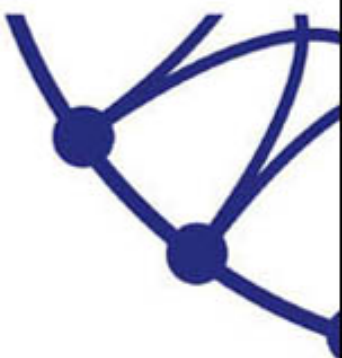# PHASETRANSITIONS

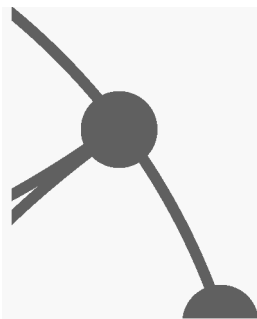## A Brief Account with Modern Applications

Moshe Gitterman
Vivian (Haim) Halpern
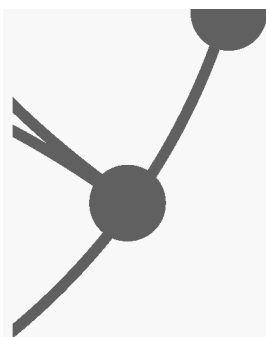
PHASETRANSITIONS
A Brief Account with Modern Applications

This page intentionally left blank

PHASETRANSITIONS
A Brief Account with Modern Applications

Moshe Gitterman
Vivian (Haim) Halpern

Bar-Ilan University, Israel

World Scientific

PHASE TRANSITIONS
A Brief Account with Modern Applications

Contents

This page intentionally left blank

Preface

This book is based on a short graduate course given by one of us (M.G) at New York University and at Bar-Ilan University, Israel. The decision to publish these lectures as a book was made, after some doubts, for the following reason. The theory of phase transitions, with excellent agreement between theory and experiment, was developed some forty years ago culminating in Wilson's Nobel prize and the Wolf prize awarded to Kadanoff, Fisher and Wilson. In spite of this, new books on phase transitions appear each year, and each of them starts with the justification of the need for an additional book. Following this tradition we would like to underline two main features that distinguish this book from its predecessors.

Firstly, in addition to the five pillars of the modern theory of phase transitions (Ising model, mean field, scaling, renormalization group and universality) described in Chapters 2–5 and in Chapter 7, we have tried to describe somewhat more extensively those problems which are of major interest in modern statistical mechanics. Thus, in Chapter 6 we consider the superfluidity of helium and its connection with the Bose–Einstein condensation of alkali atoms, and also the general theory of superconductivity and its relation to the high temperature superconductors, while in Chapter 7 we treat the x–y model associated with the theory of vortices in superconductors. The short description of percolation and of spin glasses in Chapter 8 is complemented by the presentation of the small world phenomena, which also involve short and long range order. Finally, we consider in Chapter 9 the applications of critical phenomena to self-organized

criticality in scale-free non-equilibrium systems. While each of these topics has been treated individually and in much greater detail in different books, we feel that there is a lot to be gained by presenting them all together in a more elementary treatment which emphasizes the connection between them. In line with this attempt to combine the traditional, well-established issues with the recently published and not yet so widely known and more tentative topics, our fairly short list of references consists of two clearly distinguishable parts, one related to the classical theory of the sixties and seventies and the other to the developments in the past few years. In the index, we only list the pages where a topic is discussed in some detail, and if the discussion extends over more than one page then only the first page is listed.

We hope that simplicity and brevity are the second characteristic property of this book. We tried to avoid those problems which require a deep knowledge of specialized topics in physics and mathematics, and where this was unavoidable we brought the necessary details in the text. It is desirable these days that every scientist or engineer should be able to follow the new wide-ranging applications of statistical mechanics in science, economics and sociology. Accordingly, we hope that this short exposition of the modern theory of phase transitions could usefully be a part of a course on statistical physics for chemists, biologists or engineers who have a basic knowledge of mathematics, statistical mechanics and quantum mechanics. Our book provides a basis for understanding current publications on these topics in scientific periodicals. In addition, although students of physics who intend to do their own research will need more basic material than is presented here, this book should provide them with a useful introduction to the subject and overview of it.

Mosh Gitterman & Vivian (Haim) Halpern
January 2004

Chapter 1

Phases and Phase Transitions

In discussing phase transitions, the first thing that we have to do is to define a phase. This is a concept from thermodynamics and statistical mechanics, where a phase is defined as a homogeneous system. As a simple example, let us consider instant coffee. This consists of coffee powder dissolved in water, and after stirring it we have a homogeneous mixture, i.e., a single phase. If we add to a cup of coffee a spoonful of sugar and stir it well, we still have a single phase — sweet coffee. However, if we add ten spoonfuls of sugar, then the contents of the cup will no longer be homogeneous, but rather a mixture of two homogeneous systems or phases, sweet liquid coffee on top and coffee-flavored wet sugar at the bottom.

In the above example, we obtained two different phases by changing the composition of the system. However, the more usual type of phase transition, and the one that we will consider mostly in this book, is when a single system changes its phase as a result of a change in the external conditions, such as temperature, pressure, or an external magnetic or electric field. The most familiar example from everyday life is water. At room temperature and normal atmospheric pressure this is a liquid, but if its temperature is reduced to below 0°C it will change into ice, a solid, while if its temperature is raised to above 100°C it will change into steam, a gas. As one varies both the temperature and pressure, one finds a line of points in the pressure–temperature diagram, Fig. 1.1, along which two phases can exist in equilibrium, and this is called the coexistence curve.
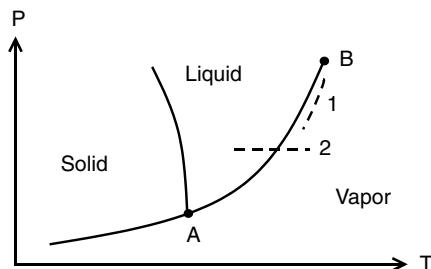
Fig. 1.1   The phase diagram for water.

We now consider in more detail the change of phase when water boils, in order to show how to characterize the different phases, instead of just using the terms solid, liquid or gas. Let us examine the density $\rho(T)$ of the system as a function of the temperature T. The type of phase transition that occurs depends on the experimental conditions. If the temperature is raised at a constant pressure of 1 atmosphere (thermodynamic path 2 in Fig. 1.1), then initially the density is close to $1\,\mathrm{g/cm^3}$, and when the system reaches the phase transition line (at the temperature of 100°C) a second (vapor) phase appears with a much lower density, of order $0.001\,\mathrm{g/cm^3}$, and the two phases coexist. After crossing this line, the system fully transforms into the vapor phase. This type of phase transition, with a discontinuity in the density, is called a first order phase transition, because the density is the first derivative of the thermodynamic potential. However, if both the temperature and pressure are changed so that the system remains on the coexistence curve AB (thermodynamic path 1 in Fig. 1.1), one has a two-phase system all along the path until the critical point B ($T_c = 374$°C, $p_c = 220$ atm.) is reached, when the system transforms into a single ("fluid") phase. The critical point is the end-point of the coexistence curve, and one expects some anomalous behavior at such a point. This type of phase transition is called a second order one, because at the critical point B the density is continuous and only a second derivative of the thermodynamic potential, the thermal expansion coefficient, behaves anomalously. Anomalies in thermodynamical quantities are the hallmarks of a phase transition.

Phase transitions, of which the above is just an everyday example, occur in a wide variety of conditions and systems, including some in fields such as economics and sociology in which they have only recently been recognized as such. The paradigm for such transitions, because of its conceptual simplicity, is the paramagnetic–ferromagnetic transition in magnetic systems. These systems consist of magnetic moments which at high temperatures point in random directions, so that the system has no net magnetic moment. As the system is cooled, a critical temperature is reached at which the moments start to align themselves parallel to each other, so that the system acquires a net magnetic moment (at least in the presence of a weak magnetic field which defines a preferred direction). This can be called an order–disorder phase transition, since below this critical temperature the moments are ordered while above it they are disordered, i.e., the phase transition is accompanied by symmetry breaking. Another example of such a phase transition is provided by binary systems consisting of equal numbers of two types of particle, A and B. For instance, in a binary metal alloy with attractive forces between atoms of different type, the atoms are situated at the sites of a crystal lattice, and at high temperatures the A and B atoms will be randomly distributed among these sites. As the temperature is lowered, a temperature is reached below which the equilibrium state is one in which the positions of these atoms alternate, so that most of the nearest neighbors of an A atom are B atoms and vice versa.

The above transitions occur in real space, i.e., in that of the spatial coordinates. Another type of phase transition, of special importance in quantum systems, occurs in momentum space, which is often referred to as k-space. Here, the ordering of the particles is not with respect to their position but with respect to their momentum. One example of such a system is superfluidity in liquid helium, which remains a liquid down to $0\,\mathrm{K}$ (in contrast to all other liquids, which solidify at sufficiently low temperatures and high pressures) but at around $2.2\,\mathrm{K}$ suddenly loses its viscosity and so acquires very unusual flow properties. This is a result of the fact that the particles tend to be in a state with zero momentum, $k = 0$, which is an ordering in k-space. Another well-known example is superconductivity. Here, at

sufficiently low temperatures electrons near the Fermi surface with opposite momentum link up to form pairs which behaves as bosons with zero momentum. Their motion is without any friction, and since the electrons are charged this motion results in an electric current without any external voltage.

Phase transitions occur in nature in a great variety of systems and under a very wide range of conditions. For instance, the paramagnetic–ferromagnetic transition occurs in iron at around $1000\,\mathrm{K}$, the superfluidity transition in liquid helium at $2.2\,\mathrm{K}$, and Bose–Einstein condensation of atoms at $10^{-7}\,\mathrm{K}$. In addition to this wide temperature range, phase transitions occur in a wide variety of substances, including solids, classical liquids and quantum fluids. Therefore, phase transitions must be a very general phenomenon, associated with the basic properties of many-body systems. This is one reason why the theory of phase transitions is so interesting and important. Another reason is that thermodynamic functions become singular at phase transition points, and these mathematical singularities lead to many unusual properties of the system which are called "critical phenomena". These provide us with information about the real nature of the system which is not otherwise apparent, just as the behavior of a poor man who suddenly wins a million-dollar lottery can show much more about his real character than one might deduce from his everyday behavior. A third reason for studying phase transitions is scientific curiosity. For instance, how do the short-range interactions between a magnetic moment and its immediate neighbors lead to a long-range ordering of the moments in a ferromagnet, without any sudden external impetus? A similar question was raised (but not answered) by King Solomon some 3000 years ago, when it was written (Proverbs 30, 27). "The locusts have no king, yet they advance together in ranks".

