 1, \dots, n-1, \quad k = 1, 2, \dots, \kappa, \quad (17.3.31)$$

$$T^{(k)}T^{(k')} = \delta^{kr}(k - k')T^{(k')}, \quad k = 1, 2, \dots, \kappa. \quad (17.3.32)$$

With $T^{(k)} := R(T^{(k)})$, it is obvious that formulae like (17.3.30)–(17.3.32) apply with operators replaced by their representation matrices. From (17.3.23) and (17.3.30)

$$T^{(k)} = T^{(k)}(1) \oplus T^{(k)}(2) \oplus \dots \oplus T^{(k)}(\gamma), \quad (17.3.33)$$

where

$$T^{(k)}(\alpha) := \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)}) Q^{(s)}(\alpha), \quad k = 1, 2, \dots, \kappa. \quad (17.3.34)$$

Then

$$W^\dagger(\alpha) T^{(k)}(\alpha) W(\alpha) = \bigoplus_{k'=1}^{\kappa} \tilde{T}^{(k)}(\alpha; k'), \quad (17.3.35)$$

where, from (17.3.24), (17.3.34) and the orthogonality formula (17.3.6),

$$\tilde{T}^{(k)}(\alpha; k') = \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)}) \tilde{Q}^{(s)}(\alpha; k') = \delta^{kr}(k - k') I^{(k)}(\alpha), \quad (17.3.36)$$

$I^{(k)}(\alpha)$ being a unit matrix of dimension $m(\alpha; k)$. From (17.3.35) and (17.3.36),

$$T^{(k)}(\alpha) W(\alpha) = W(\alpha) J^{(k)}(\alpha), \quad (17.3.37)$$

where $J^{(k)}(\alpha)$ is a $p(\alpha)$ -dimensional diagonal matrix with zeros along the diagonal except for $m(\alpha; k)$ ones in the rows corresponding to the irreducible representation $\Theta(k)$. From (11.3.1), (17.3.23) and (17.3.24),

$$\tilde{Q}^{(s)}(\alpha) [W(\alpha)]^\dagger \begin{pmatrix} \tau_1(\alpha) \\ \tau_2(\alpha) \\ \vdots \\ \tau_{p(\alpha)}(\alpha) \end{pmatrix} = [W(\alpha)]^\dagger \begin{pmatrix} Q^{(s)} \tau_1(\alpha) \\ Q^{(s)} \tau_2(\alpha) \\ \vdots \\ Q^{(s)} \tau_{p(\alpha)}(\alpha) \end{pmatrix}, \quad s = 0, 1, \dots, n-1, \quad (17.3.38)$$

and, from the block structure of $\tilde{Q}^{(s)}$ given by (17.3.24), the effect of $[W(\alpha)]^\dagger$ is to form linear combinations of the states of $\mathfrak{E}(\alpha)$, which transform according to

representations each of which reduces to a multiple of a single irreducible representation. Thus a group of $w(\alpha; k) = m(\alpha; k)e(k)$ successive rows of $[W(\alpha)]^\dagger$ (columns of $W(\alpha)$) can be identified with the irreducible representation $\Theta(k)$. It follows from (17.3.37) that these columns are eigenvectors of $T^{(k)}(\alpha)$ with eigenvalue one and all the other columns of $W(\alpha)$ are eigenvectors of $T^{(k)}(\alpha)$ with eigenvalue zero. However, from (17.3.32) and (17.3.33), each of the columns of $T^{(k')}(\alpha)$ is an eigenvector of $T^{(k)}(\alpha)$ with eigenvalue one if $k' = k$ and zero if $k' \neq k$. The columns of the matrix $W(\alpha)$ corresponding to the representation $\Theta(k)$ are, therefore, obtained by finding a subset of $w(\alpha; k)$ of the $p(\alpha)$ columns of $T^{(k)}(\alpha)$ which satisfied the required orthonormality. This is possible since the dimension of the space spanned by the columns is the number of non-zero eigenvalues which, from (17.3.35), is $w(\alpha; k)$. The simplest way to do this is to find all the functions

$$\psi^{(k)}(\alpha; m) := T^{(k)} \tau_m(\alpha) = \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)}) Q^{(s)} \tau_m(\alpha). \quad (17.3.39)$$

A selection is now made from the sets of coefficients of these linear sums of states to satisfy the orthonormality condition. Let the members of this set of columns be $(W_{1j}^{(k)}(\alpha), W_{2j}^{(k)}(\alpha), \dots, W_{p(\alpha)j}^{(k)}(\alpha))$, $j = 1, 2, \dots, w(\alpha; k)$.

For the case of an irreducible representation for which $e(k) = 1$ it follows from Theorem 17.3.8 that $w(\alpha; k) = 0$ or 1. In the latter case it follows, from (17.3.30) and (17.3.31), that

$$\begin{aligned} Q^{(s)} \psi^{(k)}(\alpha; m) &= \frac{e(k)}{n} \sum_{s'=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s')}) [Q^{(s)}]^{-1} Q^{(s')} \tau_m(\alpha) \\ &= \chi^{(k)}(Q^{(s)}) \psi^{(k)}(\alpha; m). \end{aligned} \quad (17.3.40)$$

From Theorem 17.3.7 it follows that, given any $\tau_{m'}(\alpha)$, there exists an operator $Q^{(s')}$ such that $\tau_{m'}(\alpha) = Q^{(s')} \tau_m(\alpha)$. Then, from (17.3.31), (17.3.39) and (17.3.40),

$$\psi^{(k)}(\alpha; m') = \chi^{(k)}(Q^{(s')}) \psi^{(k)}(\alpha; m). \quad (17.3.41)$$

So, to within a multiplicative factor, all the functions $\psi^{(k)}(\alpha; m)$ are the same. In two cases it is straightforward to compute the column elements explicitly:

17.3.3.1 For the One-Dimensional Representation

$\chi(Q^{(s)}) = 1$, $s = 0, 1, \dots, n-1$ and

$$\psi^{(1)}(\alpha; m) = \frac{1}{n} \sum_{s=0}^{n-1} Q^{(s)} \tau_m(\alpha) = \frac{1}{p(\alpha)} \sum_{j=1}^{p(\alpha)} \tau_j(\alpha). \quad (17.3.42)$$

Thus, on normalization,

$$W_{i1}^{(1)}(\alpha) = \frac{1}{\sqrt{p(\alpha)}}, \quad i = 1, 2, \dots, p(\alpha). \quad (17.3.43)$$

17.3.3.2 For a Regular Representation

$p(\alpha) = n$ and $\tau_{s+1}(\alpha) = Q^{(s)}\tau_1(\alpha)$, giving

$$\psi^{(k)}(\alpha; m) = \frac{e(k)}{n} \sum_{s=0}^{n-1} \bar{\chi}^{(k)}(Q^{(s)}[Q^{(m-1)}]^{-1})\tau_1(\alpha). \quad (17.3.44)$$

Regarding the states as a set of orthonormal vectors, it follows from (17.3.8) that

$$\psi^{(k)}(\alpha; m) \cdot \overline{\psi^{(k')}(\alpha; m')} = \delta^{kr} (k - k') \frac{e(k)}{n} \chi^{(k)}(Q^{(m-1)}[Q^{(m'-1)}]^{-1}). \quad (17.3.45)$$

In this case $w(\alpha; k) = e(k)^2$ for all irreducible representations. For a one-dimensional representation we take

$$W_{i1}^{(k)}(\alpha) := \frac{\bar{\chi}^{(k)}(Q^{(i-1)})}{\sqrt{n}}. \quad (17.3.46)$$

When $e(k) > 1$, an orthonormal set of columns can be obtained by choosing a subset of the group operators such that $\chi^{(k)}(Q^{(s)}[Q^{(s')}]^{-1}) = 0$ for all $s \neq s'$.

17.3.4 Symmetry Groups

At the beginning of Sect. 17.3 we discussed the symmetry group of the Hamiltonian formed as a direct product of a group applied to the lattice, to the states and possibly to an external parameter. Suppose that $\mathcal{F}$ and $\mathcal{G}$ are such groups with $\mathcal{Q} = \mathcal{F} \times \mathcal{G}$ being the symmetry group of the Hamiltonian. When constructing the representation $\mathcal{R}(\mathcal{Q})$ of $\mathcal{Q}$ from the representations $\mathcal{R}(\mathcal{F})$ and $\mathcal{R}(\mathcal{G})$ of $\mathcal{F}$ and $\mathcal{G}$ respectively it must be noted that this can be achieved either by setting

$$\mathbf{R}(\mathbf{F}^{(i)}\mathbf{G}^{(j)}) = \mathbf{R}(\mathbf{G}^{(j)}\mathbf{F}^{(i)}) = \mathbf{R}(\mathbf{F}^{(i)}) \otimes \mathbf{R}(\mathbf{G}^{(j)}) \quad (17.3.47)$$

or

$$\mathbf{R}(\mathbf{F}^{(i)}\mathbf{G}^{(j)}) = \mathbf{R}(\mathbf{G}^{(j)}\mathbf{F}^{(i)}) = \mathbf{R}(\mathbf{G}^{(j)}) \otimes \mathbf{R}(\mathbf{F}^{(i)}). \quad (17.3.48)$$

Since, in general,

$$\mathbf{R}(\mathbf{F}^{(i)}) \otimes \mathbf{R}(\mathbf{G}^{(j)}) \neq \mathbf{R}(\mathbf{G}^{(j)}) \otimes \mathbf{R}(\mathbf{F}^{(i)}), \quad (17.3.49)$$

(17.3.47) and (17.3.48) correspond to different representations of $\mathcal{R}(\mathcal{Q})$. It is important to choose *in advance* a particular group ordering for the representation and we shall make the choice given by (17.3.47). The irreducible representations $\Theta(k', k'')$ of $\mathcal{Q}$ are given by taking Kronecker products $\Theta_{\mathcal{F}}(k') \otimes \Theta_{\mathcal{G}}(k'')$ of all the irreducible representations $\Theta_{\mathcal{F}}(k')$ and $\Theta_{\mathcal{G}}(k'')$ of $\mathcal{F}$ and $\mathcal{G}$. By taking the trace of (17.3.47) for irreducible representations it follows from the property of Kronecker matrix products that

$$\chi^{(k', k'')}(F^{(i)} G^{(j)}) = \chi_{\mathcal{F}}^{(k')}(F^{(i)}) \chi_{\mathcal{G}}^{(k'')}(G^{(j)}). \quad (17.3.50)$$

The character table of $\mathcal{Q}$ can, therefore, be constructed from those of $\mathcal{F}$ and $\mathcal{G}$ as shown in the particular case in Example 17.3.4.

17.3.4.1 Periodic Boundary Conditions

Consider a two-dimensional lattice where periodic boundary conditions are applied in one direction by wrapping the lattice on a cylinder. The translational symmetry of the lattice in the direction in which periodic boundary conditions have been applied becomes rotational symmetry, with the cyclic group $\mathcal{C}_n$ of rotations about the axis of the cylinder, where n is either the number of rows of sites along the direction of the axis of the cylinder or some divisor of that number. Since the group is commutative each element is a separate conjugacy class and it follows from Theorem 17.3.3 that the group has n irreducible representations and, from (17.3.9), that they are one-dimensional. If $\chi(C_n) = \omega_n$ for an irreducible representation, $\chi(C_n^n) = \omega_n^n = 1$. The characters of C_n in the n irreducible representations are the complex n -th roots of unity and from them we can construct by multiplication all the characters of the other elements of $\mathcal{C}_n$.

Example 17.3.1 The character tables for $\mathcal{C}_2$ and $\mathcal{C}_6$ are

$\mathcal{C}_2$	I	C_2
A	1	1
B	1	-1

$\mathcal{C}_6$	I	C_6	C_6^2	C_6^3	C_6^4	C_6^5
A	1	1	1	1	1	1
B	1	-1	1	-1	1	-1
E_1	1	ω_6	ω_6^2	ω_6^3	ω_6^4	ω_6^5
E_1	1	ω_6^5	ω_6^4	ω_6^3	ω_6^2	ω_6
E_2	1	ω_6^2	ω_6^4	1	ω_6^2	ω_6^4
E_2	1	ω_6^4	ω_6^2	1	ω_6^4	ω_6^2

where $\omega_6 := \exp(2\pi i/6)$ and the representations E_i and $\bar{E}_i$, $i = 1, 2$ have conjugate complex characters. For the spin- $\frac{1}{2}$ Ising model the spin-inversion group $\mathcal{S}_2^{(m)} := \{I, Q^{(m)}\}$, discussed in Sect. 3.5.1, is, of course, isomorphic to $\mathcal{C}_2$.

In the special case of a square lattice with one lattice direction consisting of rings of n sites and the other lattice direction along the cylinder there will also be n reflectional

symmetries¹⁸ $\Sigma_1, \Sigma_2, \dots, \Sigma_n$, about planes containing the axis of the cylinder. When n is odd these reflections are about planes which cut the surface of the cylinder along a line of sites and along a line equidistant between two lines of sites. When n is even the reflections fall into two sets: those about planes which cut the surface of the cylinder along two lines of sites, which we denote by Σ_i , with i odd and those about planes which cut the surface of the cylinder along two lines equidistant between two lines of sites, which we denote by Σ_i , with i even. The group is now $\mathcal{C}_{nv}$, which is not commutative and, for $n > 2$, the rotations no longer fall into separate conjugacy classes. However, for n odd all the reflections are in one conjugacy class and for n even they fall into two conjugacy classes according to whether the index of the reflection is even or odd.

Example 17.3.2 The character table for $\mathcal{C}_{2v}$ is

$\mathcal{C}_{2v}$	I	C_2	Σ_1	Σ_2
A_1	1	1	1	1
A_2	1	1	-1	-1
B_1	1	-1	1	-1
B_2	1	-1	-1	1

In general, of course, the group $\mathcal{C}_{2v}$ is not identical to $\mathcal{C}_2$. However, for the situation considered here where we have two sites in a ring, Σ_1 leave the two sites unchanged and Σ_2 interchanges them. Effectively $\Sigma_1 \equiv I$ and $\Sigma_2 \equiv C_2$.¹⁹ The character tables for $\mathcal{C}_{3v}$ and $\mathcal{C}_{4v}$ are

$\mathcal{C}_{3v}$	Q_1	Q_2	Q_3
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0

$\mathcal{C}_{4v}$	Q_1	Q_2	Q_3	Q_4	Q_5
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	1	-1	1	-1
B_2	1	1	-1	-1	1
E	2	-2	0	0	0

where, for $\mathcal{C}_{3v}$ the conjugacy classes are $Q_1 := I$, $Q_2 := \{C_3, C_3^2\}$ and $Q_3 := \{\Sigma_1, \Sigma_2, \Sigma_3\}$ and for $\mathcal{C}_{4v}$ the conjugacy classes are $Q_1 := I$, $Q_2 := C_4^2$, $Q_3 := \{C_4, C_4^3\}$, $Q_4 := \{\Sigma_1, \Sigma_3\}$ and $Q_5 := \{\Sigma_2, \Sigma_4\}$.

Example 17.3.3 The spin- $\frac{1}{2}$ Ising model on a square lattice with two rows wrapped on a cylinder has symmetry group $\mathcal{S}_2^{(m)} \times \mathcal{C}_2$, where, as indicated above, $\mathcal{S}_2^{(m)}$ is isomorphic to $\mathcal{C}_2$. The conjugacy classes, irreducible representations and character table for $\mathcal{C}_2$ are given in Example 17.3.1. The irreducible representations, in terms

¹⁸ We are considering the square lattice for simplicity. Reflectional symmetries can arise for other lattices.

¹⁹ Another way to think about it is that, if we use the permutation representation, it can contain only the irreducible representations A_1 and B_1 . There is a caveat to these remarks. If the microstates on the lattice sites have some directional properties (like arrows) then the reflections need not reduce to the identity and a rotation.

of the products of the representations of S_2 and C_2 , are $A := A \otimes A$, $B_1 := B \otimes A$, $B_2 := A \otimes B$, and $B_3 := B \otimes B$. The conjugacy classes each contain one element.

$S_2^{(m)} \times C_2$	1	$Q^{(m)}$	C_2	$Q^{(m)}C_2$
A	1	1	1	1
B ₁	1	-1	1	-1
B ₂	1	1	-1	-1
B ₃	1	-1	-1	1

Example 17.3.4 The 3-state Potts model on a square lattice with two rows wrapped on a cylinder has symmetry group $S_3 \times C_2$. The character table is constructed using Table 5.1 for the character table of S_3 and the character table for C_2 given in Example 17.3.1. The irreducible representations, in terms of the products of the representations of S_3 and C_2 , are $A := A \otimes A$, $B_1 := B \otimes A$, $B_2 := A \otimes B$, $B_3 := B \otimes B$, $E_1 := E \otimes A$ and $E_2 := E \otimes B$. The conjugacy classes Q_1, Q_2, Q_3 are those of S_3 followed by the identity operator for C_2 and Q_4, Q_5, Q_6 are similarly constructed with the identity of C_2 replaced by C_2 .

$S_3 \times C_2$	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6
A	1	1	1	1	1	1
B ₁	1	1	-1	1	1	-1
B ₂	1	1	1	-1	-1	-1
B ₃	1	1	-1	-1	-1	1
E ₁	2	-1	0	2	-1	0
E ₂	2	-1	0	-2	1	0

Example 17.3.5 The eight-vertex model on a square lattice with N rows wrapped on a cylinder has symmetry group $S_2^{(e)} \times S_N$, where $S_2^{(e)}$ corresponds to the change of sign of edge variables in a row (reversal of arrow directions) and S_N corresponds to all permutations of a row of edge variables (which leaves invariant the number of up and down arrow directions). As we have seen S_2 is isomorphic to C_2 and S_3 is isomorphic to $\mathcal{D}_3$ and C_{3v} . However, the symmetric group S_4 , of order 24, is larger than C_{3v} . The usual way to distinguish its conjugacy classes is in terms of cycles, thus, for example $\langle 2, 1, 1 \rangle$ denotes one two-cycle combined with two one-cycles. The conjugacy classes are then: $Q_1 := \langle 1, 1, 1, 1 \rangle = I$, with one element; $Q_2 := \langle 2, 1, 1 \rangle$, with six elements, including Σ_1 and Σ_3 and with each of Σ_2 and Σ_4 broken into two separate two-cycles; $Q_3 := \langle 2, 2 \rangle$, with three elements Σ_2 , Σ_4 and Σ_1 combined with Σ_3 ; $Q_4 := \langle 3, 1 \rangle$, with eight elements consisting of all three-cycles; $Q_5 := \langle 4 \rangle$, with six elements including C_4 , C_4^2 and C_4^3 . The character table is

S_4	Q_1	Q_2	Q_3	Q_4	Q_5
A	1	1	1	1	1
B	1	-1	1	1	-1
E	2	0	2	-1	0
T ₁	3	-1	-1	0	1
T ₂	3	1	-1	0	-1

17.4 The Conformal Group

We confine our attention to the d -dimensional Euclidean space of the vectors

$$\mathbf{r} := \begin{pmatrix} r^{(1)} \\ \vdots \\ r^{(d)} \end{pmatrix} \quad (17.4.1)$$

measured in units of the lattice spacing. The line element $d\mathbf{r}$ is given by

$$[d\mathbf{r}]^2 = \mathbf{dr} \cdot \mathbf{dr} := [dr^{(1)}]^2 + [dr^{(2)}]^2 + \cdots + [dr^{(d)}]^2. \quad (17.4.2)$$

The *conformal group* is the set of transformations $\mathbf{r} \rightarrow \tilde{\mathbf{r}}$ which satisfy conditions of the form

$$[d\tilde{\mathbf{r}}]^2 = \Omega(\mathbf{r})[d\mathbf{r}]^2, \quad (17.4.3)$$

for some scalar fields $\Omega(\mathbf{r})$. Suppose that a transformation is given by

$$\tilde{r}^{(i)} = R^{(i)}(\mathbf{r}) \quad i = 1, 2, \dots, d. \quad (17.4.4)$$

Then, for the condition (17.4.3) to be satisfied, it must be the case that

$$\Omega(\mathbf{r}) \delta^{\mathbf{kr}}(j - k) = \sum_{i=1}^d \frac{\partial R^{(i)}}{\partial r^{(j)}} \frac{\partial R^{(i)}}{\partial r^{(k)}}, \quad (17.4.5)$$

and $[\Omega(\mathbf{r})]^{\frac{d}{2}}$ is the Jacobian of the transformation (17.4.4). It is clear that the following transformations satisfy (17.4.5):

(i) *Translations* where

$$R^{(i)}(\mathbf{r}) = r^{(i)} + c^{(i)}, \quad i = 1, 2, \dots, d, \quad (17.4.6)$$

for some constants $c^{(i)}$ and $\Omega(\mathbf{r}) = 1$.

(ii) *Rotations* where

$$R^{(i)}(\mathbf{r}) = \omega_i^{(1)} r^{(1)} + \cdots + \omega_i^{(d)} r^{(d)}, \quad i = 1, \dots, d, \quad (17.4.7)$$

the vectors

$$\boldsymbol{\omega}_i := \begin{pmatrix} \omega_i^{(1)} \\ \vdots \\ \omega_i^{(d)} \end{pmatrix} \quad (17.4.8)$$

are orthonormal and $\Omega(\mathbf{r}) = 1$.

(iii) *Dilatations* where

$$R^{(i)}(\mathbf{r}) = \lambda^{-1} r^{(i)}, \quad i = 1, 2, \dots, d, \quad (17.4.9)$$

and $\Omega(\mathbf{r}) = \lambda^{-2}$.

(iv) *Inversions* where

$$R^{(i)}(\mathbf{r}) = r^{(i)} / r^2, \quad i = 1, 2, \dots, d, \quad (17.4.10)$$

$r = |\mathbf{r}|$ and $\Omega(\mathbf{r}) = 1/r^4$.

It can be shown (Ginsparg 1990) that, for $d > 2$, the conformal group is finite-dimensional and generated by the transformations (i)–(iv). Alternatively, inversion can be replaced as a generator of the group by the *special conformal transformation*

$$R^{(i)}(\mathbf{r}) = \frac{r^{(i)} + c^{(i)} r^2}{1 + 2\mathbf{r} \cdot \mathbf{c} + c^2 r^2}, \quad (17.4.11)$$

for any constant vector $\mathbf{c} = (c^{(1)}, \dots, c^{(d)})$, with $\Omega(\mathbf{r}) = (1 + 2\mathbf{r} \cdot \mathbf{c} + c^2 r^2)^{-2}$. This can be seen to be composed of a translation, an inversion and then a further translation.

For $d = 2$ the notation $\mathbf{r} := (x, y)$ is used, with the transformation (17.4.4) given by

$$\tilde{x} = X(x, y), \quad \tilde{y} = Y(x, y). \quad (17.4.12)$$

Equation (17.4.5) is satisfied if

$$\frac{\partial X}{\partial x} = \frac{\partial Y}{\partial y}, \quad \frac{\partial X}{\partial y} = -\frac{\partial Y}{\partial x}. \quad (17.4.13)$$

These are the Cauchy–Riemann conditions for

$$Z(z) := X(x, y) + iY(x, y) \quad (17.4.14)$$

to be an analytic function of the complex variable $z := x + iy$. In this case, therefore, the conformal group is infinite-dimensional. From (17.4.5) the Jacobian of the transformation $z \rightarrow Z(z)$ is

$$\Omega(z) = |Z'(z)|^2. \quad (17.4.15)$$

17.5 Some Transformations in the Complex Plane

In the discussion of Ruelle's theorem in Sect. 8.4.1 a fundamental role is played by the polynomial $\xi(T, \{\mathfrak{Z}_\kappa(\mathbf{r})\}; \mathcal{Y}_\kappa)$ given by Eq. (8.4.9), where $\mathcal{Y}_\kappa$ is a member of the covering $\mathcal{Y}_\mathcal{N}$ of the lattice $\mathcal{N}$. When $\mathcal{Y}_\kappa$ consists of two lattice sites $\mathbf{r}$ and $\mathbf{r}'$

$$\xi(T, \{\mathfrak{Z}_\kappa(\mathbf{r})\}; \mathcal{Y}_\kappa) = f(\mathfrak{Z}, \mathfrak{Z}') := c_0 + c_1(\mathfrak{Z} + \mathfrak{Z}') + c_2\mathfrak{Z}\mathfrak{Z}', \quad (17.5.1)$$

where the real positive numbers c_0 , c_1 and c_2 are the Boltzmann weights when the pair of sites is occupied by 0, 1 and 2 particles respectively and $\mathfrak{Z} := \mathfrak{Z}_\kappa(\mathbf{r})$, $\mathfrak{Z}' := \mathfrak{Z}_\kappa(\mathbf{r}')$. We need to obtain two closed sets $\mathbb{X}$ and $\mathbb{X}'$ in $\mathbb{C}_\mathfrak{Z}$, not containing 0, such that $f(\mathfrak{Z}, \mathfrak{Z}') \neq 0$ when $\mathfrak{Z} \notin \mathbb{X}$ and $\mathfrak{Z}' \notin \mathbb{X}'$. In the Ising model spin formulation $c_0 = c_2 \neq c_1$ and in the lattice fluid formulation $c_0 = c_1 = 1$, $c_2 \neq 1$. In the former case we may, without loss of generality, set $c_0 = c_2 = 1$ and the two cases are related by a multiplicative transformation of the variables $\mathfrak{Z}$ and $\mathfrak{Z}'$.

We, therefore, begin by considering

$$f(\mathfrak{Z}, \mathfrak{Z}') := 1 + c(\mathfrak{Z} + \mathfrak{Z}') + \mathfrak{Z}\mathfrak{Z}', \quad (17.5.2)$$

where $f(\mathfrak{Z}, \mathfrak{Z}') = 0$ corresponds to the mapping

$$\mathfrak{Z}' = -(1 + c\mathfrak{Z})/(c + \mathfrak{Z}). \quad (17.5.3)$$

With the change of variables $\zeta := \mathfrak{Z} + c$, $\zeta' := \mathfrak{Z}' + c$, this becomes

$$\zeta' = (c^2 - 1)/\zeta. \quad (17.5.4)$$

Excluding the case $c = 1$, which corresponds to no interaction for any site occupation, the fixed points

$$\zeta^{(\pm)} = \pm\sqrt{c^2 - 1} \quad (17.5.5)$$

of (17.5.4) lie on the real or imaginary axis of the complex plane $\mathbb{C}_\zeta$ of ζ according as $c > 1$ or $c < 1$, respectively. It is now not difficult to establish the following result:

Theorem 17.5.1 *Let $\mathcal{V}$ be the family of circles in $\mathbb{C}_\zeta$ passing through $\zeta = \zeta^{(\pm)}$.²⁰ Any circle $\mathbb{V} \in \mathcal{V}$ is divided into two open arcs $\mathbb{V}^{(\pm)}$ by the fixed points. Every $\zeta \neq \zeta^{(\pm)}$ in $\mathbb{C}_\zeta$ belongs to exactly one circle $\mathbb{V} \in \mathcal{V}$ and, if $\zeta \in \mathbb{V}^{(\pm)}$, then $\zeta' \in \mathbb{V}^{(\mp)}$.*

It follows that every circle $\mathbb{V} \in \mathcal{V}$ is invariant under the mapping (17.5.4). Since the arcs $\mathbb{V}_1^{(\pm)}$ of any other $\mathbb{V}_1 \in \mathcal{V}$ lie one in the interior and one in the exterior of $\mathbb{V}$, it follows that (17.5.4) is a mapping between the interior and exterior of $\mathbb{V}$. In $\mathbb{C}_\mathfrak{z}$ any circle through the points

$$\mathfrak{z}^{(\pm)} = -c \pm \sqrt{c^2 - 1} \quad (17.5.6)$$

is invariant under (17.5.3), with (17.5.6) as fixed points. The mapping is between interior and exterior points of the circle. We consider the two cases $c < 1$ and $c > 1$ separately:

(a) $c < 1$: In this case the fixed points are

$$\mathfrak{z}^{(\pm)} = -c \pm i\sqrt{1 - c^2}. \quad (17.5.7)$$

The circle

$$\mathbb{V}(x) := \{\mathfrak{z} : |2c\mathfrak{z} + x| = \sqrt{4c^2(1 - x) + x^2}\} \quad (17.5.8)$$

passes through the points (17.5.7) for any real x . Apart from $\mathbb{V}(1)$, which passes through the origin, any circle $\mathbb{V}(x)$ can be used to define the regions $\mathbb{X}$ and $\mathbb{X}'$. Thus we see that $f(\mathfrak{z}, \mathfrak{z}') \neq 0$, when $\mathfrak{z} \notin \mathbb{X}$ and $\mathfrak{z}' \notin \mathbb{X}$, if

$$\mathbb{X} := \begin{cases} \{\mathfrak{z} : |2c\mathfrak{z} + x| \leq \sqrt{4c^2(1 - x) + x^2}\}, & \forall \quad x > 1, \\ \{\mathfrak{z} : |2c\mathfrak{z} + x| \geq \sqrt{4c^2(1 - x) + x^2}\}, & \forall \quad x < 1. \end{cases} \quad (17.5.9)$$

In particular the unit circle $|\mathfrak{z}| = 1$ corresponds to $x = 0$ in (17.5.8) and gives a region in the second category of (17.5.9). The regions in the first category correspond to discs, with centres on the negative real axis $\mathbb{R}^{(-)}$ of $\mathbb{C}_\mathfrak{z}$, and lying entirely in $\Re(\mathfrak{z}) < 0$. The angles made with $\mathbb{R}^{(-)}$ by tangents from the origin to the boundary of the disc are $\pm \arccos(2c\sqrt{x - 1}/x)$. These angles take a minimum absolute value when $x = 2$, when the points of contact of the tangents and the boundary are the fixed points (17.5.7) lying on $|\mathfrak{z}| = 1$.

(b) $c > 1$: In this case the fixed points (17.5.6) lie on $\mathbb{R}^{(-)}$. The circle

$$\mathbb{V}(y) := \{\mathfrak{z} : |\mathfrak{z} + c + iy| = \sqrt{y^2 + c^2 - 1}\} \quad (17.5.10)$$

²⁰ Including the limiting circle consisting of the axis through the fixed points.

passes through the points (17.5.6) for any real y . The origin is exterior to all the circles of this family and thus the definition (17.5.9) can be replaced by

$$\mathbb{Y} := \{\mathfrak{Z} : |\mathfrak{Z} + c + iy| \leq \sqrt{y^2 + c^2 - 1}\}, \quad \forall y. \quad (17.5.11)$$

The angles made with $\mathbb{R}^{(-)}$ by tangents from the origin to the boundary of the disc are $\arccos(c/\sqrt{y^2 + c^2}) \pm \arccos(1/\sqrt{y^2 + c^2})$. The larger of these angles take its minimum value when $y = 0$.

The other case of interest for (17.5.1) is

$$f(\mathfrak{Z}, \mathfrak{Z}') := 1 + \mathfrak{Z} + \mathfrak{Z}' + c\mathfrak{Z}\mathfrak{Z}'. \quad (17.5.12)$$

This can be expressed in the form

$$f(\mathfrak{Z}, \mathfrak{Z}') = 1 + (1/\sqrt{c})(\sqrt{c}\mathfrak{Z} + \sqrt{c}\mathfrak{Z}') + (\sqrt{c}\mathfrak{Z})(\sqrt{c}\mathfrak{Z}'), \quad (17.5.13)$$

and the previous results apply with c replaced by $1/\sqrt{c}$ and $\mathfrak{Z}$ replaced by $\sqrt{c}\mathfrak{Z}$. The two cases are:

(a) $c < 1$: From (17.5.11) and (17.5.13), $f(\mathfrak{Z}, \mathfrak{Z}') \neq 0$ when $\mathfrak{Z} \notin \mathbb{Y}'$ and $\mathfrak{Z}' \notin \mathbb{Y}'$ if

$$\mathbb{Y}' := \{\mathfrak{Z} : |c\mathfrak{Z} + 1 + iy| \leq \sqrt{y^2 + 1 - c}\}, \quad \forall y. \quad (17.5.14)$$

The angles made with $\mathbb{R}^{(-)}$ by tangents from the origin to the boundary of the disc are $\arccos(1/\sqrt{y^2 + 1}) \pm \arccos(\sqrt{c}/\sqrt{y^2 + 1})$. The larger of these angles take its minimum value when $y = 0$.

(b) $c > 1$: From (17.5.9), $f(\mathfrak{Z}, \mathfrak{Z}') \neq 0$ when $\mathfrak{Z} \notin \mathbb{X}'$ and $\mathfrak{Z}' \notin \mathbb{X}'$ if

$$\mathbb{X}' := \begin{cases} \{\mathfrak{Z} : |2\mathfrak{Z} + x| \leq \sqrt{4c^{-1}(1-x) + x^2}\}, & \forall x > 1, \\ \{\mathfrak{Z} : |2\mathfrak{Z} + x| \geq \sqrt{4c^{-1}(1-x) + x^2}\}, & \forall x < 1. \end{cases} \quad (17.5.15)$$

The angles made with $\mathbb{R}^{(-)}$ by tangents from the origin to the boundary of the discs in the first category are $\pm \arccos(2\sqrt{x-1}/x\sqrt{c})$. These angles take a minimum absolute value when $x = 2$.

17.6 Algebraic Functions

Definition 17.6.1 An algebraic function $\lambda(w)$ is the solution of an equation

$$\mathcal{F}(\lambda, w) = 0, \quad \text{where} \quad (17.6.1)$$

$$\mathcal{F}(\lambda, w) := f_n(w)\lambda^n + f_{n-1}(w)\lambda^{n-1} + \cdots + f_0(w). \quad (17.6.2)$$

The coefficients $f_j(w)$, $j = 0, 1, \dots, n$, are polynomials in w with complex coefficients, having no common factor apart from a constant and $\mathcal{F}(\lambda, w)$ is not reducible²¹ to a product of polynomials in λ , which themselves have coefficients which are polynomials in w .

$\mathcal{F}(\lambda, w)$ is called the *defining polynomial* of $\lambda(w)$ and the solutions $\lambda_1(w), \dots, \lambda_n(w)$ are called the *branches* of $\lambda(w)$. A point at which the values of two or more of the branches of $\lambda(w)$ are equal is a value of w for which (17.6.1) has repeated roots. The set of such points is given by the simultaneous solution of the equations

$$\mathcal{F}(\lambda, w) = 0, \quad \frac{\partial \mathcal{F}(\lambda, w)}{\partial \lambda} \equiv \mathcal{F}'(\lambda, w) = 0, \quad (17.6.3)$$

which is obtain by solving the equation

$$\text{Res}\{\mathcal{F}, \mathcal{F}'; w\} = 0 \quad (17.6.4)$$

for the *resultant* $\text{Res}\{\mathcal{F}, \mathcal{F}'; w\}$ of $\mathcal{F}(\lambda, w)$ and $\mathcal{F}'(\lambda, w)$.²² Since in this case $f_n(w)$ is a factor in the determinant equation, (17.6.4) can be replaced by

$$\text{Dis}\{\mathcal{F}; w\} = 0, \quad (17.6.5)$$

where

$$\text{Dis}\{\mathcal{F}; w\} := (-1)^{\frac{1}{2}n(n-1)} [f_n(w)]^{-1} \text{Res}\{\mathcal{F}, \mathcal{F}'; w\} \quad (17.6.6)$$

is called the *discriminant* (Hille 1962). Given that the coefficients $f_j(w)$ are polynomials of degree m at most, it is clear that $\Delta(w) := \text{Dis}\{\mathcal{F}; w\}$ is a polynomial in w of degree not exceeding $2m(n-1)$.

Algebraic functions feature in the methods described in this book in two way. Firstly is in the analysis of partition function zeros in Sect. 11.4 and secondly in the algebraic approximant method for the analysis of series expansions in Sect. 14.6.4 (see also Sect. 18.5.4). Most of this appendix is concerned with the former. However the importance for the latter resides in the relationship between differential and algebraic approximants. That the algebraic function $\lambda(w)$ is the solution of a linear differential equation with polynomial coefficients was observed and analysed in the nineteenth century (see Bostan et al. 2007, for references). More recently it has become a topic of some interest to computer science. The following result can be established (Cormier et al. 2002; Bostan et al. 2007):

²¹ The irreducibility of $\mathcal{F}(\lambda, w)$ is not always taken as part of the definition of an algebraic function (Bliss 1966). We shall however, following Hille (1962), assume this to be included.

²² The *resultant* of two polynomials is given by their Sylvester determinant (Bliss 1966).

Theorem 17.6.1 When $\Delta(w) \neq 0$ the algebraic function $\lambda(w)$, defined by Eqs. (17.6.1) and (17.6.2), is the solution of a linear differential equation

$$c_s(w)[\Delta(w)]^s \frac{d^s \lambda(w)}{dw^s} + \cdots c_1(w)\Delta(w) \frac{d\lambda(w)}{dw} + c_0(w)\lambda(w) = 0, \quad (17.6.7)$$

where $s \leq n$.

In fact $\lambda(w)$ can be shown to also be the solution of an inhomogeneous differential equation.

Example 17.6.1 For the quadratic case, $n = 2$, the example given by Cormier et al. (2002, p. 367) is

$$\begin{aligned} f_2(w) &= 1, & f_1(w) &= -2w(1 + 2w), \\ f_0(w) &= -w^2(w^5 - w^3 - 4w^2 - 4w - 1), \end{aligned} \quad (17.6.8)$$

for which

$$\Delta(w) = 4w^5(w^2 - 1) \quad (17.6.9)$$

and (17.6.7) is satisfied with $s = 2$ and

$$\begin{aligned} c_2(w) &= 6w^3 + 5w^2 - 2w - 3, \\ c_1(w) &= -2w^4(54w^5 + 35w^4 - 76w^3 - 54w^2 + 14w + 15), \\ c_0(w) &= 8w^8(w^2 - 1)(84w^5 + 35w^4 - 120w^3 - 54w^2 + 20w + 15). \end{aligned} \quad (17.6.10)$$

The quadratic case has also been studied by Hunter and Baker (1979) who showed that $\lambda(w)$ satisfies the inhomogeneous equation

$$a_1(w)\Delta(w) \frac{d\lambda(w)}{dw} + a_0(w)\lambda(w) = b(w), \quad (17.6.11)$$

where

$$\begin{aligned} a_0(w) &= 2w^4(5 - 7w^2), & a_1(w) &= 1, \\ b(w) &= -2w^5(6w^3 + 5w^2 - 2w - 3). \end{aligned} \quad (17.6.12)$$

Eliminating the inhomogeneous terms between (17.6.11) and its derivative will, of course, yield the homogeneous second-order equation given by (17.6.7) and (17.6.10).

We now turn to material relevant to Sect. 11.4 and a theorem proved by Hille (1962):

Theorem 17.6.2 *The branches $\lambda_1(w), \lambda_2(w), \dots, \lambda_n(w)$ of the algebraic function $\lambda(w)$ defined by equations (17.6.1) and (17.6.2) are analytic functions in any simply-connected region of the complex plane which excludes (i) the zeros of $f_n(w)$, (ii) solutions of (17.6.5), (iii) the point at infinity.*

In general the zeros of $f_n(w)$ will be poles of one or more branches of $\lambda(w)$. However, in the application of the theory of this section in Sect. 11.4, the polynomials of interest have constant leading coefficients so the algebraic functions have no infinities. The solutions of (17.6.5) are points where two or more of the branches of $\lambda(w)$ have the same value. Suppose that w_0 is such a point and that the branches of $\lambda(w)$ are labelled in such a way that $\lambda_1(w), \lambda_2(w), \dots, \lambda_p(w)$, ($1 < p \leq n$) is the complete set of branches for which

$$\lambda_1(w_0) = \lambda_2(w_0) = \dots = \lambda_p(w_0). \quad (17.6.13)$$

There are now three possibilities:

- (i) The branches which satisfy (17.6.13) can be ordered in such a way that, for small positive real r ,

$$\lambda_1(w_0 + r \exp[2\ell i\pi]) = \lambda_{\ell+1}(w_0 + r), \quad (17.6.14)$$

where ℓ is an integer modulo p . It is clear that $\lambda_1(w)$ is an analytic function of the variable $(w - w_0)^{1/p}$ in a neighbourhood of w_0 and the power-series expansion

$$\sum_{q=0}^{\infty} c_q (w - w_0)^{q/p} \quad (17.6.15)$$

is applicable to all the branches $\lambda_1(w), \lambda_2(w), \dots, \lambda_p(w)$ in a neighbourhood of w_0 , when $(w - w_0)^{1/p}$ takes its p different values. The point w_0 is a *branch-point* of $\lambda(w)$ with *cycle number* p . To distinguish this case from case (iii) below, such a branch-point is called *monocyclic*.

- (ii) The branches which satisfy (17.6.13) also satisfy the condition

$$\lambda_{\ell}(w_0 + r) = \lambda_{\ell}(w_0 + r \exp[2i\ell\pi]), \quad \ell = 1, 2, \dots, p, \quad (17.6.16)$$

for any integer r . This is a case where each of the branches is analytic at w_0 , but they happen to have the same value there. Joyce (1988b) has referred to points of this type as *apparent singular points*.

- (iii) The branches which satisfy (17.6.13) divide into distinct subsets containing $p_1, p_2, \dots, p_k$ members where

$$\sum_{s=1}^k p_s = p, \quad (17.6.17)$$

and, within the s th subset, the members have a cycle structure like that given by (17.6.14) and a power-series expansion like (17.6.15), with cycle number p_s . This can be thought of as a situation where k branch-points happen to coincide. We can also include within this case a combination of coincident branch-points and analytic components, where the latter correspond to instances with $p_s = 1$. A branch-point of this type is called *k-fold multicyclic*.