1.1   Classification of Phase Transitions

The description and analysis of phase transitions requires the use of thermodynamics and statistical physics, and so we will now summarize the thermodynamics of a many-body system [1]. In

thermodynamics each state of a system is defined by some characteristic energy. If the state of the system is defined by its temperature T and its pressure P or volume V, this energy is called the free energy. One part of this energy is the energy E of the system at zero temperature, while the other part depends on the temperature and the entropy S of the system. If the independent variables are the temperature and pressure, then the relevant thermodynamic potential is the Gibbs free energy $G = E - TS + PV$, while if they are the temperature and volume it is the Helmholtz free energy $F = E - TS$. The differentials of these free energies for a simple system are

$$dG = -SdT + VdP, \qquad dF = -SdT - PdV. \qquad (1.1)$$

If the system has a magnetic moment there is an extra term $-MdH$ in the above expressions, and if the number N of particles is variable we must add the term μdN, where μ is the chemical potential. Then the first derivatives of the free energy give us the values of physical properties of the system such as the specific volume ($V/N = [1/N]\partial G/\partial P$), entropy ($S = -\partial G/\partial T$) and magnetic moment ($M = -\partial G/\partial H$), while its second partial derivatives give properties such as the specific heat ($C_p = T\partial S/\partial T = -T\partial^2 G/\partial T^2$), the compressibility and the magnetic susceptibility of the system.

Let us now consider the effect on the free energy G of changing an external parameter, for instance the temperature. Such a change cannot introduce a sudden change in the energy of the system, because of the conservation of energy. Hence, if we consider the free energy per unit volume, g, of a system with a fixed number of particles and write $G = gV$, there are only two possibilities. Either the change δG in G arises from a change in the free energy density $g, \delta G = V \delta g$, or it comes from a change in the volume $V, \delta G = g \delta V$. When the properties of a system change as a result of a phase transition, they can undergo a small change δg all over the system at once or initially only in some parts δV of it, as shown in Fig. 1.2. If the new phase appears as $\delta G = g\delta V$, so that it appears only in parts δV of the system, then it requires the formation of stable nuclei, namely of regions of the new phase large enough for them to grow rather than to shrink. Since the energy consists of a negative volume term
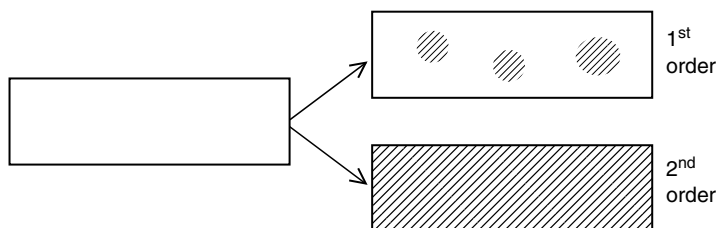
Fig. 1.2   The two different possibilities for the change δG in the free energy associated with 1st order and 2nd order phase transitions.

and a positive surface one, which for a spherical nucleus of radius r are proportional to r^3 and r^2 respectively, this critical size r_c is that for which the volume term equals the surface one, so that for $r > r_c$ growth of the nucleus leads to a decrease in its energy. Because of the need for nucleation, the first phase can coexist with the second phase, in a metastable state, even beyond the critical temperature for the phase transition. This is a first order phase transition. The best known manifestations of such a transition are superheating and supercooling.

In the other case, where the phase transition occurs simultaneously throughout the system, $\delta G = V \delta g$. Although the difference δg between the properties of these phases is small, the old phase which occupied the whole volume cannot exist, even as a metastable state, on the other side of the critical point, and it is replaced there by a new phase. These two phases are associated with different symmetries. For instance, in the paramagnetic state of a magnetic system there is no preferred direction, while in the ferromagnetic state there is a preferred direction, that of the total magnetic moment. In this case, the critical point is the end-point of the two phases, and so there must be some sudden change there, i.e., some discontinuity in their properties. This is an example of a second order phase transition.

Phase transitions are classified, as proposed by Ehrenfest, by the order of the derivative of the free energy which becomes discontinuous (or, in modern terms, exhibits a singularity) at the phase transition temperature. In a first order phase transition, a first derivative becomes discontinuous. A common example of this is the transition

from a liquid to a gas when water boils, where the density shows a discontinuity. In a second order phase transition, on the other hand, properties such as the density or magnetic moment of the system are continuous, but their derivatives (which correspond to the second derivatives of the free energy), such as the compressibility or the magnetic susceptibility, are discontinuous. In this book, we will be concerned mainly with second order phase transitions, with which are associated many unusual properties.

1.2   Appearance of a Second Order Phase Transition

Before proceeding to a detailed mathematical analysis, it is worthwhile to consider qualitatively an example of how a second order phase transition can occur. Accordingly, we will now discuss the mean field theory of the paramagnetic–ferromagnetic phase transition in magnetic materials, originally proposed by Pierre Weiss some 100 years ago, in 1907 [2]. This consists of the sudden ordering of the magnetic moments in a system as the temperature is lowered to below a critical temperature T_c. He suggested that these materials consist of particles each of which has a magnetic dipole moment μ. For N such particles, the maximum possible magnetic moment of the system is $M_0 = N\mu$, when the moments of all the particles are aligned. Such a state is possible at $T = 0\,\mathrm{K}$, when there is no thermal energy to disturb the orientation of the moments. In the presence of a small magnetic field $\mathbf{H}$, the energy of a dipole of moment $\boldsymbol{\mu}$ is $-\boldsymbol{\mu} \cdot \mathbf{H}$. For the sake of simplicity, we consider only two possible orientations of the dipoles, parallel and anti-parallel to the field, or up and down, and denote by N_+ and N_- respectively the number of dipoles in these two orientations at any given temperature. Similar results can be obtained if one allows the moments to adopt arbitrary orientations with respect to the field, but the analysis is slightly more complicated. Then the total magnetic moment of the system in the direction of the field is $M = (N_+ - N_-)\mu$, and its energy is $E = -MH$. The main assumption of Weiss was that there is some internal magnetic field acting on each of the dipoles, and that this

field is proportional to M/M_0. This is a very reasonable assumption if the internal field on a given particle is due to the magnetic moments of the surrounding particles. Of course, it is an approximation to assume that each particle experiences the same magnetic field, and we will consider more refined theories later. We therefore write the effective field acting on each dipole (in the absence of an external field) in the form $H_m = CM/M_0$. According to Boltzmann's law, which was already known by then, the number of particles $N_\pm$ with moments pointing up and down at temperature T is proportional to $\exp(\mp\mu H_m/kT)$. Here and throughout the book, we denote the Boltzmann constant by k. It readily follows that

$$\frac{M}{M_0} = \frac{N_+ - N_-}{N_+ + N_-} = \tanh\left(\frac{T_c}{T}\frac{M}{M_0}\right) \qquad (1.2)$$

where $T_c = C\mu/k$. As can be seen from Fig. 1.3, if $T_c/T < 1$ then this equation only has the trivial solution $M = 0$, since for small arguments $\tanh(x) \simeq x$, and so there is no spontaneous magnetic moment if $T > T_c$. On the other hand, if $T < T_c$ then the equation has two solutions. An examination of the effect of a small change in the internal field shows that the solution with $M > 0$ is the stable one, i.e., the system has a spontaneous magnetic moment and so is ferromagnetic. The critical temperature T_c at which this transition from paramagnetism to ferromagnetism takes place is given by
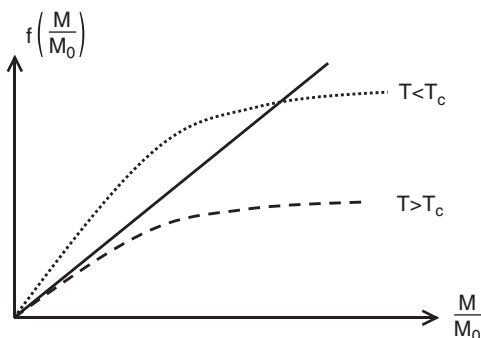
Fig. 1.3   Graphical solution of Eq. (1.2) for the magnetization M in the Weiss mean field model.

$T_c = C\mu/k$, the famous Weiss equation. At the time that Weiss wrote his paper, quantum mechanics, electronic spin and exchange effects had not been discovered, and so he could only estimate the strength of the internal field from the known dipole-dipole interactions, and this led to an estimate of T_c of around $1\,\mathrm{K}$. Weiss knew that for iron T_c is around $1000\,\mathrm{K}$, and so wrote bravely at the end of his paper [2] that his theory does not agree with experiments but future research will have to explain this discrepancy of three orders of magnitude. In spite of this discrepancy, his paper was accepted for publication, and we now know that his ideas of the nature of the paramagnetic–ferromagnetic phase transition are qualitatively correct.

1.3   Correlations

For a second order phase transition, a second derivative of the free energy diverges as the phase transition is approached. For instance, the magnetic susceptibility $\partial M/\partial H = -\partial^2 G/\partial H^2$ tends to infinity as $T \to T_c$. Now according to the fluctuation–dissipation theorem [1], the magnetic susceptibility is proportional to the integral over all space of the average of the product of the magnetic moment at two points distance r apart, which describes the correlation between the magnetic moments at these points,

$$\frac{\partial M}{\partial H} \sim \int \langle M(0)M(r) \rangle \, d\tau. \qquad (1.3)$$

In general, the magnetic moment (spin) at any site tends to align the spin at an adjacent site in the same direction as itself, so as to lower the energy. However, this tendency is opposed by that of the entropy, so that far from the critical point there is a finite correlation length ξ such that

$$\langle M(0)M(r) \rangle \sim \exp(-r/\xi).$$

Here, the correlation length ξ has the following physical meaning. If one forces a particular spin to be aligned in some specified direction, the correlation length measures how far away from that spin the other spins tend to be aligned in this direction. In the disordered

state, the spin at a given point is influenced mainly by the nearly random spins on the adjacent points, so that the correlation length is very small. As the phase transition is approached, the system has to become "prepared" for a fully ordered state, and so the "order" must extend to larger and larger distances, i.e., ξ has to grow. However, the divergence of the integral in Eq. (1.3) as $T \to T_c$ implies that at the critical point the correlation function cannot decrease exponentially with distance r, but rather must decay at best as an inverse power of r,

$$\langle M(0)M(r)\rangle \sim r^{-\gamma}, \quad \gamma \le 3. \tag{1.4}$$

This is a point of great physical significance. It means that near the critical point not only do we not have any small energy parameter, since the critical temperature is of the same order of magnitude as the interaction energy, but also we do not have any typical length scale since the correlation length diverges on approaching the critical point. In other words, all characteristic lengths are equally important near the critical point, which makes this problem extremely complicated. A similar situation of various characteristic lengths arises in the problem of the motion of water in an ocean, but these are associated with different phenomena. The Angstroms–micron length scale is appropriate for studying the interactions between water molecules, but one must take into account lengths of order of meters for studying the tides and the kilometer length scale for studying the ocean streams. This is in contrast to the situation near critical points, where one cannot perform such a separation of different length scales.

We now return to the question raised at the beginning of this chapter, namely how one can obtain long-range correlations from short-range interactions. In mathematical terms, the question is how an exponentially decaying correlation can transfer the mutual influence of different atoms located far away from each other. A qualitative answer to this question has been given by Stanley [3]. The correlations between two particles far apart do indeed decay exponentially. However, the number of paths between these two particles along which the correlations occur increase exponentially. The exponents of these two exponential functions, one positive and one

negative, compensate each other at the critical point, and this leads to the long-range power law correlations. By contrast, for a one-dimensional system the exponential increase corresponding to the number of different paths is replaced by unity, and so the negative exponent leads to the absence of ordering and so to no phase transition for non-zero temperatures.

Curiously enough, in the Red Army of the former Soviet Union the order given by a officer standing in front of the line of soldiers was "Attention! Look at the chest of the fourth man!". For some unclear reason, they decided that the correlation length is equal to four, and the soldiers will be ordered in a straight line if each one will align with his fourth neighbor in the row.

1.4   Conclusion

Phase transitions are very general phenomena which occur in a great variety of systems under very different conditions. They can be divided into first-order and second-order transitions depending on which derivatives of the free energies have anomalies at the transition. The existence of phase transitions, as such, was established a hundred years ago in the framework of the mean field theory. Three major factors which present severe difficulties for the theoretical description of phase transitions are the non-analyticity of the thermodynamic potentials, the absence of small parameters, and the equal importance of all length scales.

This page intentionally left blank

The Ising Model

We now consider a microscopic approach to phase transitions, in contrast to the phenomenological approach used in the previous chapter. In this approach, following Gibbs, we start with the interaction between particles. The first step is then to calculate the mechanical energy of the system E_n in each state n of all the particles, a problem which in general is far from trivial for a system of 10^{23} particles. In the framework of classical and quantum mechanics we must then calculate the partition function

$$Z = \sum_n \exp\left(-\frac{E_n}{kT}\right), \qquad Z = \text{Tr}\left[\exp\left(-\frac{H}{kT}\right)\right] \qquad (2.1)$$

respectively where H is the system Hamiltonian, and the Helmholtz free energy F of the system,

$$F = -kT \ln Z. \qquad (2.2)$$

For a large enough system, one can replace the summation over n in Eq. (2.1) by an integration over phase space, so that for non-interacting particles the integral over the coordinates of all the N particles equals V^N. It follows that $F = -NkT \ln(CV)$, where C is independent of V. On using Eq. (1.1), we find that

$$P = -\left(\frac{\partial F}{\partial V}\right)_T = NkT/V,$$

which is just the equation of state of an ideal gas. This equation involves only two degrees of freedom, instead of the around 10^{23} degrees of freedom of the original system. In this way, one can proceed from a detailed microscopic description of the system to a simple thermodynamic description. However, this method only applies to systems in equilibrium, while for non-equilibrium systems there is no unique approach, a point that will be considered in Chapter 9. Even for systems in equilibrium this description is only simple in principle, but not in practice since it involves first calculating the mechanical energy of the numerous different possible many-particle states and then a summation (or integration for continuous variables) over all the possible states. Only for some special simple systems it is possible to perform the calculations exactly, but it is very instructive to do so for such systems and examine the results that are obtained.

Let us mention that before the seminal work of Onsager [4], it was not at all clear whether statistical mechanics is able to describe the phenomena of phase transitions, i.e., how the "innocent" expression involving T in Eq. (2.1) will lead to non-trivial singularities at some specific temperature. The answer lies in the fact that the singularities appear only for a system of infinite size (the thermodynamic limit) which has an infinite number of configurations. It is just the infinite number of terms in the sum which appears in Eq. (2.1) that can lead to singularities.

One of the simplest model systems is the so-called Ising model, which we will now examine. This model, which will be discussed extensively in this book, is based on the following three assumptions:

(1) The objects (which we call particles) are located on the sites of a crystal lattice.
(2) Each particle i can be in one of two possible states, which we call the particle's spin S_i, and we choose $S_i = \pm 1$.
(3) The energy of the system is given by

$$E = -J \sum_{(i,j)nn} S_i S_j \qquad (2.3)$$

where J is a constant and the sum is over all pairs of adjacent particles, i.e., over all pairs of nearest neighbors i and j. Thus in a linear lattice each particle interacts only with its two nearest neighbors, in a square lattice with its four nearest neighbors, and in a simple cubic lattice with its six nearest neighbors.

In spite of its simplicity, the Ising model is used in many applications where an object can be in one or two states, such as sites occupied by A or B atoms, sites containing a particle or a hole, two possible conformations, and even votes for one of two political parties in elections. One can generalize the first two basic assumptions of the Ising model. A larger number of possible states is taken into account in the so-called Potts model [5], and one can consider not only interactions between nearest neighbors but also interactions between second nearest neighbors, third nearest neighbors, etc. These generalizations make the problem much more complicated. On the other hand, the lattice approximation seems to be of no importance for phase transition problems since, as we will see later, much longer distances (of order of the correlation length) are important in phase transitions.

The personal story of Ising is also of interest [6]. In 1924–1926, Ising was a doctoral student of the famous German physicist Wilhelm Lenz, who suggested to him this model. In his thesis, Ising showed that the system does not exhibit a phase transition in one dimension, which is correct. He also showed that there is no phase transition in two dimensions, which was shown much later (by Onsager in 1944) to be incorrect. After completing his thesis he started to work as a high-school teacher, and with the rise to power of the Nazis he went to Luxembourg. We next hear of him in 1948 in the USA, where he taught at some small university. He died in 1990, and during his whole career published only two scientific papers, one on his thesis work and the other entitled "Goethe as a physicist". However, on his arrival in America in 1948 he found big placards announcing a conference on the Ising model, and this model is still being extensively studied as a paradigm of a simple system exhibiting a well-defined phase transition.

2.1 1D Ising model

We consider first the one-dimensional (1D) Ising model, and examine how to calculate the partition function. Let us assume that the system contains N particles, for which the partition function has the following form

$$Z_N = \sum_{S_j=\pm1} \exp\left[-\frac{J}{kT}(S_1S_2 + S_2S_3 + \cdots + S_{N-1}S_N)\right]. \quad (2.4)$$

For a system with $N + 1$ particles, the sum contains one additional term, so that

$$Z_{N+1} = Z_N \sum_{S_N,S_{N+1}=\pm1} \exp\left(-\frac{J}{kT}S_NS_{N+1}\right). \quad (2.5)$$

Since $S_NS_{N+1} = \pm1$, it follows that $Z_{N+1} = 2\cosh(J/kT)Z_N$, and so (since $Z_1 = 2$), by induction,

$$Z_N = 2^N \left(\cosh \frac{J}{kT}\right)^{N-1}. \quad (2.6)$$

Hence if $N \gg 1$ the free energy is

$$G = -kT \ln Z = -NkT \ln \left(2\cosh \frac{J}{kT}\right). \quad (2.7)$$

Since this is a monotonic function of T, with no singularity except at $T = 0$, the 1D Ising system cannot exhibit a phase transition at any finite temperature.

Another interesting general type of system is the one in which the interactions between the particles are very weak but of infinite range,

$$\varphi(r) = - \lim_{\gamma\to0} \gamma \exp(-\gamma r),$$

in contrast to the strong short-range interactions of the Ising model. For this model there is a phase transition even in one dimension, and we will discuss it in Chapter 7. In fact, it is not necessary that there should be interactions between all the spins in a system for a phase transition to occur. Even a few random long-range interactions

combined with short-range interactions as in the Ising model are sufficient to produce a phase transition, as in the small world model which we discuss in Chapter 8.

2.2   2D Ising model

The exact solution of the 2D Ising model was presented by Onsager in 1944 in a very long and complicated paper [4], and even after a few simplifications, a recent version of the proof took seven pages in the very concisely written classic text of Landau and Lifshitz [1]. However, we will now present some qualitative arguments for analyzing the phase transitions in the 2D Ising model.

Since we want to analyze the role of the energy E and the entropy S in order-disorder phase transitions, it is convenient to consider the Helmholz free energy $F = E - TS$. At all temperatures the stable state corresponds to the minimum value of F. At low temperatures the energy E plays the leading role, and its minima correspond to ordered states. At high temperatures the entropy S dominates F, and so the minima of F are reached when the entropy S is a maximum. Therefore, at some intermediate temperature $T = T_c$ the ordering influence of the energy and disordering influence of the entropy are balanced, and one can qualitatively estimate the critical temperature T_c as

$$T_c \approx \frac{\Delta E}{\Delta S}. \tag{2.8}$$

In order to apply this equation to the 2D Ising model, we consider the two competing states of the Ising lattice shown in Fig. 2.1 [7]. The question arises as to whether the fully ordered state 1 is able to transform spontaneously into state 2, which contains an island of opposite spins with perimeter of length L. For this to happen, the free energy of state 2 must be lower than that of state 1. To examine when this occurs, let us consider the changes in energy and in entropy when such an island of opposite spins is formed. Since at each site on the perimeter of the island the energy is raised by $2J$, the energy required to form the island is $\Delta E = (2J)L$. On the other
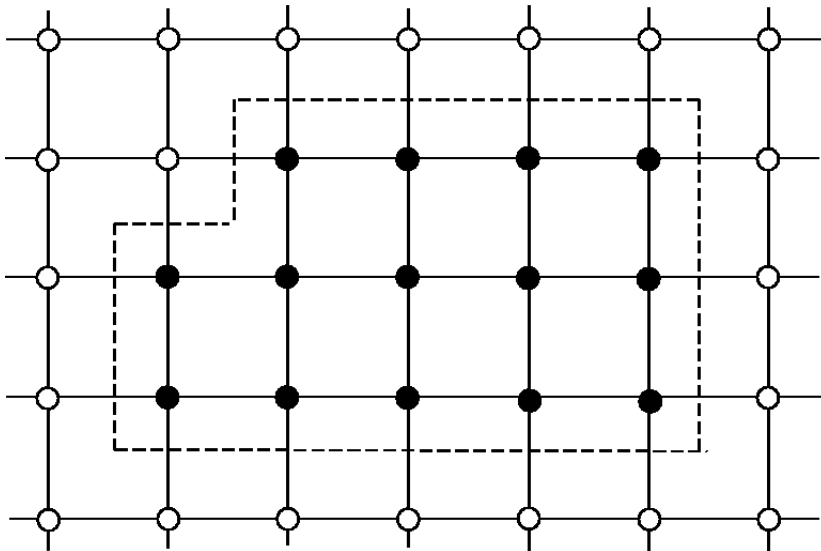
Fig. 2.1   2D Ising square lattice, with an island of opposite spins (indicated by black solid dots) of perimeter L. (B. Liu and M. Gitterman, *Am. J. Phys.* **71** 806 (2003). Copyright 2003 by the American Physical Society.)