The reason why Theorem 17.6.2 contains the restriction to simply-connected regions is now clear. In a multi-connected region around a branch-point some of the branches of $\lambda(w)$ will not be single-valued. The usual way to define a region within which the branches of $\lambda(w)$ are analytic is to define *cuts* in the complex plane, which prevent the construction of any curve along which a branch of $\lambda(w)$ passes continuously from one of its values to another. Within such a cut plane each branch of $\lambda(w)$ will be analytic and if we take n copies of the complex plane and join them in a suitable way along the edges of the cuts a *Riemann surface* of n sheets can be constructed on which $\lambda(w)$ is itself analytic.²³

The set of cuts needed to construct the Riemann surface for any algebraic function is not unique and we shall describe one of the methods for doing this based on the idea of *connection curves* introduced by Wood (1987). These are defined by all values of w which are simultaneous solutions of the equations

$$\mathcal{F}(\lambda, w) = 0, \quad \mathcal{F}(\varpi\lambda, w) \equiv \mathcal{F}^{(\varpi)}(\lambda, w) = 0, \quad \forall \varpi = \exp(i\vartheta), \quad (17.6.18)$$

where $\vartheta \in [0, 2\pi)$. The connection curves are then given by

$$\text{Res}\{\mathcal{F}, \mathcal{F}^{(\varpi)}; w\} = 0. \quad (17.6.19)$$

It is clear that the resultant of $\mathcal{F}(\lambda, w)$ and $\mathcal{F}^{(\varpi)}(\lambda, w)$ is identically zero when $\varpi = 1$ with

$$\begin{aligned} \text{Res}\{\mathcal{F}, \mathcal{F}^{(\varpi)}; w\} &= \text{Res}\{\mathcal{F}, \mathcal{F}^{(\varpi)} - \mathcal{F}; w\} \\ &= (\varpi - 1)^n \text{Res}\left\{\mathcal{F}, \frac{\mathcal{F}^{(\varpi)} - \mathcal{F}}{\varpi - 1}; w\right\} \end{aligned} \quad (17.6.20)$$

and

$$\frac{\mathcal{F}_n^{(\varpi)}(\lambda, w) - \mathcal{F}(\lambda, w)}{\varpi - 1} = \lambda \mathcal{G}^{(\varpi)}(\lambda, w), \quad (17.6.21)$$

where

$$\begin{aligned} \mathcal{G}^{(\varpi)}(\lambda, w) &= f_n(w) \phi^{(n)}(\varpi) \lambda^{n-1} + f_{n-1}(w) \phi^{(n-1)}(\varpi) \lambda^{n-2} \\ &\quad + \cdots + f_1(w), \end{aligned} \quad (17.6.22)$$

²³ For an account of the theory of Riemann surfaces see Beardon (1984).

$$\phi^{(s)}(\varpi) = \frac{\varpi^s - 1}{\varpi - 1}. \quad (17.6.23)$$

The determinant $\text{Res}\{\mathcal{F}, \frac{\mathcal{F}^{(\varpi)} - \mathcal{F}}{\varpi - 1}; w\}$ has only one non-zero element in the last column, which is $f_0(w)$ in the n th row. So

$$\text{Res}\{\mathcal{F}, \mathcal{F}^{(\varpi)}; w\} = -f_0(w)(1 - \varpi)^n \text{Res}\{\mathcal{F}, \mathcal{G}^{(\varpi)}; w\}. \quad (17.6.24)$$

The connection curves are thus given by

$$\text{Res}\{\mathcal{F}, \mathcal{G}^{(\varpi)}; w\} = 0, \quad (17.6.25)$$

and, since

$$\lim_{\varpi \rightarrow 1} \phi^s(\varpi) = s, \quad (17.6.26)$$

it follows, from (17.6.22), that

$$\lim_{\varpi \rightarrow 1} \text{Res}\{\mathcal{F}, \mathcal{G}^{(\varpi)}; w\} = \text{Res}\{\mathcal{F}, \mathcal{F}'; w\}. \quad (17.6.27)$$

The solutions of (17.6.5), which are the branch-points and the apparent singular points, correspond to points on the connection curves when $\varpi = 1$.

17.7 Determinants of Cyclic Matrices

Let $\mathbf{Z}$ be an N -dimensional square matrix with elements $z(\mathbf{j}, \boldsymbol{\ell})$, indexed by two-dimensional vectors of the form

$$\mathbf{j} := (j_1, j_2), \quad j_1 = 1, 2, \dots, N_1, \quad j_2 = 1, 2, \dots, N_2, \quad (17.7.1)$$

with $N = N_1 N_2$. $\mathbf{Z}$ is taken to be *cyclic*, meaning that

$$\begin{aligned} z(\mathbf{j}, \boldsymbol{\ell}) &= z(\boldsymbol{\ell} - \mathbf{j}), \\ z(\mathbf{j} + \mathbf{n}) &= z(\mathbf{j}), \quad \mathbf{n} = (n_1 N_1, n_2 N_2), \quad \forall \text{ integers } n_1 \text{ and } n_2. \end{aligned} \quad (17.7.2)$$

For $\boldsymbol{\ell} := (\ell_1, \ell_2)$, the notation $z(\boldsymbol{\ell}) := z(\ell_1, \ell_2)$ is used. Let $\boldsymbol{\Omega}$ be the N -dimensional square matrix with

$$\langle \mathbf{j} | \boldsymbol{\Omega} | \boldsymbol{\ell} \rangle = \omega(\mathbf{j}, \boldsymbol{\ell}) := \frac{1}{\sqrt{N}} \exp \left[2\pi i \left(\frac{j_1 \ell_1}{N_1} + \frac{j_2 \ell_2}{N_2} \right) \right], \quad (17.7.3)$$

where vectors $\mathbf{j}$ and $\boldsymbol{\ell}$ are of the form (17.7.1). It is not difficult to show that $\boldsymbol{\Omega}$ is unitary and

Theorem 17.7.1 $\mathbf{Z}$ is diagonalized by the transformation $\mathbf{\Omega}^{-1}\mathbf{Z}\mathbf{\Omega}$ and

$$\text{Det}\{\mathbf{Z}\} = \prod_{\{\mathbf{k}\}} \zeta(\mathbf{k}), \quad (17.7.4)$$

where

$$\zeta(\mathbf{k}) := \sum_{\ell_1=1}^{N_1} \sum_{\ell_2=1}^{N_2} z(\ell_1, \ell_2) \exp \left[2\pi i \left(\frac{\ell_1 k_1}{N_1} + \frac{\ell_2 k_2}{N_2} \right) \right]. \quad (17.7.5)$$

Proof The matrix $\mathbf{Z}' = \mathbf{\Omega}^{-1}\mathbf{Z}\mathbf{\Omega}$ has the elements

$$\begin{aligned} z'(\mathbf{j}, \mathbf{k}) &= \sum_{\{\ell\}} \sum_{\{\ell'\}} \overline{\omega(\ell, \mathbf{j})} z(\ell' - \ell) \omega(\ell', \mathbf{k}) \\ &= \sqrt{N} \sum_{\{\ell\}} \overline{\omega(\ell, \mathbf{j})} \omega(\ell, \mathbf{k}) \sum_{\{\ell'\}} z(\ell' - \ell) \omega(\ell' - \ell, \mathbf{k}). \end{aligned} \quad (17.7.6)$$

Since $z(\mathbf{j})\omega(\mathbf{j}, \mathbf{k}) = z(\mathbf{j} + \mathbf{n})\omega(\mathbf{j} + \mathbf{n}, \mathbf{j})$, for any of the vectors $\mathbf{n}$ defined in (17.7.2), $\ell' - \ell$ can be replaced by ℓ' in the second summation, giving, from (17.7.4) and (17.7.5),

$$z'(\mathbf{j}, \mathbf{k}) = \sqrt{N} \sum_{\{\ell\}} \overline{\omega(\ell, \mathbf{j})} \omega(\ell, \mathbf{k}) \sum_{\{\ell'\}} z(\ell') \omega(\ell', \mathbf{k}) = \delta^{\text{Kr}}(\mathbf{j} - \mathbf{k}) \zeta(\mathbf{k}). \quad (17.7.7)$$

(17.7.4) then follows from $\text{Det}\{\mathbf{Z}\} = \text{Det}\{\mathbf{Z}'\}$. $\square$

These results can be generalized by replacing each element $z(\mathbf{j}, \ell)$ of $\mathbf{Z}$ by a $\eta \times \eta$ dimensional submatrix, or block, $z(\mathbf{j}, \ell)$ with

$$z(\mathbf{j}, \ell) = z(\ell - \mathbf{j}), \quad z(\mathbf{j} + \mathbf{n}) = z(\mathbf{j}). \quad (17.7.8)$$

Such a matrix $\mathbf{Z}$ is called a *cyclic block matrix*. With $\omega(\mathbf{j}, \ell)$ replaced by $\omega(\mathbf{j}, \ell) = \omega(\mathbf{j}, \ell) \mathbf{I}_\eta$, where $\mathbf{I}_\eta$ is the unit matrix of dimension η , the blocks $z'(\mathbf{j}, \ell)$ of $\mathbf{Z}' = \mathbf{\Omega}^{-1}\mathbf{Z}\mathbf{\Omega}$ are given by

$$z'(\mathbf{j}, \mathbf{k}) = \delta^{\text{Kr}}(\mathbf{j} - \mathbf{k}) \zeta(\mathbf{k}), \quad (17.7.9)$$

where

$$\zeta(\mathbf{k}) := \sum_{\{\ell\}} z(\ell) \exp \left[2\pi i \left(\frac{\ell_1 k_1}{N_1} + \frac{\ell_2 k_2}{N_2} \right) \right]. \quad (17.7.10)$$

Equation (17.7.4) is then replaced by

$$\text{Det}\{\mathbf{Z}\} = \prod_{\{\mathbf{k}\}} \text{Det}\{\zeta(\mathbf{k})\}. \quad (17.7.11)$$

Now consider the limit $N_s \rightarrow \infty$, for $s = 1, 2$. In (17.7.10) the vector $\mathbf{k}$ is replaced by the vector $\boldsymbol{\theta} := (\theta_1, \theta_2)$, where $\theta_s := 2\pi i k_s / N_s$, $s = 1, 2$. In the limit $N_s \rightarrow \infty$, θ_s becomes a continuous variable in the range $[0, 2\pi]$. From (17.7.11)

$$\lim_{N_1 \rightarrow \infty} \lim_{N_2 \rightarrow \infty} \frac{\ln[\text{Det}\{\mathbf{Z}\}]}{N_1 N_2} = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \ln[\text{Det}\{\boldsymbol{\zeta}(\theta_1, \theta_2)\}]. \quad (17.7.12)$$

Chapter 18

Analysis

18.1 Fourier Transforms in d Dimensions

It is convenient to use the generic term *transform* to refer to both the discrete (series) and continuous (integral) cases. Here the former is related to lattices with attention confined to a hypercubic lattice. The corresponding restriction in the continuous case is to a hypercubic volume.

18.1.1 Discrete Finite Lattices

Consider a d -dimensional hypercubic lattice $\mathcal{N}$ with lattice spacing a . From (16.2.1) the sites of the lattice are

$$\mathbf{r} = a[n^{(1)}\hat{\mathbf{r}}^{(1)} + n^{(2)}\hat{\mathbf{r}}^{(2)} + \dots + n^{(d)}\hat{\mathbf{r}}^{(d)}], \quad (18.1.1)$$

where $\hat{\mathbf{r}}^{(\ell)}$, $\ell = 1, 2, \dots, d$ are the orthogonal unit axis vectors. Periodic boundary condition are applied as defined in Sect. 16.2.1, so that $n^{(\ell)}$ is an integer modulo $N^{(\ell)}$ with the number of lattice sites N given by (16.2.5). Let $f(\mathbf{r})$ be a function defined at each point of $\mathcal{N}$. Because of the periodic boundary conditions

$$f(\mathbf{r}) = f(\mathbf{r} + \mathbf{r}_L), \quad (18.1.2)$$

where

$$\mathbf{r}_L := a[m^{(1)}N^{(1)}\hat{\mathbf{r}}^{(1)} + m^{(2)}N^{(2)}\hat{\mathbf{r}}^{(2)} + \dots + m^{(d)}N^{(d)}\hat{\mathbf{r}}^{(d)}], \quad (18.1.3)$$

for any integers $m^{(1)}, \dots, m^{(d)}$. The Fourier transform $f^*(\mathbf{k})$ of $f(\mathbf{r})$ is given by¹

¹ It is sometimes convenient to use the notation $\mathcal{F}(f(\mathbf{r}))$ to denote the Fourier transform.

$$f^*(\mathbf{k}) := \sum_{\{\mathbf{r}\}} f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}), \quad (18.1.4)$$

$$f(\mathbf{r}) = \frac{1}{N} \sum_{\{\mathbf{k}\}} f^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (18.1.5)$$

with the wave-vectors $\mathbf{k}$ given by

$$\mathbf{k} := \frac{2\pi}{a} \left[\frac{\kappa^{(1)}}{N^{(1)}} \hat{\mathbf{r}}^{(1)} + \frac{\kappa^{(2)}}{N^{(2)}} \hat{\mathbf{r}}^{(2)} + \cdots + \frac{\kappa^{(d)}}{N^{(d)}} \hat{\mathbf{r}}^{(d)} \right], \quad (18.1.6)$$

where in (18.1.5) $\kappa^{(\ell)}$ ranges over all distinct integer values modulo $N^{(\ell)}$, for $\ell = 1, 2, \dots, d$. From (16.2.6)

$$f^*(\mathbf{k}) = f^*(\mathbf{k} + \mathbf{r}_R), \quad (18.1.7)$$

where $\mathbf{r}_R$ is any reciprocal lattice vector. When $d = 1$, it follows from (16.2.1), (18.1.3)–(18.1.6) that a function $f(n)$ with periodicity N has transform $f^*(\kappa)$ given by

$$f^*(\kappa) = \sum_{n=0}^{N-1} f(n) \exp\left(-\frac{2\pi i \kappa n}{N}\right), \quad (18.1.8)$$

$$f(n) = \frac{1}{N} \sum_{\kappa=0}^{N-1} f^*(\kappa) \exp\left(\frac{2\pi i \kappa n}{N}\right). \quad (18.1.9)$$

The periodic distribution

$$\widetilde{\delta^{\text{Kr}}}(N; n) := \sum_{j=-\infty}^{\infty} \delta^{\text{Kr}}(n - Nj) = \begin{cases} 1, & \text{if } n \text{ is a multiple of } N, \\ 0, & \text{otherwise,} \end{cases} \quad (18.1.10)$$

can be transformed, using (18.1.8) and (18.1.9), into

$$\widetilde{\delta^{\text{Kr}}}(N; n) = \frac{1}{N} \sum_{\kappa=0}^{N-1} \exp\left(\frac{2\pi i \kappa n}{N}\right), \quad (18.1.11)$$

with $\widetilde{\delta^{\text{Kr}}}(N; n) = \delta^{\text{Kr}}(n)$, if $0 \leq n < N$.

18.1.2 A Continuous Finite Volume

The case where $\mathbf{r} \in \mathcal{V}$, a hyper-cubic volume $\mathcal{V}$ of edge-lengths $L^{(1)}, \dots, L^{(d)}$ with periodic boundary conditions and

$$|\mathcal{V}| = \tilde{V} = \prod_{i=1}^d L^{(i)}, \quad (18.1.12)$$

can be obtained as the limit of the lattice case when $a \rightarrow 0$, $N_i \rightarrow \infty$ with $N^{(i)}a = L^{(i)}$, for $i = 1, 2, \dots, d$. With $f^*(\mathbf{k})$ replaced by $Nf^*(\mathbf{k})$, Eq. (18.1.4) is replaced by an integral according to the prescription

$$\frac{1}{N} \sum_{\{\mathbf{r}\}} = \frac{a^d}{\tilde{V}} \sum_{\{\mathbf{r}\}} \rightarrow \frac{1}{\tilde{V}} \int_{\mathcal{V}} d\mathcal{V}, \quad (18.1.13)$$

giving, in place of (18.1.4) and (18.1.5),

$$f^*(\mathbf{k}) := \frac{1}{\tilde{V}} \int_{\mathcal{V}} f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathcal{V}, \quad (18.1.14)$$

$$f(\mathbf{r}) = \sum_{\{\mathbf{k}\}} f^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (18.1.15)$$

where the definition of the values of $\mathbf{k}$ given by (18.1.6) is replaced by

$$\mathbf{k} := 2\pi \left[\frac{\kappa^{(1)}}{L^{(1)}} \hat{\mathbf{r}}^{(1)} + \frac{\kappa^{(2)}}{L^{(2)}} \hat{\mathbf{r}}^{(2)} + \dots + \frac{\kappa^{(d)}}{L^{(d)}} \hat{\mathbf{r}}^{(d)} \right], \quad k^{(i)} = 0, \pm 1, \pm 2, \dots \quad (18.1.16)$$

Taking the gradient of (18.1.15) gives

$$\nabla f(\mathbf{r}) = i \sum_{\{\mathbf{k}\}} \mathbf{k} f^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (18.1.17)$$

and then taking the divergence of (18.1.17) gives

$$\nabla^2 f(\mathbf{r}) = - \sum_{\{\mathbf{k}\}} k^2 f^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}). \quad (18.1.18)$$

Thus

$$k^2 f^*(\mathbf{k}) = - \frac{1}{\tilde{V}} \int_{\mathcal{V}} \nabla^2 f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathcal{V}. \quad (18.1.19)$$

For the case $d = 1$,

$$f^*(n) = \frac{1}{L} \int_{-L/2}^{L/2} f(r) \exp\left(-\frac{2\pi i k r}{L}\right) dr, \quad (18.1.20)$$

$$f(r) = \sum_{\kappa=-\infty}^{\infty} f^*(\kappa) \exp\left(\frac{2\pi i k r}{L}\right). \quad (18.1.21)$$

In this instance we shall also need the convolution formula

$$\mathcal{F}\left(\int_{-L/2}^{L/2} f(s)g(r-s)ds\right) = L f^*(n) g^*(n). \quad (18.1.22)$$

The periodic distribution

$$\widetilde{\delta^D}(L; r) := \sum_{j=-\infty}^{\infty} \delta^D(r - jL) \quad (18.1.23)$$

of Dirac delta functions can be transformed using (18.1.20) and (18.1.21) into

$$\widetilde{\delta^D}(L; r) = \frac{1}{L} \sum_{\kappa=-\infty}^{\infty} \exp\left(\frac{2\pi i k r}{L}\right), \quad (18.1.24)$$

and $\widetilde{\delta^D}(L; r) = \delta^D(r)$ if $0 \leq r < L$. With the change of variable $r := xL/(2\pi)$, (18.1.20) and (18.1.21) give the usual Fourier series formulae for a function periodic over the interval $[-\pi, \pi]$. The condition for the convergence of the series (18.1.15) and (18.1.21) is that $f(r)$ is piecewise smooth on every finite interval in all its component variables (Courant and Hilbert 1962).

Suppose that $\varphi(x)$ is a continuous and continuously differentiable function of x so that the series

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n + y), \quad \sum_{n=-\infty}^{\infty} \frac{\partial \varphi(2\pi n + y)}{\partial y} \quad (18.1.25)$$

converge absolutely and uniformly for $-\pi \leq y < \pi$. Then the second series in (18.1.25) is the derivative of the first which can be transformed using (18.1.20) and (18.1.21) to give

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n + y) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \varphi(2\pi n + x) \exp[im(y-x)] dx. \quad (18.1.26)$$

Setting $y = 0$ and using the transform

$$\sum_{n=-\infty}^{\infty} \int_{-\pi}^{\pi} \varphi(2\pi n + x) \exp(-imx) dx = \int_{-\infty}^{\infty} \varphi(x) \exp(-imx) dx$$

gives

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x) \exp(-imx) dx, \quad (18.1.27)$$

which is the *Poisson summation formula* (Courant and Hilbert 1962).

18.1.3 A Continuous Infinite Volume

Suppose that $f^*(\mathbf{k})$ in Sect. 18.1.2 is replaced by $f^*(\mathbf{k})/\tilde{V}$ and consider the limiting situation $L^{(i)} \rightarrow \infty$, for $i = 1, 2, \dots, d$. The sum in (18.1.15) becomes an integral according to the prescription

$$\frac{1}{\tilde{V}} \sum_{\{\mathbf{k}\}} = \frac{1}{(2\pi)^d} \prod_{i=1}^d \left(\frac{2\pi}{L_i} \right) \sum_{\{\mathbf{k}\}} \rightarrow \frac{1}{(2\pi)^d} \int_{\mathcal{K}} d\mathcal{K}, \quad (18.1.28)$$

giving

$$f^*(\mathbf{k}) := \int_{\mathcal{R}} f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathcal{R}, \quad (18.1.29)$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^d} \int_{\mathcal{K}} f^*(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) d\mathcal{K}, \quad (18.1.30)$$

where $\mathcal{R}$ and $\mathcal{K}$ are the infinite d -dimensional spaces of the vectors $\mathbf{r}$ and $\mathbf{k}$ respectively. For $d = 1$ this becomes

$$f^*(k) := \int_{-\infty}^{\infty} f(r) \exp(-ikr) dr, \quad (18.1.31)$$

$$f(r) = \frac{1}{(2\pi)} \int_{-\infty}^{\infty} f^*(k) \exp(ikr) dk. \quad (18.1.32)$$

The function f is dependent only on the magnitude r of the vector $\mathbf{r}$ if and only if f^* is a function only of the magnitude k of $\mathbf{k}$. In this case, for $d > 1$, the integrals in (18.1.29) and (18.1.30) can be expressed in polar variables in the form

$$f^*(k) = \int_0^\infty dr f(r)r \int_0^\pi d\theta \mathcal{S}_{d-1}(r\sin\theta) \exp(-ikr\cos\theta), \quad (18.1.33)$$

$$f(r) = \frac{1}{(2\pi)^d} \int_0^\infty dk f^*(k)k \int_0^\pi d\theta \mathcal{S}_{d-1}(k\sin\theta) \exp(irk\cos\theta), \quad (18.1.34)$$

where

$$\mathcal{S}_d(q) := 2\pi^{\frac{d}{2}} q^{d-1} / \Gamma\left(\frac{d}{2}\right), \quad (18.1.35)$$

is the surface area of a d -dimensional hypersphere of radius q .² To evaluate the integrals (18.1.33) and (18.1.34) we need the *Poisson integral formula*

$$\int_0^\pi \sin^{2\nu}\theta \exp(\pm iz\cos\theta) d\theta = \pi^{\frac{1}{2}} \Gamma\left(\nu + \frac{1}{2}\right) (2/z)^\nu J_\nu(z), \quad \Re(\nu) > -\frac{1}{2}, \quad (18.1.36)$$

for the *Bessel function of the first kind* $J_\nu(z)$. Substituting from (18.1.35) into (18.1.33) and (18.1.34) and using (18.1.36) we have, for all $d > 1$,

$$f^*(k) = k \int_0^\infty f(r) \left(\frac{2\pi r}{k}\right)^{\frac{d}{2}} J_{\frac{d}{2}-1}(kr) dr, \quad (18.1.37)$$

$$f(r) = r \int_0^\infty f^*(k) \left(\frac{k}{2\pi r}\right)^{\frac{d}{2}} J_{\frac{d}{2}-1}(kr) dk. \quad (18.1.38)$$

Since $J_{-\frac{1}{2}}(z) = (2/z\pi)^{\frac{1}{2}} \cos(z)$ these equations become (18.1.31) and (18.1.32) when $d = 1$.

18.1.4 Integrals Involving Bessel Functions

For Fourier transforms like (18.1.37) and (18.1.38) we need to evaluate integrals containing Bessel functions of the first kind. Cases of this occur in Sects. 3.6.1 and 5.6.3 with the integrals to be solved contained in the definitions of $\mathfrak{F}(\ell_0, k_0; x)$ and $\mathfrak{G}(k_0; x, y; z)$, given respectively by (3.6.24) and (5.6.34).

For real x and y it can be shown that

$$\lim_{k_0 \rightarrow \infty} \mathfrak{G}(k_0; x, y; 0) = \Omega(x, y), \quad \text{for } 1 - y < x < \frac{5}{2}, \quad (18.1.39)$$

² In one dimension the conditions that $f(\mathbf{r})$ and $f^*(\mathbf{k})$ are dependent only on the magnitudes of the vectors reduce to them being even functions. $\mathcal{S}_0(q)$ has just two points and can be defined as $\delta^D(q)$, when (18.1.33) and (18.1.34) reduce to (18.1.31) and (18.1.32).

where

$$\Omega(x, y) := \frac{2^{x-2} \Gamma\left(\frac{1}{2}y + \frac{1}{2}x - \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}y - \frac{1}{2}x + \frac{3}{2}\right)}, \quad (18.1.40)$$

and that the corresponding limit exists for (real) $z \neq 0$ when $-1 - y < x < \frac{5}{2}$. In particular

$$\mathfrak{G}(\infty; x, x-1; z) = z^{x-1} K_{x-1}(z), \quad \text{for } x < \frac{5}{2}, \quad (18.1.41)$$

where $K_x(z)$ is the *Bessel function of complex argument*, for which we shall need the information that

$$K_x(z) \simeq (\pi/2z)^{\frac{1}{2}} \exp(-z) \quad \forall x, \quad \text{as } |z| \rightarrow \infty, \quad (18.1.42)$$

$$K_0(z) \simeq -\ln(z/2), \quad K_1(z) \simeq 1/z, \quad \text{as } |z| \rightarrow 0, \quad (18.1.43)$$

with (18.1.42) exact for $x = \pm \frac{1}{2}$. To complete the requirements of Sect. 5.6.3 we need an asymptotic expression for $\mathfrak{G}(k_0; x, y; z)$ for $x \geq \frac{5}{2}$. A series with ascending powers of z can be generated by iterating

$$\mathfrak{G}(k_0; x, y; z) = \mathfrak{G}(k_0; x, y; 0) + z^2 \mathfrak{G}(k_0; x-2, y; z) \quad (18.1.44)$$

until, whatever the value of $x > \frac{5}{2}$, the last term $z^{2n} \mathfrak{G}(k_0; x-2n, y; z)$ is such that $0 < x-2n < \frac{5}{2}$. The most singular term in this series is the first, giving

$$\mathfrak{G}(k_0; x, y; z) \simeq \lim_{k_0 \rightarrow \infty} \mathfrak{G}(k_0; x, y; 0). \quad (18.1.45)$$

The range of (18.1.39) can be extended with

$$\begin{aligned} \mathfrak{G}(k_0; x, y; 0) &= \Omega(x, y) + k_0(x+y-3)J_y(k_0) + S_{x-3, y-1}(k_0) \\ &\quad - k_0 J_{y-1}(k_0) S_{x-2, y}(k_0), \quad \text{for } 1-y < x, \end{aligned} \quad (18.1.46)$$

where $S_{x, y}(z)$ is a *Lommel function*. Using the asymptotic forms

$$J_y(z) \simeq \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{1}{2}\pi y - \frac{1}{4}\pi\right), \quad S_{x, y}(z) \simeq z^{x-1}, \quad (18.1.47)$$

as $z \rightarrow \infty$, gives

$$\begin{aligned} \mathfrak{G}(k_0; x, y; 0) &= \Omega(x, y) - (x+y-3)\sqrt{\frac{2k_0}{\pi}} \cos\left(k_0 - \frac{1}{2}\pi[y-1]\right) + k_0^{x-4} \\ &\quad + k_0^{x-\frac{5}{2}} \sqrt{\frac{2}{\pi}} \cos\left(x - \frac{1}{4}\pi[2y-1]\right), \quad \text{for } 1-y < x. \end{aligned} \quad (18.1.48)$$

From (3.6.23), interest in $\mathfrak{F}(\ell_0, k_0; x)$ can be confined to the values $x = -2, -1, 0, 1, \dots$ for which, from (18.1.46) and (18.1.47),

$$\begin{aligned}\mathfrak{F}(0, k_0; x) &= \frac{k_0^{x+1}}{x+1} - \mathfrak{G}(k_0; x+2, 0; 0) \\ &\simeq \frac{k_0^{x+1}}{x+1} - \Omega(x+2, 0) - k_0^{x-\frac{7}{2}} \cos(k_0 - \tfrac{1}{4}), \quad x = 0, 1, \dots\end{aligned}\quad (18.1.49)$$

For the remaining cases the range of integration is divided into parts to give

$$\begin{aligned}\mathfrak{F}(\ell_0, k_0; -1) &= \int_{\ell_0}^{k_0} \frac{1 - J_0(k)}{k} dk \\ &= \int_{\ell_0}^1 \frac{1 - J_0(k)}{k} dk + \ln(k_0) - \int_1^{k_0} \frac{J_0(k)}{k} dk,\end{aligned}\quad (18.1.50)$$

$$\begin{aligned}\mathfrak{F}(\ell_0, k_0; -2) &= \int_{\ell_0}^{k_0} \frac{1 - J_0(k)}{k^2} dk \\ &= \int_{\ell_0}^1 \frac{1 - J_0(k)}{k^2} dk + \frac{k_0 - 1}{k_0} - \int_1^{k_0} \frac{J_0(k)}{k^2} dk,\end{aligned}\quad (18.1.51)$$

where

$$\lim_{\ell_0 \rightarrow 0} \int_{\ell_0}^1 \frac{1 - J_0(k)}{k} dk = \frac{1}{8} {}_2F_3(1, 1; 2, 2, 2; -\tfrac{1}{4}) \approx 0.1212, \quad (18.1.52)$$

$$\lim_{k_0 \rightarrow \infty} \int_1^{k_0} \frac{J_0(k)}{k} dk = \ln(2) - \gamma + \frac{1}{8} {}_2F_3(1, 1; 2, 2, 2; -\tfrac{1}{4}) \approx 0.2371, \quad (18.1.53)$$

$$\begin{aligned}\lim_{\ell_0 \rightarrow 0} \int_{\ell_0}^1 \frac{1 - J_0(k)}{k^2} dk &= \lim_{k_0 \rightarrow \infty} \int_1^{k_0} \frac{J_0(k)}{k^2} dk \\ &= J_1(1) {}_1F_2(1; \tfrac{3}{2}, \tfrac{3}{2}; -\tfrac{1}{4}) - \tfrac{1}{3} J_0(1) {}_1F_2(1; \tfrac{3}{2}, \tfrac{5}{2}; -\tfrac{1}{4}) \\ &\quad + 2J_0(1) - J_1(1) - 1 \approx 0.2449,\end{aligned}\quad (18.1.54)$$

${}_pF_q(\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q; z)$ is the *generalized hypergeometric function* and $\gamma \approx 0.5772$ is the *Euler-Mascheroni constant*.

18.1.5 Lattice Green's Functions

The inverse Fourier transform

$$G(d; \mathbf{r}) := \frac{1}{L^d} \sum_{\{\mathbf{k}\}} \frac{[\exp(i\mathbf{k} \cdot \mathbf{r}) - 1]}{k^2} \quad (18.1.55)$$

is a *Green's function* for the d -dimensional hypercubic lattice $\mathcal{N}$. For convergence in the continuum limit the wave-vector magnitude is confined to the range $\ell_1 < k < k_1$. Then, from (18.1.34) and (18.1.38),

$$\begin{aligned} G(d; r) &= \frac{1}{(2\pi)^d} \int_{\ell_1}^{k_1} dk k^{-1} \int_0^\pi d\theta \mathcal{S}_{d-1}(k \sin \theta) [\exp(irk \cos \theta) - 1] \\ &= \frac{1}{2\pi} \int_{\ell_1}^{k_1} k^{-1} \left(\frac{k}{2\pi r} \right)^{\frac{d}{2}-1} J_{\frac{d}{2}-1}(kr) dk \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\ell_1}^{k_1} k^{d-3} \left\{ \frac{J_{\frac{d}{2}-1}(kr)}{(kr)^{\frac{d}{2}-1}} - \frac{1}{2^{\frac{d}{2}-1} \Gamma(\frac{d}{2})} \right\} dk. \end{aligned} \quad (18.1.56)$$

Using (3.6.24),

$$G(2; r) = -\frac{1}{2\pi} \mathfrak{F}\left(\frac{2\pi r}{R}, \frac{2\pi r}{a}; -1\right), \quad (18.1.57)$$

where for use in the two-dimensional XY model this formula has been confined to the range $a \leq r \leq R$, so that $\ell_1 := 2\pi/R$, $k_1 := 2\pi/a$. At $r = R$ it is convenient to impose continuity by setting $G(2; r) := G(2; R)$ for $r \geq R$. A similar strategy for $r < a$ however leads, as we see, to difficulties in application to the XY model. We therefore let $G(2; r) := G(2; a) - \mu/4\pi^2$, for $r < a$ and some, as yet to be determined μ . Thus, from (18.1.50), (18.1.52) and (18.1.53), for large R and small a ,

$$G(2; r) = \begin{cases} -\frac{1}{4\pi^2} \{\mu + 2\pi \ln(\pi) + 2\pi\gamma\}, & r \leq a, \\ -\frac{1}{2\pi} \left\{ \ln\left(\frac{r}{a}\right) + \ln(\pi) + \gamma \right\}, & a \leq r \leq R, \\ -\frac{1}{2\pi} \left\{ \ln\left(\frac{R}{a}\right) + \ln(\pi) + \gamma \right\}, & r \geq R. \end{cases} \quad (18.1.58)$$

18.2 Doubly-Periodic and Quasi-Periodic Functions

Definition 18.2.1 A function $f(u)$, defined for $u \in \mathbb{C}_u$, is periodic, period ω , if

$$f(u + \omega) = f(u). \quad (18.2.1)$$

If $f(u)$ has periods $\omega_1, \omega_2, \dots, \omega_n$, then it is clearly periodic with period $m_1\omega_1 + m_2\omega_2 + \dots + m_n\omega_n$ for any integers $m_1, m_2, \dots, m_n$. The period ω is *primitive* if $f(u)$ is not periodic, period ω/m , for some integer $m > 1$. Clearly every period is a sum of multiples of primitive periods. The following theorem is due to Jacobi:

Theorem 18.2.1 *There does not exist a nonconstant function with more than two primitive periods. There is a nonconstant function with two given primitive periods if and only if the ratio of these periods is not real.*

Since $u \in \mathbb{C}_u$ is representable as a vector this theorem is equivalent to the result that there are only two independent vectors in a plane and any point in the plane can be given in terms of two independent vectors.

Definition 18.2.2 The class of functions with primitive periods ω and ω' is *doubly-periodic* and is denoted by $\text{DP}(\omega, \omega')$. The *period parallelogram* $\mathbb{W}(\omega, \omega')$ for members of this class has vertices³ $0, \omega, \omega + \omega', \omega' \in \mathbb{C}_u$.

The values of $f(u) \in \text{DP}(\omega, \omega')$ are repeated in parallelograms obtained by periodic translations of $\mathbb{W}(\omega, \omega')$. Before proceeding further we extend the definition of double-periodicity to:

Definition 18.2.3 The class $\text{DP}(\omega, \omega'; \tau, \tau')$ of functions satisfying

$$f(u + \omega) = \tau f(u), \quad f(u + \omega') = \tau' f(u), \quad (18.2.2)$$

where τ and τ' take possible values ± 1 will now be called doubly-periodic.

For this extended definition we retain the same period parallelogram but the values of $f(u) \in \text{DP}(\omega, \omega'; \tau, \tau')$ are repeated with alternating signs, if τ and τ' are equal to -1 , in parallelograms obtained by periodic translations of $\mathbb{W}(\omega, \omega')$. Clearly if, for example, $f(u) \in \text{DP}(\omega, \omega'; -1, 1)$ then $f(u) \in \text{DP}(2\omega, \omega'; 1, 1) = \text{DP}(2\omega, \omega')$ and $f(u)$ is said to be *antiperiodic* in ω and *periodic* in ω' . It is further the case that if $f(u) \in \text{DP}(\omega, \omega'; \tau, \tau')$ then $f'(u)/f(u) \in \text{DP}(\omega, \omega')$.

Members of the class $\text{DP}(\omega, \omega'; \tau, \tau')$ have a number of important properties.

Theorem 18.2.2 If $f(u) \in \text{DP}(\omega, \omega'; \tau, \tau')$ is analytic⁴ in and on the boundary of $\mathbb{W}(\omega, \omega')$ it is a constant.

This is a straightforward consequence of Liouville's theorem.⁵

Theorem 18.2.3 If $f(u) \in \text{DP}(\omega, \omega'; \tau, \tau')$ is meromorphic with exactly n poles β_j , $j = 1, \dots, n$ inside $\mathbb{W}(\omega, \omega')$ and none on its boundary and m zeros α_j , $j = 1, 2, \dots, m$ inside $\mathbb{W}(\omega, \omega')$ with none on the boundary,⁶ then $m = n$ and

$$\sum_{j=1}^n \alpha_j = \sum_{j=1}^n \beta_j + \frac{1}{4}(4\ell + 1 - \tau')\omega + \frac{1}{4}(4\ell' + 1 - \tau)\omega', \quad (18.2.3)$$

for some integers ℓ and ℓ' .⁷

³ We assume without loss of generality that this ordering of vertices is anticlockwise round the parallelogram.

⁴ See the footnote on p. 161 for the usage of the term 'analytic' for complex-valued functions. Clearly a doubly-periodic function which is analytic in and on the boundary of $\mathbb{W}(\omega, \omega')$ is an entire function.

⁵ Liouville's theorem in complex variable theory (not to be confused with his proof of the conservation of measure under a Hamiltonian flow which plays an important role in statistical mechanics) states that, if a function is entire and bounded, it is constant.

⁶ Here and elsewhere in this section it is assumed that multiple poles and zeros are counted according to their multiplicity and that, if necessary, the period parallelogram is displaced slightly so that no poles or zeros lie on its boundary.

⁷ A version of the second part of this theorem is proved by Akhiezer (1990, pp. 8–9).

Proof Let $\mathcal{W}$ be the contour around the boundary of the parallelogram $\mathbb{W}(\omega_1, \omega_2)$ in the anticlockwise direction. It is divided into four sequential parts $\mathcal{W}_i, i = 1, \dots, 4$ corresponding to the sides of $\mathbb{W}(\omega_1, \omega_2)$, where $\mathcal{W}_1$ is the edge from 0 to ω . Then, from Cauchy's theorem,

$$2\pi i(m - n) = \int_{\mathcal{W}} \frac{f'(u)}{f(u)} du = \sum_{i=1}^4 \int_{\mathcal{W}_i} \frac{f'(u)}{f(u)} du = 0, \quad (18.2.4)$$

since, because of the periodicity, the contribution to the contour integral from $\mathcal{W}_1$ cancels with $\mathcal{W}_3$ and $\mathcal{W}_2$ with $\mathcal{W}_4$. So $m = n$. In a similar way, using

$$\tau = \exp\left(\pm \frac{i\pi[4\ell + 1 - \tau]}{2}\right), \quad (18.2.5)$$

for any integer ℓ ,

$$\begin{aligned} 2\pi i \left\{ \sum_{j=1}^n (\alpha_j - \beta_j) \right\} &= \sum_{i=1}^4 \int_{\mathcal{W}_i} \frac{u f'(u)}{f(u)} du = -\omega' \ln \left\{ \frac{f(\omega)}{f(0)} \right\} + \omega \ln \left\{ \frac{f(\omega')}{f(0)} \right\} \\ &= \frac{1}{2} \omega \pi i (4\ell + 1 - \tau') + \frac{1}{2} i \omega' \pi i (4\ell' + 1 - \tau). \end{aligned} \quad (18.2.6)$$

□

Theorem 18.2.4 *If $f(u)$ and $g(u) \in \text{DP}(\omega_1, \omega_2; \tau_1, \tau_2)$ are meromorphic with exactly the same poles and zeros then $f(u) = cg(u)$ for some constant c .*

Proof Consider

$$G(u) := \frac{f'(u)}{f(u)} - \frac{g'(u)}{g(u)} \in \text{DP}(\omega_1, \omega_2; 1, 1). \quad (18.2.7)$$

It follows from Theorem 18.2.3, and Cauchy's theorem that both $G(u)$ and $uG(u)$ are analytic and from Theorem 18.2.2 that they are constants. So $G(u) = 0$. □

As we have seen all entire functions in $\text{DP}(\omega, \omega'; \tau, \tau')$ are constant. However, this is not the case for all members (v and v' not both zero) of the larger class of functions defined by:

Definition 18.2.4 The class $\text{QP}(\omega, \omega'; \tau, \tau'; v, \mu)$ of entire functions for which

$$f(u + \omega) = \tau f(u), \quad f(u + \omega') = \tau' \mathfrak{A}(\omega, \omega'; v, \mu; u) f(u), \quad (18.2.8)$$

with *automorphy factor*

$$\mathfrak{A}(\omega, \omega'; \nu, \mu; u) := \exp\left(-\frac{\pi i(\nu\omega' + 2\mu u)}{\omega}\right), \quad (18.2.9)$$

where ν and μ are integers is called *quasi-periodic*.

Analogues of the theorems for doubly-periodic functions can be obtained with proofs following those for Theorems 18.2.3 and 18.2.4 modified by

$$\frac{f'(u + \omega')}{f(u + \omega')} = \frac{f'(u)}{f(u)} - \frac{2\mu\pi i}{\omega}. \quad (18.2.10)$$

Theorem 18.2.5 *If $f(u) \in \text{QP}(\omega, \omega'; \tau, \tau'; \nu, \mu)$ has exactly m zeros α_j , $j = 1, 2, \dots, m$ inside $\mathbb{W}(\omega_1, \omega_2)$ with none on the boundary, then*

$$m = \mu, \quad (18.2.11)$$

and

$$\sum_{j=1}^m \alpha_j = \frac{1}{4}\omega(4\ell + 1 - \tau' - 2\mu) + \frac{1}{4}\omega'(4\ell' + 1 - \tau - 2\mu), \quad (18.2.12)$$

for some integers ℓ and ℓ' .

Theorem 18.2.6 *If $f(u)$ and $g(u) \in \text{QP}(\omega, \omega'; \tau, \tau'; \nu, \mu)$ have exactly the same zeros then $f(u) = cg(u)$ for some constant c .*

It is clear that, if $f(u) \in \text{QP}(\omega, \omega'; \tau, \tau'; \nu, \mu)$ and $g(u) \in \text{QP}(\omega, \omega'; \bar{\tau}, \bar{\tau}'; \bar{\nu}, \bar{\mu})$, then $f(u)g(u) \in \text{QP}(\omega, \omega'; \tau\bar{\tau}, \tau'\bar{\tau}'; \nu + \bar{\nu}, \mu + \bar{\mu})$. Equations (18.2.11) and (18.2.12) for the two functions sum to give that for their product.⁸

18.3 Elliptic Integrals and Functions

For comprehensive presentations of the material in this section the reader is referred to Hancock (1958), Lawden (1989) and Akhiezer (1990).