hand, the number W of microscopic possibilities for creating such an island is approximately 3^L, since from each site one can continue the perimeter in three different directions, and so the change in entropy $\Delta S \approx \ln(3^L N)$, where the factor N arises from an approximate evaluation of the number of ways in which the initial point can be chosen. This estimate is also approximate because we neglect the necessity to come back to the initial site after L steps, and also the requirement that the perimeter cannot cross itself. In this approximation, the change in the free energy of the system is

$$\Delta F = \Delta E - T\Delta S = L[2J - kT \ln 3] - kT \ln N \qquad (2.9)$$

and for L of order N the term $\ln N$ in this equation can be neglected. Therefore, a disordered state is favored if $T > T_c$, where $kT_c = 2J/\ln 3$. This is not a bad approximation to the exact result found by Onsager [4],

$$kT_c = 2J/\ln(1 + \sqrt{2}),$$

and our analysis shows clearly how the phase transition arises as a result of the competition between energy and entropy. Incidentally, for the 1D case instead of an island of perimeter L one has a change in the direction of the spin at just one point, so that $\Delta E = 2J$. On the other hand, since this point can be anywhere, $W = N$, so that

$$\Delta F = \Delta E - T\Delta S = 2J - kT\ln N,$$

which only becomes negative when $kT_c = 2J/\ln N$, so that $T_c \to 0$ as $N \to \infty$. This is in agreement with our previous conclusion that for an infinite 1D system, a phase transition can only occur at $T = 0$.

There is an important general conclusion from our above analysis of the 2D Ising model. Both our calculations and the exact one of Onsager show that kT_c is close to J, which is the only energy parameter appearing in our problem. Thus, there is no small parameter, which is not a happy situation for a physicist, since the theory of many-body problems usually relies on expressing properties as a power series in a small parameter, such as the ratio of the average potential energy to the average kinetic energy per particle for a weakly non-ideal gas, and the opposite ratio for solids. This is exactly the reason for the absence of a general rigorous theory of liquids, where these two average energies are of the same order of magnitude. It is also one reason why it took so many years to make progress in solving the problem of phase transitions.

Following the little-known article of Svrakic [8], we now present a very simple argument to "derive" exactly the result of Onsager for the 2D Ising model. Instead of considering the whole system and islands of arbitrary size, let us consider only an elementary cell containing just 4 sites. Since on each site the spin can be up or down, there are only $4^2 = 16$ possible spin configurations in such a cell, as shown in Fig. 2.2, and the energies of each can readily be calculated. As can be seen, there are 2 types of "ordered" states, with energies $\pm 4J$, each of which is doubly degenerate (spins up or down), and 12 "disordered" states with energy zero. The names "ordered" and "disordered" are connected with the fact that by putting together either of the first two "ordered" configurations one can construct an isotropic infinite two-dimensional Ising lattice, while this cannot be
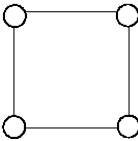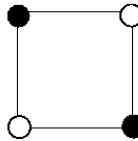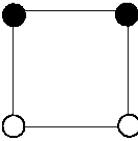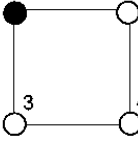
Fig. 2.2 Different configurations of an elementary cell of a 2D Ising square lattice divided into four groups, when the energy E and number of configurations N are shown for each group. Groups a and b contain non-degenerate configurations while degeneracies are inherent in groups c and d. (B. Liu and M. Gitterman, *Am. J. Phys.* **71** 806 (2003). Copyright 2003 by the American Physical Society.)

done for the "disordered" configurations. In accordance with the idea of competition between ordered and disordered states, we conjecture that the phase transition occurs when the partition function of the ordered states,

$$2 \exp\left(\frac{4J}{kT}\right) + 2 \exp\left(\frac{-4J}{kT}\right),$$

equals that of the disordered ones, 12, i.e., when $x^2 + x^{-2} = 6$, where $x = \exp(2J/kT)$. The solution of this equation is $x = 1 + \sqrt{2}$, so that $kT_c = 2J/\ln(1 + \sqrt{2})$, which quite surprisingly gives the exact result of Onsager. However, this method as presented above does not work for a three-dimensional Ising lattice.

2.3 3D Ising model

It is interesting to apply the method described above to the Ising model for three-dimensional lattices, and to find the critical temperature of these lattices by considering just one elementary cell [9]. All the possible $2^8 = 256$ configurations of an elementary cell for a simple cubic lattice are shown in Fig. 2.3, where, as in Fig. 2.2 for the 2D lattice, N represents the number of equivalent configurations

associated with the appropriate group. Let us consider first the different configurations of the 2D Ising lattice shown in Fig. 2.2. We call the configurations of groups a and b "non-degenerate" in the sense that all neighboring states, i.e., those that can be obtained from a
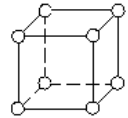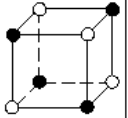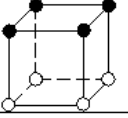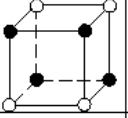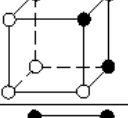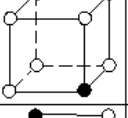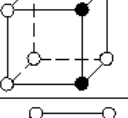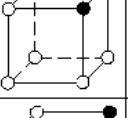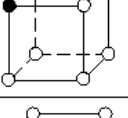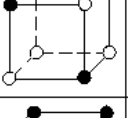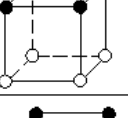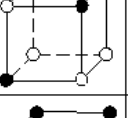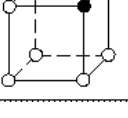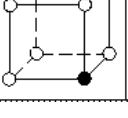
	configuration	N	E	W_G	W_N		configuration	N	E	W_G	W_N
a		2	-12 J	0	1	b		2	12 J	0	1
c		6	-4 J	0	1	d		6	4 J	0	1
e		8	0	0	1	f		8	0	0	1
g		24	0	1	0	h		24	0	1	0
i		16	-6 J	$\frac{2}{8}$	$\frac{6}{8}$	j		16	6 J	$\frac{2}{8}$	$\frac{6}{8}$
k		24	-4 J	$\frac{4}{8}$	$\frac{4}{8}$	l		24	4 J	$\frac{4}{8}$	$\frac{4}{8}$
m		48	-2 J	$\frac{6}{8}$	$\frac{2}{8}$	n		48	2 J	$\frac{6}{8}$	$\frac{2}{8}$

Fig. 2.3 Different configurations of the elementary cell of a 3D Ising cubic lattice with N the number of equivalent configurations in each group, and E their energy. Groups a–f contain non-degenerate (stable) configurations, while the configurations in groups g and h are degenerate (unstable). Groups i–n make a contribution to the partition function of both degenerate and non-degenerate states, which are characterized by the fractional statistical weights W_G and W_N respectively. (B. Liu and M. Gitterman, *Am. J. Phys.* **71** 806 (2003). Copyright 2003 by the American Physical Society.)

given one by the flipping of a single spin, have an energy different from that of the original one. By contrast, all the configurations contained in group c do not change their energy if one of the spins is flipped. Hence, these configurations have adjacent states of the same energy, and with respect to this property we call these configurations "degenerate". The situation is slightly more complicated for the configurations belonging to group d. Here, flipping spin 1 will change the energy by $4J$, while flipping spin 4 will lead to a change in energy of $-4J$. Due to the symmetry of the latter two states with respect to the original one with $E = 0$, we place this "quasi-degenerate" state in the "degenerate" group. We note that flipping spins 2 or 3 will not change the energy. Hence, all configurations of group d are related, along with those of group c, to "degenerate" configurations. Our conjecture consists of the statement that the phase transition occurs when the partition function Z_{ndg} of non-degenerate states is equal to that Z_{dg} of degenerate states,

$$Z_{\mathrm{ndg}} = Z_{\mathrm{dg}}, \qquad (2.10)$$

i.e., we identify the non-degenerate states with "ordered" configurations and the degenerate states with "disordered" ones.

Since for the two-dimensional Ising lattice our criterion (2.10) coincides with that of Svrakic [8], it clearly results in the exact Onsager solution for the critical temperature. However, in contrast to [8], our criterion (2.10) can be applied to three-dimensional lattices as well. We now examine Fig. 2.3 in detail. Here, groups a–f belong to the non-degenerate category, because a flip of the spin of any site will change the energy of the cell, and no quasi-degenerate states are observed in these configurations. Groups g and h belong to the degenerate category. A flip of each spin of group g results in the change of energy of $2J, -2J, 2J, -2J, 2J, -2J, 2J, -2J$, while a flip of each spin of group h change the energy by $6J, -6J, 6J, -6J, 2J, -2J, 2J, -2J$, and these two groups are totally quasi-degenerate. However, groups i–n contribute to both degenerate and non-degenerate states. Let us consider, as an example, group i. The flip of each of the eight spins changes the energy by $-6J, 6J, 6J, 6J, 6J, 2J, 2J, 2J$, respectively. The

first two excited states are symmetric, and, according to our classification, they are quasi-degenerate, while the remaining six states are non-degenerate. The numbers shown in Fig. 2.3 are the statistical weight of degenerate W_G and non-degenerate W_N states, and so for group i they have the fractional values 2/8 and 6/8 respectively. Analogous analyses can easily be performed for each of the groups j–n, and the final result for the solution of Eq. (2.10) is $kT_c/J = 4.277$ [9] which is much closer to the numerical result $kT_c/J = 4.511$ [10] than that, $kT_c/J = 2.030$, obtained by the method of Ref. [8].

In conclusion, we have shown that the critical temperature of two- and three-dimensional Ising models can be obtained by simple physical arguments based on the compromise between configurations which are "degenerate" or "non-degenerate" with respect to a single spin-flip excitation. A challenging problem for the interested reader is to expand our calculations to other elementary cells of two-dimensional lattices (say, 3–3 square, 3–2 rectangle, etc.) and to other types of three-dimensional lattices (BCC, FCC, etc.). For the latter, the findings have to be compared with the precise numerical results which are 6.235 and 9.792 for the BCC and FCC lattices, respectively [10]. The results of such a comparison are not known *a priori* since it is not clear how well this method will work in the general case.

2.4   Conclusion

The microscopical analysis of phase transitions can be performed only for simple models such as the Ising model. The results for these show that for non-zero temperature a phase transition does not appear in a one-dimensional system but such a transition does occur in two dimensions. A qualitative explanation of the onset of a phase transition as an order–disorder transition is that it is a result of competition between the ordering tendency of the energy and the disordering tendency of the entropy. With some simplified conjectures, such considerations can even lead to quantitatively correct results.

This page intentionally left blank

Mean Field Theory

As we mentioned in Chapter 1, the simplest way to allow for interactions between particles is the mean field approximation, in which one assumes that the environment of each particle corresponds to the average state of the system and determines this average state self-consistently.

In order to understand the nature of Weiss' mean field approximation described in Chapter 1, it is useful to examine the application of this approximation to the Ising model. In this model, the energy of a state of the system with a given arrangement $\{S_1, S_2, \ldots, S_N\}$ of the N spins S_j, each of which can have the value ± 1, situated on a lattice of sites, is

$$E = -J \sum_{(i,j)nn} S_i S_j. \tag{3.1}$$

In Weiss' mean field approximation for our system,

$$E_{mf} = -\sum_i S_i H_m \tag{3.2}$$

where H_m is the effective magnetic field. A comparison of Eqs. (3.1) and (3.2) shows that $H_m = J\langle S_j \rangle$, so that

$$E_{mf} = -J \sum_{(i,j)nn} S_i \langle S_j \rangle. \tag{3.3}$$

Now, Eq. (3.1) can be written in the form

$$E = -J \sum_{(i,j)nn} S_i \langle S_j \rangle - J \sum_{(i,j)nn} S_i(S_j - \langle S_j \rangle)]. \qquad (3.4)$$

The first term in this equation is the expression that we derived for the mean field approximation, while the second term arises from the fluctuations of each spin from its mean value. Hence the mean field approximation is equivalent to ignoring fluctuations. It is obvious that as the number of dimensions of the system increases, and with it the number of nearest neighbors of each site, the importance of these fluctuations decreases, and they are negligible for a system of sufficiently high dimensions. As we will see later on, "sufficiently high" for the Ising systems means $d > 4$. This is the physical reason why for each system there is an upper critical dimension above which the mean field approximation is valid.

3.1 Landau Mean Field Theory

We now consider the thermodynamic approach to mean field theory proposed by Landau. He suggested [1] introducing an extra parameter η, which is called an order parameter, to distinguish the phase of the system, so that the Gibbs free energy G is now a function of three parameters, $G = G(P, T, \eta)$, where $\eta = 0$ in a disordered state found, as a rule, at high temperatures, $T > T_c$, and $\eta \neq 0$ for the ordered state, which then occurs for $T < T_c$. Since the free energy of the system in equilibrium is uniquely determined by P and T, this order parameter must be a function of them, $\eta = \eta(P, T)$. The first assumption of Landau's mean field theory is that an order parameter can be defined such that $\partial G/\partial \eta = 0$ and $\partial^2 G/\partial \eta^2 > 0$, which is the natural condition for minima of the free energy. His second assumption is based on the fact that at the phase transition $\eta = 0$. Therefore, close to the phase transition point he assumed that G can be expanded as a power series in η,

$$G(P, T, \eta) = G_0(P, T) + \alpha \eta + A \eta^2 + \beta \eta^3 + B \eta^4 + \cdots \qquad (3.5)$$

where all the coefficients are functions of P and T. The first assumption, $\partial G/\partial \eta = 0$, can only be satisfied for all η if $\alpha = 0$. For a magnetic system in which the magnetic moment vector is the order parameter, symmetry arguments show that it is not possible to form a scalar from η^3, and so $\beta = 0$. Incidentally, for different systems the order parameter can have different numbers of components. For instance, in the liquid–gas phase transition the order parameter is the difference in densities between the two phases, which is a scalar. In superfluidity and superconductivity, it is the wave function ψ which has two components (its real and imaginary parts, or its amplitude and phase). For the anisotropic Heisenberg ferromagnet, in which M has three independent components, this will also be true of η.

Here we restrict our attention to systems in which η has just one component, and for which close to the phase transition point the free energy can be written in the form

$$G(P,T,\eta) = G_0(P,T) + A\eta^2 + B\eta^4. \tag{3.6}$$

For the sake of simplicity, we consider systems at a constant pressure equal to the critical one, $P = P_c$, and so omit P from the arguments. Then for arbitrary η the requirement that $\partial G/\partial \eta = 0$ means that $2A\eta + 4B\eta^3 = 0$, so that $\eta = 0$ for $T > T_c$, and there is an additional solution $\eta^2 = -A/(2B)$ for $T < T_c$. From the second requirement $\partial^2 G/\partial \eta^2 > 0$ it follows that for $T > T_c$, $A > 0$ while $A < 0$ for $T < T_c$. It follows that $A(T_c) = 0$, and the simplest form for A that satisfies these conditions is $A = a(T - T_c)$, with $a > 0$. One then finds that the order parameter

$$\eta \sim \sqrt{T_c - T} \tag{3.7}$$

and the free energy

$$G(T,\eta) = G_0(T) + a(T - T_c)\eta^2 + B\eta^4. \tag{3.8}$$

Here the term in η^4 is required since the term in η^2 vanishes at the critical point, but the temperature dependence of its coefficient B can be ignored, so that we can write $B = B(T_c, P_c)$.

We now consider the implications of Eq. (3.8) for the free energy. From the formula for the entropy

$$S = -\frac{\partial G}{\partial T} = S_0(T) - a\eta^2,$$

it follows that

$$
\begin{aligned}
S &= S_0(T), & T > T_c, \\
S &= S_0(T) + \frac{a^2(T - T_c)}{2B}, & T < T_c.
\end{aligned}
\tag{3.9}
$$

The specific heat $C = T(\partial S/\partial T)$, and so

$$
\begin{aligned}
C &= C_0(T), & T > T_c, \\
C &= C_0(T) + \frac{a^2 T_c}{2B}, & T < T_c.
\end{aligned}
\tag{3.10}
$$

In Eqs. (3.9) and (3.10), $S_0(T)$ and $C_0(T)$ arise from the regular part $G_0(P, T)$ of the free energy. The resulting jump in the specific heat leads to a λ shape of the curve of $C(T)$, and so this transition is also known as a λ-transition.

Finally, let us consider a magnetic system, for which we identify η^2 with the square of the magnetization M^2, and examine the effect of an external magnetic field H, so that

$$G(T, H, M) = G_0 + AM^2 + BM^4 - MH. \tag{3.11}$$

Since we require that $\partial G/\partial M = 0$, while at $T = T_c$ we found that $A = 0$, it follows that at the critical temperature $M^3 = H/(4B)$, i.e.,

$$M \sim H^{\frac{1}{3}}, \qquad T = T_c. \tag{3.12}$$

The magnetic susceptibility is $\chi = \partial M/\partial H = 1/(2A + 12BM^2)$, so that on substituting the above values for M^2 we find that the magnetic susceptibility χ is given by

$$
\begin{aligned}
1/\chi &= 2a(T_c - T), & T < T_c, \\
1/\chi &= 4a(T - T_c), & T > T_c.
\end{aligned}
\tag{3.13}
$$

It follows that the magnetic susceptibility becomes infinite at $T = T_c$.

In all the above consideration we assumed that in Eq. (3.5) the coefficient $\beta(P, T)$ vanishes because of symmetry requirements. If this

is not the case, all the above calculations are correct provided that in addition $\beta(P_c, T_c) = 0$ at the critical point, as well as $A(P_c, T_c) = 0$. However, these two conditions mean that the critical point is an isolated point in the T–P plane — a case which is of no special interest.

We note that all the results obtained in Eqs. (3.7)–(3.13) for the temperature dependence of the thermodynamic parameters do not involve the spatial dimensions. This is not correct, as we will see later on. Such a failing should not surprise us, since the mean field approach is based, in particular, on the assumption of analyticity, which is at least questionable for a function which may have singularities.

3.2   First Order Phase Transitions in Landau Theory

So far, we have considered only continuous (second order) phase transitions. However, the Landau mean field theory is also able to describe first order phase transitions, where there is a jump in the order parameter. There are two different ways of doing this:

1. By adding a cubic term to the Landau expansion (3.8), one obtains

$$G = G_0 + a(T - T_c)\eta^2 + C\eta^3 + B\eta^4. \qquad (3.14)$$

The assumption that G as function of η has a minimum, $\partial G/\partial \eta = 0$, leads to

$$\eta_1 = 0, \qquad \eta_2 = -\frac{3C}{8B} \pm \sqrt{\left(\frac{3C}{8B}\right)^2 - \frac{a(T - T_c)}{2B}}. \qquad (3.15)$$

A temperature T_0 exists such that

$$G(\eta_1, T_0) = G(\eta_2, T_0). \qquad (3.16)$$

When the system approaches the temperature T_0 from above, the order parameter jumps discontinuously from η_1 to η_2, which means that a first order phase transition occurs there.

2. A first order phase transition can be obtained for negative values of the coefficient B in the Landau expansion (3.8). Then, for a minimum

of $G(M)$ to exist, one has to consider the sixth order term in the expansion of $G(M)$,

$$G = G_0 + a(T - T_c)M^2 - |B|M^4 + DM^6. \qquad (3.17)$$

The conditions for a minimum of Eq. (3.17) lead to

$$M_1 = 0, \qquad M_2 = \frac{1}{\sqrt{3D}}\big[|B| + \sqrt{B^2 - 3aD(T - T_c)}\big]^{\frac{1}{2}}. \qquad (3.18)$$

By using a procedure identical to that used for the analysis of Eq. (3.16), we conclude that a first order phase transition occurs in this case as well.