18.3.1 Elliptic Integrals

Elliptic integrals of the first and second kind, are

⁸ $\frac{1}{2}(1 - \tau\bar{\tau})$ and $\frac{1}{2}(2 - \tau - \bar{\tau})$ have the same parity.

$$F(\theta|k) := \int_0^\theta \frac{d\omega}{\sqrt{1 - k^2 \sin^2(\omega)}}, \quad E(\theta|k) := \int_0^\theta \sqrt{1 - k^2 \sin^2(\omega)} d\omega, \quad (18.3.1)$$

respectively, where k is the *elliptic modulus* which for the moment we suppose to be real and in the range $(0, 1)$. With θ replaced by $i\theta$, formulae (18.3.1) become

$$F(i\theta|k) := i \int_0^\theta \frac{d\omega}{\sqrt{1 + k^2 \sinh^2(\omega)}}, \quad E(i\theta|k) := i \int_0^\theta \sqrt{1 + k^2 \sinh^2(\omega)} d\omega. \quad (18.3.2)$$

We note for use in Chap. 12 that, if

$$G(\theta|k) := \int_0^\theta \ln \left[1 + \sqrt{1 - k^2 \sin^2(\omega)} \right] d\omega, \quad (18.3.3)$$

then

$$k \frac{dG(\theta|k)}{dk} = \theta - F(\theta|k). \quad (18.3.4)$$

The *complete elliptic integrals of the first and second kind*, are

$$\begin{aligned} \mathcal{K}(k) &:= F\left(\frac{1}{2}\pi|k\right) = \int_0^{\frac{\pi}{2}} \frac{d\omega}{\sqrt{1 - k^2 \sin^2(\omega)}}, \\ \mathcal{E}(k) &:= E\left(\frac{1}{2}\pi|k\right) = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2(\omega)} d\omega, \end{aligned} \quad (18.3.5)$$

with $\mathcal{K}(0) = \mathcal{E}(0) = \frac{1}{2}\pi$, $\mathcal{E}(1) = 1$. $F(\theta|k)$ diverges as $\theta \rightarrow \infty$ along the real axis. However, from (18.3.2),

$$\mathcal{K}'(k) := -i F(i\infty|k) = \int_0^\infty \frac{d\omega}{\sqrt{1 + k^2 \sinh^2(\omega)}} \quad (18.3.6)$$

exists and, using the substitution $\sinh(\omega) = \tan(\gamma)$, it is not difficult to show that

$$\mathcal{K}'(k) = \mathcal{K}(k'), \quad \text{where} \quad k' := \sqrt{1 - k^2} \quad (18.3.7)$$

is the *complementary modulus*, which like k is either non-negative or purely imaginary with a non-negative imaginary part. From (18.3.1) and (18.3.5), by differentiating under the integral sign,

$$\frac{d\mathcal{K}(k)}{dk} = \frac{\mathcal{E}(k) - k'^2 \mathcal{K}(k)}{kk'^2}, \quad \frac{d\mathcal{E}(k)}{dk} = \frac{\mathcal{E}(k) - \mathcal{K}(k)}{k}. \quad (18.3.8)$$

From (18.3.7) and (18.3.8) it can be verified that $\mathcal{K}(k)$ and $\mathcal{K}'(k)$ are independent solutions of the differential equation

$$\frac{d}{dk} \left(k k'^2 \frac{df}{dk} \right) = k f. \quad (18.3.9)$$

Although the functions $\mathcal{K}(k)$ and $\mathcal{K}'(k)$ have been defined only for $k \in (0, 1)$ the solutions of (18.3.9) provide analytic continuations of the functions into the region $|k| < 1$. The further extension into the rest of the complex k plane is achieved by using the integral formula (18.3.5). It can be shown that $\mathcal{K}(k)$ is analytic apart from logarithmic singularities at $k = \pm 1$ and consequently, from (18.3.7), that $\mathcal{K}'(k)$ is analytic apart from a logarithmic singularity at $k = 0$. These singularities are encapsulated by the formula

$$\mathcal{K}'(k) = \frac{2}{\pi} \mathcal{K}(k) \ln \left(\frac{4}{k} \right) + w(k^2), \quad (18.3.10)$$

where $w(z)$ has a power series expansion in z with radius of convergence one. $\mathcal{K}(k)$ is a single-valued function in the complex k plane with branch-cuts on the real axis in the intervals $(-\infty, -1]$ and $[1, \infty)$. It is real-valued on the interval $(-1, 1)$ of the real axis and on the whole of the imaginary axis. The *nome* of k is given by

$$\mathcal{Q}(k) := \exp[-\pi \mathcal{K}(k')/\mathcal{K}(k)] \quad (18.3.11)$$

and, from (18.3.7) and (18.3.8),

$$\frac{d\mathcal{Q}(k)}{dk} = \frac{\mathcal{Q}(k) \pi^2}{2\mathcal{K}^2(k) k k'^2}. \quad (18.3.12)$$

Henceforth, whenever it is convenient, the more compact notation $\mathcal{K} := \mathcal{K}(k)$, $\mathcal{K}' := \mathcal{K}(k')$ and $\mathcal{Q} := \mathcal{Q}(k)$ is used.

The complete elliptic integral $\mathcal{K}$ and the moduli k and k' can be expressed in the forms

$$\begin{aligned} \mathcal{K} &= \frac{\pi}{2} \prod_{n=1}^{\infty} (1 - \mathcal{Q}^{2n})^2 (1 + \mathcal{Q}^{2n-1})^4, \\ k &= 4\mathcal{Q}^{\frac{1}{2}} \prod_{n=1}^{\infty} \left(\frac{1 + \mathcal{Q}^{2n}}{1 + \mathcal{Q}^{2n-1}} \right)^4, \quad k' = \prod_{n=1}^{\infty} \left(\frac{1 - \mathcal{Q}^{2n-1}}{1 + \mathcal{Q}^{2n-1}} \right)^4. \end{aligned} \quad (18.3.13)$$

18.3.2 Jacobi Theta Functions

One way to define the *Jacobi theta functions* $H(u|k)$ and $\Theta(u|k)$ is⁹

$$H(u|k) := 2Q^{\frac{1}{4}} \sin\left(\frac{\pi u}{2\mathcal{K}}\right) \prod_{n=1}^{\infty} \left(1 - Q^{2n}\right) \left\{1 - 2Q^{2n} \cos\left(\frac{\pi u}{\mathcal{K}}\right) + Q^{4n}\right\}, \quad (18.3.14)$$

$$\Theta(u|k) := \prod_{n=1}^{\infty} \left(1 - Q^{2n}\right) \left\{1 - 2Q^{2n-1} \cos\left(\frac{\pi u}{\mathcal{K}}\right) + Q^{4n-2}\right\}, \quad (18.3.15)$$

from which it is clear that they are entire with simple zeros, that $H(u|k)$ and $\Theta(u|k)$ are, respectively, odd and even functions of u and, using (18.3.13), that

$$H(0|k) = 0, \quad \Theta(\mathcal{K}|k) = \sqrt{\frac{2\mathcal{K}}{\pi}}, \quad \frac{H(\mathcal{K}|k)}{\Theta(\mathcal{K}|k)} = \sqrt{k}, \quad \frac{\Theta(0|k)}{\Theta(\mathcal{K}|k)} = \sqrt{k'}. \quad (18.3.16)$$

It is also useful to know that

$$H'(0|k) = \frac{\pi}{2\mathcal{K}} H(\mathcal{K}|k) \Theta(\mathcal{K}|k) \Theta(0|k), \quad (18.3.17)$$

$$H(2u|k) = \frac{2H(u|k) \Theta(u|k) H(u + \mathcal{K}|k) \Theta(u + \mathcal{K}|k)}{H(\mathcal{K}|k) \Theta(\mathcal{K}|k) \Theta(0|k)}. \quad (18.3.18)$$

It can be shown that

$$H(u + \mathfrak{U}(2\ell, \ell'|k)|k) = (-1)^{\ell} (-i)^{\ell'} \mathfrak{F}_{\ell'}(u + 2\ell\mathcal{K}|k) \times \begin{cases} H(u|k), & \ell' \text{ even,} \\ \Theta(u|k), & \ell' \text{ odd,} \end{cases} \quad (18.3.19)$$

$$\Theta(u + \mathfrak{U}(2\ell, \ell'|k)|k) = (-i)^{\ell'} \mathfrak{F}_{\ell'}(u + 2\ell\mathcal{K}|k) \times \begin{cases} \Theta(u|k), & \ell' \text{ even,} \\ H(u|k), & \ell' \text{ odd,} \end{cases} \quad (18.3.20)$$

⁹ The other two theta functions, which are avoided in this text, are $H_1(u|k) := H(u + \mathcal{K}|k)$ and $\Theta_1(u|k) := \Theta(u + \mathcal{K}|k)$. In the interests of clarity it should be observed that the notation $\vartheta_i(u)$, $i = 1, \dots, 4$, used in many texts for the theta functions, is related to this notation by $H(u|k) := \vartheta_1(\frac{\pi u}{2\mathcal{K}}|Q)$, $H_1(u|k) := \vartheta_2(\frac{\pi u}{2\mathcal{K}}|Q)$, $\Theta_1(u|k) := \vartheta_3(\frac{\pi u}{2\mathcal{K}}|Q)$, $\Theta(u|k) := \vartheta_4(\frac{\pi u}{2\mathcal{K}}|Q)$.

where

$$\mathfrak{U}(\ell, \ell'|k) := \ell \mathcal{K} + \ell' i \mathcal{K}', \quad (18.3.21)$$

$$\mathfrak{F}_\ell(u|k) := \exp\left(\frac{-i\pi\ell(\ell i \mathcal{K}' + 2u)}{4\mathcal{K}}\right) = \mathcal{Q}^{-\frac{1}{4}\ell^2} \exp\left(\frac{-\pi\ell i u}{2\mathcal{K}}\right), \quad (18.3.22)$$

with the special cases

$$H(u \pm i\mathcal{K}'|k) = \mp i \mathfrak{F}_1(\pm u|k) \Theta(u|k), \quad \Theta(u \pm i\mathcal{K}'|k) = \mp i \mathfrak{F}_1(\pm u|k) H(u|k), \quad (18.3.23)$$

$$H(u \pm 2\mathcal{K}|k) = -H(u|k), \quad \Theta(u \pm 2\mathcal{K}|k) = \Theta(u|k), \quad (18.3.24)$$

$$H(u \pm 2i\mathcal{K}'|k) = -\mathfrak{F}_2(\pm u|k) H(u|k), \quad \Theta(u \pm 2i\mathcal{K}'|k) = -\mathfrak{F}_2(\pm u|k) \Theta(u|k). \quad (18.3.25)$$

We also have

$$\begin{aligned} H(iu|k) &= i \mathfrak{X}(u|k) H(u|k'), & H(iu + \mathcal{K}|k) &= \mathfrak{X}(u|k) \Theta(u|k'), \\ \Theta(iu|k) &= \mathfrak{X}(u|k) H(u + \mathcal{K}'|k'), & \Theta(iu + \mathcal{K}|k) &= \mathfrak{X}(u|k) \Theta(u + \mathcal{K}'|k'), \end{aligned} \quad (18.3.26)$$

where

$$\mathfrak{X}(u|k) := \sqrt{\frac{\mathcal{K}}{\mathcal{K}'}} \exp\left(\frac{\pi u^2}{4\mathcal{K}\mathcal{K}'}\right), \quad (18.3.27)$$

in which the constant is evaluated using (18.3.16).

According to the notation of Sect. 18.2,

$$\begin{aligned} H(u|k) &\in \text{QP}(2\mathcal{K}, 2i\mathcal{K}'; -1, -1; 1, 1), & \Theta(u|k) &\in \text{QP}(2\mathcal{K}, 2i\mathcal{K}'; 1, -1; 1, 1), \\ H(u + \mathcal{K}|k) &\in \text{QP}(2\mathcal{K}, 2i\mathcal{K}'; -1, 1; 1, 1), & \Theta(u + \mathcal{K}|k) &\in \text{QP}(2\mathcal{K}, 2i\mathcal{K}'; 1, 1; 1, 1), \end{aligned} \quad (18.3.28)$$

in which, from (18.3.14) and (18.3.15), for $k \in [0, 1]$:

- (i) On the real- u axis and on $\Re\{u\} = \mathcal{K}$ both $H(u|k)$ and $\Theta(u|k)$ are real-valued.
- (ii) On the imaginary- u axis $H(u|k)$ is purely imaginary and $\Theta(u|k)$ is real-valued.

Theorem 18.2.5 predicts a single zero for each elliptic function in the period parallelogram $\mathbb{W}(2\mathcal{K}, 2i\mathcal{K}')$ (which now becomes a rectangle)¹⁰ and

$$\begin{aligned} H(u|k) & \text{ has zeros at } u = \mathcal{U}(2\ell, 2\ell'|k), \\ \Theta(u|k) & \text{ has zeros at } u = \mathcal{U}(2\ell, 2\ell' + 1|k), \end{aligned} \quad \text{for all integers } \ell \text{ and } \ell'. \quad (18.3.29)$$

This then leads to the important theorem:

Theorem 18.3.1 *For some complex quantities $\alpha_j, \beta_j, j = 1, 2, \dots, r$, lying in the period rectangle $\mathbb{W}(2\mathcal{K}, 2i\mathcal{K}')$,*

$$W(u|k) := \exp\left(-\frac{\pi i \aleph u}{2\mathcal{K}}\right) \prod_{j=1}^r \frac{H(u - \alpha_j|k)}{H(u - \beta_j|k)} \quad (18.3.30)$$

is, to within a multiplicative quantity possibly depending on k , the unique member of $\text{DP}(2\mathcal{K}, 2i\mathcal{K}'; \tau, \tau')$ with zeros $\alpha_j, j = 1, 2, \dots, r$ and poles $\beta_j, j = 1, 2, \dots, r$ if

$$\aleph = \frac{1}{2}(4\ell' + 1 - \tau), \quad \sum_{j=1}^r (\alpha_j - \beta_j) = i\aleph\mathcal{K}' + \frac{1}{2}\mathcal{K}(4\ell + 1 - \tau'), \quad (18.3.31)$$

for some integers ℓ and ℓ' .

Proof From (18.3.24) and (18.3.25),

$$\begin{aligned} W(u + 2\mathcal{K}|k) &= \exp(-\pi i \aleph) W(u|k), \\ W(u + 2i\mathcal{K}'|k) &= \exp\left(\frac{\pi[\aleph\mathcal{K}' + i\sum_{j=1}^m (\alpha_j - \beta_j)]}{\mathcal{K}}\right) W(u|k). \end{aligned} \quad (18.3.32)$$

Using (18.2.5), the conditions that $W(u|k) \in \text{DP}(2\mathcal{K}, 2i\mathcal{K}'; \tau, \tau')$ are (18.3.31). That it is, to within a multiplicative quantity possibly dependent on k , the unique member of $\text{DP}(2\mathcal{K}, 2i\mathcal{K}'; \tau, \tau')$ with these zeros and poles follows from Theorem 18.2.4. $\square$

The uniqueness Theorems 18.2.4 and 18.2.6 can be used to establish identities between theta functions. In particular

$$\begin{aligned} H(v + u|k)H(t + u|k)\Theta(t + v|k)\Theta(0|k) \\ = H(v|k)H(t|k)\Theta(u|k)\Theta(v + t + u|k) \\ + \Theta(v|k)\Theta(t|k)H(u|k)H(v + t + u|k), \end{aligned} \quad (18.3.33)$$

¹⁰ In the case of $\Theta(u|k)$ it must be displaced slightly so that the zero at $u = i\mathcal{K}'$ does not lie on the boundary.

$$\begin{aligned}
& \Theta(u+v|k) \Theta(t+u|k) \Theta(t+v|k) \Theta(0|k) \\
&= \Theta(v|k) \Theta(t|k) \Theta(u|k) \Theta(v+t+u|k) \\
&+ H(v|k)H(t|k)H(u|k)H(v+t+u|k),
\end{aligned} \tag{18.3.34}$$

$$\begin{aligned}
& H(\mathcal{K}+u|k)H(\mathcal{K}-v|k) + \Theta(\mathcal{K}+u|k) \Theta(\mathcal{K}-v|k) \\
&= 2\Omega^{\frac{1}{4}}[H(\mathcal{K}|k) \Theta(\mathcal{K}|k)]^{-1} \\
&\times H(\tfrac{1}{2}[i\mathcal{K}'+u+v] + \mathcal{K}|k)H(\tfrac{1}{2}[i\mathcal{K}'-u-v] + \mathcal{K}|k) \\
&\times H(\tfrac{1}{2}[i\mathcal{K}'+u-v] + \mathcal{K}|k)H(\tfrac{1}{2}[i\mathcal{K}'-u+v] + \mathcal{K}|k).
\end{aligned} \tag{18.3.35}$$

18.3.3 Jacobi Elliptic Functions

The *Jacobi amplitude* $\text{am}(u|k)$, defined by

$$F(\text{am}(u|k)|k) = u, \tag{18.3.36}$$

is the inverse of the elliptic integral of the first kind. It is clear that it is an odd function of u and, from (18.3.1)–(18.3.7), it can be shown that

$$\begin{aligned}
& \text{am}(u|0) = u, & \text{am}(u|1) &= \arcsin(\tanh(u)), \\
& \text{am}(i u|k) = \arcsin(i \tan(\text{am}(u|k'))), \\
& \lim_{\epsilon \rightarrow \pm 0} \text{am}(\epsilon + 2\mathcal{K}|k) = \mp\pi, & \lim_{\epsilon \rightarrow \pm 0} \text{am}(\epsilon + 2i\mathcal{K}'|k) &= \pm\pi.
\end{aligned} \tag{18.3.37}$$

One way to define the *Jacobi elliptic functions* $\text{sn}(u|k)$, $\text{cn}(u|k)$ and $\text{dn}(u|k)$ is

$$\begin{aligned}
& \text{sn}(u|k) := \sin(\text{am}(u|k)), \\
& \text{cn}(u|k) := \cos(\text{am}(u|k)) = \sqrt{1 - \text{sn}^2(u|k)}, \\
& \text{dn}(u|k) := \frac{d \text{am}(u|k)}{du} = \sqrt{1 - k^2 \text{sn}^2(u|k)},
\end{aligned} \tag{18.3.38}$$

giving the useful relations

$$\text{am}(u|k) = -i \ln[\text{cn}(u|k) + i \text{sn}(u|k)], \quad \text{am}(u|k) = \int_0^u \text{dn}(v|k) dv \tag{18.3.39}$$

and the derivatives

$$\begin{aligned}
& \text{sn}'(u|k) = \text{cn}(u|k)\text{dn}(u|k), & \text{cn}'(u|k) &= -\text{sn}(u|k)\text{dn}(u|k), \\
& \text{dn}'(u|k) = -k^2 \text{sn}(u|k)\text{cn}(u|k).
\end{aligned} \tag{18.3.40}$$

From (18.3.37),

$$\operatorname{sn}(u|0) = \sin(u), \quad \operatorname{cn}(u|0) = \cos(u), \quad \operatorname{dn}(u|0) = 1. \quad (18.3.41)$$

We shall henceforth, and in Chap. 12, use, whenever convenient, a useful notation which was invented by Glaisher to express reciprocals and quotients of elliptic functions (Whittaker and Watson 1927, Sect. 22.122). For reciprocals the letters of the function are reversed and for quotients the first letters of the numerator and denominator are used to denote the quantity. Thus, for example,

$$\operatorname{ns}(u|k) := \frac{1}{\operatorname{sn}(u|k)}, \quad \operatorname{sc}(u|k) := \frac{\operatorname{sn}(u|k)}{\operatorname{cn}(u|k)}. \quad (18.3.42)$$

In the spirit of Glaisher we also find it useful to define

$$\operatorname{cds}(u|k) := \frac{\operatorname{cn}(u|k)\operatorname{dn}(u|k)}{\operatorname{sn}(u|k)}. \quad (18.3.43)$$

An alternative way to defined the elliptic functions is¹¹

$$\begin{aligned} \operatorname{sn}(u|k) &:= \frac{H(u|k)}{\sqrt{k} \Theta(u|k)}, \quad \operatorname{cn}(u|k) := \sqrt{\frac{k'}{k}} \frac{H(u + \mathcal{K}|k)}{\Theta(u|k)}, \\ \operatorname{dn}(u|k) &:= \frac{\sqrt{k'} \Theta(u + \mathcal{K}|k)}{\Theta(u|k)}, \end{aligned} \quad (18.3.44)$$

thus making it clear:

- (i) from (18.3.28), that $\operatorname{sn}(u|k) \in \operatorname{DP}(2\mathcal{K}, 2i\mathcal{K}'; -1, 1)$,
 $\operatorname{cn}(u|k) \in \operatorname{DP}(2\mathcal{K}, 2i\mathcal{K}'; -1, -1)$ and $\operatorname{dn}(u|k) \in \operatorname{DP}(2\mathcal{K}, 2i\mathcal{K}'; 1, -1)$;
- (ii) that, for real $u \in (0, \mathcal{K})$, $\operatorname{sn}(u|k)$, $\operatorname{cn}(u|k)$ and $\operatorname{dn}(u|k)$ are real-valued and positive, with $\operatorname{sn}(u|k)$ monotonically increasing and $\operatorname{cn}(u|k)$ and $\operatorname{dn}(u|k)$ monotonically decreasing;
- (iii) that, for real $u \in (0, \mathcal{K}')$, $-i \operatorname{sn}(iu|k)$, $\operatorname{cn}(iu|k)$ and $\operatorname{dn}(iu|k)$ are real-valued, positive and monotonically increasing.

From (18.3.19)–(18.3.22),

$$\operatorname{sn}(u \pm \mathcal{U}(2\ell, \ell'|k)|k) = (-1)^\ell \times \begin{cases} \operatorname{sn}(u|k), & \ell' \text{ even,} \\ k^{-1} \operatorname{ns}(u|k), & \ell' \text{ odd,} \end{cases} \quad (18.3.45)$$

$$\operatorname{cn}(u \pm \mathcal{U}(2\ell, \ell'|k)|k) = (-1)^\ell (i)^{-\ell'} \times \begin{cases} \operatorname{cn}(u|k), & \ell' \text{ even,} \\ k^{-1} \operatorname{ds}(u|k), & \ell' \text{ odd,} \end{cases} \quad (18.3.46)$$

¹¹ It should be noted, for use in Chap. 12, that the expression for $\operatorname{sn}(u|k)$, but not $\operatorname{cn}(u|k)$ and $\operatorname{dn}(u|k)$, takes the same form in terms of Baxter's modified theta functions.

$$\operatorname{dn}(u \pm \mathfrak{U}(2\ell, \ell'|k)|k) = (i)^{-\ell'} \times \begin{cases} \operatorname{dn}(u|k), & \ell' \text{ even}, \\ \operatorname{cs}(u|k), & \ell' \text{ odd}, \end{cases} \quad (18.3.47)$$

with special cases

$$\operatorname{sn}(u - \mathfrak{U}(\ell, \ell'|k)|k) = (-1)^\ell \operatorname{sn}(u + \mathfrak{U}(\ell, \ell'|k)|k), \quad (18.3.48)$$

$$\operatorname{sn}(u \pm 2\mathfrak{U}_0|k) = \operatorname{sn}(u|k), \quad (18.3.49)$$

$$\operatorname{sn}\left(u + \frac{1}{2}i\mathcal{K}'|k\right) \operatorname{sn}\left(u - \frac{1}{2}i\mathcal{K}'|k\right) = 1/k, \quad (18.3.50)$$

$$\operatorname{cds}\left(u + \frac{1}{2}i\mathcal{K}'|k\right) = -\operatorname{cds}\left(u - \frac{1}{2}i\mathcal{K}'|k\right). \quad (18.3.51)$$

Integrating the last of Eqs. (18.3.47) and setting the constant using (18.3.37) gives

$$\operatorname{am}(u + 2\mathcal{K}|k) = -\operatorname{am}(u + 2i\mathcal{K}'|k) = \operatorname{am}(u|k) \pm \pi, \quad (18.3.52)$$

where the upper a lower signs correspond to the left and right of the discontinuity at $u = 0$. From (18.3.26) and (18.3.44),

$$\operatorname{sn}(iu|k) = i \operatorname{sc}(u|k'), \quad \operatorname{cn}(iu|k) = \operatorname{nc}(u|k'), \quad \operatorname{dn}(iu|k) = \operatorname{dc}(u|k'), \quad (18.3.53)$$

and, from (18.3.16), (18.3.25) and (18.3.44),

$$\begin{aligned} \operatorname{sn}(0|k) &= 0, & \operatorname{cn}(0|k) &= 1, & \operatorname{dn}(0|k) &= 1, \\ \operatorname{sn}\left(\frac{1}{2}i\mathcal{K}'|k\right) &= \frac{i}{\sqrt{k}}, & \operatorname{cn}\left(\frac{1}{2}i\mathcal{K}'|k\right) &= \sqrt{1+k^{-1}}, & \operatorname{dn}\left(\frac{1}{2}i\mathcal{K}'|k\right) &= \sqrt{1+k}, \end{aligned} \quad (18.3.54)$$

and hence, from (18.3.45)–(18.3.47),

$$\operatorname{sn}(i\mathcal{K}'|k) = \operatorname{cn}(i\mathcal{K}'|k) = \operatorname{dn}(i\mathcal{K}'|k) = \infty, \quad (18.3.55)$$

but

$$\operatorname{cs}(i\mathcal{K}'|k) = -i, \quad \operatorname{ds}(i\mathcal{K}'|k) = -ik. \quad (18.3.56)$$

All three elliptic functions are meromorphic with simple poles at the roots of $\Theta(u|k)$, that is when $u = \mathfrak{U}(2\ell, 2\ell' + 1|k)$, for all integer values of ℓ and ℓ' . From (18.3.38), $\operatorname{cn}(\mathfrak{U}(2\ell, 2\ell')|k) = 1$ and it follows from (18.3.39) that the zeros of $\operatorname{am}(u|k)$ coincide with those of $\operatorname{sn}(u|k)$ which are at $u = \mathfrak{U}(2\ell, 2\ell'|k)$.

From (18.3.33), (18.3.34) and (18.3.44),

$$\operatorname{sn}(u \pm v|k) = \frac{\operatorname{sn}(u|k)\operatorname{cn}(v|k)\operatorname{dn}(v|k) \pm \operatorname{sn}(v|k)\operatorname{cn}(u|k)\operatorname{dn}(u|k)}{1 - k^2\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)}, \quad (18.3.57)$$

$$\operatorname{cn}(u \pm v|k) = \frac{\operatorname{cn}(u|k)\operatorname{cn}(v|k) \mp \operatorname{sn}(u|k)\operatorname{sn}(v|k)\operatorname{dn}(u|k)\operatorname{dn}(v|k)}{1 - k^2\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)}, \quad (18.3.58)$$

$$\operatorname{dn}(u \pm v|k) = \frac{\operatorname{dn}(u|k)\operatorname{dn}(v|k) \mp k^2\operatorname{sn}(u|k)\operatorname{sn}(v|k)\operatorname{cn}(u|k)\operatorname{cn}(v|k)}{1 - k^2\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)}, \quad (18.3.59)$$

from which it follows that

$$\operatorname{sn}(u + v|k)\operatorname{sn}(u - v) = \frac{\operatorname{sn}^2(u|k) - \operatorname{sn}^2(v|k)}{1 - k^2\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)}, \quad (18.3.60)$$

$$\operatorname{dn}(u \pm v|k) - \operatorname{dn}(u|k)\operatorname{dn}(v|k) = \mp k^2\operatorname{cn}(u \pm v|k)\operatorname{sn}(u|k)\operatorname{sn}(v|k), \quad (18.3.61)$$

$$\operatorname{dn}(u + v|k)\operatorname{sn}(v|k) = \operatorname{sn}(u + v|k)\operatorname{cn}(u|k)\operatorname{dn}(v|k) - \operatorname{cn}(u + v|k)\operatorname{sn}(u|k), \quad (18.3.62)$$

$$\operatorname{cn}(u|k)\operatorname{dn}(u|k)\operatorname{sn}(2u|k) = \operatorname{sn}(u|k)[\operatorname{dn}(2u|k) + \operatorname{cn}(2u|k)], \quad (18.3.63)$$

$$\begin{aligned} & \frac{\operatorname{cn}(u + v|k) \mp i \operatorname{sn}(u + v|k)}{\operatorname{dn}(u + v|k) \pm i k \operatorname{sn}(u + v|k)} \\ &= \frac{[\operatorname{cn}(u|k) \mp i \operatorname{sn}(u|k)\operatorname{dn}(v|k)][\operatorname{cn}(v|k) \mp i \operatorname{sn}(v|k)\operatorname{dn}(u|k)]}{[\operatorname{dn}(u|k) \pm i k \operatorname{sn}(u|k)\operatorname{cn}(v|k)][\operatorname{dn}(v|k) \pm i k \operatorname{sn}(v|k)\operatorname{cn}(u|k)]}, \end{aligned} \quad (18.3.64)$$

$$\operatorname{cds}^2(u|k) - \operatorname{cds}^2(v|k) = \frac{[\operatorname{sn}^2(v|k) - \operatorname{sn}^2(u|k)][1 - k^2\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)]}{\operatorname{sn}^2(u|k)\operatorname{sn}^2(v|k)}, \quad (18.3.65)$$

$$\operatorname{sn}^2(u|k) = \frac{1 - \operatorname{cn}(2u|k)}{1 + \operatorname{dn}(2u|k)}, \quad \operatorname{cn}^2(u|k) = \frac{\operatorname{cn}(2u|k) + \operatorname{dn}(2u|k)}{1 + \operatorname{dn}(2u|k)}, \quad (18.3.66)$$

$$\operatorname{dn}^2(u|k) = \frac{\operatorname{cn}(2u|k) + \operatorname{dn}(2u|k)}{1 + \operatorname{cn}(2u|k)},$$

$$\begin{aligned} k \operatorname{sn}^2\left(u \pm \frac{1}{2} i \mathcal{K}'|k\right) &= \frac{-\operatorname{cn}(2u|k) \pm i \operatorname{sn}(2u|k)}{\operatorname{dn}(2u|k) \pm i k \operatorname{sn}(2u|k)}, \\ k \operatorname{cn}^2\left(u \pm \frac{1}{2} i \mathcal{K}'|k\right) &= \frac{\operatorname{dn}(2u|k) + k \operatorname{cn}(2u|k)}{\operatorname{cn}(2u|k) \pm i \operatorname{sn}(2u|k)}, \end{aligned} \quad (18.3.67)$$

$$\operatorname{dn}^2\left(u \pm \frac{1}{2} i \mathcal{K}'|k\right) = \frac{\operatorname{dn}(2u|k) + k \operatorname{cn}(2u|k)}{\operatorname{dn}(2u|k) \pm i k \operatorname{sn}(2u|k)},$$

$$\operatorname{cds}\left(u \pm \frac{1}{2} i \mathcal{K}'|k\right) = \mp i [\operatorname{dn}(2u|k) + k \operatorname{cn}(2u|k)]. \quad (18.3.68)$$

From (18.3.1), (18.3.37) and (18.3.38)

$$\int_0^{2\mathcal{K}} du \, \operatorname{dn}^2(u|\mathbf{k}) = 2\mathcal{E}. \quad (18.3.69)$$

Also

$$\int_0^{2\mathcal{K}} du \int_0^{2\mathcal{K}} dv \, \operatorname{dn}(u|\mathbf{k}) \operatorname{dn}(u-v|\mathbf{k}) \operatorname{dn}(v|\mathbf{k}) = 2\mathbf{k}'^2 \mathcal{K}^2 + \frac{1}{2}\pi^2, \quad (18.3.70)$$

$$\int_0^{2\mathcal{K}} du \, \operatorname{dn}(u|\mathbf{k}) \operatorname{dn}\left(u + \frac{1}{2}i\mathcal{K}'|\mathbf{k}\right) = \frac{1}{2}\sqrt{1+\mathbf{k}}[\pi + 2(1-\mathbf{k})\mathcal{K}], \quad (18.3.71)$$

which are a little more difficult to prove and need the use of (18.3.59).

Suppose that θ , α and $\mathbf{k}$ are related by

$$\sinh(\theta) = \frac{i}{\mathbf{k} \operatorname{sn}(i\alpha|\mathbf{k})}. \quad (18.3.72)$$

Since, from (18.3.38), $\theta = -i \operatorname{am}(i[\mathcal{K}' - \alpha]|\mathbf{k})$, it follows, from (18.3.2), (18.3.6), (18.3.7) and (18.3.36), that, if $0 < \mathbf{k} < 1$ and θ is real,

$$\alpha = \int_{\theta}^{\infty} \frac{d\omega}{\sqrt{1 + \mathbf{k}^2 \sinh^2(\omega)}} \quad (18.3.73)$$

is a real, but not unique, solution for α .

18.3.4 Transformations in the Elliptic Modulus

It is often useful to transform the modulus of elliptic functions using combinations of the *complementary and Landen transformations*, $\mathcal{C}$ and $\mathcal{L}$ respectively, given by

$$\mathbf{k} \rightarrow \mathcal{C}(\mathbf{k}) := \sqrt{1 - \mathbf{k}^2}, \quad \mathbf{k} \rightarrow \mathcal{L}(\mathbf{k}) := \frac{1 - \mathbf{k}}{1 + \mathbf{k}}, \quad \text{with } \mathcal{C}\mathcal{L}(\mathbf{k}) = \frac{2\sqrt{\mathbf{k}}}{1 + \mathbf{k}}. \quad (18.3.74)$$

It is clear that these transformations are self-inverse, but that they do not commute.

The complementary modulus $\mathbf{k}' := \mathcal{C}(\mathbf{k})$ has already been encountered and in Chap. 12 we shall make use of four elliptic moduli $\mathbf{k}_0, \dots, \mathbf{k}_3$ related by Landen

transformations. With the compact notation $\mathcal{K}_j := \mathcal{K}(k_j)$, $\mathcal{K}'_j := \mathcal{K}(k'_j)$, $\mathcal{Q}_j := \mathcal{Q}(k_j)$, $\mathcal{E}_j := \mathcal{E}(k_j)$, $j = 0, \dots, 3$:

$$(i) \quad k_1 := \frac{1 - k'_0}{1 + k'_0} = \mathcal{L} \mathcal{C}(k_0), \quad (18.3.75)$$

for which

$$\begin{aligned} (1 + k'_0)(1 + k_1) &= 2, \\ \mathcal{K}_1 &= \frac{1}{2}(1 + k'_0)\mathcal{K}_0, \quad \mathcal{K}'_1 = (1 + k'_0)\mathcal{K}'_0, \\ \mathcal{K}'_1/\mathcal{K}'_0 &= 2\mathcal{K}_1/\mathcal{K}_0, \quad \mathcal{Q}_1 = \mathcal{Q}_0^2. \end{aligned} \quad (18.3.76)$$

For k_0 and k_1 we shall also need the relationship

$$\mathcal{E}_1 = \frac{\mathcal{E}_0 + k'_0\mathcal{K}_0}{1 + k'_0} \quad (18.3.77)$$

between complete elliptic integrals of the second kind.

$$(ii) \quad k_2 := \mathcal{C} \mathcal{L}(k_0) = \frac{2\sqrt{k_0}}{1 + k_0}, \quad (18.3.78)$$

for which

$$\begin{aligned} (1 + k_0)(1 + k'_2) &= 2, \\ \mathcal{K}_2 &= (1 + k_0)\mathcal{K}_0, \quad \mathcal{K}'_2 = \frac{1}{2}(1 + k_0)\mathcal{K}'_0, \\ \mathcal{K}'_2/\mathcal{K}'_0 &= \mathcal{K}_2/2\mathcal{K}_0, \quad \mathcal{Q}_2 = \mathcal{Q}_0^{\frac{1}{2}}. \end{aligned} \quad (18.3.79)$$

$$(iii) \quad k_3 := \mathcal{L} \mathcal{C} \mathcal{L}(k_1) = \frac{1 + k_1 - 2\sqrt{k_1}}{1 + k_1 + 2\sqrt{k_1}}. \quad (18.3.80)$$

This is the *Bazhanov–Mangazeev* transformation (Bazhanov and Mangazeev 2007) for which

$$(1 - k_1)(1 - k_3) = 4\sqrt{k_1 k_3}, \quad \mathcal{K}'_3/\mathcal{K}_3 = 4\mathcal{K}_1/\mathcal{K}'_1. \quad (18.3.81)$$

For any u_0 let

$$u_1 := \frac{u_0}{1 + k_1} = \frac{1}{2}u_0(1 + k'_0) \quad \text{so that} \quad \frac{u_0}{\mathcal{K}_0} = \frac{u_1}{\mathcal{K}_1}. \quad (18.3.82)$$

Then elliptic functions are related by

$$\begin{aligned}\operatorname{sn}(u_0|k_0) &= \frac{(1+k_1)\operatorname{sn}(u_1|k_1)}{1+k_1\operatorname{sn}^2(u_1|k_1)}, & \operatorname{cn}(u_0|k_0) &= \frac{\operatorname{cn}(u_1|k_1)\operatorname{dn}(u_1|k_1)}{1+k_1\operatorname{sn}^2(u_1|k_1)}, \\ \operatorname{dn}(u_0|k_0) &= \frac{1-k_1\operatorname{sn}^2(u_1|k_1)}{1+k_1\operatorname{sn}^2(u_1|k_1)}.\end{aligned}\tag{18.3.83}$$

18.3.5 The Modified Amplitude Function

$$\operatorname{Am}(u|k) := -i \ln \left[i\sqrt{k} \operatorname{sn}\left(u - \frac{1}{2}i\mathcal{K}'|k\right) \right], \tag{18.3.84}$$

is the *modified amplitude function* (Baxter 1982b, p. 462). It is clear that it is not meromorphic since it has logarithmic branch-points at the zeros and poles of $\operatorname{sn}(u - \frac{1}{2}i\mathcal{K}')$, that is, when $u = 2(2\ell + 1)\mathcal{K} + (2\ell' \pm \frac{1}{2})i\mathcal{K}'$ for all integers ℓ and ℓ' . From (18.3.45),

$$\begin{aligned}\operatorname{Am}(u + 2\mathcal{K}|k) &= \operatorname{Am}(u|k) \pm \pi, \\ \operatorname{Am}(u + i\mathcal{K}'|k) &= -\operatorname{Am}(u|k) \pm \pi, \quad \operatorname{Am}(u + 2i\mathcal{K}'_0|k) = \operatorname{Am}(u|k),\end{aligned}\tag{18.3.85}$$

where the choice of signs, reflects the branch-point structure.

For real values of u , $\operatorname{Am}(u|k)$ is real and the choice of signs can be made so that its values lie in the range $[-\pi, \pi]$. Then it monotonically increases from $-\pi$ to π as u increases from $2(2\ell - 1)\mathcal{K}$ to $2(2\ell + 1)\mathcal{K}$, with zeros at $4\ell\mathcal{K}$.

For the imaginary axis consider the function $-i\operatorname{Am}(iu|k_0)$. For real values of u this is real and monotonically increasing from $-\infty$ to ∞ as u increases from $(2\ell' - \frac{1}{2})\mathcal{K}'$ to $(2\ell' + \frac{1}{2})\mathcal{K}'$, with zeros at $2\ell'\mathcal{K}'$. Otherwise on the real axis it is complex. From the elliptic moduli k_0 and k_1 related by the Landen transformation given in Sect. 18.3.4

$$\operatorname{cn}(u|k_0) + i\operatorname{sn}(u|k_0) = i\sqrt{k_1} \operatorname{sn}\left(u_1 - \frac{1}{2}i\mathcal{K}'_1|k_1\right) \tag{18.3.86}$$

and, from (18.3.39),

$$\operatorname{am}(u|k_0) = \operatorname{Am}(u_1|k_1), \tag{18.3.87}$$

$$i \operatorname{am}(i[\mathcal{K}'_0 - u]|k_0) = -i \operatorname{Am}(i[u_1 - \frac{1}{2}i\mathcal{K}'_1]|k_1) = \ln[-i\sqrt{k_1} \operatorname{sn}(iu_1|k_1)]. \tag{18.3.88}$$

The latter function is real and monotonically increasing from $-\infty$ to ∞ as u increases from zero to $2\mathcal{K}'_0$ with a zero at $u = \mathcal{K}'_0$. It also follows from (18.3.53), (18.3.83) and (18.3.88) that

$$\exp(i \operatorname{am}(i[\mathcal{K}'_0 - u]|k_0)) = \begin{cases} -i[1 - \operatorname{dn}(i u|k_0)]/[k_0 \operatorname{sn}(i u|k_0)], \\ [\operatorname{dn}(u|k'_0) - \operatorname{cn}(u|k'_0)]/[k_0 \operatorname{sn}(u|k'_0)], \\ -i\sqrt{k_1} \operatorname{sn}(i u_1|k_1), \\ \sqrt{k_1} \operatorname{sn}(u_1|k'_1). \end{cases} \quad (18.3.89)$$

18.3.6 Nome Series

The product formulae for the theta functions can be used to obtain *nome series* for the elliptic functions. In particular

$$\operatorname{dn}(u|k) = \frac{\pi}{2\mathcal{K}} + \frac{2\pi}{\mathcal{K}} \sum_{n=1}^{\infty} \frac{Q^n}{1 + Q^{2n}} \cos \left[\frac{n\pi u}{\mathcal{K}} \right]. \quad (18.3.90)$$

Substituting $u = 0$ into (18.3.90) and using (18.3.54) gives

$$\sum_{n=1}^{\infty} \frac{Q^n}{1 + Q^{2n}} = \frac{\mathcal{K}}{2\pi} - \frac{1}{4}. \quad (18.3.91)$$

By substituting from (18.3.90) into (18.3.69) and noting that the cosines in (18.3.90) are orthogonal over the range of this integral, it follows that

$$\sum_{n=1}^{\infty} \frac{Q^{2n}}{(1 + Q^{2n})^2} = \frac{\mathcal{E} \mathcal{K}}{2\pi^2} - \frac{1}{8} \quad (18.3.92)$$

and, by replacing k by k_1 in (18.3.71), substituting from (18.3.90) and performing the integration, it can be shown that

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{Q_0^{3n} (1 + Q_0^{2n})}{(1 + Q_0^{4n})^2} &= \frac{1}{4\pi^2} \mathcal{K}_1 \sqrt{1 + k_1} [\pi + 2(1 - k_1)\mathcal{K}_1] - \frac{1}{4} \\ &= \frac{1}{4\pi^2} \mathcal{K}_0 (\pi + 2k'_0 \mathcal{K}_0) \sqrt{\frac{1 + k'_0}{2}} - \frac{1}{4}. \end{aligned} \quad (18.3.93)$$

18.3.7 Special Results and Functions for Chap. 12

- The functions

$$\begin{aligned}
 P_H(v, u|k) &:= H(i[u + v]|k)H(i[u - v]|k), \\
 P_\Theta(v, u|k) &:= \Theta(i[u + v]|k) \Theta(i[u - v]|k), \\
 Q(v, u|k) &:= H(i[u + v]|k) \Theta(i[u - v]|k),
 \end{aligned} \tag{18.3.94}$$

have the properties

$$\begin{aligned}
 P_H(v + \mathcal{K}', u|k) &= \mathfrak{F}_1(2iv|k)P_\Theta(v, u|k), \\
 P_\Theta(v + \mathcal{K}', u|k) &= \mathfrak{F}_1(2iv|k)P_H(v, u|k), \\
 Q(v + \mathcal{K}', u|k) &= \mathfrak{F}_1(2iv|k)Q(v, u|k),
 \end{aligned} \tag{18.3.95}$$

$$P_H(v - \tfrac{1}{2}\mathcal{K}', u + \tfrac{1}{2}\mathcal{K}'|k) = i \mathfrak{F}_1(i[u - v]|k)Q(v, u|k), \tag{18.3.96}$$

and, from (18.2.8), (18.2.9) and (18.3.22), $P_H(v, u|k) \in \text{QP}(2i\mathcal{K}, 2\mathcal{K}'; 1, 1; -2, -2)$, $P_\Theta(v, u|k) \in \text{QP}(2i\mathcal{K}, 2\mathcal{K}'; 1, 1; -2, -2)$ and $Q(v, u|k) \in \text{QP}(2i\mathcal{K}, 2\mathcal{K}'; -1, 1; -2, -2)$, with respect to u .

- The formula

$$\ln \left[\frac{\Theta(u + v|k)}{\Theta(u - v|k)} \right] = \sum_{n=1}^{\infty} \frac{2 \sin(n\pi u/\mathcal{K}) \sin(n\pi v/\mathcal{K})}{n \sinh(n\pi \mathcal{K}'/\mathcal{K})} \tag{18.3.97}$$

is obtained by taking the logarithm of (18.3.15) expanding the individual logarithmic factors and summing over m . It then follows in a straightforward way that

$$\ln \left[\frac{\Theta(u + v|k) \Theta(u + \mathcal{K} + v|k)}{\Theta(u - v|k) \Theta(u + \mathcal{K} - v|k)} \right] = \sum_{n=1}^{\infty} \frac{2 \sin(2n\pi u/\mathcal{K}) \sin(2n\pi v/\mathcal{K})}{n \sinh(2n\pi \mathcal{K}'/\mathcal{K})} \tag{18.3.98}$$

and

$$\begin{aligned}
 \ln \left[\frac{H(u + v|k)}{H(u - v|k)} \right] &= \sum_{n=1}^{\infty} \frac{2 \sin(n\pi u/\mathcal{K}) \sin(n\pi v/\mathcal{K}) \cosh(n\pi \mathcal{K}'/\mathcal{K})}{n \sinh(n\pi \mathcal{K}'/\mathcal{K})}, \\
 &\quad \text{if } u > v > 0.
 \end{aligned} \tag{18.3.99}$$

• The function $Y(u', u|k)$, defined by (12.2.57), can, with a slight change of notation, be expressed in terms of v and u so that

$$\begin{aligned} Y(v, u|k) &:= Y(u + v, u - v|k) = \{-\operatorname{sn}(i[\mathcal{K}' - u - v]|k) \operatorname{sn}(i[u - v]|k)\}^{-\frac{1}{2}} \\ &= \{\operatorname{sn}(i[\mathcal{K}' + u + v]|k) \operatorname{sn}(i[u - v]|k)\}^{-\frac{1}{2}} = \sqrt{\frac{kQ(v, u|k)}{Q(-v, u|k)}}, \end{aligned} \quad (18.3.100)$$

$$Y(-v, u|k) = Y(v, u + \mathcal{K}'|k), \quad (18.3.101)$$

$$Y(v, u|k) = Y(v, u + 2\mathcal{K}'|k) = Y(v, u + 2i\mathcal{K}|k), \quad (18.3.102)$$

$$\frac{Y(v, u|k)}{Q(v, u|k)} = \frac{Y(-v, u|k)}{Q(-v, u|k)} = \sqrt{\frac{k}{Q(v, u|k)Q(-v, u|k)}}. \quad (18.3.103)$$

• The function

$$\begin{aligned} A(v, \zeta|k) &:= \operatorname{Am}(\zeta + i v|k) + \operatorname{Am}(\zeta - i v|k), \\ &= -i \ln[-k \operatorname{sn}(\zeta - i v - \tfrac{1}{2}i\mathcal{K}'|k) \operatorname{sn}(\zeta + i v - \tfrac{1}{2}i\mathcal{K}'|k)], \end{aligned} \quad (18.3.104)$$

is real-valued even in v and odd in ζ , with values in the range $[-2\pi, 2\pi]$ for $-2\mathcal{K} \leq \zeta \leq 2\mathcal{K}$. Its periodicity properties are given by

$$\begin{aligned} A(v, \zeta + 2\mathcal{K}|k) &= A(v, \zeta|k) \pm 2\pi, \\ A(v, \zeta + i\mathcal{K}'|k) &= -A(v, \zeta|k) \pm 2\pi, \quad A(v, \zeta + 2i\mathcal{K}'|k) = A(v, \zeta|k). \end{aligned} \quad (18.3.105)$$

• Using the notation for rapidities given by (12.2.22), let

$$F^{(s)}(v, u|k) := \left[-\frac{\sinh(2K^{(\mathfrak{f})}(u + v))}{\sinh(2K^{(\mathfrak{b})}(u - v))} \right]^{n_2} + s \left[-\frac{\sinh(2K^{(\mathfrak{b})}(u - v))}{\sinh(2K^{(\mathfrak{f})}(u + v))} \right]^{n_2}, \quad (18.3.106)$$

for $s = \pm 1$,¹² and, from (12.2.4), in terms of elliptic functions,

$$F^{(s)}(v, u|k) = \left[-\frac{\operatorname{sn}(i[\mathcal{K}' - u - v]|k)}{\operatorname{sn}(i[u - v]|k)} \right]^{n_2} + s \left[-\frac{\operatorname{sn}(i[u - v]|k)}{\operatorname{sn}(i[\mathcal{K}' - u - v]|k)} \right]^{n_2}, \quad (18.3.107)$$

¹² This function just gives the 2^{2n_2-1} -dimensionally degenerate eigenvalues of the matrix given in (12.2.60).

from which we note that $F^{(s)}(v, u|k) \in \text{DP}(2i\mathcal{K}, 2\mathcal{K}'; 1, 1)$. From (18.3.50)

$$F^{(s)}(v, u|k) = \frac{s G^{(s)}(v, u|k)}{[P(v, u|k)]^{n_2}} \quad (18.3.108)$$

where $P(v, u|k)$ is given by (12.2.81) and

$$G^{(s)}(v, u|k) := s + [P(v, u|k)]^{2n_2}. \quad (18.3.109)$$

We are interested in the poles and zeros of $G^{(s)}(v, u|k)$. Since $s = \exp(i\pi[1-s]/2)$, we have the factorization

$$G^{(s)}(v, u|k) = \prod_{j=1}^{2n_2} g(v, u; \psi_j^{(s)}|k), \quad (18.3.110)$$

where¹³

$$\psi_j^{(s)} := \frac{\pi(4j-s-1)}{8n_2}, \quad j = 1, 2, \dots, 2n_2 \quad (18.3.111)$$

and

$$g(v, u; \psi|k) := \exp(2i\psi) - P(v, u|k) \quad (18.3.112)$$

$$= \exp\left(-\frac{\pi(\mathcal{K}' + 2v)}{2\mathcal{K}}\right) \frac{P_{\Theta}(v, u|k) \exp(2i\psi) - P_{\text{H}}(v, u|k)}{P_{\text{H}}(v + \mathcal{K}', u|k)}. \quad (18.3.113)$$

This is the ratio of two $\text{QP}(2i\mathcal{K}, 2\mathcal{K}', 1, 1; -2, -2)$ functions making it $\text{DP}(2i\mathcal{K}, 2\mathcal{K}'; 1, 1)$. It has two first-order poles in $\mathbb{W}(2i\mathcal{K}, 2\mathcal{K}')$ which coincide with the zeros of $P_{\Theta}(v, u|k)$. The zeros of $g(v, u; \psi|k)$ are determined by using (18.3.104) to re-express it in the form

$$g(v, u; \psi|k) = \exp(2i\psi) + \exp[iA(v, iu + \tfrac{1}{2}i\mathcal{K}'|k)]. \quad (18.3.114)$$

Let $\zeta(v; \psi)$ be a solution of the equation

$$A(v, \zeta|k) = 2\psi - \pi, \quad 0 \leq \psi \leq \pi. \quad (18.3.115)$$

It is straightforward to see that $\zeta(v; \psi)$ is an even function of v , and an increasing function of $0 < \psi < \pi$, with

¹³ Note that $\prod_{j=1}^{2n_2} \exp(2i\psi_j^{(s)}) = s$.

$$\zeta(v; \pi - \psi) = -\zeta(v; \psi), \quad \zeta(v; 0) = -\mathcal{K}, \quad \zeta\left(v; \frac{\pi}{2}\right) = 0, \quad \zeta(v; \pi) = \mathcal{K}. \quad (18.3.116)$$

From (18.3.105),

$$u = -i \zeta(v; \psi) - \frac{1}{2} \mathcal{K}' \quad \text{and} \quad u = i \zeta(v; \psi) + \frac{1}{2} \mathcal{K}', \quad (18.3.117)$$

are roots of $g(v; u; \psi) = 0$ in $\mathbb{W}'(2i\mathcal{K}, 2\mathcal{K}')$ and, from Theorem 18.3.1 and (18.3.95),

$$g(v, u; \psi|k) = c(v; \psi|k) \frac{H(iu - \zeta(v; \psi) + \frac{1}{2} i \mathcal{K}'|k) H(iu + \zeta(v; \psi) - \frac{1}{2} i \mathcal{K}'|k)}{P_{\Theta}(v, u|k)}, \quad (18.3.118)$$

where, from (18.3.112) and (18.3.118),

$$c(v; \psi|k) := \frac{\exp(2i\psi) \Theta(2iv|k) \Theta(0|k)}{H(iv - \zeta(v; \psi) + \frac{1}{2} i \mathcal{K}'|k) H(iv + \zeta(v; \psi) - \frac{1}{2} i \mathcal{K}'|k)}. \quad (18.3.119)$$

18.3.8 Baxter's Modified Theta Functions

The modified theta functions

$$\Phi_{\mathfrak{B}}(u|k) := \mathfrak{B}(u|k) \Phi(u|k), \quad \Phi = H, \Theta, \quad (18.3.120)$$

where

$$\mathfrak{B}(u|k) := \mathfrak{E}\left(\frac{m'[u - \mathcal{K}]^2}{4\mathcal{K}} \middle| k\right), \quad (18.3.121)$$

$$\mathfrak{E}(u|k) := \exp\left(\frac{i\pi u}{\mathcal{U}_{\mathfrak{B}}}\right), \quad \mathcal{U}_{\mathfrak{B}} := \mathcal{U}(2m, m'|k) \quad (18.3.122)$$

and $\mathcal{U}(\ell, \ell'|k)$ is defined in (18.3.21), were introduced by Baxter (1973a). Of course, there is no loss of generality in assuming that $2m$ and m' are coprime in the definition of Baxter's modified theta functions. However, the case where m' is even gives a non-trivially different situation in the roots of unity formula (12.3.128). $H(u|k)$ and $\Theta(u|k)$ are respectively odd and even functions, but

$$H_{\mathfrak{B}}(-u|k) = -\mathfrak{E}(m'u|k) H_{\mathfrak{B}}(u|k), \quad \Theta_{\mathfrak{B}}(-u|k) = \mathfrak{E}(m'u|k) \Theta_{\mathfrak{B}}(u|k). \quad (18.3.123)$$

From (18.3.19) and (18.3.20),

$$H_{\mathfrak{B}}(u + \mathfrak{U}(2\ell, \ell'|k)|k) = (-1)^\ell (-i)^{\ell'} \mathfrak{G}(\ell, \ell'; u|k) \times \begin{cases} H_{\mathfrak{B}}(u|k), & \text{if } \ell' \text{ even,} \\ \Theta_{\mathfrak{B}}(u|k), & \text{if } \ell' \text{ odd,} \end{cases} \quad (18.3.124)$$

$$\Theta_{\mathfrak{B}}(u + \mathfrak{U}(2\ell, \ell'|k)|k) = (-i)^{\ell'} \mathfrak{G}(\ell, \ell'; u|k) \times \begin{cases} \Theta_{\mathfrak{B}}(u|k), & \text{if } \ell' \text{ even,} \\ H_{\mathfrak{B}}(u|k), & \text{if } \ell' \text{ odd,} \end{cases} \quad (18.3.125)$$

where

$$\mathfrak{G}(\ell, \ell'; u|k) := \frac{\mathfrak{B}(u + \mathfrak{U}(2\ell, \ell'|k)) \mathfrak{F}_{\ell'}(u + 2\ell\mathcal{K}'|k)}{\mathfrak{B}(u|k)}. \quad (18.3.126)$$

The following special cases are needed in Chap. 12:

•

$$\begin{aligned} H_{\mathfrak{B}}(u + 2\mathcal{K}|k) &= -\mathfrak{C}(m'u) H_{\mathfrak{B}}(u|k), \\ \Theta_{\mathfrak{B}}(u + 2\mathcal{K}|k) &= \mathfrak{C}(m'u) \Theta_{\mathfrak{B}}(u|k), \end{aligned} \quad (18.3.127)$$

$$\begin{aligned} H_{\mathfrak{B}}(u + i\mathcal{K}'|k) &= \mathfrak{C}(m[\mathcal{K} - \tfrac{1}{2}i\mathcal{K}' - u]) \Theta_{\mathfrak{B}}(u|k), \\ \Theta_{\mathfrak{B}}(u + i\mathcal{K}'|k) &= \mathfrak{C}(m[\mathcal{K} - \tfrac{1}{2}i\mathcal{K}' - u]) H_{\mathfrak{B}}(u|k). \end{aligned} \quad (18.3.128)$$

- If m_0 is the greatest common divisor of $2m$ and m' , with

$$\bar{m} = 2m/m_0, \quad \bar{m}' = m'/m_0, \quad (18.3.129)$$

then $2(m'\ell - m\ell') = m_0(2\bar{m}'\ell - \bar{m}\ell')$. Thus, with $\ell = \bar{m}$, $\ell' = 2\bar{m}'$ and

$$\mathcal{O}_{\mathfrak{B}} := \mathfrak{U}_{\mathfrak{B}}/m_0 = \bar{m}\mathcal{K} + i\bar{m}'\mathcal{K}', \quad (18.3.130)$$

$$\begin{aligned} H_{\mathfrak{B}}(u \pm 2\mathcal{O}_{\mathfrak{B}}|k) &= (-1)^{\bar{m} + \bar{m}'} H_{\mathfrak{B}}(u|k), \\ \Theta_{\mathfrak{B}}(u \pm 2\mathcal{O}_{\mathfrak{B}}|k) &= (-1)^{\bar{m}\bar{m}'} \Theta_{\mathfrak{B}}(u|k). \end{aligned} \quad (18.3.131)$$

- If $\bar{m}$ is even and $\bar{m}'$ is odd, let $\ell = \bar{m}/2$, $\ell' = \bar{m}'$, giving

$$\begin{aligned} H_{\mathfrak{B}}(u \pm \mathcal{O}_{\mathfrak{B}}|k) &= (i)^{\bar{m}} \exp\left(\tfrac{1}{4}i\pi\bar{m}\bar{m}'\right) \Theta_{\mathfrak{B}}(u|k), \\ \Theta_{\mathfrak{B}}(u \pm \mathcal{O}_{\mathfrak{B}}|k) &= \exp\left(\tfrac{1}{4}i\pi\bar{m}\bar{m}'\right) H_{\mathfrak{B}}(u|k). \end{aligned} \quad (18.3.132)$$

- Setting $\ell = n$, $\ell' = 2n'$ with

$$2mn' - m'n = m_0, \quad \overline{m}n' - \overline{m}'n = 1, \quad (18.3.133)$$

gives

$$\mathcal{U}(2(n-1), 2n'|k) = 2(1 + \overline{m}')\mathcal{O}'_{\mathfrak{B}} - 2n'\mathcal{O}_{\mathfrak{B}}, \quad (18.3.134)$$

where

$$\mathcal{O}'_{\mathfrak{B}} := n\mathcal{K} + i n'\mathcal{K}' \quad (18.3.135)$$

and

$$\begin{aligned} \mathcal{H}_{\mathfrak{B}}(u \pm 2\mathcal{O}'_{\mathfrak{B}}|k) &= (-1)^{n+n'+nn'} \exp\left(\frac{-i\pi\{[1 \pm \overline{m}']\mathcal{O}'_{\mathfrak{B}} \pm u\}}{\mathcal{O}_{\mathfrak{B}}}\right) \mathcal{H}_{\mathfrak{B}}(u|k), \\ \Theta_{\mathfrak{B}}(u \pm 2\mathcal{O}'_{\mathfrak{B}}|k) &= (-1)^{n'+nn'} \exp\left(\frac{-i\pi\{[1 \pm \overline{m}']\mathcal{O}'_{\mathfrak{B}} \pm u\}}{\mathcal{O}_{\mathfrak{B}}}\right) \Theta_{\mathfrak{B}}(u|k). \end{aligned} \quad (18.3.136)$$

From (18.2.8) and (18.3.22) it can be seen that

$$\begin{aligned} \mathcal{H}_{\mathfrak{B}}(u|k) &\in \mathcal{QP}(2\mathcal{O}_{\mathfrak{B}}, 2\mathcal{O}'_{\mathfrak{B}}; (-1)^{\overline{m}+\overline{m}'}, (-1)^{n+n'+nn'}; 1 + \overline{m}', 1), \\ \Theta_{\mathfrak{B}}(u|k) &\in \mathcal{QP}(2\mathcal{O}_{\mathfrak{B}}, 2\mathcal{O}'_{\mathfrak{B}}; (-1)^{\overline{m}'}, (-1)^{n'+nn'}; 1 + \overline{m}', 1), \end{aligned} \quad (18.3.137)$$

which can be compared to the corresponding results (18.3.28) for the standard theta functions. It may be seen directly or from Theorem 18.2.5 that $\mathcal{H}_{\mathfrak{B}}(u|k)$ and $\Theta_{\mathfrak{B}}(u|k)$ each has one zero in $\mathbb{W}(2\mathcal{O}_{\mathfrak{B}}, 2\mathcal{O}'_{\mathfrak{B}})$ located respectively at

$$u = \frac{1}{2}\mathcal{O}_{\mathfrak{B}}(4\ell - 1 - (-1)^{n+n'+nn'}) + \frac{1}{2}\mathcal{O}'_{\mathfrak{B}}(4\ell' - 1 - (-1)^{\overline{m}+\overline{m}'}), \quad (18.3.138)$$

$$u = \frac{1}{2}\mathcal{O}_{\mathfrak{B}}(4\ell - 1 - (-1)^{n'+nn'}) + \frac{1}{2}\mathcal{O}'_{\mathfrak{B}}(4\ell' - 1 - (-1)^{\overline{m}'}), \quad (18.3.139)$$

for some choices of ℓ and ℓ' . Then corresponding to the Theorem 18.3.1 for doubly-periodic functions we have:

Theorem 18.3.2 *For some complex quantities α_j , $j = 1, 2, \dots, r$, lying in the period rectangle $\mathbb{W}(2\mathcal{O}_{\mathfrak{B}}, 2i\mathcal{O}'_{\mathfrak{B}})$,*

$$W(u|k) := \exp\left(-\frac{\pi i \Im u}{2\mathcal{O}_{\mathfrak{B}}}\right) \prod_{j=1}^r \mathcal{H}_{\mathfrak{B}}(u - \alpha_j|k) \quad (18.3.140)$$

is, to within a multiplicative quantity possibly depending on k , the unique entire member of $\mathcal{QP}(2\mathcal{O}_{\mathfrak{B}}, 2i\mathcal{O}'_{\mathfrak{B}}, \tau, \tau'; r, r)$ with zeros α_j , $j = 1, 2, \dots, r$, if

$$\begin{aligned}\mathfrak{N} &= \frac{1}{2}(4\ell' + 1 - \tau) + r\overline{m}(1 + \overline{m}'), \\ \sum_{j=1}^r \alpha_j &= \mathfrak{N} \mathcal{O}'_{\mathfrak{B}} + \frac{1}{2} \mathcal{O}_{\mathfrak{B}}(4\ell + 1 - \tau') + r\{\mathcal{O}_{\mathfrak{B}}(n + n' + nn') + \mathcal{O}'_{\mathfrak{B}}\overline{m}'\},\end{aligned}\tag{18.3.141}$$

for some integers ℓ and ℓ' .

Proof From (18.3.133) and (18.3.138)

$$\begin{aligned}W(u + 2\mathcal{O}_{\mathfrak{B}}|k) &= (-1)^{r[\overline{m} + \overline{m}'\overline{m}']} \exp(-\pi i \mathfrak{N}) W(u|k), \\ W(u + 2\mathcal{O}'_{\mathfrak{B}}|k) &= (-1)^{r[n + n' + nn']} \exp\left(-\frac{\pi i \mathfrak{N} \mathcal{O}'_{\mathfrak{B}}}{\mathcal{O}_{\mathfrak{B}}}\right) \\ &\quad \times \exp\left(\frac{-i\pi\{r[1 + \overline{m}']\mathcal{O}'_{\mathfrak{B}} + ru - \sum_{j=1}^r \alpha_j\}}{\mathcal{O}_{\mathfrak{B}}}\right) W(u|k).\end{aligned}\tag{18.3.142}$$

Using (18.2.5), the conditions that $W(u|k) \in \text{QP}(2\mathcal{O}_{\mathfrak{B}}, 2i\mathcal{O}'_{\mathfrak{B}}, \tau, \tau'; r, r)$ are (18.3.141). That it is, to within a multiplicative quantity possibly dependent on k , the unique member of $\text{QP}(2\mathcal{O}_{\mathfrak{B}}, 2i\mathcal{O}'_{\mathfrak{B}}, \tau, \tau'; r, r)$ with these zeros follows from Theorem 18.2.6. $\square$

Formulae (18.3.33) and (18.3.34) hold without change for the modified theta functions. This is not the case for (18.3.35). Now it is convenient to define

$$\mathfrak{L}(v, u|k) := H_{\mathfrak{B}}(\mathcal{K} + u|k)H_{\mathfrak{B}}(\mathcal{K} - v|k) + \Theta_{\mathfrak{B}}(\mathcal{K} + u|k)\Theta_{\mathfrak{B}}(\mathcal{K} - v|k), \tag{18.3.143}$$

when the modified version of (18.3.35) is expressible in the form

$$\mathfrak{L}(v, u|k) = \mathcal{R}(k) \mathfrak{H}(u + v|k) \mathfrak{H}(u - v|k), \tag{18.3.144}$$

where

$$\mathcal{R}(k) := 2\mathcal{Q}_1^{\frac{1}{4}}[H_{\mathfrak{B}}(\mathcal{K}|k)\Theta_{\mathfrak{B}}(\mathcal{K}|k)]^{-1} \mathfrak{E}\left(\frac{m'\mathcal{K}'^2}{4\mathcal{K}} \middle| k\right), \tag{18.3.145}$$

$$\mathfrak{H}(u|k) := H_{\mathfrak{B}}(\tfrac{1}{2}[i\mathcal{K}' + u] + \mathcal{K}|k)H_{\mathfrak{B}}(\tfrac{1}{2}[i\mathcal{K}' - u] + \mathcal{K}|k). \tag{18.3.146}$$

It is clear that

$$\mathfrak{L}(v, u|k) = \mathfrak{L}(v, u|k) = \mathfrak{L}(-u, v|k) = \mathfrak{L}(u, -v|k), \quad (18.3.147)$$

$$\mathfrak{L}(u - \mathfrak{U}_{\mathfrak{B}}, v|k) = (-1)^{mm'} \mathfrak{L}(u, v - \mathfrak{U}_{\mathfrak{B}}|k). \quad (18.3.148)$$

For some α and integers j and k

$$\frac{\mathfrak{L}(v + \alpha j, u + \alpha k|k)}{\mathfrak{L}(u + \alpha j, v + \alpha k|k)} = \frac{\mathfrak{H}(u - v - \alpha[j - k]|k)}{\mathfrak{H}(u - v + \alpha[j - k]|k)}. \quad (18.3.149)$$

Let

$$\begin{aligned} \Psi(+1, +1; v, u|k) &:= \mathfrak{H}(-2u|k) \Theta(u + v|k) \Theta(u - v|k), \\ \Psi(+1, -1; v, u|k) &:= \mathfrak{H}(-2u|k) \mathfrak{H}(u + v|k) \mathfrak{H}(u - v|k), \\ \Psi(-1, +1; v, u|k) &:= \Theta(-2u|k) \mathfrak{H}(u + v|k) \Theta(u - v|k), \\ \Psi(-1, -1; v, u|k) &:= \Theta(-2u|k) \Theta(u + v|k) \mathfrak{H}(u - v|k), \end{aligned} \quad (18.3.150)$$

with the corresponding expressions when the theta functions are replaced by their modified forms given, for $\alpha, \beta = \pm 1$, by

$$\Psi_{\mathfrak{B}}(\alpha, \beta; v, u|k) := \mathfrak{E}\left(\frac{m'[3\mathcal{K}^2 + 6u^2 + 2v^2]}{4\mathcal{K}}|k\right) \Psi(\alpha, \beta; v, u|k). \quad (18.3.151)$$

These functions have the properties

$$\begin{aligned} \Psi_{\mathfrak{B}}(\alpha, \beta; -v, u|k) &= \Psi_{\mathfrak{B}}(\alpha, \alpha\beta; v, u; k), \\ \Psi_{\mathfrak{B}}(\alpha, \beta; v, -u|k) &= -\Psi_{\mathfrak{B}}(\alpha, \alpha\beta; v, u; k), \end{aligned} \quad (18.3.152)$$

$$\begin{aligned} \Psi_{\mathfrak{B}}(\alpha, \beta; v + 2\mathcal{K}, u|k) &= \alpha \mathfrak{E}(2m'[\mathcal{K} + v]|k) \Psi_{\mathfrak{B}}(\alpha, \beta; v, u|k), \\ \Psi_{\mathfrak{B}}(\alpha, \beta; v + i\mathcal{K}', u|k) &= \mathfrak{E}(-m[i\mathcal{K}' + 2v]|k) \Psi_{\mathfrak{B}}(\alpha, -\beta; v, u|k), \\ \Psi_{\mathfrak{B}}(\alpha, \beta; v + 2i\mathcal{K}', u|k) &= \mathfrak{E}(-4m[i\mathcal{K}' + v]|k) \Psi_{\mathfrak{B}}(\alpha, \beta; v, u|k), \end{aligned} \quad (18.3.153)$$

$$\Psi_{\mathfrak{B}}(\alpha, \beta; v + \tfrac{1}{2}i\mathcal{K}', u + \tfrac{1}{2}i\mathcal{K}'|k) = -\mathfrak{E}(-m[i\mathcal{K}' + v + 3u]|k) \Psi_{\mathfrak{B}}(-\alpha, \beta; v, u|k). \quad (18.3.154)$$

In particular, the function

$$\mathfrak{h}_{\mathfrak{Z}}(v|k) := \Psi_{\mathfrak{Z}}(-1, +1; v, 0|k) = \Theta_{\mathfrak{Z}}(0|k)H_{\mathfrak{Z}}(v|k) \Theta_{\mathfrak{Z}}(v|k), \quad (18.3.155)$$

is odd and has periodicity properties given by (18.3.153). In the special case $m' = 0$ the modified theta functions reduce to their standard form and $\mathfrak{h}_{\mathfrak{Z}}(v|k)$ becomes

$$\mathfrak{h}(u|k) := \Theta(0|k)H(u|k) \Theta(-u|k), \quad (18.3.156)$$

with

$$\mathfrak{h}(u + 2\mathcal{K}|k) = -\mathfrak{h}(u) = \mathfrak{h}(-u|k), \quad (18.3.157)$$

$$\mathfrak{h}(u + i\mathcal{K}'|k) = -\frac{1}{\sqrt{Q}} \exp\left(-\frac{\pi i u}{\mathcal{K}}\right) \mathfrak{h}(u|k). \quad (18.3.158)$$

The complementary transformation is particularly useful for manipulating the function

$$\begin{aligned} \frac{\mathfrak{h}(u + \mathcal{K} - i v|k)}{\mathfrak{h}(u + \mathcal{K} + i v|k)} &= \frac{H(u + \mathcal{K} - i v|k) \Theta(u + \mathcal{K} - i v|k)}{H(u + \mathcal{K} + i v|k) \Theta(u + \mathcal{K} + i v|k)} \\ &= \exp\left(\frac{2\pi i u v}{\mathcal{K} \mathcal{K}'}\right) \frac{\Theta(v + i u|k') \Theta(v + \mathcal{K}' + i u|k')}{\Theta(v - i u|k') \Theta(v + \mathcal{K}' - i u|k')}. \end{aligned} \quad (18.3.159)$$

From (18.3.98),

$$\begin{aligned} f(v, u|k) &:= \ln \left[\frac{\mathfrak{h}(u + \mathcal{K} - i v|k)}{\mathfrak{h}(u + \mathcal{K} + i v|k)} \right] = \frac{2\pi i u v}{\mathcal{K} \mathcal{K}'} \\ &+ \sum_{n=1}^{\infty} \frac{2i \sin(2n\pi v/\mathcal{K}') \sinh(2n\pi u/\mathcal{K}')}{n \sinh(2n\pi \mathcal{K}/\mathcal{K}')}. \end{aligned} \quad (18.3.160)$$

Using a dot to denote differentiation with respect to v ¹⁴

$$\dot{f}(v, u|k) = \frac{2\pi i}{\mathcal{K}'} \sum_{n=-\infty}^{\infty} \frac{\exp(2n\pi i v/\mathcal{K}') \sinh(2n\pi u/\mathcal{K}')}{\sinh(2n\pi \mathcal{K}/\mathcal{K}')}, \quad (18.3.161)$$

which is periodic in v of period $\mathcal{K}'$. From (18.1.21), with $L = \mathcal{K}'$, we see that its Fourier transform is

$$\dot{f}^*(u, n|k) = \frac{2\pi i \sinh(2n\pi u/\mathcal{K}')}{\mathcal{K}' \sinh(2n\pi \mathcal{K}/\mathcal{K}')}. \quad (18.3.162)$$

¹⁴ The $n = 0$ term is understood as the result of taking the limit $n \rightarrow 0$.

Again, for $\alpha, \beta = \pm 1$, let

$$W_{\mathfrak{B}}(\alpha, \beta; v, u|k) := \frac{1}{2}[\Psi_{\mathfrak{B}}(\alpha, +1; v, u|k) + \beta \Psi_{\mathfrak{B}}(\alpha, -1; v, u|k)]. \quad (18.3.163)$$

Then it can be shown that the effect of the Bazhanov–Mangazeev transformation with

$$u_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - u_1]}{\mathcal{K}'_1}, \quad v_3 := \frac{2i\mathcal{K}_3[\mathcal{K}_1 - v_1]}{\mathcal{K}'_1}, \quad (18.3.164)$$

is

$$\begin{aligned} W_{\mathfrak{B}}(+1, +1; v_3, u_3|k_3) &= \Pi_{\mathfrak{B}}(v_1, u_1|k_1) W_{\mathfrak{B}}(-1, +1; v_1, u_1|k_1), \\ W_{\mathfrak{B}}(+1, -1; v_3, u_3|k_3) &= \Pi_{\mathfrak{B}}(v_1, u_1|k_1) W_{\mathfrak{B}}(+1, +1; v_1, u_1|k_1), \\ W_{\mathfrak{B}}(-1, +1; v_3, u_3|k_3) &= \Pi_{\mathfrak{B}}(v_1, u_1|k_1) W_{\mathfrak{B}}(-1, -1; v_1, u_1|k_1), \\ W_{\mathfrak{B}}(-1, -1; v_3, u_3|k_3) &= \Pi_{\mathfrak{B}}(v_1, u_1|k_1) W_{\mathfrak{B}}(+1, -1; v_1, u_1|k_1), \end{aligned} \quad (18.3.165)$$

where

$$\Pi_{\mathfrak{B}}(v_1, u_1|k_1) := \frac{\mathfrak{e}\left(\frac{m'[6u_3^2 + 2v_3^2 + 3\mathcal{K}_3^2]}{4\mathcal{K}_3}\right)|_{k_3}}{\mathfrak{e}\left(\frac{m'[6u_1^2 + 2v_1^2 + 3\mathcal{K}_1^2]}{4\mathcal{K}_1}\right)|_{k_1}} \Pi(v_1, u_1|k_1), \quad (18.3.166)$$

$$\Pi(v_1, u_1|k_1) := i\sqrt{\frac{\mathcal{K}'_1}{2\mathcal{K}_1} \frac{c(k_3)}{c(k_1)}} \exp\left(\frac{\pi[v_1(v_1 - 2\mathcal{K}_1) + 3u_1(u_1 - 2\mathcal{K}_1) + 4\mathcal{K}_1^2]}{2\mathcal{K}_1\mathcal{K}'_1}\right). \quad (18.3.167)$$

and

$$c(k) := H(\mathcal{K}(k)|k) \Theta(\mathcal{K}(k)|k) [\Theta(0|k)]^2. \quad (18.3.168)$$

18.4 The Potts Delta Function

The properties of the *Potts delta function*

$$\Delta(v; \sigma) := v \delta^{\text{Kr}}(\sigma) - 1, \quad (18.4.1)$$

are important for the derivation of high-temperature series (Sect. 14.2.2).¹⁵ In particular for the zero- μ series it is necessary to compute expressions of the form

$$\Sigma(e, k) := \sum_{\{\sigma\}} \Pi(e, k), \quad (18.4.2)$$

where¹⁶

$$\begin{aligned} \Pi(e, k) := & \nu^{-n} \Delta(v; \sigma(\mathbf{r}_1) - \sigma(\mathbf{r}'_1)) \Delta(v; \sigma(\mathbf{r}_2) - \sigma(\mathbf{r}'_2)) \\ & \times \Delta(v; \sigma(\mathbf{r}_3) - \sigma(\mathbf{r}'_3)) \cdots \Delta(v; \sigma(\mathbf{r}_e) - \sigma(\mathbf{r}'_e)) \end{aligned} \quad (18.4.3)$$

is a product of Δ -functions on the e edges $(\mathbf{r}_1, \mathbf{r}'_1), (\mathbf{r}_2, \mathbf{r}'_2), \dots, (\mathbf{r}_e, \mathbf{r}'_e)$ of a graph $(e, k) \in \mathcal{N}$ and the summation is over all the state variables. In the case of a $\mu \neq 0$ series (18.4.2) is modified to

$$\Sigma(e, k; n', k') := \sum_{\{\sigma\}} \Pi(e, k) \times \Lambda\{n', k'\}, \quad (18.4.4)$$

where¹⁷

$$\Lambda\{n', k'\} := \Delta(v; \sigma(\mathbf{r}_1)) \Delta(v; \sigma(\mathbf{r}_2)) \Delta(v; \sigma(\mathbf{r}_3)) \cdots \Delta(v; \sigma(\mathbf{r}_{n'})) \quad (18.4.5)$$

is a product of Δ -functions at n' vertices $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n'}$ of $\mathcal{N}$ in a configuration denoted by $\{n', k'\}$ and the summation is over all the state variables for (e, k) and $\{n', k'\}$. These contributions to $\Sigma(e, k; n', k')$ will be referred to as *single-vertex terms*.

Before considering the general case we note that when $\nu = 2$ (the ‘Ising case’)

$$\sum_{\sigma(\mathbf{r}_0)=0}^1 \prod_{k=1}^n \Delta(2; \sigma(\mathbf{r}_k) - \sigma(\mathbf{r}_0)) = (-1)^m [1 + (-1)^n], \quad (18.4.6)$$

where m is the number of $\sigma(\mathbf{r}_k) = 0$. Thus $\Sigma(e, k) = 0$ if (e, k) has odd vertices and $\Sigma(e, k; n', k') = 0$ if (e, k) has a single-vertex term on an even vertex or no single-vertex term on a odd vertex. Hence the name *magnetic* used to denote a graph with odd vertices (Sect. 16.4.3).¹⁸

Unfortunately the situation for general ν is more complicated and the aim of this section is to determine some conditions under which $\Sigma(e, k)$ and $\Sigma(e, k; n', k')$ are

¹⁵ In all that follows σ (and all its primed and subscripted versions) is a Potts variable with integer values in the range $[0, \nu - 1]$.

¹⁶ This is a repetition of (14.2.37).

¹⁷ This is a repetition of (14.2.42).

¹⁸ These observations are useful because they predict a factor of $\nu - 2$ in all the general computations of cases which would yield zero for $\nu = 2$.

zero; this being equivalent to a determination of graphs which can be excluded from the computation of high-temperature series for $\mu = 0$ and $\mu \neq 0$ respectively. For this analysis we need the following results

$$\sum_{\sigma=0}^{v-1} \Delta(v; \sigma_1 - \sigma) = 0, \quad (18.4.7)$$

$$\sum_{\sigma=0}^{v-1} \Delta(v; \sigma_1 - \sigma) \Delta(v; \sigma_2 - \sigma) = v \Delta(v; \sigma_1 - \sigma_2), \quad (18.4.8)$$

$$\begin{aligned} \sum_{\sigma=0}^{v-1} \Delta(v; \sigma_1 - \sigma) \Delta(v; \sigma_2 - \sigma) \Delta(v; \sigma_3 - \sigma) \\ = v [\Delta(v; \sigma_1 - \sigma_2) \Delta(v; \sigma_2 - \sigma_3) - \Delta(v; \sigma_1 - \sigma_3)] \end{aligned} \quad (18.4.9)$$

and, more generally,¹⁹

$$\sum_{\sigma=0}^{v-1} \prod_{k=1}^n \Delta(v; \sigma_k - \sigma) = v B_n, \quad (18.4.10)$$

where

$$B_n := \sum_{k=1}^{n-1} (-1)^{k+1} A_{k,n} \quad n > 1, \quad B_1 = 0, \quad (18.4.11)$$

$$A_{k,n} := \prod_{i=k+1}^n \Delta(v; \sigma_i - \sigma_k), \quad k < n, \quad A_{n,n} = 1. \quad (18.4.12)$$

In the special cases when

$$(a) \quad \sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma',$$

$$\sum_{\sigma=0}^{v-1} [\Delta(v; \sigma' - \sigma)]^n = v(v-1) Q(v; n-1), \quad (18.4.13)$$

$$(b) \quad \sigma_1 = \sigma_2 = \cdots = \sigma_{n-1} = \sigma', \sigma_n = \sigma'',$$

¹⁹ This result is derived by proving by induction that

$$\prod_{k=1}^n \Delta(v; \sigma_k - \sigma) = B_n + \sum_{k=1}^n (-1)^{k+1} A_{k,n} \Delta(v; \sigma_k - \sigma)$$

and then summing over the values of σ .

$$\sum_{\sigma=0}^{v-1} [\Delta(v; \sigma' - \sigma)]^{n-1} \Delta(v; \sigma'' - \sigma) = v \Delta(v; \sigma' - \sigma'') Q(v; n-1), \quad (18.4.14)$$

where

$$Q(v; n) := \frac{(v-1)^n - (-1)^n}{v}. \quad (18.4.15)$$

18.4.1 The $\mu = 0$ Case

We now derive a number of properties of the sum (18.4.2) which allow us to restrict the number of graphs which need to be considered in the high-temperature zero- μ expansion in Sect. 14.2.2.

Property A.1: $\Sigma(e, k) = 0$ if (e, k) has any vertices of valency one. The summation over a vertex of valency one involves only one Δ -function and the property follows from (18.4.7).

Property A.2: Summations in (18.4.2) over vertices of valency two in (e, k) can be performed independently of the other summations, yielding from (18.4.8) a factor v . This means that in most cases vertices of valency two can be removed and ignored. An exception to this is the case where (e, k) is a simple cycle (Example 18.4.1, below).

Property A.3: Let

$$\Sigma_1(e, k | \sigma(\mathbf{r}_1) = j) := v \sum_{\{\sigma\}_1} \Pi(e, k), \quad (18.4.16)$$

where the summation is over all the vertex variables of (e, k) except one, which, without loss of generality, we can take to be $\sigma(\mathbf{r}_1)$, which is fixed at a particular value $j \in [0, v-1]$. Since the terms on the right-hand side of (18.4.2) transform into each other under the symmetry group $\mathcal{S}_v$, they form a set of v equal groups labelled by the value of $\sigma(\mathbf{r}_1)$. Thus

$$\Sigma_1(e, k) := \Sigma_1(e, k | \sigma(\mathbf{r}_1) = j) = \Sigma(e, k), \quad (18.4.17)$$

for any vertex of (e, k) and any state j . This means that by the time the last summation is performed in (18.4.2) all the vertex variables have disappeared and the summation just yields a factor of v .

Property A.4: If (e, k) contains a type A articulation point (Sect. 16.1.1 and Fig. 16.1) then the two parts of the graph joined at the articulation point can be treated as separate components for the purposes of the summation (18.4.2). This follows from (18.4.17) by choosing the

vertex for which the summation is omitted as the articulation point. Example 18.4.3 illustrates this property since the same result would be achieved for a two-component graph consisting of two simple cycles.

Property A.5: The summation $\Sigma(e, k) = 0$ for any graph containing a type B or C articulation point. In this case the graph has one or more bridges. If summations are performed over all variables of a part of the graph connected to the rest by a bridge, but omitting the vertex at the end of the bridge, the bridge will be left as a ‘dangling’ edge with a vertex of valency one. The result then follows from Property A.1.

As observed in Sect. 16.1.1 any general graph consists of a collection of stars linked by articulation points. We are interested in determining the graphs for which summations $\Sigma(e, k)$ make a non-zero contribution to the high-temperature zero- μ series described in Sect. 14.2.2. According to the properties listed here graphs with vertices of valency one and those with articulation points of type B and C can be omitted since they make a zero contribution. The parts of a graph linked by articulation points of type A can be treated as separate components as long this is taken into account when calculating the lattice constants. The final conclusion of this discussion is a result due to Domb (1974a, b) that the only terms which need to be considered in the high-temperature zero- μ series are those due to star graphs.

Example 18.4.1 Let the graph be a simple cycle for which all the vertices are of valency two. If all except two are removed, using (18.4.8), the cycle is reduced to a double edge between these vertices and, applying (18.4.13), $\Sigma(e, k) = (v - 1)$.

Example 18.4.2 Let the graph consist of two simple cycles with one common edge. If all the edges are removed, using (18.4.8), except the shared edge the graph is reduced to a triple edge between these vertices and, applying (18.4.13), $\Sigma(e, k) = (v - 1)(v - 2)$. The graph has two odd vertices so, as indicated above, we expected the factor $v - 2$.