3.3 Landau Theory Supplemented with Fluctuations

As we showed earlier in this chapter, the difference between the results of mean field theory and the exact results are due to fluctuations. We now generalize the Landau theory by taking into account these fluctuations. In order to do this, the first step is to use the quasi-hydrodynamical approach, and replace the discrete lattice by a continuum, so that the free energy G and the order parameter, M in the case of ferromagnetism, become functions of the continuous coordinate r, $G = G(r)$ and $M = M(r)$. Then the global expansion (3.8) has to be replaced by a local one for $G(r)$, while the global free energy and magnetic moment are obtained by integration over the whole d-dimensional system.

For a non-homogeneous system the expansions of the free energy will contain not only the thermodynamic variables but also their derivatives. Assuming that the inhomogeneities are small, one can retain in the expansions only the simple gradient term $(\nabla M)^2$, and neglect higher derivatives. There is no need to multiply this term by a constant since the length scale can be chosen to make this constant unity. Thus, the simplest possible form for the free energy density $G(r)$ is

$$G(r) = G_0 + AM^2 + BM^4 + (\nabla M)^2. \qquad (3.19)$$

According to Boltzmann's formula, the probability for a fluctuation $P(M)$ is proportional to $\exp[-(G - G_0)/kT]$, and so

$$P(M) \sim \exp\left[-\int d\tau \frac{AM^2 + (\nabla M)^2}{kT}\right]. \tag{3.20}$$

Here, we ignore the term BM^4 since this was only introduced to prevent $G - G_0$ vanishing at the critical point, where $A = 0$, while the extra term $(\nabla M)^2$ ensures that this does not happen. Thus, we use a Gaussian approximation, i.e., one that involves only quadratic terms in M and ∇M.

In order to calculate the above integral, it is convenient to use the Fourier transform of $M(r)$,

$$M(\mathbf{r}) = \int M_{\mathbf{K}} \exp(i\mathbf{K} \cdot \mathbf{r}) \, d^d K. \tag{3.21}$$

The radial distribution function $g(r)$ of $M(\mathbf{r})$ is determined by the coefficients $M_{\mathbf{K}}$ [1],

$$g(r) = \frac{\int M(\mathbf{R})M(\mathbf{R} + \mathbf{r}) \, d\mathbf{R}}{\int M(\mathbf{R})M(\mathbf{R}) \, d\mathbf{R}} = \int \langle M_{\mathbf{K}}^2 \rangle \exp(i\mathbf{K} \cdot \mathbf{r}) \, d^d K. \tag{3.22}$$

According to the Gaussian approximation (3.20), the probability of a fluctuation described by a given set $\{M_{\mathbf{K}}\}$ of Fourier coefficients is

$$P\{M_{\mathbf{K}}\} \sim \exp\left[-\frac{(A + K^2)M_{\mathbf{K}}^2}{kT}\right] \equiv \exp\left(-\frac{M_{\mathbf{K}}^2}{2\langle M_{\mathbf{K}}^2 \rangle}\right), \tag{3.23}$$

where $\langle M_{\mathbf{K}}^2 \rangle = kT/[2(A + K^2)]$. On substituting this expression in Eq. (3.22), we find that

$$g(r) = \int \frac{kT \exp(i\mathbf{K} \cdot \mathbf{r})}{2(A + K^2)} d^d K. \tag{3.24}$$

It follows, e.g. from integration in the complex plane and the residue theorem, that in three dimensions

$$g(r) \sim \frac{\exp(-r\sqrt{A})}{r} \tag{3.25}$$

so that the correlation length

$$\xi = \frac{1}{\sqrt{A}} \sim (T - T_c)^{-1/2}. \tag{3.26}$$

At the critical point $A = 0$, and so at $T = T_c$ the correlation function becomes

$$g(r) \sim \frac{1}{r}. \tag{3.27}$$

3.4   Critical Indices

In the theory of phase transitions, it is very important to describe the behavior of various properties of the system in the vicinity of the critical point. For a differentiable function with a singularity at the critical point, such as $a \sim |T - T_c|^x$, this behavior is characterized by the index x, which is called a critical index. If $x \neq 0$, this critical index is given by $x = \ln(a)/\ln|T - T_c|$, while if $x = 0$ there are two possibilities. In one case, a becomes a constant at the critical point, with the possibility of different values of this constant on the two sides of the transition, which is a jump singularity. In the other case, a exhibits a logarithmic singularity, $a \sim \ln|T - T_c|$. For other properties of the system near and at the critical point, the power exponents of these dependencies are called critical indices. In Table 4.1 in the next chapter, we define a set of critical indices and present their values for the mean field model calculated above, as well as the values of the critical indices for some other models which we analyze there.

3.5   Ginzburg Criterion

Since the difference between the results of mean field theory and the exact ones is due to the fluctuations in the order parameter, we expect that the mean field approximation will be accurate when these fluctuations are sufficiently small. Ginzburg proposed [11] that the mean field theory is applicable when the fluctuations are small in comparison with the thermodynamical values. On using Eq. (3.23) and the fact that $M^2 \sim A$, one finds after integrating over angles

that this criterion is fulfilled if

$$\int_0^{K_{\max}} kT \frac{K^{d-1}}{A+K^2} dK < |A|. \tag{3.28}$$

If the dependence of the integral on its upper limit can be neglected, it follows from the substitution $y = K\sqrt{A}$ that this condition becomes

$$(A^{d/2}/A) \int f(y)dy \ll A,$$

and so $A^{d/2} \ll A^2$. Hence, since A is small in the region of the critical point, the mean field approximation is valid if $d > 4$, so that the upper critical dimension is four. For $d = 4$ there are only logarithmic corrections to the mean field results.

While the above estimates are simple but crude, they still provide us with useful information. If one takes into account the coefficients in the inequality (3.28), and substitutes their values for different types of phase transitions, qualitative estimates can be obtained of the region around the critical point where the mean field theory is valid. It turns out that for the gas–liquid phase transition the mean field theory can be applied up to within ten percent of the critical temperature, while for the superconducting transitions the mean field theory gives the correct results everywhere in the experimentally accessible vicinity of the critical point. The latter result is connected with the large value of the correlation length in low temperature superconductors.

3.6 Wilson's ϵ-Expansion

So far, we have considered the local value $\eta(\mathbf{r})$ associated with the points $\mathbf{r}$ of a lattice, which are a distance a apart and can be regarded as a continuum, so that in our previous analysis we often used integrals rather than sums. Such a coarse-graining procedure does not lead to any problems in hydrodynamics, where a is the only characteristic length, as soon as the condition $r > a$ is satisfied. However, as Wilson pointed out [12], this coarse-graining procedure becomes

problematic in the theory of phase transitions, since near the critical point the correlation length ξ appears as an additional characteristic length. Hence, the simple coarse-graining procedure for $r > a$ becomes unjustified unless in addition $r > \xi$, and one has to find some way to treat the range $a < r < \xi$. This region is very important near the critical point, since ξ tends to infinity as the critical point is approached.

In order to deal with this problem, Wilson suggested that, in contrast to the situation for hydrodynamics, the coefficients A and B in Eq. (3.19) depend on the size of the coarse-graining region, which we denote by L. In order to find these dependencies let us consider how the free energy changes when one goes from a region of size L, with $a < L < \xi$, to a region of size $L + \Delta L$. Such a comparison of the free energies in two regions is, in fact, the basis of the renormalization group approach which we will consider in detail in Chapter 5.

Let us consider a value of ΔL so small that it introduces just one additional mode into a system. If we label the modes by their Fourier components K, this means [1] that $V\, 4\pi K^{d-1} \Delta K/(2\pi)^3 = 1$, or since $K \sim 1/L$ that, for $V = 1$, $\Delta L \sim L^{d+1}$. Let us now write the free energy $G_{L+\Delta L}(P,T)$ in the Landau form (3.19),

$$G_{L+\Delta L}(M) = \int d\tau [G_0 + A_{L+\Delta L}M^2 + B_{L+\Delta L}M^4 + (\nabla M)^2] \quad (3.29)$$

and compare it with that obtained from $G_L(P,T)$ supplemented by an additional fluctuating mode mM_1 with a scaling factor m:

$$\begin{aligned} G_L(M + mM_1) = \int d\tau \big[&G_0 + A_L(M + mM_1)^2 \\ &+ B_L(M + mM_1)^4 + [\nabla(M + mM_1)]^2 \big]. \end{aligned} \quad (3.30)$$

The contribution of the additional mode mM_1 will be taken into account not by the "hydrodynamic" average but by the correct "Boltzmann" average. Thus, one finds that

$$\exp\left(-\frac{G_{L+\Delta L}}{kT}\right) = \int_{-\infty}^{\infty} dm \exp\left(-\frac{G_L(M + mM_1)}{kT}\right). \quad (3.31)$$

We restrict our attention to the Gaussian distribution, and so leave in the exponent of the right hand side of Eq. (3.31) only terms quadratic in m, and require that for the fluctuating mode

$$\int M_1(r)dr = 0, \qquad \int M_1^2(r)dr = 1. \tag{3.32}$$

For small ΔL one can write

$$A_{L+\Delta L} = A_L + \frac{dA}{dL}\Delta L, \qquad B_{L+\Delta L} = B_L + \frac{dB}{dL}\Delta L. \tag{3.33}$$

The substitution of Eqs. (3.32) and (3.33) into Eq. (3.31) leads to the equation

$$\frac{dA_L}{dL}\Delta L M^2 + \frac{dB_L}{dL}\Delta L M^4$$
$$= \text{Const} + \ln(1 + A_L L^2 + 6B_L L^2 M^2). \tag{3.34}$$

On expanding the logarithm in (3.34) in a power series and equating the coefficients of M^2 and M^4, we find that

$$\begin{aligned}
\frac{dA}{dL}\Delta L &= 3BL^2 - 3ABL^4, \\
\frac{dB}{dL}\Delta L &= -9B^2L^4.
\end{aligned} \tag{3.35}$$

Since $\Delta L \sim L^{d+1}$, this gives in terms of the parameter $\epsilon \equiv 4 - d$,

$$A \sim L^{-\frac{\epsilon}{3}}, \qquad B \sim L^{-\epsilon}. \tag{3.36}$$

Equations (3.36) must be valid for $a < L < \xi$, and so we substitute in them $L = \xi$. On using Eq. (3.26) for ξ and $A = a(T - T_c)$, one finds from Eq. (3.36) that $M \sim [(T - T_c)/T_c]^\beta$ and $\xi \sim (|T - T_c|/T_c)^{-\nu}$ where

$$\beta = \frac{1}{2} - \frac{\epsilon}{6\left(1 - \frac{\epsilon}{6}\right)}, \qquad \nu = \frac{1}{2\left(1 - \frac{\epsilon}{6}\right)}. \tag{3.37}$$

This equation gives the values of the critical indices, which were defined in the last section and are listed in Table 4.1. In this equation, the critical indices are expanded as power series in the small parameter ϵ, which is why the method is known as Wilson's ϵ-expansion. Not by chance is Wilson's article [13] called "Critical phenomena in 3.99 dimensions". If one is brave enough to set the "small" parameter

ϵ equal to unity, $\epsilon = 1$, then the one finds that for three-dimensional systems the critical indices are $\beta = 0.3$ and $\nu = 0.6$.