Example 18.4.3 Let the graph consist of two simple cycles linked at one vertex (an articulation point of type A). Then, as in Example 18.4.1, each cycle can be reduced to a double edge and the graph becomes two double edges linked at the articulation point. Using (18.4.13) $\Sigma(e, k) = (v - 1)^2$.

Example 18.4.4 Let the graph be $(12, 1)$, the block of four squares shown in Table 16.1. Then, as in the previous examples, the four corner vertices of valency two can be removed leaving a diamond of four vertices. This, using (18.4.10) with $n = 4$, is followed by a summation over the centre vertex. After summation over the resulting cycle $\Sigma(12, 1) = (v - 1)(v - 2)(v^2 - 5v + 7)$. As an example of the discussion in Sect. 14.2.3 we note that $[4, 8]$ in Table 16.1, with its vertices in the centres of the faces, is the dual of $(12, 1)$ and that the number of ways of assigning state variables to the vertices of $[4, 8]$ so that each edge contains an unlike pair with variable values in the range $\sigma = 1, 2, \dots, v - 1$ is also $(v - 1)(v - 2)(v^2 - 5v + 7)$.²⁰

²⁰ The state $\sigma = 0$ is excluded by the links to the boundary vertices.

18.4.2 The $\mu \neq 0$ Case

The properties from which it can be deduced that $\Sigma(e, k) = 0$ no longer apply to $\Sigma(e, k; n', k')$. However, it is easy to see that

Property B.1: $\Sigma(e, k; n', k') = 0$ unless $\{n', k'\}$ is identical to, or a subset of the vertex configuration of (e, k) with $n' \leq n(e, k)$. This follows from (18.4.7).

Property B.2: $\Sigma(e, k; 1, k') = 0$. This follows from (18.4.17), where the initial summation excludes the site containing the single-vertex term. The final summation is then over the single-vertex Δ -function.

These properties are now used in some examples:

Example 18.4.5 Let the graph be $(1, 1)$, the single pair of vertices connected by an edge shown in Table 16.1. It follows from Properties 16.1 and 17.1 that $\Sigma(1, 1; n', k') = 0$ if one or both of the vertices does not contain a single-vertex term. However, when both do it follows from (18.4.8) and (18.4.14) that $\Sigma(1, 1; 2, 1) = v - 1$.

Example 18.4.6 For the simple chain²¹ $(e, \mathbb{S}\mathbb{C})$, $\Sigma(e, \mathbb{S}\mathbb{C}; n', k') = 0$ unless the two ends of the chain both contain single-vertex terms and a vertex not containing a single-vertex term can be removed. So we need consider only the case where each vertex contains a single-vertex term when it follows from (18.4.13) and (18.4.14) that

$$\Sigma(e, \mathbb{S}\mathbb{C}; e + 1, k') = (v - 1)(v - 2)^{e-1}. \quad (18.4.18)$$

Example 18.4.7 Let the graph (e, k) be a simple cycle of e edges with two vertices containing single-vertex terms. The cycle can be reduced to a double edge with a single-vertex term at each end giving, from (18.4.14), $\Sigma(e, k; 2, k') = (v - 1)(v - 2)$.

Example 18.4.8 Consider now the simple cycle (e, k) with every vertex occupied by a single vertex term. From (18.4.9), (18.4.13) and (18.4.14)

$$\Sigma(e, k; e, 1) = (v - 1)(v - 2)^{e-1} - \Sigma(e - 1, k; e - 1, 1), \quad (18.4.19)$$

which can be iterated to give

$$\Sigma(e, k; e, 1) = (v - 2) \left[(v - 2)^{e-1} - (-1)^{e-1} \right]. \quad (18.4.20)$$

Example 18.4.9 Let the graph consist of two simple cycles linked by one bridge (and two type B articulation points). Each cycle can be reduced to a double edge and summing over the vertices at the free ends using (18.4.13) gives a factor of $(v - 1)^2$.

²¹ See the notation defined in Sect. 16.4.4.

We are left with the bridge. This has reduced the problem to Example 18.4.5. It follows that the result will be zero unless both ends of the bridge hold a single-vertex term in which case $\Sigma(e, k; 2, 1) = (\nu - 1)^3$. A similar result arises when the graph consists of a simple cycle attached at one vertex to an edge terminating at a vertex of valency one (see the graph (5, 1) of Table 16.2). The only difference here is that one cycle and one factor of $\nu - 1$ is missing. Again the result is zero unless both ends of the dangling edge contain single-vertex terms, in which case $\Sigma(e, k; 2, 1) = (\nu - 1)^2$.

18.5 Padé, Differential and Algebraic Approximants

Here we work with the function $F(w)$, with power series (14.1.1), and $H(w)$, with power series

$$H(w) = \sum_{j=0}^{\infty} h_j w^j, \quad (18.5.1)$$

for which, as for $F(w)$, the series truncated at w^ℓ is denoted by $H_\ell(w)$.

18.5.1 Padé Approximants

A Padé approximant is a rational function of a variable w constructed as an approximation to a function $F(w)$ so that the power series of $F(w)$ and the power series of the approximant agree up to a specified degree in w . First introduced in the eighteenth century, they were presented in a systematic way using *Padé tables* (see below) by the French mathematician Henri Padé (1863–1953) in his thesis of 1892.²²

Definition 18.5.1 The function²³

$$H(w) := [n/m] := \frac{P_n^{(m)}(w)}{Q_m^{(n)}(w)}, \quad (18.5.2)$$

where

$$P_n^{(m)}(w) := \sum_{j=0}^n p_j w^j, \quad Q_m^{(n)}(w) := \sum_{j=0}^m q_j w^j, \quad \text{with } q_0 = 1, \quad (18.5.3)$$

²² For an historical account of the development of Padé approximants see Brezinski (1996, pp. 311–313) and for a comprehensive account of the modern theory see Baker and Graves-Morris (1996).

²³ This is the Baker definition (Baker and Graves-Morris 1996, p. 21). The condition $q_0 \neq 0$ originally replaced $q_0 = 1$.

are coprime, is the n - m Padé approximant to the function $F(w)$ if²⁴

$$F(w) - [n/m] = O(w^{n+m+1}). \quad (18.5.4)$$

This definition does not, of course, imply the existence of $[n/m]$, for all n and m for a given $F(w)$. To explore this question we note that (18.5.4) and (18.5.2) are equivalent to

$$F(w)Q_m^{(n)}(w) - P_n^{(m)}(w) = O(w^{n+m+1}), \quad (18.5.5)$$

$$H(w)Q_m^{(n)}(w) - P_n^{(m)}(w) = 0, \quad (18.5.6)$$

which means that

$$F_\ell(w) = H_\ell(w), \quad (18.5.7)$$

$$\ell := n + m \quad (18.5.8)$$

and the coefficients in $P_n^{(m)}(w)$ and $Q_m^{(n)}(w)$ are obtained, if they exist, by solving the $n + m + 1$ linear equations

$$f_0 = p_0, \quad (18.5.9)$$

$$\sum_{i=0}^m f_{k-i} q_i = p_k, \quad k = 1, \dots, n, \quad (18.5.10)$$

$$\sum_{i=0}^m f_{k-i} q_i = 0, \quad k = n+1, \dots, n+m, \quad (18.5.11)$$

using the known coefficients in $F_\ell(w)$, with $f_k := 0$ if $k < 0$.

Initially we shall suppose that $f_0 \neq 0$:

Let

$$X(n/m; w) := \begin{pmatrix} 1 & w & w^2 & \cdots & w^m \\ f_{n+1} & f_n & f_{n-1} & \cdots & f_{n-m+1} \\ f_{n+2} & f_{n+1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n+m} & f_{n+m-1} & \cdots & \cdots & f_{n+m-m} \end{pmatrix}, \quad (18.5.12)$$

²⁴ Notation: $g(w) = O(w^k)$ if $g(w)/w^k$ tends to a finite limit as $w \rightarrow 0$. If $g(w) = O(w^k)$ then $g(w) = O(w^{k'})$ for $k' \leq k$.

and denote by $X_k(n/m)$ the matrix obtained by deleting the first row and $(k+1)$ th column of $X(n/m; w)$. Thus

$$X_0(n/m) := \begin{pmatrix} f_n & f_{n-1} & \cdots & f_{n-m+1} \\ f_{n+1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ f_{n+m-1} & \cdots & \cdots & f_n \end{pmatrix}. \quad (18.5.13)$$

From (18.5.12) and (18.5.13),²⁵

$$X_0(0/m) = (f_0)^m \neq 0, \quad (18.5.14)$$

$$X_m(n/m) = X_0(n+1/m), \quad (18.5.15)$$

$$X(n/m; w) = \sum_{j=0}^m (-w)^j X_j(n/m), \quad X_0(n/0) := 1. \quad (18.5.16)$$

Equations (18.5.9)–(18.5.11) can be expressed in the matrix form

$$\mathbf{p} = \mathbf{f}_n^{(0)} + X_0(-1/n+1, m) \mathbf{q}, \quad (18.5.17)$$

$$X_0(n/m) \mathbf{q} = -\mathbf{f}_{n+m}^{(n+1)}, \quad (18.5.18)$$

where

$$\mathbf{p} := \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \mathbf{q} := \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix}, \quad \mathbf{f}_j^{(k)} := \begin{pmatrix} f_k \\ f_{k+1} \\ \vdots \\ f_j \end{pmatrix}. \quad (18.5.19)$$

Thus it can be concluded that:

Theorem 18.5.1 *There exists a unique n - m Padé approximant to $F(w)$ if $X_0(n/m) \neq 0$.*

In that case, from (18.5.12), (18.5.16)–(18.5.18),

²⁵ Adopting the notation that for any square matrix $\mathbf{W}$, $W := \text{Det}\{\mathbf{W}\}$.

$$q_k = (-1)^k \frac{X_k(n/m)}{X_0(n/m)}, \quad k = 0, 1, \dots, m, \quad (18.5.20)$$

$$Q_m^{(n)}(w) = \frac{X(n/m; w)}{X_0(n/m)}, \quad (18.5.21)$$

$$p_k = \frac{1}{X_0(n/m)} \sum_{j=0}^m (-1)^j f_{k-j} X_j(n/m),$$

$$= \frac{1}{X_0(n/m)} \begin{vmatrix} f_k & f_{k-1} & f_{k-2} & \cdots & f_{k-m} \\ f_{n+1} & f_n & f_{n-1} & \cdots & f_{n-m+1} \\ f_{n+2} & f_{n+1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n+m} & f_{n+m-1} & \cdots & \cdots & f_n \end{vmatrix}, \quad k = 0, 1, \dots, n, \quad (18.5.22)$$

$$P_n^{(m)}(w) = \frac{1}{X_0(n/m)} \begin{vmatrix} F_n(w) & w F_{n-1}(w) & \cdots & w^m F_{n-m}(w) \\ f_{n+1} & f_n & \cdots & f_{n-m+1} \\ \vdots & \vdots & \cdots & \vdots \\ f_{n+m} & f_{n+m-1} & \cdots & f_n \end{vmatrix}. \quad (18.5.23)$$

Formulae (18.5.20)–(18.5.23) provide an ansatz for calculating Padé approximants when $X_0(n/m) \neq 0$. However, it is necessary to know, and straightforward to verify, that:

Theorem 18.5.2 *The polynomials $Q_m^{(n)}(w)$ and $P_n^{(m)}(w)$, calculated using (18.5.21) and (18.5.23) respectively are coprime.*

The array

$$\begin{array}{ccccccc} [0/0] & [0/1] & [0/2] & [0/3] & \cdots & & \\ [1/0] & [1/1] & [1/2] & [1/3] & \cdots & & \\ [2/0] & [2/1] & [2/2] & [2/3] & \cdots & & \\ [3/0] & [3/1] & [3/2] & [3/3] & \cdots & & \\ \vdots & \vdots & \vdots & \vdots & \cdots & & \end{array} \quad (18.5.24)$$

is called the *Padé table* of $F(w)$ and the X_0 -table is just the corresponding array of values of $X_0(n/m)$. This latter table can tell us a lot about the Padé table.

In particular, of course, a zero entry in the X_0 -table corresponds to a place in the Padé table where a *unique* approximant does not exist. In this respect the following result is important (Baker and Graves-Morris 1996, pp. 24–27):

Theorem 18.5.3 *Given that sufficient elements of the X_0 -table have been calculated and that $f_0 \neq 0$, then every zero entry, if such exists, lies either in a square block of zero entries bounded by non-zero entries or in the infinite segment bounded above and to the left by non-zero entries and extending to infinity below and to the right.*

Using this result for the X_0 -table we now consider the Padé table. First note, from (18.5.15), (18.5.20) and (18.5.22), that

$$\begin{aligned} p_0 &= f_0, & p_n &= \frac{X_0(n/m+1)}{X_0(n/m)}, \\ q_0 &= 1, & q_m &= (-1)^m \frac{X_0(n+1/m)}{X_0(n/m)}. \end{aligned} \tag{18.5.25}$$

When $p_n \neq 0$ and $q_m \neq 0$, $[n/m]$ is said to be of *full nominal degree* and thus a unique $[n/m]$ of full nominal degree exists if $X_0(n/m) \neq 0$, $X_0(n/m+1) \neq 0$ and $X_0(n+1/m) \neq 0$. With reference to the X_0 -table this will be true for all points (n, m) except those lying inside a square of zero values or on the left-hand and top edges of such a square. This leads to *Padé's principal theorem* (Baker and Graves-Morris 1996, pp. 27–28).

Theorem 18.5.4 *Given a $k \times k$ square of zero entries in the X_0 -table, surrounded by non-zero entries and with top left-hand corner $(n+1, m+1)$:*

- (i) $[n, m]$ exists and is of full nominal degree.
- (ii) $[n+r/m+s] = [n, m]$ for $r \geq 0$, $s \geq 0$ and $r+s \leq k$.
- (iii) $[n+r/m+s]$ does not exist for $1 \leq r \leq k$, $1 \leq s \leq k$, $r+s > k$.

(ii) and (iii) correspond, respectively, to the situations where there is an infinite number of solutions to the linear equations (18.5.10) and (18.5.11), and where (18.5.10) and (18.5.11) are inconsistent.

Corollary 18.5.1 *The X_0 -table contains an infinite segment as defined in Theorem 18.5.3 if and only if $F(w)$ is a rational function.*

Proof (of corollary) Consider the limit $k \rightarrow \infty$ for the square in Theorem 18.5.4. This gives an infinite segment of zero entries in the X_0 -table with top left-hand corner $(n+1, m+1)$, $[n/m]$ of full nominal degree and $[n+r/m+s] = [n/m]$, for all $r \geq 0$ and $s \geq 0$. From (18.5.4),

$$F(w) - [n/m] = O(w^{n+r+m+s+1}) \tag{18.5.26}$$

and taking the limit $(r+s) \rightarrow \infty$ it follows that $F(w)$ is the rational function $[n/m]$. Conversely if $F(w)$ is a rational function of coprime polynomials with numerator

and denominator degrees n and m respectively, then it is given exactly by its Padé approximant $[n/m]$ and in the Padé table $[n + r/m + s] = [n/m]$, for all $r \geq 0$ and $s \geq 0$. $\square$

A function $F(w)$ is said to be *normal* if its X_0 -table contains no zero entries.

Consider now the case $f_0 = \cdots f_{s-1} = 0$:

It is straightforward to show that:

Theorem 18.5.5 *If*

$$F(w) = Cw^s G(w), \quad \text{where} \quad G(w) := \sum_{j=0}^{\infty} g_j w^j \quad \text{and} \quad a_0 \neq 0, \quad (18.5.27)$$

then

$$[n/m]^{(F)} = Cw^s [n - s/m]^{(G)}, \quad n \geq s, \quad (18.5.28)$$

where Padé approximants for F and G are distinguished by superscripts.

18.5.2 Dlog Padé Approximants

Using the transformation (14.6.12), let

$$G(w) := \mathcal{DL}\{F(w)\}, \quad \text{where} \quad G(w) := \sum_{j=0}^{\infty} g_j w^j. \quad (18.5.29)$$

This transformation is effected by solving the linear equations

$$\sum_{j=0}^k f_{k-j} g_j = (k+1) f_{k+1}, \quad k = 0, 1, \dots \quad (18.5.30)$$

Suppose that all the coefficients of $F_\ell(w)$ are known. Then the first ℓ of these linear equations can be expressed in the form

$$X_0(0/\ell) \mathbf{g} = \mathbf{f}', \quad (18.5.31)$$

where

$$\mathbf{g} := \begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ g_{\ell-1} \end{pmatrix}, \quad \mathbf{f}' := \begin{pmatrix} f_1 \\ 2f_2 \\ \vdots \\ nf_\ell \end{pmatrix}. \quad (18.5.32)$$

Since $X_0(0/\ell) = f_0^\ell$ the coefficients of $G_{\ell-1}(w)$ are derivable from the coefficients of $F_\ell(w)$ by inverting the lower-triangular matrix $X_0(0/\ell)$.

Definition 18.5.2 The n - m Dlog Padé approximant for $F(w)$, denoted by $[n/m]_{\text{DL}}$, is the ‘ordinary’ n - m Padé approximant for $G(w)$.

The uniqueness of the Dlog Padé follows from Theorem 18.5.1. The method of calculation follows that given in Sect. 18.5.1 with $F_\ell(w)$ replaced by $G_{\ell-1}(w)$. Let

$$[n/m]_{\text{DL}} := \frac{\mathring{P}_n^{(m)}(w)}{\mathring{Q}_m^{(n)}(w)}, \quad (18.5.33)$$

where

$$\mathring{P}_n^{(m)}(w) := \sum_{j=0}^n \mathring{p}_j w^j, \quad \mathring{Q}_m^{(n)}(w) := \sum_{j=0}^m \mathring{q}_j w^j, \quad \text{with } \mathring{q}_0 = 1. \quad (18.5.34)$$

Then

$$F'(w) \mathring{Q}_m^{(n)}(w) - F(w) \mathring{P}_n^{(m)}(w) = O(w^{n+m+1}). \quad (18.5.35)$$

Given that $H(w)$ satisfies

$$H'(w) \mathring{Q}_m^{(n)}(w) - H(w) \mathring{P}_n^{(m)}(w) = 0, \quad (18.5.36)$$

(18.5.35) and (18.5.36) replace (18.5.5) and (18.5.6) for ordinary approximants, but now the recurrence relations derived from (18.5.35) for terms of degree w^ℓ , for $\ell = 0, 1, \dots, n+m$, contain f_ℓ , for $\ell = 0, 1, \dots, n+m+1$. So the solution $H(w)$ of the first-order differential equation (18.5.36), with $H(0) = F(0)$, is such that

$$F_\ell(w) = H_\ell(w), \quad (18.5.37)$$

$$\ell := n + m + 1. \quad (18.5.38)$$

18.5.3 Differential Approximants

The coefficients in the expansion of (18.5.36) satisfy the recurrence relationships

$$\sum_{i=0}^m \dot{q}_i h_{k-1} - \sum_{i=0}^n (k+1-i) \dot{p}_i h_{k+1-i} = 0, \quad k = 0, 1, \dots \quad (18.5.39)$$

Given that, apart from $\dot{q}_0 = 1$, the $n+m+1$ coefficients of $\dot{Q}_m^{(n)}(w)$ and $\dot{P}_n^{(m)}(w)$ are unknown they can be found using the identity (18.5.37). After which (18.5.39) can be used to successively determine further coefficients in the power series of $H(w)$. Here we have seen a change of interest from the Dlog Padé approximant $[n/m]_{\text{DL}}$ to $H(w)$, which analytically continues the known $F_\ell(w)$ and

- (a) has coefficients which satisfy the recurrence relations (18.5.39) for all k ,
- (b) satisfies the differential equation (18.5.36).

These represent the two strands in the development of the differential approximant method²⁶ by Joyce (1972) (see also Joyce and Guttman 1973; Guttman 1989).

Definition 18.5.3 The function $H(w)$ which satisfies

$$\sum_{i=0}^k A_i(w) \mathcal{D}^i \{H(w)\} = B(w), \quad \text{where } \mathcal{D} := w \frac{d}{dw}, \quad (18.5.40)$$

and

$$\begin{aligned} A_i(w) &:= \sum_{j=0}^{m_i} a_{ij} w^j, \quad i = 0, 1, \dots, k, \quad \text{with } a_{k0} := 1, \\ B(w) &:= \sum_{j=0}^n b_j w^j, \quad \text{with } B(w) \neq 0, \end{aligned} \quad (18.5.41)$$

is the k th order inhomogeneous differential approximant to $F(w)$, denoted by $[n/m_0, \dots, m_k]_{\text{DA}}$, if

$$\sum_{i=0}^k A_i(w) \mathcal{D}^i \{F(w)\} - B(w) = O(w^{\eta+1}), \quad (18.5.42)$$

$$\eta := k + n + \sum_{i=0}^k m_i. \quad (18.5.43)$$

²⁶ Initially called the ‘recurrence relation method’.

In the corresponding homogeneous case, $B(w) = 0$, (18.5.43) is replaced by

$$\eta := k - 1 + \sum_{i=0}^k m_i. \quad (18.5.44)$$

It can, therefore, be conveniently denoted by $[-1/m_0, \dots, m_k]_{\text{DA}}$.²⁷

The recurrence relations corresponding to (18.5.40) are, for the inhomogeneous case,

$$\sum_{i=0}^k \sum_{j=0}^{m_i} (s-j)^i a_{ij} h_{s-j} = \begin{cases} b_s, & 0 \leq s \leq n, \\ 0, & s > n, \end{cases} \quad (18.5.45)$$

and, for the homogeneous case,

$$\sum_{i=0}^k \sum_{j=0}^{m_i} (s-j)^i a_{ij} h_{s-j} = 0. \quad (18.5.46)$$

Those for $s = 0, 1, \dots, \eta$, needed to determine the $\eta + 1$ unknown parameters in (18.5.42), contain h_j , $j = 0, 1, \dots, \eta$. For the homogeneous case a_{00} is zero if $f_0 \neq 0$ and can be chosen arbitrarily if $f_0 = 0$.

The solution $H(w)$ of the k -order differential equation (18.5.40), with

$$\left(\frac{d^s H(w)}{dw^s} \right)_{w=0} := s! f_s, \quad s = 0, 1, \dots, k-1, \quad (18.5.47)$$

is the differential approximant $[n/m_0, \dots, m_k]_{\text{DA}}$ for $F(w)$. It can be obtained directly using one of the algebraic computing packages or by Frobenius' method. For the latter we need to determine the singularities of (18.5.40). First note that

$$\mathcal{D}^i = \sum_{r=1}^i \theta(i, r) w^r \frac{d^r}{dw^r}, \quad (18.5.48)$$

where the coefficients $\theta(i, r)$ are given by the recursion equations

$$\begin{aligned} \theta(i, 1) &= 1, & \theta(i, i) &= 1, \\ \theta(i, r) &= \theta(i-1, r-1) + r\theta(i-1, r), & r &= 2, \dots, i-1. \end{aligned} \quad (18.5.49)$$

Substituting into (18.5.40) gives

²⁷ $[0/m_0, \dots, m_k]_{\text{DA}}$ is the case $B(w) = b_0$.

$$w^k A_k(w) \frac{d^k H(w)}{dw^k} + \sum_{r=0}^{k-1} w^r \mathring{A}_r(w) \frac{d^r H(w)}{dw^r} = B(w), \quad (18.5.50)$$

where

$$\mathring{A}_r(w) := \sum_{i=r}^k A_i(w) \theta(i, r), \quad r = 1, \dots, k-1, \quad \mathring{A}_0(w) := A_0(w). \quad (18.5.51)$$

The roots of $A_k(w)$ (all non-zero), 0 and ∞ are the *singularities* of (18.5.40). If w' is a root of $w^k A_k(w)$ of degree τ then it is a *regular singularity* if $(w - w')^{k-r-\tau} w^r \mathring{A}_r(w)$ is regular at $w = w'$, for $r = 0, \dots, k-1$. This is certainly the case for $w = 0$ and ∞ , and it is not difficult to show that:

Theorem 18.5.6 $(w - w')^{k-r-\tau} \mathring{A}_r(w)$ is regular, for $r = 0, \dots, k-1$, at $w = w'$ if and only if $(w - w')^{k-r-\tau} A_r(w)$ is regular, for $r = 0, \dots, k-1$, at $w = w'$.

Corollary 18.5.2 A zero $w = w'$ of $A_k(w)$ of degree τ is a *regular singularity* of (18.5.40) if $A_r(w)$, for all $k > r > k - \tau$, has $w = w'$ as a zero of degree at least $r + \tau - k$.

The equation is said to be of *Fuchsian type* if all its singularities are regular and this can be determined from $A_r(w)$, $r = 0, \dots, k$, without the need to compute $\mathring{A}_r(w)$, $r = 0, \dots, k-1$.

Suppose that $w = w'$ is a root of $A_k(w)$ of degree one and not a root of $A_r(w)$, $r = 0, \dots, k-1$; then, from Corollary 18.5.2, it is a regular singularity. Let the solution in a neighbourhood of $w = w'$ be

$$H(w) \simeq (w - w')^{-\gamma} U(w) + V(w), \quad (18.5.52)$$

where $U(w)$ and $V(w)$ are regular in a neighbourhood of $w = w'$ with expansions

$$U(w) := \sum_{j=0}^{\infty} u_j (w - w')^j, \quad V(w) := \sum_{j=0}^{\infty} v_j (w - w')^j. \quad (18.5.53)$$

Then²⁸

$$\begin{aligned} \frac{d^r H(w)}{dw^r} &= \sum_{j=0}^{\infty} \frac{\Gamma(j-1-\gamma)}{\Gamma(j-r+1-\gamma)} u_j (w - w')^{j-r+1-\gamma} \\ &\quad + \sum_{j=0}^{\infty} \frac{\Gamma(j-1)}{\Gamma(j-r+1)} v_j (w - w')^{j-r+1}. \end{aligned} \quad (18.5.54)$$

²⁸ The Γ -functions take care of the zero terms since, for any integer n , if $1 < n \leq r-1$, $\frac{\Gamma(n-1)}{\Gamma(n-r+1)} = 0$. Also if $n \leq 1$, $\frac{\Gamma(n-1)}{\Gamma(n-r+1)} = \frac{\Gamma(1-n)}{\Gamma(r-n-1)}$.

Substituting into (18.5.50), using (18.5.49) and (18.5.51), and equating the coefficient of the term of order $(w - w')^{2-k-\gamma}$ to zero gives

$$\gamma = 1 - k + \frac{A_{k-1}(w')}{w' A'_k(w')}. \quad (18.5.55)$$

This formula gives a singularity except when γ is a negative integer. Then the first term in (18.5.52) can be absorbed into the second, resetting γ to zero and indicating that $H(w)$ is regular at $w = w'$.

A test function, given by Hunter and Baker (1979) and quoted by Guttman (1989) serves as an illustration.

Example 18.5.1 Suppose that

$$\begin{aligned} F_9(w) := & 2 + \frac{1}{2}w + \frac{19}{8}w^2 + \frac{97}{48}w^3 + \frac{961}{384}w^4 + \frac{10363}{3840}w^5 + \frac{135199}{46080}w^6 \\ & + \frac{2026897}{645120}w^7 + \frac{34459681}{10321920}w^8 + \frac{654728563}{185794560}w^9. \end{aligned} \quad (18.5.56)$$

Then the maximum second-order homogeneous approximant has $\eta = 1 + m_0 + m_1 + m_2 = 9$. Let $m_0 := 3$, $m_1 := 3$ and $m_2 := 2$. With $a_{20} := 1$ it is now straightforward to solve the system of ten linear equations yielding the differential approximant equation in the form

$$\begin{aligned} \left[\frac{2}{5}w^2 - \frac{7}{5}w + 1\right]\mathcal{D}^2\{H(w)\} + \left[\frac{2}{5}w^3 - \frac{6}{5}w^2 + \frac{3}{10}w - 1\right]\mathcal{D}\{H(w)\} \\ + \left[\frac{3}{5}w^3 - \frac{21}{10}w^2\right]H(w) = 0. \end{aligned} \quad (18.5.57)$$

The roots of $A_2(w)$ are $w = w_1 := 1$, $w = w_2 := \frac{5}{2}$ and, from (18.5.55), $\gamma_1 = \frac{3}{2}$, $\gamma_2 = -2$; $w = 1$ is a regular singularity and $w = \frac{5}{2}$ is a regular point. The solution to (18.5.57) with $H(0) := 2$ and $H'(0) := \frac{1}{2}$ is easily shown to be

$$H(w) = (1 - w)^{-\frac{3}{2}} + \exp(-w). \quad (18.5.58)$$

It should also be noted that (18.5.57) is equivalent to the quadratic recurrence relationship

$$\begin{aligned} (s + 3)(s + 2)h_{s+3} - \frac{1}{10}(s + 2)(14s + 25)h_{s+2} \\ + \frac{1}{10}(4s^2 - 4s - 29)h_{s+1} + \frac{1}{5}(2s + 3)h_s = 0 \end{aligned} \quad (18.5.59)$$

between the coefficients of the power series for $H(w)$. For $s = 1, \dots, 6$ this recurrence relationship is also satisfied by the coefficients of $F_9(w)$.

For a discussion of the case of confluent singularities, of biased approximants and of the derivation of the leading amplitude the reader is referred to Guttman (1989, pp. 89–94).

18.5.4 Algebraic Approximants

Differential approximants represent one way to extend Padé approximants. Another was proposed by Shafer (1974) who generalized (18.5.5) and (18.5.6) to

$$C_2(w)\{F(w)\}^2 + C_1(w)F(w) + C_0(w) = O(w^{n_0+n_1+n_2+2}), \quad (18.5.60)$$

$$C_2(w)\{H(w)\}^2 + C_1(w)H(w) + C_0(w) = 0, \quad (18.5.61)$$

where

$$C_i(w) := \sum_{j=0}^{n_i} c_{ij} w^j, \quad i = 0, 1, 2. \quad (18.5.62)$$

He also suggested the generalization to an s -degree algebraic equation, so that:

Definition 18.5.4 The function $H(w)$ which satisfies

$$\sum_{i=0}^s C_i(w)\{H(w)\}^i = 0, \quad (18.5.63)$$

where $C_i(w)$, given by extending the range of (18.5.62) to $i = 0, \dots, s$, are the *Hermite-Padé polynomials*, is the s th order algebraic approximant to $F(w)$, denoted by $[n_0, \dots, n_s]_{AA}$, if

$$\sum_{i=0}^s C_i(w)\{F(w)\}^i = O(w^{\eta+1}), \quad (18.5.64)$$

$$\eta := s - 1 + \sum_{i=0}^s n_i. \quad (18.5.65)$$

Substituting from (18.5.1) and (18.5.62) into (18.5.63) gives²⁹

$$c_{0j} + \omega_j = 0 \quad (18.5.66)$$

where

$$\omega_j := \sum_{i=1}^s \sum_{r=0}^{n_i} \sum_{k_1=0}^{\infty} \cdots \sum_{k_i=0}^{\infty} \delta^{\text{Kr}}(r + k_1 + \cdots + k_i - j) c_{ir} \prod_{m=1}^i h_{k_m}. \quad (18.5.67)$$

The $\eta + 1$ equations with $0 \leq j \leq \eta$ are satisfied by $h_k := f_k$, which can be solved to determine the $\eta + 1$ constants $\{c_{ij}\}$. The problem now is to solve the s th degree

²⁹ With $c_{0j} := 0$ for $j > n_0$.

equation for $H(w)$. This is straightforward for the quadratic case and possible, but complicated, for the cubic and quartic cases. The alternative is to derive recurrence relations giving the successive coefficients h_j for $j = \eta, \eta + 1, \dots$. Since $n_0 < \eta$, $\omega_j = 0$ for $j > \eta$. The largest h_{k_m} in ω_j is h_j and this will occur only in terms where all the other $h_{k_m} = f_0$. These terms are given by $h_j \tau$, where

$$\tau := \sum_{i=1}^s i c_{i0} f_0^{i-1}. \quad (18.5.68)$$

If all h_m are known for $m = 0, \dots, j-1$,

$$\theta_j := \omega_j - h_j \tau \quad (18.5.69)$$

is known, and thus

$$h_j = -\theta_j / \tau. \quad (18.5.70)$$

This formula was derived by Homeier (2011) who gave examples of error-checking, by determining the succession of coefficients h_j using fewer than the known number of f_j . An example is provided by the work of Joyce (1988b) on the hard-hexagon model (Sect. 14.6.4.2).

Example 18.5.2 Suppose that

$$\begin{aligned} F_{13}(w) := & -\frac{1}{2} + \frac{15}{4}w - \frac{15}{8}w^2 - \frac{85}{16}w^3 - \frac{15}{32}w^4 + \frac{3315}{64}w^5 + \frac{52185}{128}w^6 \\ & + \frac{709515}{256}w^7 + \frac{9543585}{512}w^8 + \frac{129548515}{1024}w^9 + \frac{1777942185}{2048}w^{10} \\ & + \frac{24646968315}{4096}w^{11} + \frac{344714623985}{8192}w^{12} + \frac{4859013288915}{6384}w^{13}. \end{aligned} \quad (18.5.71)$$

Then the maximum second-degree approximant has $\eta = 1 + m_0 + m_1 + m_2 = 13$. Let $m_0 := 4$, $m_1 := 4$ and $m_2 := 4$. With $c_{20} := 1$ it is now straightforward to solve the system of fourteen linear equations, yielding the algebraic approximant equation

$$\begin{aligned} (1 - \tfrac{3}{2}w)(1 - w)^3 [H(w)]^2 + \tfrac{1}{2}(1 - 12w + 45w^2 - 66w^3 + 33w^4)H(w) \\ + \tfrac{1}{2}w^3(1 - 3w) = 0. \end{aligned} \quad (18.5.72)$$

The singular points are given by the roots $w = 1$ and $w = \frac{2}{3}$ of $C_2(w)$ and the roots $w = \frac{1}{10}(5 \pm \sqrt{5})$ and $w = \frac{1}{6}(3 \pm \sqrt{5})$ of the discriminant (Sect. 17.6). The radius of the circle of convergence is $w = w_c := \frac{1}{10}(5 - \sqrt{5})$. Singular points of quadratic algebraic functions have cycle number two. By making the change of variable $x := \sqrt{w - w_c}$ and substituting the Taylor expansion for $H(x^2 + w_c)$ into (18.5.72), it can be shown that

$$H(w) \simeq H_0 + H_3(w - w_c)^{\frac{3}{2}}, \quad (18.5.73)$$

in a neighbourhood of w_c , where

$$H_0 := H(w_c) = -\frac{1}{2}(11 - 5\sqrt{5}), \quad H_3 := \sqrt{\frac{76875\sqrt{5} - 171875}{2}}. \quad (18.5.74)$$

Of course, in this quadratic case, direct solutions

$$H(w) = -\frac{(1 - 12w + 45w^2 - 66w^3 + 33w^4) \pm (1 - 5w + 5w^2)^{\frac{3}{2}}(1 - 9w + 9w^2)^{\frac{1}{2}}}{2(2 - 3w)(1 - w)^3}. \quad (18.5.75)$$

of (18.5.72) are easily obtained, together with the singularities and their exponents.

References and Author Index

A seminal resource for statistical mechanics is the series of twenty volumes *Phase Transitions and Critical Phenomena* published by Academic Press, London. The first six volumes¹ which appeared in the years 1972–1975 were edited by C. Domb and M.S. Green and, after the death of Professor Green, Volumes 7–20, which appeared in the years 1983–2001, were edited by C. Domb and J. Lebowitz. Many articles from these volumes are referenced below, where for brevity Volume n of this series is denoted by $\text{PTCP}(n)$.

At certain points in the text there is the need to use the definitions of particular mathematical quantities or the results of extensive manipulations. The canonical sources for these are Abramowitz and Segun (1965) and Gradshteyn and Ryzhik (1980). However, most are now easily available on the internet or by using computer algebra packages. So specific references are omitted.

Page numbers where the references are cited in the text are given in square brackets.

- Abraham, D. B., & Heilmann, O. J. (1980). Interacting dimers on the simple cubic lattice as a model for liquid crystals. *Journal of Physics A: Mathematical and General*, 13, 1051–1062. [253].
- Abramowitz, M., & Segun, I. A. (1965). *Handbook of mathematical functions*. New York, USA: Dover. [757].
- Adler, J., Brandt, A., Janke, W., & Shmulyan, S. (1995). Three-state Potts model on the triangular lattice. *Journal of Physics A: Mathematical and General*, 28, 5117–5129. [188].
- Adler, J., & Enting, I. G. (1984). The two-dimensional spin-1 Ising system and related models. *Journal of Physics A: Mathematical and General*, 17, 2233–2345. [555].
- Adler, J., Enting, I. G., & Privman, V. (1983). Two three-state Potts models: The universality of confluent corrections to scaling. *Journal of Physics A: Mathematical and General*, 16, 1967–1973. [551, 553].
- Affleck, I. (1986). Universal term in the free energy at a critical point and the conformal anomaly. *Physical Review Letters*, 56, 746–748. [162].

¹ Volume 4 was never published, but Volume 5 appeared in two parts 5a and 5b.

- Aharony, A. (1974). Two-scale-factor universality and the ε expansion. *Physical Review B*, 9, 2107–2109. [151].
- Akhiezer, N. I. (1990). *Elements of elliptic functions* (Vol. 79). Translations of Mathematical Monographs. Providence, Rhode Island: American Mathematical Society. [712, 714].
- Alexander, S. (1975). Lattice gas transitions of He on grafoil. A continuous transition with cubic terms. *Physics Letters A*, 54, 353–354. [490, 605, 615].
- Amit, D. J. (1978). *Field theory, the renormalization group and critical phenomena*. New York, USA: McGraw-Hill. [167, 520].
- Anderson, G. R., & Wheeler, J. C. (1978a). Directionality dependence of lattice models for solutions with closed-loop coexistence curves. *Journal of Chemical Physics*, 69, 2082–2088. [331].
- Anderson, G. R., & Wheeler, J. C. (1978b). Theory of lower critical solution points in aqueous mixtures. *Journal of Chemical Physics*, 69, 3403–3413. [332].
- Andrews, T. (1869). The Bakerian Lecture: On the continuity of the gaseous and liquid states of matter. *Philosophical Transactions of the Royal Society*, 159, 575–590. [171].
- Arisue, H., & Tabata, K. (1997). Low-temperature series for the square lattice Potts model by the improved finite-lattice method. *Journal of Physics A: Mathematical and General*, 30, 3313–3327. [518, 553].
- Arrott, A., & Noakes, J. E. (1967). Approximate equation of state for nickel near its critical temperature. *Physical Review Letters*, 19, 786–789. [149].
- Asano, T. (1968). Generalization of the Lee-Yang theorem. *Progress of Theoretical Physics*, 40, 1328–1336. [254].
- Asano, T. (1970). Theorems on the partition functions of the Heisenberg ferromagnets. *Journal of the Physical Society of Japan*, 29, 350–359. [254, 274].
- Ashcroft, N. W., & Mermin, N. D. (1976). *Solid state physics*. New York, USA: Holt-Saunders. [530].
- Ashkin, J., & Teller, E. (1943). Statistics of two-dimensional lattices with four components. *Physical Review*, 64, 178–184. [53].
- Au-Yang, H., & Fisher, M. E. (1975). Bounded and inhomogeneous Ising models. II Specific heat scaling function for a strip. *Physical Review B*, 11, 3469–3487. [156].
- Au-Yang, H., & Perk, J. H. H. (1995). The chiral Potts models revisited. *Journal of Statistical Physics*, 78, 17–78. [59].
- Babujian, H. M. (1982). Exact solution of the one-dimensional isotropic Heisenberg chain with arbitrary spin s . *Physics Letters*, 90A, 479–482. [32].
- Baker, G. A., Jr. (1961). Application of the Padé approximant method to the investigation of some magnetic properties of the Ising model. *Physical Review*, 124, 768–774. [519, 559].
- Baker, G. A., Jr. (2000). Defects and the convergence of Padé approximants. *Acta Applicandae Mathematica*, 61, 37–52. [561].
- Baker, G. A., Jr. (2005). Counter-examples to the Baker-Gammel-Wills conjecture and patchwork convergence. *Journal of Computational and Applied Mathematics*, 179, 1–14. [561].
- Baker, G. A., Jr., & Gaunt, D. S. (1967). Ising-model critical indices below the critical temperature. *Physical Review*, 155, 545–552. [338].
- Baker, G. A., Jr., & Graves-Morris, P. R. (1996). *Padé approximants* (2nd ed.). Cambridge, UK: C.U.P. [743, 747].
- Baker, G. A., Jr., & Hunter, D. L. (1973). Methods of series analysis. II Generalized and extended methods with application to the Ising model. *Physical Review B*, 7, 3377–3392. [556, 560].
- Barber, M. N. (1983). Finite-size scaling. In *PTCP(8)* (pp. 145–266). [153, 157, 158, 609].
- Barber, M. N., & Fisher, M. E. (1973a). Critical phenomena in systems of finite thickness. I The spherical model. *Annals of Physics*, 77, 1–78. [156].
- Barber, M. N., & Fisher, M. E. (1973b). Critical phenomena in systems of finite thickness. III Specific heat of an ideal boson film. *Physical Review A*, 8, 1124–1135. [156].
- Barbosa, M. A., & Henriques, V. B. (2008). Frustration and anomalous behaviour in the Bell-Lavis model of liquid water. *Physical Review E: Statistical, Nonlinear, and Soft Matter*, 77, 051204–1–051204-15. [59].

- Barker, J. A., & Fock, W. (1953). Theory of upper and lower critical solution temperatures. *Discussions of the Faraday Society*, 15, 188–195. [332].
- Bartis, J. T., & Hall, C. K. (1974). A lattice model of gas-gas equilibria in binary mixtures. *Physica*, 78, 1–21. [332].
- Batchelor, M. T. (2007). The Bethe ansatz after 75 years. *Physics Today*, 60(1), 36–40. [32].
- Baxter, R. J. (1972a). One-dimensional anisotropic Heisenberg chain. *Annals of Physics*, 70, 323–337. [33, 439].
- Baxter, R. J. (1972b). Partition function of the eight-vertex lattice model. *Annals of Physics*, 70, 193–228. [61, 257, 391, 434, 438, 439, 442, 450, 452, 456, 459, 476, 477, 478].
- Baxter, R. J. (1973a). Eight-vertex model in lattice statistics and one-dimensional anisotropic Heisenberg chain. I. Some fundamental eigenvectors. *Annals of Physics*, 76, 1–24. [257, 434, 442, 452, 731].
- Baxter, R. J. (1973b). Eight-vertex model in lattice statistics and one-dimensional anisotropic Heisenberg chain. II. Equivalence to a generalized ice-type lattice model. *Annals of Physics*, 76, 25–47. [257, 434, 442].
- Baxter, R. J. (1973c). Eight-vertex model in lattice statistics and one-dimensional anisotropic Heisenberg chain. III. Eigenvectors of the transfer matrix and Hamiltonian. *Annals of Physics*, 76, 48–71. [257, 434, 442].
- Baxter, R. J. (1973d). Potts model at the critical temperature. *Journal of Physics C: Solid State Physics*, 6, L445–L448. [79, 187, 489].
- Baxter, R. J. (1973e). Spontaneous staggered polarization of the F model. *Journal of Physics C: Solid State Physics*, 6, L94–L96. [490].
- Baxter, R. J. (1973f). Spontaneous staggered polarization of the F model. *Journal of Statistical Physics*, 9, 145–182. [490].
- Baxter, R. J. (1974). Ising model on a triangular lattice with three-spin interactions. II Free energy and correlation length. *Australian Journal of Physics*, 27, 369–381. [379].
- Baxter, R. J. (1978). Solvable eight-vertex model on an arbitrary lattice. *Philosophical Transactions of the Royal Society A*, 289, 315–346. [23, 60, 62, 63, 64, 257, 384, 385, 391, 393, 394, 395, 431, 634, 635].
- Baxter, R. J. (1980). Hard hexagons: Exact solution. *Journal of Physics A: Mathematical and General*, 13, L61–L70. [282].
- Baxter, R. J. (1981). Rogers-Ramanujan identities in the hard hexagon model. *Journal of Statistical Physics*, 26, 427–452. [282].
- Baxter, R. J. (1982a). Critical antiferromagnetic square lattice Potts model. *Proceedings of the Royal Society*, 383, 43–54. [80].
- Baxter, R. J. (1982b). *Exactly solved models in statistical mechanics*. London, UK: Academic Press. [33, 35, 61, 62, 149, 230, 282, 300, 385, 394, 395, 402, 407, 421, 434, 435, 450, 452, 477, 481, 490, 491, 514, 602, 726].
- Baxter, R. J. (1986). Free-fermion, checkerboard and Z-invariant lattice models in statistical mechanics. *Proceedings of the Royal Society of London. Series A*, 404, 1–33. [71, 72, 257, 395, 399, 402, 421, 635].
- Baxter, R. J. (1988). Free energy of the solvable chiral Potts model. *Journal of Statistical Physics*, 52, 639–667. [59].
- Baxter, R. J. (1990). Chiral Potts model: Eigenvalues of the transfer matrix. *Physics Letters A*, 146, 110–114. [59].
- Baxter, R. J. (2005a). Derivation of the order parameter of the chiral Potts model. *Physical Review Letters*, 94, 130602. [59].
- Baxter, R. J. (2005b). The order parameter of the chiral Potts model. *Journal of Statistical Physics*, 120, 1–36. [59].
- Baxter, R. J. (2006). The challenge of the chiral Potts model. *Journal of Physics: Conference Series*, 42, 11–24. [59].
- Baxter, R. J., Bazhanov, V. V., & Perk, J. H. H. (1990). Functional relations for transfer matrices of the chiral Potts model. *International Journal of Modern Physics B*, 4, 803–870. [59].

- Baxter, R. J., & Enting, I. G. (1978). 339th solution of the Ising model. *Journal of Physics A: Mathematical and General*, 11, 2463–2473. [256].
- Baxter, R. J., Enting, I. G., & Tsang, S. K. (1980). Hard-square lattice gas. *Journal of Statistical Physics*, 22, 465–489. [282].
- Baxter, R. J., Kelland, S. B., & Wu, F. Y. (1976). Equivalence of the Potts model or Whitney polynomial with an ice-type model. *Journal of Physics A: Mathematical and General*, 9, 397–406. [75, 78].
- Baxter, R. J., Perk, J. H. H., & A-Yang, H. (1988). New solutions of the star-triangle relations for the chiral Potts model. *Physics Letters A*, 128, 138–142. [59].
- Baxter, R. J., Temperley, H. N. V., & Ashley, S. E. (1978). Triangular Potts model at its transition temperature, and related problems. *Proceedings of the Royal Society A*, 358, 535–559. [187].
- Baxter, R. J., & Wu, F. Y. (1973). Exact solution of an Ising model with three-spin interactions on a triangular lattice. *Physical Review Letters*, 31, 1294–1303. [379, 561].
- Baxter, R. J., & Wu, F. Y. (1974). Ising model on a triangular lattice with three-spin interactions. I The eigenvalue problem. *Australian Journal of Physics*, 27, 357–367. [379].
- Bazhanov, V. V., & Mangazeev, V. V. (2007). Analytic theory of the eight-vertex model. *Nuclear Physics B*, 775, 225–282. [442, 725].
- Beardon, A. F. (1984). *A primer on Riemann surfaces*. LMS Lecture Note Series (Vol. 78). Cambridge, UK: C.U.P. [699].
- Bell, G. M. (1972). Statistical mechanics of water: Lattice model with directed bonding. *Journal of Physics C: Solid State Physics*, 5, 889–905. [80, 246, 247, 248, 251, 252, 254].
- Bell, G. M. (1974a). Ising ferrimagnetic models: I. *Journal of Physics C: Solid State Physics*, 7, 1174–1188. [90].
- Bell, G. M. (1974b). Ising ferrimagnetic models II. *Journal of Physics C: Solid State Physics*, 7, 1189–1205. [319, 322].
- Bell, G. M., & Lavis, D. A. (1970). Two-dimensional bonded lattice fluids. II orientable molecule model. *Journal of Physics A: Mathematical and General*, 3, 568–581. [59, 80, 246, 252, 254].
- Bell, G. M., Miggins, J., & Taylor, R. A. G. (1978). Second-order phase changes in phospholipid monolayer at the oil-water interface and related phenomena. *Journal of the Chemical Society, Faraday Transactions II*, 74, 223–234. [515].
- Bell, G. M., & Sallouta, H. (1975). An interstitial model for fluid water: Accurate and approximate results. *Molecular Physics*, 29, 1621–1637. [332, 337, 338, 339].
- Bell, G. M., & Salt, D. W. (1976). Three-dimensional lattice model for the water/ice system. *Journal of the Chemical Society, Faraday Transactions II*, 72, 76–86. [251].
- Berezinskii, V. L. (1971). Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. Classical systems. *Soviet Physics JETP*, 32, 493–500. [50].
- Berezinskii, V. L. (1972). Destruction of long-range order in one-dimensional and two-dimensional systems with a continuous symmetry group II. Quantum systems. *Soviet Physics JETP*, 34, 610–616. [50].
- Berge, C. (1962). *The theory of graphs*. London: Methuen. [619, 621].
- Berlin, T. H., & Kac, M. (1952). The spherical model of a ferromagnet. *Physical Review*, 86, 821–835. [34, 35].
- Bethe, H. A. (1931). Theorie der Metalle. Erster Teil. Eigenwerte und Eigenfunktionen der lineären atomischen Kette [English translation in Mattis (1993), pp. 689–716]. *Zeitschrift für Physik*, 71, 205–226. [32, 169, 229].
- Bethe, H. A. (1935). Statistical theory of superlattices. *Proceedings of the Royal Society A*, 150, 552–575. [169, 229].
- Betts, D. D. (1974). The XY model. In *PTCP(3)* (pp. 569–652). [560].
- Bhanot, G., Creutz, M., Glässer, U., Horvath, I., Lacki, J., Schilling, K., et al. (1993). Low-temperature expansions for Potts models. *Physical Review B*, 48, 6183–6191. [563].

- Bidaux, R., Carrara, P., & Vivet, B. (1967). Antiferromagnetisme dans un champ magnetique I. traitement de champ moleculaire. *Journal of Physics and Chemistry of Solids*, 28, 2453–2469. [226].
- Binder, K. E. (1983). Critical behaviour at surfaces. In *PTCP(8)* (pp. 1–144). [153].
- Binder, K. E. (1986). *Monte Carlo methods in statistical physics*. Topics in Current Physics (Vol. 7). Berlin: Springer. [616].
- Binder, K. E. (1987). Theory of first-order transitions. *Reports on Progress in Physics*, 50, 783–859. [175].
- Binder, K. E., & Landau, D. P. (1984). Finite-size scaling at first-order phase transitions. *Physical Review B*, 30, 1477–1485. [153].
- Binney, J. J., Dowrick, N. J., Fisher, A. J., & Newman, M. E. J. (1993). *The theory of critical phenomena*. Oxford, UK: O.U.P. [25, 520].
- Bishop, D. J., & Reppy, J. D. (1978). Study of the superfluid transition in two-dimensional ^4He films. *Physical Review Letters*, 40, 1727–1730. [49].
- Björnberg, J. E., & Grimmett, G. R. (2009). The phase transition of the quantum Ising model is sharp. *Journal of Statistical Physics*, 136, 231–273. [31].
- Black, J. L., & Emery, V. J. (1981). Critical properties of two-dimensional models. *Physical Review B*, 23, 429–432. [490].
- Bliss, G. A. (1966). *Algebraic functions*. New York, USA: Dover. [696].
- Blöte, H. W. J., Cardy, J. L., & Nightingale, M. P. (1986). Conformal invariance, the central charge and universal finite-size amplitudes at criticality. *Physical Review Letters*, 56, 742–745. [162].
- Blöte, H. W. J., & Nightingale, M. P. (1982). Critical behaviour of the two-dimensional Potts model with a continuous number of states; a finite-size scaling analysis. *Physica A*, 112, 405–465. [490, 613].
- Blöte, H. W. J., Nightingale, M. P., & Derrida, B. (1981). Critical exponents of two-dimensional Potts and bond percolation models. *Journal of Physics A: Mathematical and General*, 14, L45–L49. [613, 615].
- Blume, M., Emery, V. J., & Griffiths, R. B. (1971). Ising model for the λ transition and phase separation on $\text{He}^3\text{-He}^4$ mixtures. *Physical Review A*, 4, 1071–1077. [44, 92].
- Bostan, A., Chyzak, F., Lecerf, G., Salvy, B., & Schost, E. (2007). Differential equations for algebraic functions. In C. W. Brown (Ed.), *Proceedings of the 2007 international symposium on symbolic and algebraic computing* (pp. 25–32). ACM Press. [696].
- Boukraa, S., Guttman, A. J., Hassani, S., Jensen, I., Maillard, J.-M., Nickel, B., & Zenine, N. (2008). Experimental mathematics on the magnetic susceptibility of the square lattice Ising model. *Journal of Physics A: Mathematical and Theoretical*, 41, 455202 (51 pp). [563].
- Bragg, W. L., & Williams, E. (1934). The effect of thermal agitation on atomic arrangement in alloys. *Proceedings of the Royal Society A*, 145, 699–730. [44, 169].
- Brak, R., & Guttmann, A. J. (1990). Algebraic approximants: A new method of series analysis. *Journal of Physics A: Mathematical and General*, 23, L1331–L1337. [519, 564].
- Brascamp, H. J., & Kunz, H. (1974). Zeros of the partition function for the Ising model in the complex temperature plane. *Journal of Mathematical Physics*, 15, 65–66. [254].
- Bretz, M. (1977). Ordered helium films on highly uniform graphite—finite-size effects, critical parameters and the three-state Potts model. *Physical Review Letters*, 38, 501–505. [606].
- Brézin, E. (1982). An investigation of finite-size scaling. *Journal de Physique*, 43, 15–22. [609].
- Brézin, E. (1983). Finite-size scaling. *Annals of the New York Academy of Sciences*, 410, 339–349. [157].
- Brezinski, C. (1996). Extrapolation algorithms and Padé approximations: A historical survey. *Applied Numerical Mathematics*, 20, 299–318. [743].
- Brezinski, C. (2000). Convergence acceleration during the 20th century. *Journal of Computational and Applied Mathematics*, 122, 1–21. [558].
- Briggs, K. M., Enting, I. G., & Guttmann, A. J. (1994). Series studies of the Potts model. II: Bulk series for the square lattice. *Journal of Physics A: Mathematical and General*, 27, 1503–1523. [523, 553, 565, 566].

- Brush, S. G. (1967). History of the Lenz-Ising model. *Reviews of Modern Physics*, 39, 883–893. [37].
- Buckingham, M. J. (1972). Thermodynamics. In *PTCP(2)* (pp. 1–38). [3, 124, 133].
- Burkhardt, T. W. (1976a). Application of Kadanoff's lower-bound renormalization transformation to the Blume-Capel model. *Physical Review B*, 14, 1196–1201. [609].
- Burkhardt, T. W. (1976b). Kadanoff's lower-bound renormalization transformation. *Physical Review B*, 13, 3187–3191. [609].
- Burkhardt, T. W., & Eisenriegler, E. (1978). Renormalization group approach to surface critical behaviour in the Ising model. *Physical Review B*, 17, 318–323. [609].
- Burkhardt, T. W., & Knops, H. J. F. (1977). Renormalization group results for the Blume-Capel model in two and three dimensions. *Physical Review B*, 15, 1602–1605. [609].
- Burkhardt, T. W., Knops, H. J. F., & den Nijs, M. (1976). Renormalization group results for the three-state Potts model. *Journal of Physics A: Mathematical and General*, 9, L179–L181. [609].
- Burkhardt, T. W., & Swendsen, R. H. (1976). Critical temperatures of the spin- s Ising model. *Physical Review B*, 13, 3071–3073. [613].
- Burkhardt, T. W., & van Leeuwen, J. M. J. E. (1982). *Real-space renormalization*. Topics in Current Physics (Vol. 30). Berlin: Springer. [520, 615].
- Butera, P., & Comi, M. (1994). High-temperature study of the Kosterlitz-Thouless phase transition in the XY model on the triangular lattice. *Physical Review B*, 50, 3052–3057. [560].
- Butera, P., Comi, M., & Guttmann, A. J. (1993). Critical exponents of the three-dimensional classical plane rotator model on the SC lattice from a high temperature series analysis. *Physical Review B*, 48, 13987–13990. [49, 560].
- Butera, P., Comi, M., & Marchesini, G. (1989). Critical behaviour of the two-dimensional XY model: An analysis of extended high-temperature series. *Physical Review B*, 40, 534–545. [560].
- Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. *Journal of Philosophy*, 96, 348–373. [1].
- Camp, W. J., & Van Dyke, J. P. (1975). Isotropic magnets in two dimensions. *Journal of Physics C: Solid State Physics*, 8, 336–352. [560].
- Cardy, J. (2005). SLE for theoretical physicists. *Annals of Physics*, 318, 81–118. [163].
- Cardy, J. L. (1984). Conformal invariance and universality in finite-size scaling. *Journal of Physics A: Mathematical and General*, 17, L385–L387. [159, 162, 593].
- Cardy, J. L. (1987). Conformal invariance. In *PTCP(11)* (pp. 55–126). [3, 160, 162].
- Cardy, J. L. (1988). *Finite-size scaling*. Current Physics-Sources and Comments (Vol. 2). Amsterdam, New York: North-Holland. [153].
- Cardy, J. L. (1996). *Scaling and renormalization in statistical physics*. Cambridge, UK: C.U.P. [30, 34, 53, 490, 520].
- Cardy, J. L., Nauenberg, M., & Scalapino, D. J. (1980). Scaling theory of the Potts-model multicritical point. *Physical Review B*, 22, 2560–2568. [491].
- Chaikin, P. M., & Lubensky, T. C. (1995). *Principles of condensed matter physics*. Cambridge, UK: C.U.P. [49, 53, 593, 627].
- Chao, N.-C. (1982). Duality in three-dimensions. *Journal of Physics C: Solid State Physics*, 15, L1263–L1267. [631].
- Christe, P., & Henkel, M. (1993). *Introduction to conformal invariance and its applications to critical phenomena* (Vol. 16). Lecture Notes in Physics. Berlin: Springer. [3, 159].
- Coleman, S. (1973). There are no Goldstone bosons in two dimensions. *Communications in Mathematical Physics*, 31, 259–264. [31].
- Conway, J. H., Curtis, R. T., Norton, S. P., Parker, R. A., & Wilson, R. A. (1985). *Atlas of finite groups*. Oxford, UK: O.U.P. [681].
- Corless, R., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, 5, 329–359. [318].
- Cormier, O., Singer, M. F., Trager, B. M., & Ulmer, F. (2002). Linear differential operators for polynomial equations. *Journal of Symbolic Computation*, 34, 355–398. [696, 697].

- Costa-Santos, R. (2006). Geometrical aspects of the Z-invariant Ising model. *European Physical Journal B: Condensed Matter and Complex Systems*, 53, 85–90. [634, 635].
- Courant, R., & Hilbert, D. (1962). *Methods of Mathematical Physics*. New York, USA: Interscience. [706, 707].
- Darboux, G. (1878). Mémoire sur l'approximation des fonctions de très-grand nombres, et sur une classe étendue de développements en série. *Journal de Mathématiques Pures et Appliquées*, 3, 5–56. [556].
- de Gennes, P. G. (1972). Exponents for the excluded volume problem as derived by the Wilson method. *Physics Letters A*, 38, 339–340. [34].
- de Gennes, P. G., & Prost, J. (1993). *The physics of liquid crystals*. Oxford, UK: O.U.P. [2].
- de Neef, T. (1975). Some applications of series expansions in magnetism. Ph.D. Thesis, Technische Wetenschappen aan de Technische Hogeschool, Eindhoven. [518, 524, 543].
- de Neef, T., & Enting, I. G. (1977). Series expansions from the finite-lattice method. *Journal of Physics A: Mathematical and General*, 10, 801–805. [230, 518, 553].
- de Podesta, M. (2013). Redefining temperature. *Physics World*, 26(8), 28–32. [5].
- den Nijs, M. P. M. (1979). A relationship between the temperature exponents of the eight-vertex and q -state Potts models. *Journal of Physics A: Mathematical and General*, 12, 1857–1868. [490].
- des Cloizeaux, J., & Pearson, J. J. (1962). Spin-wave spectrum of the antiferromagnetic linear chain. *Physical Review*, 128, 2131–2135. [32].
- Devaney, R. L. (1989). *Chaotic dynamic systems*. California, USA: Addison-Wesley. [570].
- Dingle, R. B. (1973). *Asymptotic expansions: Their derivation and interpretation*. London: Academic Press. [556].
- Dobrushin, R. L. (1965). Existence of a phase transition in two-dimensional and three-dimensional Ising models. *Theory of Probability and Its Applications*, 10, 193–213. [253].
- Dobrushin, R. L. (1968a). Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and Its Applications*, 2, 292–301. [253, 261, 265].
- Dobrushin, R. L. (1968b). The problem of uniqueness of a Gibbsian random field and the problem of phase transitions. *Functional Analysis and Its Applications*, 2, 302–312. [253, 265].
- Domb, C. (1960). On the theory of cooperative phenomena in crystals. *Philosophical Magazine Supplement*, 9, 149–361. [301, 429, 579, 583, 584, 605].
- Domb, C. (1974a). Configurational studies of the Potts models. *Journal of Physics A: Mathematical and General*, 7, 1335–1348. [54, 55, 532, 741].
- Domb, C. (1974b). Graph theory and embeddings. In *PTCP(3)* (pp. 1–95). [532, 540, 646, 741].
- Domb, C. (1974c). Ising model. In *PTCP(3)* (pp. 357–484). [222, 223].
- Domb, C. (1976). Finite-cluster partition functions for the d -vector model. *Journal of Physics A: Mathematical and General*, 9, 983–998. [543].
- Domb, C., & Guttman, A. J. (1970). Low-temperature series for the Ising model. *Journal of Physics C: Solid State Physics*, 3, 1652–1660. [558].
- Domb, C., & Hunter, D. L. (1965). On the critical behaviour of ferromagnets. *Proceedings of the Physical Society*, 86, 1147–1151. [3, 315].
- Domb, C., & Sykes, M. F. (1956). On metastable approximations in cooperative assemblies. *Proceedings of the Physical Society A*, 235, 247–259. [518].
- Domb, C., & Sykes, M. F. (1957a). The calculation of lattice constants in crystal statistics. *Philosophical Magazine*, 2, 733–749. [518].
- Domb, C., & Sykes, M. F. (1957b). On the susceptibility of a ferromagnetic above the Curie point. *Proceedings of the Royal Society A*, 240, 214–228. [518].
- Duxbury, P. M., Yeomans, J., & Beale, P. D. (1984). Wavevector scaling and the phase diagram of the chiral clock model. *Journal of Physics A: Mathematical and General*, 17, L179–L184. [59].
- Ehrenfest, P. (1933). Phase changes in the ordinary and extended sense classified according to the corresponding singularity of the thermodynamic potential. *Proceedings of the National Academy of Sciences Amsterdam*, 36, 153–157. [2].
- Eisenberg, D., & Kauzmann, W. (1969). *The structure and properties of water*. Oxford, UK: Clarendon Press. [2, 61, 247].

- Ellis, R. S., Otto, P. T., & Touchette, H. (2005). Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model. *Annals of Applied Probability*, 15, 2203–2254. [44].
- English, P. S., Hunter, D. L., & Domb, C. (1979). Extension of the high-temperature free energy series for the classical vector model of ferromagnetism in general spin dimensionality. *Journal of Physics A: Mathematical and General*, 12, 2111–2130. [543].
- Enting, I. G. (1978a). Generalized Möbius functions for rectangles on the square lattice. *Journal of Physics A: Mathematical and General*, 11, 563–568. [648].
- Enting, I. G. (1978b). Some algebraic techniques for obtaining low-temperature series expansions. *Australian Journal of Physics*, 31, 515–522. [551].
- Enting, I. G. (1980a). Series analysis for the three-state Potts model. *Journal of Physics A: Mathematical and General*, 13, L133–L136. [553].
- Enting, I. G. (1980b). Universality in a generalised Potts model. *Journal of Physics A: Mathematical and General*, 13, L409–L412. [555].
- Enting, I. G. (1996). Series expansions from the finite lattice method. *Nuclear Physics B: Proceedings Supplements*, 47, 180–187. [518, 533].
- Enting, I. G., & Baxter, R. J. (1977). A special series expansion technique for the square lattice. *Journal of Physics A: Mathematical and General*, 10, L117–L119. [649].
- Enting, I. G., & Baxter, R. J. (1980). An investigation of the high-field series expansions for the square lattice Ising model. *Journal of Physics A: Mathematical and General*, 13, 3723–3734. [524, 565].
- Enting, I. G., Guttmann, A. J., & Jensen, I. (1994). Low-temperature series expansions for the spin-1 Ising model. *Journal of Physics A: Mathematical and General*, 27, 6987–7006. [555].
- Enting, I. G., & Wu, F. Y. (1982). Triangular lattice Potts models. *Journal of Statistical Physics*, 28, 351–373. [184, 555].
- Erez, A., & Meir, Y. (2010). Thermal phase transition in two-dimensional disordered superconductors. *Europhysics Letters*, 91(47003), 1–5. [49].
- Essam, J. W. (1972). Percolation and cluster size. In *PTCP(2)* (pp. 197–270). [318].
- Essam, J. W., & Fisher, M. E. (1963). Padé approximant studies of the lattice gas and Ising ferromagnet below the critical point. *Journal of Chemical Physics*, 38, 802–812. [133].
- Essam, J. W., & Garelick, H. (1967). Critical behaviour of a soluble model of dilute ferromagnetism. *Proceedings of the Physical Society*, 92, 136–149. [318].
- Essam, J. W., & Sykes, M. F. (1963). The crystal statistics for the diamond lattice. *Physica*, 29, 378–388. [338].
- Fabricsius, K. (2007). A new q -matrix in the eight-vertex model. *Journal of Physics A: Mathematical and Theoretical*, 40, 4075–4086. [442, 450, 456].
- Fabricsius, K., & McCoy, B. M. (2003). New developments in the eight-vertex model. *Journal of Statistical Physics*, 111, 323–327. [434, 442, 450, 469].
- Fabricsius, K., & McCoy, B. M. (2004). Functional equations and fusion matrices for the eight-vertex model. *Publication of RIMS Kyoto University*, 40, 905–932. [469].
- Fabricsius, K., & McCoy, B. M. (2005). New developments in the eight-vertex model II: Chains of odd length. *Journal of Statistical Physics*, 120, 37–70. [434, 442, 472].
- Fabricsius, K., & McCoy, B. M. (2007). The tq equation of the eight-vertex model for complex elliptic roots of unity. *Journal of Physics A: Mathematical and Theoretical*, 40, 14893–14926. [442, 450, 452, 456, 465].
- Fabricsius, K., & McCoy, B. M. (2009). New q matrices and their functional equations for the eight-vertex model at elliptic roots of unity. *Journal of Statistical Physics*, 134, 643–668. [442, 452].
- Falicov, L. M. (1966). *Group theory and its applications*. Chicago, USA: University of Chicago Press. [676, 677].
- Fan, C. (1972). On the critical properties of the Ashkin-Teller model. *Physics Letters A*, 39, 136–136. [53].
- Fan, C., & Wu, F. Y. (1970). General lattice statistical model of phase transitions. *Physical Review B*, 2, 723–733. [71, 492].

- Felderhof, B. U. (1973a). Direct diagonalization of the transfer matrix of the zero-field free-fermion model. *Physica*, 65, 421–451. [71].
- Felderhof, B. U. (1973b). Diagonalization of the transfer matrix of the free-fermion model II. *Physica*, 66, 279–297. [71].
- Feldmann, H., Guttman, A. J., Jensen, I., Schrock, R., & Tsai, S.-H. (1998). Study of the Potts model on the honeycomb and triangular lattices: Low-temperature series and partition function zeros. *Journal of Physics A: Mathematical and General*, 31, 2287–2310. [554].
- Ferdinand, A. E., & Fisher, M. E. (1969). Bounded and inhomogeneous Ising models. I Specific heat anomaly on a finite lattice. *Physical Review*, 185, 832–846. [156].
- Ferer, M., Moore, M. A., & Wortis, M. (1973). Critical indices and amplitudes of classical planar models in finite field for temperatures greater than t_c . *Physical Review*, 8, 5205–5212. [151].
- Ferer, M., & Velgakis, M. J. (1983). High-temperature critical behaviour of two-dimensional planar models: A series investigation. *Physical Review B*, 27, 314–325. [560].
- Fields, J. L. (1967). A uniform treatment of Darboux's method. *Archive for Rational Mechanics and Analysis*, 27, 289–305. [557].
- Fiore, C. E., Szortyka, M. M., Barbosa, M. C., & Henriques, V. B. (2009). Liquid polymorphism, order-disorder transitions and anomalous behaviour: A Monte Carlo study of the Bell-Lavis model for water. *Journal of Chemical Physics*, 131, 9–164506. [60].
- Fischer, P., Lebech, B., Meier, G., Rainford, B. D., & Vogt, O. (1978). Magnetic phase transitions of CeSb: I Zero applied magnetic field. *Journal of Physics C: Solid State Physics*, 11, 345–364. [2].
- Fisher, M. E. (1959). Transformations of Ising model. *Physical Review*, 113, 969–981. [312].
- Fisher, M. E. (1961). Statistical mechanics of dimers on a plane lattice. *Physical Review*, 124, 1664–1672. [257, 503, 504].
- Fisher, M. E. (1963). Perpendicular susceptibility of the Ising model. *Journal of Mathematical Physics*, 4, 124–135. [223].
- Fisher, M. E. (1965). The nature of critical points. In W. Brittin (Ed.), *Lectures in theoretical physics* (Vol. 7c, pp. 1–157). Boulder, USA: University of Colorado Press. [254, 374].
- Fisher, M. E. (1966). On the dimer solution of planar Ising models. *Journal of Mathematical Physics*, 7, 1776–1781. [86].
- Fisher, M. E. (1967). Theory of equilibrium critical phenomena. *Reports on Progress in Physics*, 30, 615–730. [2, 156].
- Fisher, M. E. (1968). Renormalization of critical exponents by hidden variables. *Physical Review*, 176, 257–272. [96, 318].
- Fisher, M. E. (1969). Rigorous inequalities for critical-point correlation exponents. *Physical Review*, 180, 594–600. [135].
- Fisher, M. E. (1971). The theory of critical point singularities. In M. Green (Ed.), *Critical Phenomena, Proceedings of the 51st Enrico Fermi Summer School, Varenna, Italy* (pp. 1–99). London, UK: Academic Press. [153, 156].
- Fisher, M. E. (1983). Scaling, universality and renormalization group theory. In F. Hahne (Ed.), *Critical phenomena* (pp. 1–139). Berlin: Springer. [29, 139, 147, 149].
- Fisher, M. E., & Au-Yang, H. (1979). Inhomogeneous differential approximants for power series. *Journal of Physics A*, 12, 1677–1692. [563].
- Fisher, M. E., & Barber, M. N. (1972). Scaling theory for finite-size effects in the critical region. *Physical Review Letters*, 28, 1516–1519. [153, 380].
- Fisher, M. E., & Berker, A. N. (1982). Scaling for first-order transitions in thermodynamic and finite systems. *Physical Review B*, 26, 2507–2513. [105, 134, 153].
- Fisher, M. E., & Ferdinand, A. E. (1967). Interfacial, boundary and size effects at critical points. *Physical Review Letters*, 19, 169–172. [153, 156, 380].
- Fisher, M. E., & Privman, V. (1985). First-order transitions breaking $O(n)$ symmetry: Finite-size scaling. *Physical Review B*, 32, 447–464. [153].
- Fisher, M. E., & Scesney, P. E. (1970). Visibility of critical-exponent renormalization. *Physical Review A*, 2, 825–835. [317].

- Fisher, M. E., & Sykes, M. F. (1962). Antiferromagnetic susceptibilities on the simple cubic and body-centred cubic Ising lattices. *Physica*, 28, 939–956. [222].
- Fleming, P. D., & Gibbs, J. H. (1974). An adaption of the lattice gas to the water problem. *Journal of Statistical Physics*, 10, 157–173 & 351–378. [252].
- Forder, H. G. (1949). The Euler-Maclauren formula. *The Mathematical Gazette*, 33, 172–176. [426].
- Fowler, D. (1972). The Riemann-Hugoniot catastrophe and van der Waals equation. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 1–7). Edinburgh: Edinburgh University Press. [659].
- Fox, P. F., & Guttman, A. J. (1973). Low-temperature critical behaviour of the Ising model with $s > \frac{1}{2}$. *Journal of Physics C: Solid State Physics*, 6, 913–930. [518].
- Friedan, D., Qiu, Z., & Shenker, S. (1984). Conformal invariance, unitarity and critical exponents in two dimensions. *Physical Review Letters*, 52, 1575–1578. [490].
- Frobenius, G. (1912). *Über Matrizen aus nicht negativen Elementen*. Berlin: Sitzungsberichte Preussische Akademie der Wissenschaft (pp. 456–477). [673].
- Gaite, J. (1992). Symmetry in catastrophe theory, phase diagrams and two-dimensional conformal field theories. *Journal of Physics A: Mathematical and General*, 25, 3051–3062. [185, 659].
- Gaite, J., Margalef-Roig, J., & Miret-Artés, S. (1998). Analysis of a three-component model phase diagram by catastrophe theory. *Physical Review B*, 57, 13527–13534. [659].
- Gammel, J. L. (1973). Review of two recent generalizations of the Padé approximant. In P. R. Graves-Morris (Ed.), *Padé approximants and their applications* (pp. 3–9). London: Academic Press. [563].
- Gantmacher, F. R. (1959). *Applications of the theory of matrices*. New York, USA: Interscience. [673, 674].
- Gaunt, D. S., & Fisher, M. E. (1965). Hard-sphere lattice gases. I Plane-square lattice. *Journal of Chemical Physics*, 43, 2840–2863. [282].
- Gaunt, D. S., & Guttman, A. J. (1965). Asymptotic analysis of coefficients. In *PTCP(3)* (pp. 181–243). [555].
- Georges, A., Hansel, D., Le Doussal, P., & Maillard, J. M. (1986). The vicinity of disorder varieties: A systematic expansion. *Journal of Physics A: Mathematical and General*, 19, 1001–1005. [326].
- Georgii, H.-O. (1988). *Gibbs measures and phase transitions*. Berlin, New York: Walter de Gruyter. [263].
- Gibbs, J. W. (1902). *Elementary principles in statistical mechanics*. Yale University Press. Reprinted by New York: Dover (1960). [14].
- Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York: Wiley. [659].
- Ginsparg, P. (1990). Applied conformal field theory. In E. Brezin & J. Zinn-Justin (Eds.), *Fields, strings and critical phenomena. Proceedings of 49th Summer School in Theoretical Physics*, Les Houches, 1988 (pp. 6–168). Amsterdam, New York: North-Holland. [692].
- Ginzburg, V. L. (1960). Some remarks on phase transitions of the second kind and the microscopic theory of ferroelectric materials. *Soviet Physics JETP*, 2, 1824–1834. [204].
- Ginzburg, V. L., & Landau, L. D. (1950). On the theory of superconductivity. In D. ter Haar (Ed.), *Collected papers of L.D. Landau* (pp. 546–568). Reprinted in (1965), Oxford, UK: Pergamon. [167].
- Glasser, M. L. (1969). Evaluation of the partition functions for some two-dimensional ferroelectric models. *Physical Review*, 184, 539–542. [488].
- Goldenfeld, N. (1992). *Lectures on phase transitions and the renormalization group*. Reading, Mass: Addison-Wesley. [520, 593].
- Goodwin, E. T. (1949). The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(x)e^{-x^2} dx$. *Proceedings of the Cambridge Philosophical Society*, 45, 241–245. [426].
- Gradshteyn, I. S., & Ryzhik, I. M. (1980). *Table of integrals, series, and products* (4th ed.). London, UK: Academic Press. [757].
- Green, H. S., & Hurst, C. A. (1964). *Order-disorder phenomena*. London: Interscience. [497].
- Green, M. S., Vicentini-Missoni, M., & Levelt-Sengers, J. M. H. (1967). Scaling-law equation of state for gases in the critical region. *Physical Review Letters*, 18, 1113–1117. [149].

- Grest, G. S. (1981). Monte Carlo study of the antiferromagnetic Potts model on frustrated lattices. *Journal of Physics A: Mathematical and General*, 14, L217–L221. [158].
- Griffiths, R. B. (1964). Peierls proof of spontaneous magnetization in a two-dimensional Ising model. *Physical Review A*, 136, 437–439. [253].
- Griffiths, R. B. (1965). Ferromagnets and simple fluids near the critical point: Some thermodynamic inequalities. *Journal of Chemical Physics*, 43, 1958–1968. [125].
- Griffiths, R. B. (1970). Thermodynamics near the two-fluid critical mixing point in $\text{He}^3\text{-He}^4$. *Physical Review Letters*, 24, 715–717. [94].
- Griffiths, R. B. (1972). Rigorous results and theorems. In *PTCP(1)* (pp. 7–109), UK. [260].
- Griffiths, R. B. (1981). Mathematical properties of renormalization group transformations. *Physica A*, 106, 59–69. [573].
- Griffiths, R. B., & Pearce, P. A. (1978). Position-space renormalization group transformations: Some proofs and some problems. *Physical Review Letters*, 41, 917–920. [573].
- Griffiths, R. B., & Pearce, P. A. (1979). Mathematical properties of position-space renormalization group transformations. *Journal of Statistical Physics*, 20, 499–545. [573].
- Griffiths, R. B., & Wheeler, J. C. (1970). Critical points in multicomponent systems. *Physical Review A*, 2, 1047–1064. [136].
- Guggenheim, E. A. (1935). The statistical mechanics of regular solutions. *Proceedings of the Royal Society A*, 148, 304–312. [229, 654].
- Guggenheim, E. A., & McGlashan, M. L. (1951). Statistical mechanics of regular mixtures. *Proceedings of the Royal Society A*, 206, 335–353. [229, 230, 248, 654].
- Gunton, J. D., San Miguel, M., & Sahai, P. S. (1983). Dynamics of first-order transitions. In *PTCP(8)* (pp. 269–466). [175].
- Guttmann, A. J. (1969). Determination of critical behaviour in lattice statistics from series expansions III. *Journal of Physics C: Solid State*, 2, 1900–1907. [558].
- Guttmann, A. J. (1975). On the recurrence relation method of series analysis. *Journal of Physics A: Mathematical and General*, 8, 1081–1088. [556].
- Guttmann, A. J. (1978). The critical behaviour of two-dimensional isotropic spin systems. *Journal of Physics A: Mathematical and General*, 11, 545–553. [560].
- Guttmann, A. J. (1989). Asymptotic analysis of power-series expansions. In *PTCP(13)* (pp. 1–234). [518, 519, 522, 555, 558, 562, 750, 753].
- Guttmann, A. J., & Enting, I. G. (1988). The number of convex polygons on the square and honeycomb lattices. *Journal of Physics A*, 21, L467–L474. [562].
- Guttmann, A. J., & Enting, I. G. (1993). Series studies of the Potts model: I The simple cubic Ising model. *Journal of Physics A: Mathematical and General*, 26, 807–821. [555, 563].
- Guttmann, A. J., & Enting, I. G. (1994). Series studies of the Potts model: III The 3-state model on the simple cubic lattice. *Journal of Physics A: Mathematical and General*, 27, 5801–5812. [555, 566].
- Guttmann, A. J., & Joyce, G. S. (1972). On a new method of series analysis in lattice statistics. *Journal of Physics A: Mathematical and General*, 5, L81–L84. [519, 559].
- Hancock, H. (1958). *Lectures on the theory of elliptic functions*. New York: Dover. [714].
- Hankey, A., & Stanley, H. E. (1972). Systematic application of generalized homogeneous functions to static scaling, dynamic scaling and universality. *Physical Review B*, 6, 3515–3542. [3, 96, 99, 100].
- Hankey, A., Stanley, H. E., & Chang, T. S. (1972). Geometric predictions of scaling at tricritical points. *Physical Review Letters*, 29, 278–281. [3, 142].
- Heilmann, O. J. (1972). Existence of phase transitions in certain lattice gases with repulsive potential. *Lettere Al Nuovo Cimento*, 3, 95–98. [253].
- Heilmann, O. J. (1980). Existence of an ordered phase for the repulsive lattice gas on the fcc lattice. *Journal of Physics A: Mathematical and General*, 13, 1803–1810. [253].
- Heilmann, O. J., & Huckaby, D. A. (1979). Phase transitions in lattice gas models for water. *Journal of Statistical Physics*, 20, 371–383. [254, 264, 265].

- Heilmann, O. J., & Lieb, E. H. (1972). Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25, 190–232. [254].
- Hemmer, P. C., & Stell, G. (1970). Fluids with several phase transitions. *Physical Review Letters*, 24, 1284–1287. [43].
- Henkel, M. (1999). *Conformal invariance and critical phenomena*. Berlin: Springer. [159, 162].
- Hijmans, J., & de Boer, J. (1955a). An approximation method for order-disorder problems I. *Physica*, 21, 471–484. [230, 234, 235, 236].
- Hijmans, J., & de Boer, J. (1955b). An approximation method for order-disorder problems II. *Physica*, 21, 485–498. [230].
- Hijmans, J., & de Boer, J. (1955c). An approximation method for order-disorder problems III. *Physica*, 21, 499–516. [230, 241].
- Hilhorst, H. J., Schick, M., & van Leeuwen, J. M. J. (1978). Differential form of real-space renormalization: Exact results for two-dimensional Ising models. *Physical Review Letters*, 40, 1605–1608. [588].
- Hilhorst, H. J., Schick, M., & van Leeuwen, J. M. J. (1979). Exact renormalization group equations for the two-dimensional Ising model. *Physical Review B*, 19, 2749–2763. [588].
- Hille, E. (1959). *Analytic function theory* (Vol. 1). New York, USA: Chelsea. [556].
- Hille, E. (1962). *Analytic function theory* (Vol. 2). New York, USA: Chelsea. [696, 697].
- Hindermann, A., Kunz, H., & Wu, F. Y. (1978). Exact results for the Potts model in two dimensions. *Journal of Statistical Physics*, 19, 623–632. [306].
- Ho, J. T., & Litster, J. D. (1969). Magnetic equation of state of CrBr_3 near the critical point. *Physical Review Letters*, 22, 603–606. [149].
- Hohenberg, P. C. (1967). Existence of long-range order in one and two dimensions. *Physical Review*, 158, 383–386. [31].
- Homeier, H. H. H. (2011). Series prediction based on algebraic approximants. *ISRN Applied Mathematics*, 2011, 7 (Article ID 958968). [755].
- Houlrik, J. M., & Knak Jensen, S. J. (1986). Phase diagram of the three-state chiral clock model studied by Monte Carlo renormalization-group calculations. *Physical Review B*, 34, 325–329. [59].
- Houtappel, R. M. F. (1950). Order-disorder in hexagonal lattices. *Physica*, 16, 425–455. [257, 318].
- Hsu, S.-C., & Gunton, J. D. (1977). Study of the cumulant expansion for renormalization group transformations on two- and three-dimensional Ising models. *Physical Review B*, 15, 2688–2693. [599].
- Hu, Z.-N., & Hou, B.-Y. (1996). Three-dimensional vertex model in statistical mechanics from the Baxter-Bazhanov model. *Journal of Statistical Physics*, 82, 633–655. [60].
- Huang, H. Y., Popkov, V., & Wu, F. Y. (1997). Exact solution of a three-dimensional dimer system. *Physical Review Letters*, 78, 409–412. [256].
- Hulthén, L. (1938). Über das austauschproblem eines kristalles. *Arkiv för Matematik, Astronomi och Fysik*, 26A, 1–106. [32].
- Hunter, D. L., & Baker, G. A., Jr. (1973). Method of series analysis. I. Comparison of current methods used in the theory of critical phenomena. *Physical Review B*, 7, 3346–3376. [556].
- Hunter, D. L., & Baker, G. A., Jr. (1979). Methods of series analysis, III Integral approximant methods. *Physical Review B*, 19, 3808–3821. [563, 597, 753].
- Huse, D. A. (1981). Simple three-state model with infinitely many phases. *Physical Review B*, 24, 5180–5194. [59].
- Husimi, K., & Syôzi, I. (1950). The statistics of honeycomb and triangular lattice I. *Progress of Theoretical Physics*, 5, 177–186. [257].
- Iooss, G., & Joseph, D. D. (1989). *Elementary stability and bifurcation theory*. Berlin: Springer. [662].
- Isakov, S. N. (1984). Nonanalytic features of the first order phase transition in the Ising model. *Communications in Mathematical Physics*, 95, 427–443. [565].
- Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31, 253–258. [37, 256].