The above approximate calculation, which gives the ϵ-corrections to the mean field critical indices, is a good introduction to the renormalization group approach, where this transformation leading from G_L to $G_{L+\Delta L}$ is applied repeatedly.

3.7   Conclusion

The simplest phenomenological mean field description of phase transitions is obtained by introducing an order parameter η and assuming that close to the critical point the free energy G can be expanded in this order parameter. The requirement that $G(\eta)$ be a minimum enables us to find the critical indices which define the behavior of the thermodynamical parameters near the critical point. As it follows from comparison of Weiss' mean free theory with the Ising model, the main approximation of the mean field theory is the neglect of fluctuations. According to the Ginzburg criterion, this is justified for space dimensions larger than four (the upper critical dimension). The inclusion of fluctuations in the mean field theory is achieved by adding to the free energy the simplest gradient terms, and this enables us to find the temperature dependence of the correlation length and the form of the correlation function at the critical point.

Another type of correction to the mean field theory is associated with the increase of the correlation length on approaching the critical point. These corrections are taking into account by introducing the spatial dependence of the coefficients in an expansion $G = G(\eta)$. This leads to the dependence of the critical indices on the parameter $\epsilon = 4 - d$, which defines the distance from the upper critical dimension $d = 4$. This ϵ-expansion is the simplest form of a renormalization group procedure.

Chapter 4

Scaling

Scaling is a very general and well-known method for describing the response of a system to a disturbance. For physical systems, it is most readily studied in terms of dimensional analysis and the construction of dimensionless parameters, and so we consider this topic first, and start with a few simple examples.

Let us consider first the free fall (i.e., with no air resistance) of a body of mass m under the gravitational force mg. If the body starts from rest, then we expect that its velocity v after falling from a height h will be a function of m, h, and g, $v = f(m, h, g)$. We assume that v has a power-law dependence on these parameters, and so write $v = am^{\alpha}h^{\beta}g^{\gamma}$, where a is a dimensionless constant. In any equation, the dimensions of the quantities on the two sides must be equal. Since $[v] = \mathrm{LT}^{-1}, [h] = \mathrm{L}, [m] = \mathrm{M}$ and $[mg] = \mathrm{MLT}^{-2}$, it follows that

$$\mathrm{LT}^{-1} = \mathrm{M}^{\alpha}\mathrm{L}^{\beta}(\mathrm{LT}^{-2})^{\gamma} \tag{4.1}$$

which has the unique solution $\alpha = 0$, $\beta = \gamma = 1/2$, so that $v = a\sqrt{gh}$. Thus, we have found the functional dependence of v on g and h without having to solve the problem, and can conclude that if the height from which the body falls is multiplied by four then its velocity will be doubled.

Another example, which is very relevant for engineering applications, is the force F acting on a sphere of radius R moving through a fluid of viscosity η with velocity v. We assume that $F = aR^{\alpha}v^{\beta}\eta^{\gamma}$,

and once again equate the dimensions of the two sides of this equation. Since $[F] = \mathrm{MLT}^{-2}, [R] = \mathrm{L}, [v] = \mathrm{LT}^{-1}$, and $[\eta] = \mathrm{ML}^{-1}\mathrm{T}^{-1}$, it follows that $\alpha = \beta = \gamma = 1$, so that $F = aRv\eta$, which is just Stokes' law. This law permits the prediction of such properties as the air resistance to the motion of cars or aeroplanes from those of model systems with smaller objects and/or lower velocities, although allowance has to be made for the fact that cars and aeroplanes are not spherical. An important point about this example, in contrast to the first one, is that we had to decide (or guess) which of the four parameters in the system, namely the radius, velocity and mass of the object and the viscosity of the medium, was irrelevant, since for mechanical (as opposed to electromagnetic) quantities there are only three independent dimensions, namely mass M, length L and time T.

Scaling is also of considerable importance for presenting theoretical predictions and experimental data for different values of the parameters on a single curve. The basis for such a procedure is that a homogeneous function $f(x, y)$ of x and y of order p is defined by the requirement that

$$f(\lambda x, \lambda y) = \lambda^p f(x, y). \tag{4.2}$$

Then, on choosing $\lambda = 1/x$, one finds that $f(1, y/x) = (1/x)^p f(x, y)$, so that

$$f(x, y) = x^p f\left(1, \frac{y}{x}\right) = x^p \psi\left(\frac{y}{x}\right). \tag{4.3}$$

Equation (4.3) means that a homogeneous function of two variables can be scaled to a function of a single variable, so that by a suitable choice of variables ($f(x, y)/x^p$ and y/x) all of the data collapse on to a single curve.

One can also consider [14] a generalized homogeneous function

$$f(\lambda^a x, \lambda^b y) = \lambda^p f(x, y) \tag{4.4}$$

which for $a = b$ reduces to a standard homogeneous function (4.2).

4.1 Relations Between Thermodynamic Critical Indices

Critical indices, which were introduced in previous chapters, define the behavior of the thermodynamic parameters near the critical point. A list of such indices, with their definitions, is given in Table 4.1. While the definitions listed there are for a magnetic system, they can easily be formulated for other types of system. For instance, for the liquid–gas system, the magnetic field H is replaced by the pressure P, and the magnetic moment M by the specific volume. The different thermodynamic parameters are not independent, since there are some thermodynamic relations between them, and these lead to connections between different critical indices. We now consider two examples of such relations.

The well-known formula [1] linking specific heats in magnetic systems, c_H and c_M (or c_p and c_v for non-magnetic systems) has the following form:

$$c_H - c_M = \frac{T(\partial M/\partial T)_H^2}{(\partial M/\partial H)_T}. \tag{4.5}$$

Table 4.1 Critical indices for the specific heat (α), order parameter (β), susceptibility (γ), magnetic field (δ), correlation length (ν) and correlation function (η). The indices α, γ, ν and α', γ', ν' refer to $T > T_c$ and $T < T_c$, respectively. The results for the mean field and 2D Ising models are exact, while those for the 3D Ising model are approximate.

Exponent	Definition	Condition	Mean field	2D Ising	3D Ising
α	$C \sim \left\| \frac{T-T_c}{T_c} \right\|^{-\alpha(-\alpha')}$	$H = 0$	0 (*jump*)	0 (ln)	~ 0.11
β	$M \sim \left(\frac{T_c-T}{T_c} \right)^{\beta}$	$H = 0; T < T_c$	0.5	0.125	~ 0.32
γ	$\chi \sim \left\| \frac{T-T_c}{T_c} \right\|^{-\gamma(-\gamma')}$	$H = 0$	1	1.75	~ 1.24
δ	$H \sim \|M\|^{\delta}$	$T = T_c$	3	15	~ 4.82
ν	$\xi \sim \left\| \frac{T-T_c}{T_c} \right\|^{-\nu(-\nu')}$	$H = 0$	0.5	1	~ 0.63
η	$g(r) \sim \frac{1}{r^{d-2+\eta}}$	$T = T_c$	0	0.25	~ 0.03

In order for a system to be stable, c_M must be positive, and so it follows from Eq. (4.5) that

$$c_H > \frac{T(\partial M/\partial T)^2_H}{(\partial M/\partial H)_T} \tag{4.6}$$

or, on substituting the critical indices from Table 4.1,

$$\tau^{-\alpha} > \tau^{2(\beta-1)+\gamma} \tag{4.7}$$

where $\tau = |T - T_c|/T_c$. Close to the critical point $\tau \ll 1$, and so Eq. (4.7) leads [15] to the following relation between the critical indices

$$\alpha + \beta + 2\gamma \geq 2. \tag{4.8}$$

Another inequality can be obtained as follows [16]. For $T_1 < T_c$,

$$G(T_c, M_1) = G(T_1, M_1) + \int_{T_1}^{T_c} \left(\frac{\partial G}{\partial T}\right)_M dT = G(T_1, M_1) - \int_{T_1}^{T_c} S dT. \tag{4.9}$$

However, since the stability condition requires that

$$c_M \sim -\left(\frac{\partial^2 G}{\partial T^2}\right)_M > 0,$$

the derivative $\left(\frac{\partial G}{\partial T}\right)$ must be a decreasing function of T, and so $S = -\left(\frac{\partial G}{\partial T}\right)_M$ must be an increasing function of T. It follows then from Eq. (4.9) that

$$G(T_c, M_1) \leq G(T_1, M_1) - (T_c - T_1)S(T_1, M_1), \tag{4.10}$$

and

$$G(T_c, M_1) \geq G(T_1, M_1) - (T_c - T_1)S(T_c, M_1). \tag{4.11}$$

For $T < T_c$, a magnetic moment $M(T)$ appears spontaneously, so that $G(T, M(T))$ and $S(T, M(T))$ are fully determined by the temperature T, and so Eq. (4.10) can be rewritten in the form

$$G(T_1) \geq G(T_c) + (T_c - T_1)S(T_1). \tag{4.12}$$

On adding Eqs. (4.12) and (4.11), we obtain

$$G(T_c, M_1) - G(T_c) \leq (T_c - T_1)[S(T_c) - S(T_1)]. \tag{4.13}$$

Now

$$G(T_c, M_1) - G(T_c) \sim M_1^{1+\delta} \sim (T_c - T_1)^{\beta(1+\delta)}$$

and

$$[S(T_c) - S(T_1)] \sim (T_c - T_1)^{1-\alpha}.$$

Thus, Eq. (4.13) leads to

$$\beta(1 + \delta) \geq 2 - \alpha. \qquad (4.14)$$

The inequalities (4.8) and (4.14) follow from thermodynamics, and so are rigorous. In order to derive additional relations between critical indices one have to use some approximations.

4.2 Scaling Relations

There are several different approximate methods for obtaining scaling relations. We shall follow the method of introducing the concept of a new "block" lattice which can be related to the original "site" lattice, as proposed by Kadanoff [17]. In this chapter we apply the method to thermodynamics, while its application to a Hamiltonian will be the starting point for the renormalization group theory considered in the next chapter.

The Kadanoff construction shown in Fig. 4.1 ("scaling hypothesis") consists of the following three stages.