- Itzykson, C. G. (1988). *Conformal invariance and applications to statistical mechanics*. Singapore: World Scientific. [159].
- Itzykson, C. G., Pearson, R. B., & Zuber, J. B. (1983). Distribution of zeros in Ising and gauge models. *Nuclear Physics B*, 220, 415–433. [379, 380].
- Izyumov, Y. A., & Skryabin, Y. N. (1988). *Statistical mechanics of magnetically ordered systems*. New York: Consultants Bureau. [32].
- Jensen, I., & Guttmann, A. J. (1996). Extrapolation procedure for the low-temperature series for the square lattice spin-1 Ising model. *Journal of Physics A: Mathematical and General*, 29, 3817–3836. [518].
- Jensen, I., Guttmann, A. J., & Enting, I. G. (1996). Low-temperature series expansions for the square lattice Ising model with $s > 1$. *Journal of Physics A: Mathematical and General*, 29, 3805–3815. [555].
- Jensen, I., Guttmann, A. J., & Enting, I. G. (1997). The Potts model on kagomé and honeycomb lattices. *Journal of Physics A: Mathematical and General*, 30, 8067–8083. [554].
- Jimbo, M., & Miwa, T. (1995). *Algebraic analysis of solvable lattice models*. Providence, Rhode Island: AMS for CBMS. [32].
- Jones, G. L. (1966). Complex temperatures and phase transitions. *Journal of Mathematical Physics*, 7, 2000–2005. [254].
- José, J. V., Kadanoff, L. P., Kirkpatrick, S., & Nelson, D. R. (1977). Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model. *Physical Review B*, 16, 1217–1241. [53].
- Josephson, B. D. (1967). Inequality for the specific heat. *Proceedings of the Physical Society*, 92, 269–275 & 276–284. [135].
- Joyce, G. S. (1966). Spherical model with long-range ferromagnetic interactions. *Physical Review*, 146, 349–358. [35].
- Joyce, G. S. (1967). Classical Heisenberg model. *Physical Review*, 155, 478–495. [35, 543].
- Joyce, G. S. (1969). Absence of ferromagnetism or antiferromagnetism in the isotropic Heisenberg model with long-range interactions. *Journal of Physics C: Solid State Physics*, 2, 1531–1533. [36].
- Joyce, G. S. (1972). Critical properties of the spherical model. In *PTCP(2)* (pp. 375–442). [750].
- Joyce, G. S. (1975a). Analytic properties of the Ising model with triplet interactions on the triangular lattice. *Proceedings of the Royal Society A*, 343, 45–62. [379, 561].
- Joyce, G. S. (1975b). On the magnetization of the triangular lattice Ising model with triplet interactions. *Proceedings of the Royal Society A*, 345, 277–293. [379, 561].
- Joyce, G. S. (1988a). Exact results for the activity and isothermal compressibility of the hard-hexagon model. *Journal of Physics A: Mathematical and General*, 21, L983–L988. [282].
- Joyce, G. S. (1988b). On the hard-hexagon model and the theory of modular functions. *Philosophical Transactions of the Royal Society A*, 325, 643–702. [282, 564, 698, 775].
- Joyce, G. S., & Bowers, R. G. (1966). Cluster series for the infinite spin Heisenberg model. *Proceedings of the Physical Society*, 88, 1053–1055. [43].
- Joyce, G. S., & Guttmann, A. J. (1973). A new method of series analysis. In P. R. Graves-Morris (Ed.), *Padé approximants and their applications* (pp. 163–167). London, UK: Academic Press. [562, 750].
- Jüngling, K. (1975). Exact solution of a nonplanar two-dimensional Ising model with short range two-spin interaction. *Journal of Physics C: Solid State Physics*, 8, L169–L171. [70, 71].
- Jüngling, K., & Obermair, G. (1974). Note on universality and the eight-vertex model. *Journal of Physics C: Solid State Physics*, 7, L363–L365. [70].
- Kac, M. (1964). The work of T. H. Berlin in statistical mechanics. a personal reminiscence. *Physics Today*, 17(10), 40–42. [45].
- Kac, M., & Thompson, C. S. (1971). Spherical model and the infinite spin dimensionality limit. *Physica Norvegica*, 5, 163–168. [35].
- Kadanoff, L. P. (1966). Scaling laws for Ising models near T_c . *Physics*, 2, 263–272. [96, 100, 182, 520].

- Kadanoff, L. P. (1969a). Correlations along a line in the two-dimensional Ising model. *Physical Review*, 188, 859–863. [120].
- Kadanoff, L. P. (1969b). Operator algebra and the determination of critical indices. *Physical Review Letters*, 23, 1430–1433. [119, 120].
- Kadanoff, L. P. (1975). Variational principles and approximate renormalization group calculations. *Physical Review Letters*, 34, 1005–1008. [3, 600, 607, 609].
- Kadanoff, L. P. (1976a). Notes on Migdal's recursion formulas. *Annals of Physics*, 100, 359–394. [599, 600, 601].
- Kadanoff, L. P. (1976b). Scaling, universality and operator algebras. In *PTCP(5a)* (pp. 1–34). [95].
- Kadanoff, L. P. (2006). Theories of matter: infinities and renormalization. See under Papers about Science. <http://jfi.uchicago.edu/leop/>. [520].
- Kadanoff, L. P. (2009). More is the same; phase transitions and mean field theories. *Journal of Statistical Physics*, 137, 777–797. [167].
- Kadanoff, L. P. (2013). Relating theories via renormalization. *Studies in History and Philosophy of Modern Physics*, 44, 22–39. [520].
- Kadanoff, L. P., Götze, W., Hamblen, D., Hecht, R., Lewis, E. A. S., Palciauskas, V. V., et al. (1967). Static phenomena near critical points: Theory and experiments. *Reviews of Modern Physics*, 19, 395–431. [96].
- Kadanoff, L. P., & Houghton, A. (1975). Numerical evaluations of the critical properties of the two-dimensional Ising model. *Physical Review B*, 11, 377–386. [572, 573, 607].
- Kadanoff, L. P., Houghton, A., & Yalabik, M. C. (1976). Variational approximations for renormalization group transformations. *Journal of Statistical Physics*, 14, 171–203. [600, 607, 609].
- Kadanoff, L. P., & Wegner, F. J. (1971). Some critical properties of the eight-vertex model. *Physical Review B*, 4, 3989–3993. [54, 67, 120, 482, 491].
- Kager, W., Nienhuis, B., & Kadanoff, L. P. (2004). Exact solutions for Loewner evolutions. *Journal of Statistical Physics*, 115, 805–821. [165].
- Kasteleyn, P. W. (1961). The statistics of dimers on a lattice I. The number of dimer arrangements on a quadratic lattice. *Physica*, 27, 1209–1225. [257, 496, 497, 504].
- Kasteleyn, P. W. (1963). Dimer statistics and phase transitions. *Journal of Mathematical Physics*, 4, 287–293. [508, 510].
- Kasteleyn, P. W. (1967). Graph theory and crystal statistics. In F. Harary (Ed.), *Graph theory and theoretical physics* (pp. 43–110). London, UK: Academic Press. [497].
- Kasteleyn, P. W., & Fortuin, C. M. (1969). Phase transitions in lattice systems with random local properties. *Journal of the Physical Society of Japan*, 26, 11–14. [56].
- Kasteleyn, P. W., & van Kranendonk, J. (1956a). Constant-coupling approximation for Heisenberg ferromagnetism. *Physica*, 22, 317–337. [229].
- Kasteleyn, P. W., & van Kranendonk, J. (1956b). Constant-coupling approximation for antiferromagnetism. *Physica*, 22, 367–385. [229].
- Kasteleyn, P. W., & van Kranendonk, J. (1956c). Constant-coupling approximation for Ising spin systems. *Physica*, 22, 387–396. [229].
- Katsura, S. (1962). Statistical mechanics of the anisotropic linear Heisenberg model. *Physical Review*, 127, 1508–1518. [33].
- Kaufman, B. (1949). Crystal statistics. II Partition function evaluated by spinor analysis. *Physical Review*, 76, 1232–1243. [257].
- Kaufman, B., & Onsager, L. (1949). Crystal statistics. III Short-range order in a binary Ising lattice. *Physical Review*, 76, 1244–1252. [257].
- Kaufman, M., & Griffiths, R. B. (1981). Exactly soluble Ising models on hierarchical lattices. *Physical Review B*, 24, 496–498. [588].
- Kenna, R. (1998). Fisher zeros and singular behaviour of the two-dimensional Potts model in the thermodynamic limit. *Journal of Physics A: Mathematical and General*, 31, 9419–9427. [255].
- Kennedy, T. (1985). Long range order in the anisotropic quantum ferromagnetic Heisenberg model. *Communications in Mathematical Physics*, 100, 447–462. [31].

- Kenyon, R., & Schlenker, J.-M. (2005). Rhombic embeddings of planar graphs with faces of degree 4. *Transactions of the American Mathematical Society*, 357, 3443–3458. [636].
- Khinchin, A. I. (1949). *Mathematical foundations of statistical mechanics*. New York: Dover. [14].
- Kikuchi, R. (1951). A theory of cooperative phenomena. *Physical Review*, 81, 988–1003. [229, 236, 241].
- Kim, D., & Joseph, R. I. (1975). High-temperature series study of the q -component Potts model in two and three dimensions. *Journal of Physics A: Mathematical and General*, 8, 891–904. [518, 543, 553].
- Kincaid, J. M., & Cohen, E. G. D. (1975). Phase diagrams of liquid helium mixtures and metamagnets: Experiment and mean field theory. *Physics Reports*, 22, 57–143. [90, 93, 225].
- Kinzel, W., & Schick, M. (1981). Phenomenological scaling approach to the triangular Ising anti-ferromagnet. *Physical Review B*, 23, 3435–3441. [611, 614, 615].
- Knops, H. J. F. (1977). Exact relation between the solid-on-solid model and the xy model. *Physical Review Letters*, 39, 766–769. [53, 609].
- Kosterlitz, J. M. (1974). The critical properties of the two-dimensional xy model. *Journal of Physics C: Solid State Physics*, 7, 1046–1060. [588, 591, 594].
- Kosterlitz, J. M., & Thouless, D. J. (1972). Long range order and metastability in two dimensional solids and superfluids. *Journal of Physics C: Solid State Physics*, 5, L124–L126. [50].
- Kosterlitz, J. M., & Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6, 1181–1203. [50, 52, 560].
- Kouvel, J. S., & Comly, J. B. (1968). Magnetic equation of state for nickel near its Curie point. *Physical Review Letters*, 20, 1237–1239. [149].
- Kouvel, J. S., & Rodbell, D. S. (1967). Magnetic equation of state for CrO_2 and nickel near their Curie points. *Physical Review Letters*, 18, 215–218. [149].
- Kramers, H. A., & Wannier, G. H. (1941a). Statistics of the two-dimensional ferromagnet Part I. *Physical Review*, 60, 252–262. [256, 283].
- Kramers, H. A., & Wannier, G. H. (1941b). Statistics of the two-dimensional ferromagnet Part II. *Physical Review*, 60, 263–276. [256].
- Kurmann, J., Thomas, H., & Müller, G. (1982). Antiferromagnetic long-range order in the anisotropic quantum spin chain. *Physica A*, 112, 235–255. [33].
- Lancaster, P., & Tismenetsky, M. (1985). *The theory of matrices* (2nd ed.). London: Academic Press. [670, 671, 672, 675].
- Landau, L. D. (1937). On the theory of phase transitions. In D. ter Haar (Ed.), *Collected papers of L.D. Landau* (pp. 193–216). Reprinted in (1965). Oxford, UK: Pergamon. [94, 176].
- Lanford, O. E., & Ruelle, D. (1969). Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics*, 13, 194–215. [261].
- Lavis, D. A. (1973). The steam-water-ice system: A two-dimensional bonded lattice model. The first-order approximation. *Journal of Physics C: Solid State Physics*, 6, 45–1530. [59].
- Lavis, D. A. (1975). Symmetry properties and the two-phase coexistence curve of a two-dimensional bonded lattice fluid. *Journal of Physics A: Mathematical and General*, 8, 51–1933. [59].
- Lavis, D. A. (1976). An exact matrix calculation for a two-dimensional model of the steam-water-ice system. Bulk and boundary properties. *Journal of Physics A: Mathematical and General*, 9, 2077–2095. [59, 153, 256, 427].
- Lavis, D. A. (1996). The derivation of the free energy of the Ising model from that of the eight-vertex model. *Reports on Mathematical Physics*, 37, 147–155. [284, 380, 484].
- Lavis, D. A. (2005). Boltzmann and Gibbs: An attempted reconciliation. *Studies in History and Philosophy of Modern Physics*, 36, 245–273. [1].
- Lavis, D. A. (2008). Boltzmann, Gibbs and the concept of equilibrium. *Philosophy of Science*, 75, 682–696. [1].
- Lavis, D. A. (2011). An objectivist account of probabilities in statistical mechanics. In C. Beisbart & S. Hartmann (Eds.), *Probabilities in physics*. Oxford: Oxford University Press (Chap. 2). [1].
- Lavis, D. A., & Bell, G. M. (1977). Thermodynamic phase changes and catastrophe theory. *Bulletin I. M. A.*, 13(2), 34–42. [176, 659].

- Lavis, D. A., & Bell, G. M. (1999a). *Statistical mechanics of lattice systems (Vol. 1): Closed-Form and Exact Solutions*. Berlin: Springer. [256, 429, 511].
- Lavis, D. A., & Bell, G. M. (1999b). *Statistical mechanics of lattice systems (Vol. 2): Exact, Series and Renormalization Group Methods*. Berlin: Springer. [294, 371, 434, 524, 537, 620, 644, 676].
- Lavis, D. A., & Christou, N. I. (1977). The dielectric constant of a bonded lattice model for water. *Journal of Physics A: Mathematical and General*, 10, 2153–2169. [59, 251].
- Lavis, D. A., & Christou, N. I. (1979). A two-dimensional bonded lattice model for water. *Journal of Physics A: Mathematical and General*, 12, 1869–1890. [252].
- Lavis, D. A., & Quinn, A. G. (1983). An Ising ferrimagnetic model on a triangular lattice. *Journal of Physics C: Solid State Physics*, 16, 347–3562. [90, 606, 613].
- Lavis, D. A., & Quinn, A. G. (1987). Transfer matrix and phenomenological renormalization methods applied to a triangular Ising ferrimagnetic model. *Journal of Physics C: Solid State Physics*, 20, 2129–2138. [90, 613, 615].
- Lavis, D. A., & Southern, B. W. (1984). Renormalization group study of a three-dimensional lattice model with directional bonding. *Journal of Statistical Physics*, 35, 489–506. [252, 254, 607].
- Lavis, D. A., Southern, B. W., & Bell, G. M. (1982). Phase transitions in monolayers at air/water and oil/water interfaces. *Journal of Physics C: Solid State Physics*, 15, 1077–1088. [606].
- Lawden, D. F. (1989). *Elliptic functions and applications*. Berlin: Springer. [714].
- Lawler, G. F. (2011). Scaling limits and the Schramm-Loewner evolution. *Probability Surveys*, 8, 442–495. [163].
- Lawler, G. F., Schramm, O., & Werner, W. (2004). Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Annals of Probability*, 32, 939–995. [166].
- Lawrie, I. D., & Sarbach, S. (1984). Theory of tricritical points. In *PTCP(9)* (pp. 2–161). [94, 139].
- Le Guennec, P. (1998). On the concept of chirality. *Journal of Mathematical Chemistry*, 23, 429–439. [57].
- Lee, T. D., & Yang, C. N. (1952). Statistical theory of equations of state and phase transitions. II Lattice gas and Ising model. *Physical Review*, 87, 410–419. [43, 254, 273].
- Leech, J. W., & Newman, D. J. (1969). *How to use groups*. London, UK: Methuen. [185, 676].
- Levelt Sengers, J. M. H. (1974). From van der Waals' equation to the scaling laws. *Physica*, 73, 73–106. [94, 152].
- Lieb, E. H. (1967a). Exact solution of the problem of the entropy of two-dimensional ice. *Physical Review Letters*, 18, 692–694. [2, 32, 67, 73, 257, 488].
- Lieb, E. H. (1967b). Residual entropy of square ice. *Physical Review*, 162, 162–171. [2, 32, 61, 73, 257, 511].
- Lieb, E. H. (1967c). Exact solution of the F model of an antiferroelectric. *Physical Review Letters*, 18, 1046–1048. [2, 32, 61, 73, 257].
- Lieb, E. H. (1967d). Exact solution of the two-dimensional Slater KDP model of a ferroelectric. *Physical Review Letters*, 19, 108–110. [2, 32, 61, 73, 257].
- Lieb, E. H. (1969). Two dimensional ice and ferroelectric models. In K. Mantappa & W. Brittin (Eds.), *Lectures in theoretical physics* (Vol. 11, pp. 329–354). New York, USA: Gordon and Breach. [74].
- Lieb, E. H. (1994). Flux phase of the half-filled band. *Physical Review Letters*, 73, 2158–2161. [639].
- Lieb, E. H., & Loss, M. (1993). Fluxes, Laplacians and Kasteleyn's theorem. *Duke Mathematical Journal*, 71, 337–363. [495, 639].
- Lieb, E. H., & Ruelle, D. (1972). A property of zeros of the partition function for Ising spin systems. *Journal of Mathematical Physics*, 13, 781–784. [254].
- Lieb, E. H., & Sokal, A. D. (1981). A general Lee-Yang theorem for one-component and multi-component ferromagnets. *Communications in Mathematical Physics*, 80, 153–179. [254].
- Lieb, E. H., & Wu, F. Y. (1972). Two-dimensional ferroelectric models. In *PTCP(1)* (pp. 331–487). [93, 96, 257, 488, 515, 516].
- Lieb, E. H., & Yngvason, J. (1998). A guide to entropy and the second law of thermodynamics. *Notices of the American Mathematical Society*, 45, 571–581. [1].

- Lieb, E. H., & Yngvason, J. (1999). The physics and mathematics of the second law of thermodynamics. *Physics Reports*, 310, 1–96 (For Fig. 8 see *ibid.* (1999), 314, 699). [1].
- Lieb, E. H., & Yngvason, J. (2000). A fresh look at entropy and the second law of thermodynamics. *Physics Today*, 53, 32–37. [1].
- Lieb, E. H., & Yngvason, J. (2013). The entropy concept for non-equilibrium states. *Proceedings of the Royal Society A*, 469, 20130408. [1].
- Lieb, E., Schultz, T., & Mattis, D. (1961). Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 16, 407–466. [32, 33].
- Lloyd, E. H. (1953). The direct product of matrices. *The Mathematical Gazette*, 37, 29–33. [675].
- Löwner, K. (1923). Untersuchungen über schlichte konforme Abbildungen des Einheitskreises I. *Mathematische Annalen*, 89, 103–121. [163, 164, 165].
- Luban, M. (1976). Generalized Landau theories. In *PTCP(5a)* (pp. 35–86). [167].
- Ma, S. K. (1976a). *Modern theory of critical phenomena*. Mass, USA: Benjamin. [106, 202, 520, 616].
- Ma, S. K. (1976b). Renormalization group by Monte Carlo methods. *Physical Review Letters*, 37, 461–464. [616].
- Martin, D. H. (1967). *Magnetism in solids*. London, UK: Iliffe. [30, 40, 90, 212, 223].
- Martin, P. P. (1991). *Potts models and related problems in statistical mechanics*. Philadelphia, USA: World Scientific. [55].
- Martin, P. P., & Maillard, J. M. (1986). Zeros of the partition function for the triangular lattice three-state Potts model. *Journal of Physics A: Mathematical and General*, 19, L547–L551. [254].
- Mattis, D. C. (1981). *The theory of magnetism I*. Berlin: Springer. [30].
- Mattis, D. C. (1985). *The theory of magnetism II*. Berlin: Springer. [223].
- Mattis, D. C. (1993). *The many-body problem: An encyclopedia of exactly solved models in one dimension*. Singapore: World Scientific. [760].
- Maxwell, J. C. (1875). On the dynamical evidence of the molecular constitution of bodies. *Nature*, 11, 357–374. (Reprinted in *The Scientific Papers of James Clerk Maxwell*, Vol. 2, pp. 418–438, W. D. Niven (Ed.), 1890, C.U.P.). [171].
- McCoy, B. M. (2010). *Advanced statistical mechanics*. Oxford: Oxford University Press. [434, 442, 456, 459, 465, 467, 469, 472].
- McCoy, B. M., & Wu, T. T. (1973). *The two-dimensional Ising model*. Cambridge, Mass, USA: Harvard U.P. [483].
- McWeeny, R. (1963). *Symmetry*. Oxford, UK: Pergamon. [676].
- Mermin, N. D. (1971a). Lattice gas with short-range pair interactions and a singular coexistence-curve diameter. *Physical Review Letters*, 26, 957–959. [43, 332, 337].
- Mermin, N. D. (1971b). Solvable model of a vapour-liquid transition with a singular coexistence-curve diameter. *Physical Review Letters*, 26, 169–172. [43, 337].
- Mermin, N. D., & Rehr, J. J. (1971). Generality of the singular diameter of the liquid-vapour coexistence curve. *Physical Review Letters*, 26, 1155–1156. [43].
- Mermin, N. D., & Wagner, H. (1966). Absence of ferromagnetism or antiferromagnetism in one- and two-dimensional isotropic Heisenberg models. *Physical Review Letters*, 17, 1133–1136. [31].
- Metcalf, B. D. (1973). Phase diagram of a nearest-neighbour triangular antiferromagnet in an external field. *Physics Letters A*, 45, 1–3. [605, 606].
- Migdal, A. A. (1975). Phase transitions in gauge and spin-lattice theory. *Soviet Physics JETP*, 42, 743–746. [600, 601].
- Minnhagen, P. (1987). The two-dimensional Coulomb gas, vortex unbinding and superfluid-superconducting films. *Reviews of Modern Physics*, 59, 1001–1066. [49, 53].
- Mittag, L., & Stephen, M. J. (1974). Mean-field theory of the many component Potts model. *Journal of Physics A: Mathematical and General*, 7, L109–L112. [187].
- Montroll, E. W. (1964). Lattice statistics. In E. Beckenbach (Ed.), *Applied combinatorial mathematics* (pp. 96–143). New York, USA: Wiley. [500, 501].
- Moore, M. A. (1969). Additional evidence for a phase transition in the plane-rotator and classical Heisenberg models for two-dimensional lattices. *Physical Review Letters*, 23, 861–863. [49].

- Mukamel, D., & Blume, M. (1974). Ising model for tricritical points in ternary mixtures. *Physical Review A*, 10, 610–617. [45].
- Mulholland, G. W., & Rehr, J. J. (1974). Coexistence curve properties of Mermin's decorated lattice gas. *Journal of Chemical Physics*, 60, 1297–1306. [332, 339, 342, 343].
- Munkres, J. R. (1984). *Elements of algebraic topology*. London, UK: Addison-Wesley. [265].
- Nagle, J. F. (1966a). Lattice statistics of hydrogen bonded crystals. I the residual entropy of ice. *Journal of Mathematical Physics*, 7, 1484–1491. [257, 488].
- Nagle, J. F. (1966b). Lattice statistics of hydrogen bonded crystals. II. The Slater KDP model and the Rys F-model. *Journal of Mathematical Physics*, 7, 1492–1496. [73].
- Nagle, J. F. (1969). Study of the F model using low-temperature series. *Journal of Chemical Physics*, 50, 2813–2818. [562].
- Nagle, J. F. (1975). Chain model theory of lipid monolayer transitions. *Journal of Chemical Physics*, 63, 1255–1261. [515].
- Nagle, J. F. (1986). Theory of lipid monolayer and bilayer chain-melting phase transitions. *Faraday Discussions of the Chemical Society*, 81, 151–162. [515].
- Nagle, J. F., & Temperley, H. N. V. (1968). Combinatorial theorem for graphs on a lattice. *Journal of Mathematical Physics*, 9, 1020–1026. [284].
- Nagle, J. F., Yokoi, C. S. O., & Bhattacharjee, S. M. (1989). Dimer modules on anisotropic lattices. In *PTCP(13)* (pp. 235–297). [86, 96, 510, 515].
- Nakono, H. (1968). Ordering in certain statistical systems of Ising spins. *Progress of Theoretical Physics*, 39, 1121–1132. [326].
- Nauenberg, M. (1975). Renormalization group solution of the one-dimensional Ising model. *Journal of Mathematical Physics*, 16, 703–705. [583].
- Nauenberg, M., & Nienhuis, B. (1974a). Critical surface for square Ising spin lattice. *Physical Review Letters*, 33, 944–946. [603].
- Nauenberg, M., & Nienhuis, B. (1974b). Renormalization group approach to the solution of general Ising models. *Physical Review Letters*, 33, 1598–1601. [571, 603].
- Nauenberg, M., & Scalapino, D. J. (1980). Singularities and scaling functions at the Potts-model multicritical point. *Physical Review Letters*, 44, 837–840. [491].
- Nelson, D. R. (1983). Defect-mediated phase transitions. In *PTCP(7)* (pp. 1–99). [53].
- Nelson, D. R., & Fisher, M. E. (1975). Soluble renormalization groups and scaling fields for low-dimensional Ising systems. *Annals of Physics*, 91, 226–274. [572, 579, 585].
- Newell, G. F. (1950). Crystal statistics of a two-dimensional triangular Ising lattice. *Physical Review*, 79, 876–882. [257].
- Newman, C. M. (1975). Inequalities for Ising models and field theories which obey the Lee-Yang theorem. *Communications in Mathematical Physics*, 41, 1–9. [254].
- Niemeijer, T., & van Leeuwen, J. M. J. (1973). Wilson theory for spin systems on a triangular lattice. *Physical Review Letters*, 31, 1411–1414. [571].
- Niemeijer, T., & van Leeuwen, J. M. J. (1974). Wilson theory for two-dimensional Ising spin systems. *Physica*, 71, 17–40. [571, 595].
- Niemeijer, T., & van Leeuwen, J. M. J. (1976). Renormalization theory for Ising-like spin systems. In *PTCP(6)* (pp. 425–505). [115, 572, 595, 603].
- Nienhuis, B. (1987). Coulomb gas formulation of two-dimensional phase transitions. In *PTCP(11)* (pp. 1–53). [53].
- Nienhuis, B., & Nauenberg, M. (1975). First-order phase transitions in renormalization group theory. *Physical Review Letters*, 35, 477–479. [105].
- Nienhuis, B., Riedel, E. K., & Schick, M. (1980). Magnetic exponents of the two-dimensional q -state Potts model. *Journal of Physics A: Mathematical and General*, 13, L189–L192. [490].
- Nightingale, M. P. (1976). Scaling theory and finite systems. *Physica A*, 83, 561–572. [609, 611, 612].
- Nightingale, M. P. (1977). Non-universality for Ising-like spin systems. *Physics Letters A*, 59, 486–488. [611].

- Nightingale, M. P., & Blöte, H. W. J. (1980). Finite-size scaling and critical point exponents of the Potts model. *Physica A*, 104, 352–357. [613].
- Nightingale, M. P., & 't Hooft, A. H. (1974). Scaling theory and logarithmic singularities. *Physica*, 77, 390–402. [3, 100, 109].
- Nishimori, H., & Griffiths, R. B. (1983). Structure and motion of the Lee-Yang zeros. *Journal of Mathematical Physics*, 24, 2637–2647. [254].
- Niss, M. (2005). History of the Lenz-Ising model 1920–1950: From ferromagnetic to cooperative phenomena. *Archive for History of Exact Sciences*, 59, 267–318. [37].
- Oitmaa, J., Hamer, C., & Zheng, W. (2006). *Series expansion methods for strongly interacting lattice models*. Cambridge UK: CUP. [530, 644].
- Onsager, L. (1944). Crystal statistics. I Two-dimensional model with order-disorder transition. *Physical Review*, 65, 117–149. [95, 256, 306, 372, 421, 423].
- Opechowski, W. (1937). On the exchange interaction in magnetic crystals. *Physica*, 4, 181–199. [517].
- O'Reilly, D. E. (1973). Scaled particle-quasilattice model of liquid water. *Physical Review A*, 7, 1659–1661. [252].
- O'Rourke, M. J., Baxter, R. J., & Bazhanov, V. V. (1995). Numerical results for the three-state critical Potts model on finite rectangular lattices. *Journal of Statistical Physics*, 78, 665–680. [254].
- Ostlund, S. (1981). Incommensurate and commensurate phases in asymmetric clock models. *Physical Review B*, 24, 398–405. [59].
- Park, H. (1994). Three-state Potts model on a triangular lattice. *Physical Review B*, 49, 881–887. [188].
- Patashinskii, A. Z., & Pokrovskii, V. L. (1966). Behaviour of ordered systems near the transition point. *Soviet Physics JETP*, 23, 292–297. [3].
- Pauling, L. (1935). The structure and entropy of ice and other crystals with some randomness of atomic arrangement. *Journal of the American Chemical Society*, 57, 2680–2684. [488].
- Pawley, G. S., Swendsen, R. H., Wallace, D. J., & Wilson, K. G. (1984). Monte Carlo renormalization group calculations of critical behaviour in the simple cubic Ising model. *Physical Review B*, 29, 4030–4040. [616].
- Pearce, P. A., & Thompson, C. J. (1977). The spherical limit for n -vector correlations. *Journal of Statistical Physics*, 17, 189–196. [35].
- Pearson, R. (1980). Conjecture for the extended Potts model magnetic eigenvalue. *Physical Review B*, 22, 2579–2580. [490].
- Peierls, R. (1936). On Ising's model of ferromagnetism. *Proceedings of the Cambridge Philosophical Society*, 32, 477–481. [253, 264].
- Penrose, O. (1979). Foundations of statistical mechanics. *Reports on Progress in Physics*, 42, 200–1937. [261].
- Perram, J. W. (1971). Interstitial models of water in the random mixing approximation. *Molecular Physics*, 20, 1077–1085. [337].
- Perron, O. (1907). Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus. *Mathematische Annalen*, 64, 11–76. [673].
- Pfeuty, P., & Toulouse, G. (1977). *Introduction to the renormalization group and to critical phenomena*. London, UK: Wiley. [104].
- Pierson, S. W. (1997). A review of the real space renormalization group analysis of the two-dimensional Coulomb gas, the Kosterlitz-Thouless-Berezinski transition and extensions to a layered vortex gas. *Philosophical Magazine B*, 76, 715–730. [588].
- Pippard, A. B. (1957). *Elements of classical thermodynamics*. Cambridge, UK: C.U.P. [1, 2].
- Polyakov, A. M. (1970). Conformal symmetry of critical fluctuations, Soviet Phys. *Soviet Physics JETP Letters*, 12, 381–383. [3, 159–161].
- Polyakov, A. M. (1987). *Gauge fields and strings*. Chur: Harwood Academic Publishers. [2].
- Popkov, V., & Nienhuis, B. (1997). The 3D multi-layered solvable six-vertex model: The phase diagram. *Journal of Physics A: Mathematical and General*, 30, 99–116. [256].

- Poston, T., & Stewart, I. N. (1997). *Catastrophe theory and its applications*. New York, USA: Dover. [659, 662].
- Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, 48, 106–169. [54, 55, 490].
- Prigogine, I. (1967). *Introduction to the thermodynamics of irreversible processes*. London: Interscience. [1].
- Privman, V., & Fisher, M. E. (1983). Finite-size effects at first-order transitions. *Journal of Statistical Physics*, 33, 385–417. [153].
- Privman, V., & Fisher, M. E. (1984). Universal critical amplitudes in finite-size scaling. *Physical Review B*, 30, 322–327. [151, 154, 158].
- Pynn, R., & Skjeltorp, A. E. (1984). *Multicritical phenomena*. New York, USA: Plenum. [94].
- Rapaport, D. C. (1972). The Ising ferromagnet with impurities: A series expansion approach, I. *Journal of Physics C: Solid State Physics*, 5, 1830–1858. [313].
- Ree, F. H., & Chesnut, D. A. (1966). Phase transition of a hard-core lattice gas. The square lattice with nearest-neighbour exclusion. *Journal of Chemical Physics*, 45, 3983–4003. [153, 354, 361].
- Rehr, J. J., Joyce, G. S., & Guttman, A. J. (1980). A recurrence technique for confluent singularity analysis of power series. *Journal of Physics A*, 13, 1587–1602. [563].
- Resnick, D. J., Garland, J. C., Boyd, J. T., Shoemaker, S., & Newrock, R. S. (1981). Kosterlitz-Thouless transition in proximity-coupled superconducting arrays. *Physical Review Letters*, 47, 1542–1545. [49].
- Riedel, E. K. (1972). Scaling approach to tricritical phase transitions. *Physical Review Letters*, 28, 675–678. [142].
- Riedel, E. K., & Wegner, F. J. (1974). Effective critical and tricritical exponents. *Physical Review B*, 9, 294–315. [105, 108].
- Roan, S.-S. (2007). The q operator and the functional relations of the eight-vertex model at root-of-unity $\eta = 2mk/n$ for odd n . *Journal of Physics A: Mathematical and Theoretical*, 40, 11019–11044. [442].
- Rohrer, H. (1975). Properties of GdAlO_3 near the spin flop bicritical point. *Physical Review Letters*, 34, 1638–1641. [93].
- Roomany, H. H., & Wyld, H. W. (1981). Finite-lattice Hamiltonian results for the phase structure of the $z(q)$ models and the q -state Potts models. *Physical Review B*, 23, 1357–1361. [615].
- Rossat-Mignod, J., Burlet, P., Bartholin, H., Vogt, O., & Lagnier, R. (1980). Specific heat analysis of the magnetic phase diagram of CeSb . *Journal of Physics C: Solid State Physics*, 13, 6381–6389. [2].
- Rowlinson, J. S. (1969). Thomas Andrews and the critical point. *Nature*, 224, 541–543. [171].
- Rowlinson, J. S. (1973). Legacy of van der Waals. *Nature*, 244, 414–417. [175].
- Rowlinson, J. S. (1979). English translation of J.D. van der Waals' "the thermodynamic theory of capillarity under the hypothesis of a continuous variation of density". *Journal of Statistical Physics*, 20, 197–244. [778].
- Rowlinson, J. S. (1988). English translation (with introductory essay) of J. D. van der Waals: On the continuity of the gaseous and liquid states of matter. *Studies in Statistical Mechanics* (Vol. 14). Amsterdam, New York: North-Holland. [175].
- Rowlinson, J., & Swinton, F. L. (1982). *Liquids and liquid mixtures*. London, UK: Butterworths. [337].
- Ruelle, D. (1969). *Statistical mechanics*. New York, USA: Benjamin. [260–261].
- Ruelle, D. (1971). Extension of the Lee-Yang circle theorem. *Physical Review Letters*, 26, 303–304. [254, 273, 275].
- Runnels, L. K., & Combs, L. L. (1971). Exact finite method of lattice statistics. I square and triangular lattice gases of hard molecules. *Journal of Chemical Physics*, 45, 2482–2492. [153, 282].
- Runnels, L. K., & Hubbard, J. B. (1972). Applications of the Yang-Lee-Ruelle theory to hard-core lattice gases. *Journal of Statistical Physics*, 6, 1–20. [281–282].

- Rushbrooke, G. S. (1963). On the thermodynamics of the critical region for the Ising model. *Journal of Chemical Physics*, 39, 842–843. [125].
- Rys, F. (1963). A two-dimensional classic configuration model. *Helvetica Physica Acta*, 36, 537–559. [73].
- Sacco, J. E., & Wu, F. Y. (1975). 32-vertex model on the triangular lattice. *Journal of Physics A: Mathematical and General*, 8, 1780–1787. [60].
- Saito, Y. (1981). Monte Carlo study of the three-state Potts model with two- and three-body interactions. *Journal of Physics A: Mathematical and General*, 15, 1885–1892. [188].
- Salas, J., & Sokal, A. D. (1997). Logarithmic corrections and finite-size scaling in the two-dimensional 4-state Potts model. *Journal of Statistical Physics*, 88, 567–615. [491].
- Sarbach, S., & Fisher, M. E. (1978). Tricriticality and the failure of scaling in the many-component limit. *Physical Review B*, 18, 2350–2363. [204].
- Saunders, P. T. (1980). *An introduction to catastrophe theory*. Cambridge, UK: C.U.P. [659].
- Savit, R. (1980). Duality in field theories and statistical systems. *Reviews of Modern Physics*, 52, 453–487. [631].
- Schäfer, L. (1976). Conformal covariance in the framework of Wilson’s renormalization group approach. *Journal of Physics A: Mathematical and General*, 9, 377–395. [160, 606].
- Schick, M. A., Walker, J. S., & Wortis, M. (1976). Antiferromagnetic triangular Ising model. *Physics Letters A*, 58, 479–480. [603, 605].
- Schick, M., & Griffiths, R. B. (1977). Antiferromagnetic ordering in the three-state Potts model. *Journal of Physics A: Mathematical and General*, 10, 2123–2131. [60, 187, 605, 606].
- Schick, M., Walker, J. S., & Wortis, M. (1977). Phase diagram of the triangular Ising model: Renormalization group calculation with application to adsorbed monolayers. *Physical Review B*, 16, 2205–2219. [603, 604, 606].
- Schramm, O. (2000). Scaling limits of loop-erased random walks and uniform spanning trees. *Israel Journal of Mathematics*, 118, 221–288. [163, 165].
- Schulman, L. S. (1973). Tricritical points and type-three phase transitions. *Physical Review B*, 7, 1960–1966. [659].
- Schulman, L. S., & Revzen, M. (1972). Phase transitions and catastrophes. *Collective Phenomena*, 1, 43–49. [659].
- Serre, J.-P. (1977). *Linear representations of finite groups*. Berlin: Springer. [676].
- Sewell, G. L. (1989). *Quantum theory of collective phenomena* (2nd ed.). Oxford, UK: O.U.P. [263].
- Shafer, R. E. (1974). On quadratic approximation. *SIAM Journal for Numerical and Analytical*, 11, 447–460. [754].
- Shinmi, M., & Huckaby, D. (1986). A lattice gas model for carbon tetrachloride. *Journal of Chemical Physics*, 84, 951–955. [252].
- Shockley, W. (1938). Theory of order for the copper gold alloy system. *Journal of Chemical Physics*, 6, 130–144. [44].
- Simon, B. (1993). *The statistical mechanics of lattice gases* (Vol. I). Princeton: Princeton University Press. [263].
- Sinha, K., & Kumar, N. (1980). *Interactions in magnetically ordered solids*. Oxford, UK: O.U.P. [30].
- Sklar, L. (1993). *Physics and chance*. Cambridge, UK: C.U.P. [1].
- Slater, J. C. (1941). Theory of the transition KH_2PO_4 . *Journal of Chemical Physics*, 9, 16–33. [73].
- Smirnov, S. (2001). Critical percolation in the plane: Conformal invariance, Cardy’s formula, scaling limits. *Comptes Rendus de l’Académie des Sciences-Series I Mathematics*, 333(2001), 239–244. [166].
- Sneddon, L. (1978). Critical properties of two-dimensional spin s Ising systems with lattice anisotropy. *Journal of Physics C: Solid State Physics*, 11, 2823–2828. [611, 613, 614].
- Sneddon, L., & Barber, M. N. (1977). Decimation transformations for two-dimensional Ising spin systems. *Journal of Physics C: Solid State Physics*, 10, 2653–2664. [574, 599].

- Southern, B. W. (1978). Kadanoff's variational renormalization group method: The Ising model on the square and triangular lattices. *Journal of Physics A: Mathematical and General*, *11*, L1–L4. [609].
- Southern, B. W., & Lavis, D. A. (1979). A model for adsorbed monolayers of orientable molecules. *Journal of Physics C: Solid State Physics*, *12*, 5333–5343. [606].
- Southern, B. W., & Lavis, D. A. (1980). Renormalization group study of a two-dimensional lattice model with directional bonding. *Journal of Physics A: Mathematical and General*, *13*, 251–262. [578, 606].
- Stahl, H. (1997). The convergence of diagonal Padé approximants and the Padé conjecture. *Journal of Computational and Applied Mathematics*, *86*, 287–296. [561].
- Stanley, H. E. (1968a). Dependence of critical properties on dimensionality of spins. *Physical Review Letters*, *20*, 589–592. [34].
- Stanley, H. E. (1968b). Spherical model as the limit of infinite spin dimensionality. *Physical Review*, *176*, 718–722. [35, 36].
- Stanley, H. E. (1974). D-vector model. In *PTCP(3)* (pp. 485–567). [34].
- Stanley, H. E. (1987). *Introduction to phase transitions and critical phenomena*. Oxford, UK: O.U.P. [133].
- Stauffer, D., Ferer, M., & Wortis, M. (1972). Universality of second-order phase transitions: The scale factor for the correlation length. *Physical Review Letters*, *29*, 345–349. [151].
- Stephenson, J. (1970). Ising model with antiferromagnetic next-nearest-neighbour coupling: Spin correlations and disorder points. *Physical Review B*, *1*, 4405–4409. [326].
- Stillinger, E. H. (1975). Theory and molecular models for water. *Advances in Chemical Physics*, *31*, 1–101. [252].
- Sutherland, B. (1967). Exact solution of a two-dimensional model for hydrogen-bonded crystals. *Physical Review Letters*, *19*, 103–104. [74].
- Sutherland, B. (1970). Two-dimensional hydrogen bonded crystals without the ice rule. *Journal of Mathematical Physics*, *11*, 3183–3186. [2, 33, 299, 441].
- Suzuki, M. (1974). New universality of critical exponents. *Progress of Theoretical Physics*, *51*, 1992–1993. [492].
- Suzuki, M. (1976). Relationship between d -dimensional quantal spin systems and $(d + 1)$ -dimensional Ising systems. *Progress of Theoretical Physics*, *56*, 1454–1469. [2, 31].
- Suzuki, M. (1977). Static and dynamic finite-size scaling theory based on the renormalization group approach. *Progress of Theoretical Physics*, *58*, 1142–1150. [609].
- Suzuki, M., & Fisher, M. E. (1971). Zeros of the partition function for the Heisenberg, ferroelectric and general Ising models. *Journal of Mathematical Physics*, *12*, 235–246. [60, 254].
- Swendsen, R. H. (1979a). Monte Carlo renormalization group. *Physical Review Letters*, *42*, 859–861. [616].
- Swendsen, R. H. (1979b). Monte Carlo renormalization group studies of the $d = 2$ Ising model. *Physical Review B*, *20*, 2080–2087. [517].
- Sykes, M. F., Essam, J. W., & Gaunt, D. S. (1965). Derivation of low-temperature expansions for the Ising model of a ferromagnet and an antiferromagnet. *Journal of Mathematical Physics*, *6*, 283–298. [517].
- Sykes, M. F., Essam, J. W., Heap, B. R., & Hiley, B. J. (1966). Lattice constant systems and graph theory. *Journal of Mathematical Physics*, *7*, 1557–1572. [517, 640, 645].
- Sykes, M. F., Gaunt, D. S., Essam, J. W., & Elliott, C. J. (1973). Derivation of low-temperature expansions for the Ising model. VI Three-dimensional lattices-temperature grouping. *Journal of Physics A: Mathematical and General*, *6*, 1507–1516. [517].
- Sykes, M. F., Gaunt, D. S., Essam, J. W., Heap, B. R., Elliott, C. J., & Mattingly, S. R. (1973). Derivation of low-temperature expansions for the Ising model. V Three-dimensional lattices-field grouping. *Journal of Physics A: Mathematical and General*, *6*, 1498–1506. [517].
- Sykes, M. F., Gaunt, D. S., Essam, J. W., & Hunter, D. L. (1973). Derivation of low-temperature expansions for the Ising model. II General theory. *Journal of Mathematical Physics*, *14*, 1060–1065. [517].

- Sykes, M. F., Gaunt, D. S., Martin, J. L., Mattingly, S. R., & Essam, J. W. (1973). Derivation of low-temperature expansions for the Ising model. IV Two-dimensional lattices-temperature grouping. *Journal of Mathematical Physics*, 14, 1071–1073. [517].
- Sykes, M. F., Gaunt, D. S., Mattingly, S. R., Essam, J. W., & Elliott, C. J. (1973). Derivation of low-temperature expansions for the Ising model. III Two-dimensional lattices-field grouping. *Journal of Mathematical Physics*, 14, 1066–1070. [517].
- Sykes, M. F., Gaunt, D. S., Roberts, P. D., & Wyles, J. A. (1972a). High temperature series for the susceptibility of the Ising model. I Two dimensional lattices. *Journal of Physics A: General Physics*, 5, 624–639. [518].
- Sykes, M. F., Gaunt, D. S., Roberts, P. D., & Wyles, J. A. (1972b). High temperature series for the susceptibility of the Ising model. II Three dimensional lattices. *Journal of Physics A: General Physics*, 5, 640–652. [518].
- Sykes, M. F., McKenzie, S., Watts, M. G., & Gaunt, D. S. (1975). Derivation of low-temperature expansions for the Ising model. IX High-field polynomials for the honeycomb-triangular system. *Journal of Physics A: Mathematical and General*, 8, 1461–1468. [517].
- Sykes, M. F., & Watts, M. G. (1975). Derivation of low-temperature series expansions for the Ising model with triplet interactions on the plane triangular lattice. *Journal of Physics A: Mathematical and General*, 8, 1469–1479. [518].
- Sykes, M. F., Watts, M. G., & Gaunt, D. S. (1975a). Derivation of low-temperature expansions for the Ising model. VII The honeycomb-triangular code system. *Journal of Physics A: Mathematical and General*, 8, 1441–1447. [517].
- Sykes, M. F., Watts, M. G., & Gaunt, D. S. (1975b). Derivation of low-temperature expansions for the Ising model. VIII Ferromagnetic and antiferromagnetic polynomials for the honeycomb-triangular system. *Journal of Physics A: Mathematical and General*, 8, 1448–1460. [517].
- Syôzi, I. (1950). The statistics of honeycomb and triangular lattice. *Progress of Theoretical Physics*, 5, 341–351. [517].
- Syoz, I. (1972). Transformation of Ising models. In *PTCP(1)* (pp. 270–329). [318, 319, 326, 631].
- Syoz, I., & Nakono, H. (1955). Statistical models of ferrimagnetism. *Progress of Theoretical Physics*, 13, 69–78. [319].
- Takhtadzhyan, L. A., & Faddeev, L. D. (1979). The quantum method of the inverse problem and the Heisenberg XYZ model. *Uspekhi Matematicheskikh Nauk*, 34, 13–63. [33].
- Temperley, H. N. V. (1950). Statistical mechanics of the two-dimensional assembly. *Proceedings of the Royal Society A*, 202, 202–207. [257].
- Temperley, H. N. V., & Fisher, M. E. (1961). Dimer problem in statistical mechanics-an exact result. *Philosophical Magazine*, 6, 1061–1063. [257, 497].
- Temperley, H. N. V., & Lieb, E. H. (1971). Relations between the percolation and colouring problems on the plane square lattice. Some exact results for the percolation problem. *Proceedings of the Royal Society*, 322, 251–280. [55, 75].
- Thibaudier, C., & Villain, J. (1972). On the three dimensional eight-vertex model. *Journal of Physics C: Solid State Physics*, 5, 3429–3437. [60].
- Thom, R. (1972). *Stabilité Structurale et Morphogénèse*. Reading, Mass: W. A. Benjamin. [659].
- Thom, R. (1974). *Modèles Mathématiques de la Morphogénèse* (10–18 ed.). Paris: UGE. [659].
- Thompson, B. V., & Lavis, D. A. (1967). The Gaussian average method applied to phase transitions in dilute magnetic alloys. *Proceedings of the Physical Society*, 91, 645–657. [45].
- Thompson, C. S. (1988). *Classical equilibrium statistical mechanics*. Oxford, UK: O.U.P. [264].
- Thomson, J. (1871). Considerations on the abrupt change at boiling or condensing in reference to the continuity of the fluid state of matter. *Proceedings of the Royal Society*, 20, 1–8. [171, 174].
- Thomson, W. T. (1904). *Baltimore lectures on molecular dynamics and the wave theory of light*. Baltimore: C.J. Clay. [57].
- Titchmarsh, E. C. (1939). *The theory of functions* (2nd ed.). Oxford, UK: O.U.P. [271].
- Tjon, J. A. (1974). Numerical study of the renormalization group equations in the four-cell approximation. *Physics Letters A*, 49, 289–290. [603].

- van der Waals, J. D. (1873). Over de Continuïteit van der Gas- en Vloeistoofstand. Ph.D. Thesis, Leiden. [For an English translation see Rowlinson (1988)]. [94, 171, 173, 175].
- van der Waals, J. D. (1893). The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density. *Verhandel Konink Akad Wet Amsterdam*, 1(8). [For an English translation see Rowlinson (1979)]. [775].
- van Enter, A., & Zegarliński, B. (1997). Non-periodic long-range order for one-dimensional pair interactions. *Journal of Physics A: Mathematical and General*, 30, 501–505. [353].
- van Leeuwen, J. M. J. (1975). Renormalization theory for spin systems. In E. Cohen (Ed.), *Fundamental problems in statistical mechanics* (Vol. 3, pp. 81–101). Amsterdam, New York: North-Holland. [573].
- Villian, J. (1975). Theory of one- and two-dimensional magnets with an easy magnetization plane. II. The planar, classical, two-dimensional magnet. *Journal de Physique*, 36, 581–590. [53].
- Wang, J.-S., Swendsen, R. H., & Kotecky, R. (1989). Antiferromagnetic Potts models. *Physical Review Letters*, 63, 109–112. [187].
- Wang, Z. F., Southern, B. W., & Lavis, D. A. (2003). Monte Carlo study of an extended 3-state Potts model on the triangular lattice. *Physical Review B*, 67, 054415–1–054415-6. [60].
- Wannier, G. H. (1950). Antiferromagnetism: The triangular Ising net. *Physical Review*, 79, 357–364. [257].
- Wegner, F. J. (1972). Duality relation between the Ashkin-Teller and eight-vertex models. *Journal of Physics C: Solid State Physics*, 5, L131–L132. [54].
- Wegner, F. J. (1973). A transformation including the weak-graph theorem and the duality transformation. *Physica*, 68, 570–578. [254, 304].
- Weres, O., & Rice, S. A. (1972). A new model for liquid water. *Journal of the American Chemical Society*, 94, 8983–9002. [252].
- Wheeler, J. C. (1975). “Exactly solvable” two component lattice solution with upper and lower critical solution temperatures. *Journal of Chemical Physics*, 62, 433–439. [331].
- Wheeler, J. C. (1977). Decorated lattice-gas models of critical phenomena in fluids and fluid mixtures. *Annual Review of Physical Chemistry*, 28, 411–443. [43, 331, 332].
- Whitehouse, J. S., Christou, N. I., Nicholson, D., & Parsonage, N. G. (1984). A grand ensemble Monte Carlo investigation of the Bell lattice model for water. *Journal of Physics A: Mathematical and General*, 17, 1671–1682. [252].
- Whitney, H. (1932). The coloring of graphs. *Annals of Mathematics*, 33, 688–718. [638].
- Whittaker, E. T., & Watson, G. N. (1927). *A course of modern analysis* (4th ed.). Cambridge, England: Cambridge University Press. [721].
- Widom, B. (1964). Degree of the critical isotherm. *Journal of Chemical Physics*, 41, 1633–1634. [133].
- Widom, B. (1965). Equation of state in the neighbourhood of the critical point. *Journal of Chemical Physics*, 43, 3898–3905. [3, 616].
- Widom, B. (1967). Plait points in two- and three-component liquid mixtures. *Journal of Chemical Physics*, 49, 3324–3333. [318].
- Widom, B., & Rowlinson, J. S. (1970). New model for the study of liquid-vapour phase transitions. *Journal of Chemical Physics*, 52, 1670–1684. [43].
- Williams, N. O., & Lavis, D. A. (1996). Finite-size scaling of branch-points in lattice models. *Physics Letters A*, 217, 275–279. [380].
- Wilson, G. L., & Bell, G. M. (1978). Lattice model for aqueous solutions of non-electrolytes. *Journal of the Chemical Society, Faraday Transactions II*, 74, 1702–1722. [251].
- Wilson, K. G. (1971). Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff picture. *Physical Review B*, 4, 3174–3183. [519].
- Wilson, K. G. (1975). The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47, 773–840. [520, 573, 586, 588].
- Wilson, R. J., & Watkins, J. J. (1990). *Graphs: An introductory approach*. London, UK: Wiley. [619, 621].

- Witten, L. (1954). A generalization of Yang and Lee's theory of condensation. *Physical Review*, 93, 1131–1135. [260].
- Wong, R., & Zhao, Y.-Q. (2005). On a uniform treatment of Darboux's method. *Constructive Approximation*, 21, 225–255. [557].
- Wood, D. W. (1985). The exact location of partition function zeros, a new method for statistical mechanics. *Journal of Physics A: Mathematical and General*, 18, L917–L921. [255, 374].
- Wood, D. W. (1987). The algebraic construction of partition function zeros: Universality and algebraic cycles. *Journal of Physics A: Mathematical and General*, 20, 3471–3493. [256, 371, 376, 379, 699].
- Wood, D. W., & Ball, J. K. (1990). On a point of order. *Journal of Statistical Physics*, 58, 599–615. [256, 376].
- Wood, D. W., & Goldfinch, M. (1980). Vertex models for the hard-square and hard-hexagon gases, and critical parameters for the scaling transformation. *Journal of Physics A: Mathematical and General*, 13, 2781–2794. [282].
- Wood, D. W., & Osbaldstun, A. H. (1982). Phase equilibria and the scaling transformation. *Journal of Physics A: Mathematical and General*, 15, 3579–3591. [615].
- Wood, D. W., & Osbaldstun, A. H. (1983). Multiple phase coexistence and the scaling transformation. *Journal of Physics A: Mathematical and General*, 16, 1019–1033. [615].
- Wood, D. W., & Turnbull, R. W. (1988). $z^2 - 11z - 1$ as an algebraic invariant for the hard-hexagon model. *Journal of Physics A: Mathematical and General*, 21, L989–L994. [256].
- Wood, D. W., Turnbull, R. W., & Ball, J. K. (1987). Algebraic approximations to the locus of partition function zeros. *Journal of Physics A: Mathematical and General*, 20, 3495–3521. [376, 379].
- Wood, D. W., Turnbull, R. W., & Ball, J. K. (1989). An observation on the partition function zeros of the hard hexagon model. *Journal of Physics A: Mathematical and General*, 22, L105–L109. [256].
- Wortis, M. (1963). Bound states of two spin waves in the Heisenberg ferromagnet. *Physical Review*, 132, 85–97. [32].
- Wu, F. Y. (1968). Remarks on the modified potassium dihydrogen phosphate model of a ferroelectric. *Physical Review*, 168, 539–543. [506].
- Wu, F. Y. (1971). Ising model with four-spin interactions. *Physical Review B*, 4, 2312–2314. [54, 67].
- Wu, F. Y. (1974). Eight-vertex model on the honeycomb lattice. *Journal of Mathematical Physics*, 15, 687–691. [60, 379].
- Wu, F. Y. (1982). The Potts model. *Reviews of Modern Physics*, 54, 235–268. [55, 184, 188, 306, 555].
- Yang, C. N., & Lee, T. D. (1952). Statistical theory of equations of state and phase transitions I Theory of condensation. *Physical Review*, 87, 404–409. [43, 254, 260, 271, 273].
- Yang, C. N., & Yang, C. P. (1966a). Ground-state energy of a Heisenberg-Ising chain. *Physical Review*, 147, 303–306. [32].
- Yang, C. N., & Yang, C. P. (1966b). One-dimensional chain of anisotropic spin-spin interactions. I. Proof of Bethe's hypothesis for ground state in a finite system. *Physical Review*, 150, 321–327. [32, 474].
- Yang, C. N., & Yang, C. P. (1966c). One-dimensional chain of anisotropic spin-spin interactions. II. Properties of the ground-state energy per lattice site for an infinite system. *Physical Review*, 150, 327–339. [32].
- Yang, C. N., & Yang, C. P. (1966d). One-dimensional chain of anisotropic spin-spin interactions III. Applications. *Physical Review*, 151, 258–264. [32].
- Yeomans, J. M. (1992). *Statistical mechanics of phase transitions*. Oxford, UK: O.U.P. [133].
- Young, A. P., & Lavis, D. A. (1979). Critical behaviour of a two-dimensional bonded lattice fluid. *Journal of Physics A: Mathematical and General*, 12, 229–243. [60, 606].
- Young, S. (1892). On the generalizations of van der Waals regarding “corresponding” temperatures, pressures and volumes. *Philosophical Magazine*, 33, 153–185. [95].

- Zener, C. (1951). Interaction between the d shells in the transition metals. *Physical Review*, 81, 440–444. [30].
- Ziman, J. M. (1979). *Models of disorder*. Cambridge, UK: C.U.P. [2, 97].
- Zollweg, J. A., & Mulholland, G. W. (1972). On the law of the rectilinear diameter. *Journal of Chemical Physics*, 57, 1021–1025. [332, 337].

Index

Symbols

$O(n)$ models, 34
 ϵ -expansion, 151, 616
 λ -point, 204
 λ -transition, 44, 49
 ν -string, 467
 ^3He – ^4He
 mixtures, 44
 n -vector model, 33, 34, 522, 543
 L -string, 467
 complete, 468
 MAPLE, 551
 MATHEMATICA, 551
 3-12
 dimer model, 514
 lattice, 86

A

Achiral, 58
 Algebraic function, 695
 apparent singular
 point, 698
 branch-point, 698
 branches of, 696
 irreducible, 370
 Antiferroelectric
 ordered state, 75
 Antiferromagnet
 critical field, 41
 Antiferromagnetism, 30
 Articulation point, 740
 Asano contraction procedure, 274
 Ashkin–Teller model, 53
 relation to eight-vertex
 model, 54
 AuCu, 44

B

Baker–Gammel–Wills conjecture, 561
 Basin of attraction, 101
 Baxter’s modified theta functions, 442
 Baxter–Wu model, 561
 Bazhanov–Mangazeev transformation, 443,
 448, 476, 725, 737
 Bell–Lavis model, 59
 Bessel function
 of the first kind, 708
 of complex argument, 709
 Bethe
 -pair method, 229
 lattice, 620
 ansatz, 32
 equation, 467
 Biased estimate, 555, 560
 Bicritical
 end-point, 92, 93, 136
 in metamagnet, 225
 point, 92
 Binary mixture, 43
 Ising model equivalence, 43
 Block-spin method, 520, 568
 Blume–Emery–Griffiths model, 44, 60, 92
 Boltzmann’s constant, 5
 Bond moving, 600
 Bond percolation, 631
 Bose
 gas, 156
 liquid, 31
 Boundary matrix, 346
 Brownian motion in SLE, 3, 165

C

- Canonical distribution, 15
- Canonical flux distribution, 495
- Carbon tetrachloride, 252
- Catalan's constant, 503
- Catastrophe theory, 177
- Cauchy–Riemann conditions, 377
- Cerium antimonide, 2
- Chain conformal system, 515
- Chemical potential, 1
- Chess board, arranging
 - dominos on, 504
- Chiral clock model, 58
- Chirality, 45, 57
- Circle theorem, 254, 273, 277, 279
- Clock model, 54
- Cluster expansion, 540
- Cluster function, 540
- Coexistence
 - p -phase, 90
 - curve, 124, 133
 - closed-loop, 326, 329
 - curvature of, 94
 - four-phase, 93
 - region, 3, 122, 124, 136
 - diameter, 43, 336
 - width, 43, 335
- Commensurate phases, 41
- Compensation point, 90
- Compensation temperature, 321
- Compressibility, 10
 - minimum in water model, 338
- Confluent singularities, 557, 558, 561
- Conformal
 - anomaly number, 162
 - group, 159
 - invariance, 3, 164, 490
- Connection curve, 370, 699, 700
- Constant-coupling method, 229
- Contour segment, 266
- Corner transfer matrix, 477
- Correlation function, 616
 - algebraic, 50
 - connected, 25
 - pair, 97
 - two-point, 134
 - fluctuation, 24
 - one-point connected, 27
 - pair, 167, 197, 198, 349, 573
 - total, 24
- Correlation length, 95, 97, 98, 134, 153, 162, 167, 197, 198, 210, 481, 609–612, 614
 - Kosterlitz–Thouless transition, 594
 - scaling form, 135
- Correlation time, 95
- Correlations
 - asymptotic decay of, 574
 - long-range, 209, 574
- Coulomb lattice gas, 53
- Couplings
 - external, 17
 - internal, 17
 - proliferation of, 572, 586
 - spatially dependent, 588
- Covering of lattice, 693
- Critical
 - curve
 - in metamagnet, 216
 - dimension
 - lower, 29
 - upper, 29, 151
 - double-point, 331, 343
 - end-point, 91, 92, 136, 606
 - in metamagnet, 225
 - exponents, 3, 94, 95, 567
 - classical, 95, 181, 209, 230
 - spherical model, 35
 - field, 605
 - in antiferromagnet, 42
 - isochamp, 130
 - isochore, 124, 133
 - isotherm, 125, 130
 - line, 93, 136, 188
 - opalescence, 95
 - point, 90, 122, 148, 183
 - first order, 105
 - ratio, 94
 - region, 3, 204
 - width, 203
 - temperature, 522
 - Gaussian model, 35
 - lower, 325, 326
 - upper, 325, 326
- Cross-block curves, 370
- Crossover, 204
 - curve, 129, 136
 - region, 108, 134
 - surface, 104
- Cumulant, 523
 - approximation
 - first-order, 596, 597
 - expansion, 596
- Curie temperature, 38, 91
 - mean-field expression, 207
 - paramagnetic, 223

Curie–Weiss law, 223

Cyclic

- group, 354
- matrix, 700
- block, 701

D

Dangerous irrelevant variable, 151

Decimation, 633

Decorated lattice gas

Mermin's, 332

Density, 8, 89, 578

fluctuations, 167

matrix, 14

maximum in water model, 338

operator, 14

Dice lattice, 628

Dimer, 81

fugacity, 82, 496

model, 23, 96, 257

and five-vertex model, 83

exponents in, 509

heat capacity of, 509

partition function, 82

superposition diagram for, 497

pure assembly, 81

Direction rule, 62

Directional bonding, 80, 607

Discriminant of a polynomial, 696

Disorder point, 326, 329

Distribution

canonical, 14, 260

constant-pressure, 11, 14

Gibbs, 14

grand-canonical, 11, 14, 259

Maxwell-Boltzmann, 14

probability function, 13

Distributions

thermodynamic equivalence of, 19

Dlog transformation, 559

double, 560

Dobrushin–Lanford–Ruelle (DLR) equilibrium condition, 263

Domain boundary, 164

Doubly-periodic function, 712

extended definition, 712

Dual lattice, 164

Duality, 255, 283

E

Easy magnetization, 30

Edge decoration, 255

transformations, 284

Edge-variable model, 22

Eigenvalue

branch-points, 372

irrelevant, 576

marginal, 576

relevant, 576

strong, 576

weak, 576

Eight-vertex model, 23, 50, 61, 96, 257, 291, 293, 516

Z-invariant, 381, 388, 435

free-energy density decomposition, 391

disorder point, 300, 326

exponents, 481

free-energy density, 475

ground states, 68

heat capacity, 480

ice, 300

entropy of, 66

kagomé lattice, 61, 395

magnetization, 477

marginal four-spin coupling, 482

rapidity lattice, 634

staggered, 65

staircase transformation, 295

symmetric, 611

symmetric group S_4

subgroup in, 295

symmetry group, 690

transition surface, 300

varying exponents, 120, 480, 491

weak-graph transformation in, 296

XYZ model equivalence, 441

Elliptic integrals, 714

complete, 715

Elliptic roots of unity condition, 456

Embedding

strong graph, 517

weak graph, 517

Energy

free, 7

internal, 5

Ensemble, 14

Enthalpy, 6, 13

Equation of state, 170

Equilibrium, 1

Equivalence classes, 356

Euler's

relationship, 534, 621, 629

transformation, 558

Euler-Mascheroni constant, 710

Exchange interaction, 30

- Exclusion models, 81
- Expansion
 - high-field, 524
- Exponent
 - classical, 149
 - critical, 182
 - crossover, 104
 - dangerous irrelevant, 110
 - effective, 108
 - field, 130, 581, 599
 - gap, 102, 127, 129, 136
 - Gaussian, 151
 - irrelevant, 101, 576
 - marginal, 50, 101, 576, 615
 - mean-field, 203
 - relevant, 101, 576
 - renormalization, 96, 317, 560
 - hidden variable in, 318
 - rounding, 156
 - scaling, 99
 - shift, 126, 129, 142, 156
 - tricritical, 139
 - strong, 128
 - temperature-dependent, 50
 - thermal, 130, 587, 599, 611, 615
 - varying, 481, 482
 - weak, 128
- Extension curve, 377

- F**
- F model, 73, 257, 562
 - staggered, 490, 613
- Face decoration, 255
- Face model, 22
- FeAl, 44
- Fermi liquid, 31
- Ferrimagnetic order, 606
- Ferrimagnetism, 2, 319
- Ferroelectric, 61
 - ordered state, 74
- Ferromagnetism, 2, 30
- Field, staggered, 93, 212
- Finite cluster, 543
- Finite lattice, 543
- Finite-cluster method, 524, 646, 647
- Finite-lattice method, 646, 647
- Finite-size
 - effects, 153
 - scaling, 490, 609
 - field, 154
- First-order approximation, 332
- First-order critical point, 134

- First-order transitions
 - line of, 91, 223
- Fisher circles, 255, 374
- Five-vertex model, 510, 511, 516
- Fixed point, 101, 492, 570
 - antiferromagnetic, 585
 - attractive, 570
 - basin of attraction of, 570, 576
 - ferromagnetic, 581, 585
 - Gaussian, 616
 - infinite
 - field, 582
 - temperature, 582
 - isolated, 576
 - marginal, 570
 - repulsive, 570
- Fixed points, line of, 50, 120, 570, 581
- Fluctuation-response function relation, 15, 26, 163, 212
- Fluctuations, 95, 167
 - harmonic, 167, 194
 - in statistical mechanics, 2
 - long wave length, 520
 - longitudinal, 195
 - low-frequency, 199
 - transverse, 195
- Fourier transform, 703
- Fractal dimension, 165
- Free energy, 18
 - density
 - dimensionless, 18
 - local, 193
 - extensivity and thermodynamics, 529
 - Helmholtz, 15
 - singularities, 299
- Free-fermion condition, 71, 399
- Free-fermion model
 - Pfaffian solution, 71
- Frobenius'
 - method, 751
 - theorem, 346
- Frozen order, in dimer
 - assembly, 506
- Fuchsian type equation, 752
- Fugacity, 260

- G**
- Gap
 - exponent, 127, 136
 - tricritical, 142
 - field, 102, 127
 - tricritical, 144

- Gaussian
 - approximation, 167, 198
 - critical exponents, 202
 - exponents, 198, 201
 - fixed point, 106, 202
 - model, 34
- Generalized hypergeometric function, 710
- Genus of manifold, 621
- Gibbs distribution, 14, 261
- Gibbs measure, 261, 262
- Gibbs phase rule, 90
- Gibbs state, 263
- Ginzburg criterion, 204
- Ginzburg–Landau
 - local free-energy density, 520
 - theory, 35, 204, 616
- Glaisher’s notation for elliptic functions, 721
- Glycerol-ethylbenzylamine, 326, 331
- Goldstone modes, 31
- Graph
 - augmented, 638
 - chain, 619
 - dual, 623
 - spanning, 637
 - strong embedding, 637
 - triangulation number, 622
 - weak embedding, 637
 - articulation point, 620
 - bridge, 620
 - Cayley tree, 620
 - complement, 85
 - component, 620
 - connected, 619
 - cut edge, 620
 - cut vertex, 620
 - cyclomatic number, 621
 - disjoint, 620
 - edge, 619
 - face, 621
 - face cycle, 621
 - independent cycle, 621
 - multi-bond, 524
 - multi-bonded, 536, 538
 - multiply-connected, 619
 - oriented, 622
 - section, 637
 - separated, 620
 - silhouette, 540
 - single-bonded, 536
 - star, 620
 - tree, 620
 - triangulation, 622
 - vertex, 619
 - vertex valency, 619
 - zero field, 85
- Ground state, 31, 263, 264
 - degeneracy, 264
 - periodic, 351
 - stable, 264
 - well-defined for low-temperature series, 522
- Group
 - conjugacy class, 678
 - representation, 678
 - canonical form, 355, 681, 682
 - character system, 679
 - character table, 681, 688
 - equivalent, 679
 - irreducible, 679
 - multiplicity, 681
 - one-dimensional symmetric, 679
 - permutation, 683
 - reducible, 679
 - regular, 683
 - unitary, 679
- Guggenheim–McGlashan method, 654
- H**
 - Hamiltonian, 13
 - symmetries of, 676
 - symmetry group, 354
 - Hard-hexagon model, 81, 282, 379, 490
 - Hard-square model, 81, 281, 378
 - Heat capacity, 10, 200
 - discontinuity in, 183
 - logarithmic divergence, 201
 - logarithmic singularity, 483
 - of modified KDP model, 511
 - square-root singularity, 515
 - Heisenberg model, 15, 517
 - classical, 34, 530
 - heat capacity, 530
 - quantum, 30, 522
 - susceptibility, 530
 - Heisenberg spin chain, 441
 - Heisenberg–Ising XXZ model, 32
 - Helium
 - adsorbed on graphite, 606
 - liquid, 204
 - Hermite–Padé polynomials, 754
 - Hierarchical lattice, 588
 - High-density series, 378
 - High-temperature series, 613, 738
 - Hopping matrix, 495
 - flux of, 638

Hydrogen bond, 246, 337
 Hyper-scaling law, Josephson, 151, 157,
 202, 204, 481
 Hyper-universality hypothesis, 151

I

Ice structures, 247
 Ice-rule, 61, 73, 516
 six-vertex model, 257
 IF model, 301
 IKDP model, 301
 Incipient phase transition, 256
 Infinite-system approach, 9
 Information, loss of, 574
 Instability, intrinsic, 223
 Interacting set, 260, 265
 Interactions, proliferating, 588
 Internal fields, 17
 Interstitial model, 337
 Intrinsically unstable, 173
 Inverse F model, 74
 Inverse KDP model, 73
 Irrelevant variable
 dangerous, 147, 203, 204, 606
 harmless, 147
 Ising model, 60, 571
 Z-invariant, 381
 θ representation, 45
 positive transfer matrix, 673
 symmetry group, 689
 anisotropic, 85
 honeycomb lattice, 513
 antiferromagnet, 603, 614
 Hamiltonian, 39
 binary mixture equivalence, 43
 bond-dilute, 96
 checkerboard, 257, 395
 free-fermion model equivalence, 71
 modified, 401
 critical, 163
 critical parameters, 38
 critical temperature, 303, 482
 honeycomb lattice, 310
 square lattice, 290, 424
 triangular lattice, 309
 duality transformation, 298
 edge-decorated, 22
 ferrimagnet, 90, 613
 ferromagnet, 96, 122
 ferromagnet/antiferromagnet one-to-one
 correspondence, 40
 free energy

 square lattice, 514
 triangular lattice, 514
 general-s, 31
 Hamiltonian, 36
 hypercubic lattice, 600
 internal energy, 485
 lattice fluid equivalence, 42
 local order, 209
 magnetization, 39, 483
 density, 37, 263
 one-dimensional, 301, 369, 574, 579
 critical exponents, 584
 partition function, 37
 quantum, 31
 short-range order in, 206
 spin- $\frac{1}{2}$, 44, 491, 517, 607, 611
 spin-1, 44, 184, 491
 square lattice, 370, 371
 pure four-spin interaction, 482
 self-duality, 376, 411
 standard square-lattice, 394
 three-dimensional, 616
 three-spin, 379
 transfer matrix calculation, 359
 triangular lattice, 71, 376
 tricritical, 163
 two-dimensional, 95, 149, 574, 597
 critical exponents, 483, 490
 heat capacity, 430
 Onsager's solution, 257, 423
 with bond-dilution, 313
 with super-exchange, 313

J

Jacobi amplitude, 720
 Jacobi elliptic function, 720
 Josephson hyper-scaling law, 135

K

Kac formula, 162
 Kastelyn's theorem, 497
 KDP model, 73, 84, 257
 modified, 511
 KHDeB method, 646, 647
 Kosterlitz-Thouless
 singularity, 560
 transition, 29, 49, 50, 530

L

Lambert W function, 318
 Landau expansion

- ferromagnetism, 668
 - metamagnetism, 668
 - Potts model, 668
 - Landau theory, 94, 149, 167, 659
 - and scaling theory, 182
 - Latent heat, 515
 - Lattice
 - 3-12, 512
 - 4-8, 515
 - brick, 515
 - cell, 265
 - close-packed, 39, 624
 - cluster, 266
 - covering, 273
 - fluid, 42
 - bonded, 264, 606
 - first-neighbour exclusion, 281
 - Hamiltonian, 42, 269, 277, 278
 - Ising model equivalence, 43
 - simple, 253
 - graph
 - cyclomatic number, 640
 - lattice constant, 640
 - partially-ordered sequence, 647
 - Green's function, 52, 711
 - loose-packed, 39, 624
 - slice, 345
 - unit, 265
 - Law of corresponding states, 3, 94, 152, 175
 - Liedermann's theorem, 501
 - Lenz-Ising model, 37
 - Liquid crystal, 2
 - Locally-ordered regions, 95
 - Loewner evolution, 164
 - Loewner's equation, 163, 165
 - Logarithmic singularity, 100, 115, 155, 379, 510
 - Lommel function, 709
 - Long-range order, quasi, 50
 - Loop-erased random walk, 163, 166
 - Lord Kelvin (William Thomson), 57
 - Lower critical point, 324
- M**
- Maclaurin method, 537
 - Magnetic field
 - longitudinal, 30
 - transverse, 30
 - Magnetic graph, 532, 738
 - Marginal
 - coupling, 491
 - parameter, 202
 - Maxithermal point, 342
 - Maxwell's equal areas rule, 171, 207
 - Maxwell-Boltzmann distribution, 14
 - Medial lattice, 164, 311
 - Mermin-Wagner theorem, 31, 34, 35, 49, 50, 530
 - Metamagnet, 90, 91
 - Metastable, 173
 - Microstate, 13
 - Migdal-Kadanoff transformation, 600
 - Modified amplitude function, 726
 - Modified KDP model, 84
 - Moment expansion, 523
 - Moment-cumulant relations, 537
 - Monolayer, 606
 - Monomer, 81
 - Monte Carlo renormalization, 616
 - Monte Carlo simulation, 616
 - Multicritical phenomena, 91
 - Multiphase coexistence, 615
- N**
- Néel temperature, 40, 92, 216, 605
 - Neville-Aitkin polynomial interpolation, 558
 - Non-commuting operators, 15
- O**
- One-hypercube approximation, 607, 608
 - Order, 2, 263
 - long-range, 363
 - parameter, 95, 96, 176, 263, 265, 523
 - symmetry group of, 96
 - short-range, 2, 97, 363
 - spatial, 2
 - Orientable molecule, 327
 - Orthogonal group $O(n)$, 676
- P**
- Padé approximant, defective, 561
 - Padé table, 746
 - Padé's principal theorem, 747
 - Pair-interaction models, 22
 - Paramagnetic state, 41
 - Particle-hole symmetry, 341
 - Partition function, 14, 347, 367
 - canonical, 260
 - grand, 259
 - zeros of, 254, 271
 - Pauli matrices, 436, 437, 445

- Peculiarities in renormalization group transformations, 573
 - Peierls argument, 264
 - Peierls method, 253
 - Periodic boundary conditions, 627
 - Perron's theorem, 346, 369, 426
 - Pfaff's theorem, 496
 - Pfaffian, 496
 - Phase, 2
 - boundary, 90
 - coexistence, 3
 - diagram, 90
 - energetically stabilized, 330
 - entropically stabilized, 330
 - separation
 - in binary alloys, 44
 - transition
 - commensurate-incommensurate, 59
 - first-order, 2, 90, 105, 183, 189, 515, 581, 584, 606, 615
 - higher-order, 3
 - in one-dimensional quantum Ising model, 353
 - incipient, 354
 - infinite-order, 516
 - K-type, 96, 379, 510, 515
 - lower temperature bound, 264
 - O-type, 96, 510, 514
 - second-order, 3, 514, 516, 615
 - Phenomenological renormalization, 353, 610
 - Plane rotator model, 34
 - Point of accumulation, 379
 - Poisson integral formula, 708
 - Poisson summation formula, 478, 707
 - Polarization, 516
 - Potassium dihydrogen phosphate, 61
 - Potts delta function, 530
 - Potts model, 355, 376, 518, 522, 613
 - ν -state, 184
 - 3-state, 163, 183, 184, 379, 491, 606, 609, 615
 - transfer matrix, 360
 - duality, 530
 - symmetry group, 690
 - antiferromagnetic square-lattice 3-state, 80
 - chiral, 58, 59
 - conjectured exponents, 490
 - critical and F model, 489
 - critical exponents, 163
 - critical temperature
 - square lattice, 289
 - duality, invariant circle, 289
 - first-order transition in, 187, 490
 - high-temperature series variable, 530
 - internal energy, 489
 - latent heat, 490
 - partition function, 524
 - planar, 54
 - self-dual, 306
 - spontaneous polarization, 490
 - staggered six-vertex model equivalence, 55
 - standard, 55
 - symmetry group, 55
 - two-state, 57
 - vector, 54
 - Power series, normal, 561
 - Pseudo-lattice, 654
 - Pseudo-spinodal line, 565
 - Pseudocritical values, 153
- Q**
- Quantum systems, 14
 - quasi-periodicity, 261
 - Quasi-chemical method, 229, 654
 - Quasi-periodic functions, 714
 - Quasi-static, 1
- R**
- Radius of convergence, 521
 - Range of interaction, 277
 - Rapidity, 388, 635, 729
 - lattice, 62
 - lines, 629
 - property, 634
 - Real-space renormalization group (RSRG), 283, 520
 - Reciprocal lattice, 627
 - Rectilinear diameter law, 43, 336
 - Recurrence relationships, 570, 579, 586, 599, 605
 - Reduction hypothesis, 120
 - Relaxation time, 174
 - Renormalization group, 120, 151, 202
 - Representation
 - coupling-density, 9, 18, 89
 - coupling-extensive variable, 9
 - field-density, 8
 - field-extensive variable, 8
 - Response function, 10, 15, 89, 137, 143, 153, 523, 579
 - Resultant of two polynomials, 696
 - Rhombic embedding, 636

Riemann surface, 699
 Rotational invariance, 27
 Ruelle's theorem, 254, 273, 274, 693

S

Sackur-Tetrode equation, 176

Scaling

and Landau theory, 182
 approach
 strong, 129, 136
 weak, 129, 136
 corrections to, 155
 correlation function, 134
 dimension, 118, 134
 exponent, 3, 101
 field, 100
 generalized, 105
 irrelevant, 100
 marginal, 120
 relevant, 100
 hypothesis, 491
 Kadanoff, 100, 101
 Nightingale–'T Hooft, 109, 578
 law, 3
 Essam–Fisher, 133, 145, 202, 317, 481
 Fisher, 135, 200, 204, 481
 Widom, 133, 145, 200, 202, 317, 481
 operator, 117
 power-law, 148
 theory, 3, 182, 482, 567
 transformation, 101

Schramm–Loewner evolution, *see* SLE, 3

Schur orthogonality relations, 680

Self-avoiding walk, 34

Self-dual Potts model, 284

Semi-group, 570

Series

n -vector model, 543
 duality variable, 532
 code method, 517
 cumulant method, 518
 Darboux's n -th approximation, 557
 finite-cluster method, 540
 finite-lattice method, 230, 518
 first-order transition, 519
 free energy (cumulant), 536
 general-spin Ising model, 529
 graph-specific contributions, 528
 high-temperature, 521
 high-temperature/low-temperature duality, 518, 534

homographic transformation, 558

late terms, 556

low-temperature, 521

N -dependence of coefficients, 525

 ground-state term, 525

 Potts model, 526

 variable, 525

low-temperature susceptibility, 517

methods, 282, 521

partial-sum method, 517

partition function (moment), 536

perturbations of the ground state, 526

second-order transition, 519

shadow method, 517

use of test functions, 556

Shading rule, 62

Shape ratio, 152

Shift exponent, 126

Simplex, 265

Sink, 570, 576, 587

Site percolation, 166

Site-variable model, 22

Six-vertex model, 23, 33, 61, 84, 474, 485

 K-type transition, 514

 antiferroelectric, 516

 eight-vertex mapping, 299

 ferroelectric, 514

 frozen state, 263

 ground state, 507

 first-order transition, 486

 frozen completely-ordered ferroelectric state, 486

 staggered, 187, 613

 XXZ model equivalence, 486

Sixteen-vertex model, 60, 282

SLE

 central charge, 165

 chordal, 164

 radial, 164

Solid-on-solid model, 53

Spanning graph, 81

Spatially fluctuating variables, 167

Spherical model, 34, 156

 phase transition in, 36

Spin waves, 530

Spin- $\frac{1}{2}$ Ising model

 line of first-order transitions, 565

Spin-wave stiffness, 593

Spinodal curve, 173

SQK model, 515

Square ice

 zero-point entropy

 Pauling's estimate, 488

Lieb's formula, 488
 Nagle's approximation, 488
 Square lattice, self-duality, 628
 Staggered field, 136
 Star, 532
 Star graphs, 741
 Star-triangle conditions
 generalized, 383
 Star-triangle transformation, 255, 283, 514, 588
 State operator, 23
 Strong embedding, 640
 Sublattice
 layered, 212
 order, 40
 Sublattices, interpenetrating, 39
 Superconductivity, 31
 Superconductor, 204
 two-dimensional, 49
 Supercooled, 174
 Superfluid transition in ^4He films, 49
 Superfluidity, 31
 Superposition polygon, 498
 Supersaturated, 174
 Surface effects, 153
 Susceptibility, 10
 Symmetry group
 one-dimensional symmetric representation, 358
 Syozi dilution model, 318

T

T matrix, 541, 646
 Temperature scales, 91
 Theory of capillarity, 94
 Thermal expansion, coefficient of, 10
 Thermodynamic limit, 2, 9, 89, 260, 524
 Thickness, 152, 609
 Thermodynamics
 reduce to statistical mechanics, 1
 TQ equation, *see* VQ equation, 450
 Transfer matrix, 153, 282, 346, 579, 611
 corner, 282
 eight-vertex model, 33, 346
 Ising model, 346
 reduction to block-diagonal form, 355
 Transformation
 eight-vertex and Ising models, 284
 Legendre, 6, 20
 lower-bound, 599, 600
 staggered six-vertex model and ν -state Potts model, 284

 upper-bound, 595, 597
 Transition
 Ehrenfesti classification, 2
 Fisher's classification, 2
 region, 2, 95
 surface, 299
 Translational invariance, 27, 260
 Tricritical exponents
 classical, 192
 Tricritical point, 91, 92, 94, 136, 188, 190, 606
 in metamagnet, 225
 Triple
 line, 93, 136, 188, 189
 point, 91
 Trivial coupling, 24, 567

U

Unbiased estimate, 555
 Universality, 153, 481
 class, 29, 95, 490, 491, 615
 hypothesis, 95, 148, 149, 152, 491
 Upper critical dimension, 203, 616

V

van der Waals equation, 20, 167
 van Hove criterion, 152, 261
 Vapour-liquid transition, 91
 Variable
 density, 5
 extensive, 5
 field, 5
 intensive, 5
 Variational approximation, 607
 Vector-invariant, 357
 Vertex models, 23
 non-negative transfer matrix, 673
 three dimensional, 60
 Villain model, 53
 Virasoro algebra
 central charge, 3, 162
 Vitali convergence theorem, 271
 Vortex distribution function, 52
 Vortex in XY model, 50
 VQ equation, 450

W

Water
 molecule, 61
 triple point, 91

Water model, 80, 337
 compressibility minimum in, 251
 density maximum in, 251
 dielectric properties, 251
 Monte Carlo methods for, 252
 RSRG methods for, 252
Water phases, 91
Water-nicotine, 326, 332
Weak embedding, 640
Weak scaling laws, 492, 594
Weak universality, 492, 564
Weak-graph transformation, 255, 284, 386, 478
Wegner transformation, 255
Weight function, 568, 571, 607
 decimation, 571, 573, 579
 Kadanoff, 599
 Kadanoff–Houghton, 572, 608
 majority-rule, 571
Weiss field equation, 209
Whitney polynomial, 56, 638

X

XXZ Heisenberg chain, 474

XXZ model, 441
XY model, 441, 560
 two-dimensional, 711
 classical, 34
 harmonic approximation, 46
 one-dimensional classical, 49
 quantum, 33
 three-dimensional classical, 49
 two-dimensional, 588
 two-dimensional classical, 49
XYZ model, 441
 eight-vertex model equivalence, 441

Y

Yang–Baxter equations, 305
Yang–Lee circle theorem, 278

Z

Zero-field graph, 532
Zero-point entropy, 605
Zeroth-order methods, 230