(1) Divide the original Ising site lattice with lattice constant of unit length into blocks of size L less than the correlation length ξ, i.e., $1 < L < \xi$.
(2) Replace the L^d spins inside each block by a single spin $\mu_i = \pm 1$, the sign of which is determined by that of the majority of the sites inside the block. Since the size L of a block is less than the correlation length ξ, we expect that most of the spins within a block will be of the same sign, so that this majority rule is a very reasonable approximation.
(3) Return to the original site lattice by dividing all lengths by L.
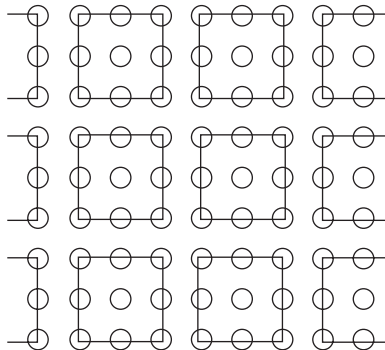
Fig. 4.1   The Kadanoff construction for a square lattice, with blocks of nine sites.

The aim of this procedure is to make an average over the small degrees of freedom, or in other words to make a coarse graining so as to reduce the number of degrees of freedom of the system. The consequences of this idea are not at all trivial. Even if we do not know the exact solution of the problem for either the site lattice or the block lattice, a comparison of these two problems can provide us with valuable information.

Since all the thermodynamic properties of a system depend only on the temperature and the external magnetic field, which determine its free energy, let us consider a site lattice in an external (dimensionless) magnetic field h and in contact with a thermal bath at temperature T, which appears in our analysis in the dimensionless form $\tau = |T - T_c|/T_c$. Let the corresponding parameters for the block lattice be h' and τ'. It is obvious that if there is no external field for the site lattice so that $h = 0$, then there is no such field for the block lattice so that $h' = 0$, and similarly if $\tau = 0$ then $\tau' = 0$. In view of this, and since the single parameter that characterizes the size of the block lattice is L, Kadanoff proposed that $\tau' = L^x \tau$ and $h' = L^y h$.

Let $f(\tau, h)$ be the free energy per site for the site lattice, and $f(\tau', h')$ the free energy per block of the block lattice. Since the free energy of a block is the sum of the free energy of the sites composing it, it follows that in d dimensions

$$L^d f(\tau, h) = f(\tau', h') = f(L^x \tau, L^y h). \tag{4.15}$$

As can be seen by substitution, a solution of this functional equation is

$$f(\tau, h) = \tau^{d/x} \psi(\tau/h^{x/y}), \tag{4.16}$$

where ψ is an arbitrary function. This property enables us to find the connection between the critical indices associated with different properties of the system without knowing the function ψ.

The magnetic moment

$$M \sim [\partial f/\partial h]_{h=0} \sim \tau^{d/x}[\tau/h^{1+x/y}]\psi'(z),$$

where $z = \tau/h^{x/y}$. The only way that this can be independent of h as $h \to 0$ is that for $z \longrightarrow \infty$, $\psi'(z) \sim z^{-1-y/x}$, in which case $M \sim \tau^{(d-y)/x}$. Hence, the critical index β defined in Table 4.1, is given by

$$\beta = \frac{d-y}{x}. \tag{4.17}$$

It then follows that the magnetic susceptibility as $h \to 0$, i.e., as $z \to \infty$,

$$\chi \sim \partial M/\partial h \sim \tau^{d/x}[\tau/h^{1+x/y}]^2\psi''(z),$$

and thus $\psi''(z) \sim z^{-2-2y/x}$, so that $\chi \sim \tau^{d/x+2-2-2y/x}$, and hence

$$\gamma = \frac{2y-d}{x}. \tag{4.18}$$

The specific heat as $h \to 0$, i.e., as $z \to \infty$, $C \sim \partial^2 f/\partial \tau^2 \sim \tau^{d/x-2}$, and so

$$\alpha = 2 - \frac{d}{x}. \tag{4.19}$$

At the critical point, $\tau = 0$, the temperature-independent magnetic moment $M \sim \tau^{d/x}[\tau/h^{1+x/y}]\psi'(z)$ requires $\psi'(z) \sim z^{-(1+d/x)}$ as $z \to 0$. A new critical index δ is defined by $M \sim h^{1/\delta}$ at $\tau = 0$. It then follows that

$$\delta = \frac{y}{d-y}. \tag{4.20}$$

In order to obtain more critical indices, let us now consider the correlation function $g(r, \tau)$ in the site and block lattices. The first step is to derive the relationship between the average spins $\langle S \rangle$ of the site lattice and $\langle \mu \rangle$ of the block lattice. To do this, we compare the average energy per block of L^d spins in an external magnetic field h in the site lattice with the energy per block in the corresponding external magnetic field $h' = L^y h$ in the block lattice

$$hL^d \langle S \rangle = hL^y \langle \mu \rangle. \tag{4.21}$$

It follows that $\langle S \rangle \sim L^{y-d} \langle \mu \rangle$, or for the correlation functions g_s and g_μ in the site and block lattices, respectively,

$$g_s(r, \tau) = L^{2(y-d)} g_\mu \left(\frac{r}{L}, L^x \tau \right). \tag{4.22}$$

It can easily be proved by substitution that the solution of the functional equation (4.22) has the following form

$$g_s(r, \tau) = r^{2(y-d)} \Psi(r\tau^{1/x}) \tag{4.23}$$

where Ψ is an arbitrary function. At the critical point, $\tau = 0$, the limiting form of $\Psi(z)$ has to be $\Psi(z) \sim$ constant and so $g_s(r, \tau) \sim r^{2(y-d)}$. On comparing this with the definition of the critical indices in Table 1, one finds that

$$2y - d = 2 - \eta. \tag{4.24}$$

The characteristic length enters Eq. (4.23) in the form $r\tau^{1/x}$, i.e., for the correlation length ξ one obtains $\xi \sim \tau^{-1/x}$, or

$$\frac{1}{x} = \nu \tag{4.25}$$

Thus, the Eqs. (4.17)–(4.20) and (4.24)–(4.25) define six relations between the critical indices. On eliminating from these relations the unknown parameters x and y, one finds the following four relations between different critical indices

$$\alpha + \beta + 2\gamma = 2, \quad \beta\delta = \beta + \gamma, \quad \gamma = \nu(2 - \eta), \quad \alpha = 2 - d\nu. \tag{4.26}$$

The above analysis makes it clear that when the critical indices are define for $T > T_c$ and $T < T_c$, the calculations can be performed in both cases so that the limiting behavior of the thermodynamic functions is symmetric with respect to T_c, i.e.

$$\alpha = \alpha', \qquad \nu = \nu', \qquad \gamma = \gamma'. \qquad (4.27)$$

Equations (4.26) and (4.27) show that, within the framework of the scaling hypothesis, there are seven relations between the nine thermodynamic critical indices, i.e., only two indices are independent. These two remaining indices will be found in the next chapter by the use of the renormalization group theory.

Since the scaling hypothesis is based on some postulates, it is worthwhile to check it using the values of critical indices listed in Table 4.1. One can immediately see that the critical relations (4.26) are satisfied for the exact Onsager solution of the two-dimensional Ising model, and (apart from the last dimensions-dependent relation in Eqs. (4.26)) for the mean field theory. It is also clear how important it was to obtain very accurate experimental values of the critical indices.

4.3 Dynamic Scaling

In addition to the above scaling parameters describing a system's static properties near the critical point, its dynamic properties also exhibit singularities, with which are associated dynamic critical indices. Since the dynamic properties describe the approach of a system to its equilibrium state, one cannot consider them in terms of the mean field expression for the static order parameter. Instead, we must consider a system with fluctuations in the order parameter, and so use the free energy used in Chapter 3 in Landau's theory supplemented with fluctuations,

$$G(\eta) = G_0 + A\eta^2 + B\eta^4 + (\nabla \eta)^2. \qquad (4.28)$$

Just as we did in Chapter 3, we ignore in what follows the term in η^4, since the $(\nabla \eta)^2$ term is sufficient to make the free energy depend on η even for $T = T_c$, when $A = 0$. For the static equilibrium

system, where obviously η is independent of time, we required that $\partial G/\partial \eta = 0$ and $\partial^2 G/\partial \eta^2 > 0$, and so we postulate that when the system is not far from equilibrium,

$$\frac{d\eta}{dt} = -\Gamma \frac{\delta G}{\delta \eta} = -\Gamma \left[\frac{\partial G}{\partial \eta} - \nabla \frac{\partial G}{\partial(\nabla \eta)} \right], \qquad (4.29)$$

namely, the Landau–Khalatnikov equation. A simple justification of this equation lies in the fact that in equilibrium both sides of this equation vanish, and it is natural to assume that they are proportional for states which are close to equilibrium. We can choose the time scale so that $\Gamma = 1/2$, and so obtain the equation

$$\frac{d\eta}{dt} = \nabla^2 \eta - A\eta. \qquad (4.30)$$

As in the previous chapter, we use the Fourier transform η_K of η, so that

$$\eta_K(t) = \eta_K(0) \exp[-t(A + K^2)] = \eta_K(0) \exp\{-t[a(T_c - T) + K^2]\} \qquad (4.31)$$

where we have used the mean field value of A as derived in the Landau theory. If one denotes by τ_K the relaxation time for mode K so that $\eta_K(t) = \eta_K(0) \exp[-t/\tau_K]$ then we see that for $K = 0$ the relaxation time $\tau_0 \sim 1/(T_c - T)$, which tends to infinity as T approaches the critical temperature T_c, while for other values of K, $\tau_K \sim 1/K^2$ as $T \to T_c$. Thus, the approach to equilibrium is very slow near the critical temperature, a phenomenon known as critical slowing down. Great care is required to ensure that measurements of thermodynamic properties of the system in this temperature region are performed on systems in equilibrium.

The above results are for the mean field approximation, but in complete analogy with Eq. (4.23) we can obtain similar results for the more general case. To do this, we write

$$\tau_K = f(K, \tau) = \tau^z \Psi(K\tau^{-\nu}) \qquad (4.32)$$

where τ is the reduced temperature and z is the dynamic scaling exponent. In order for τ_K to be proportional to a power

of τ, $\Psi(0)$ must be a constant. It then follows that for $K = 0$, $\tau_0 \sim \tau^z$. For $K \neq 0$, the argument of Ψ tends to infinity as $\tau \to 0$, i.e., as $T \to T_c$, and so τ_K can only remain finite if $\Psi(y) \sim y^{z/\upsilon}$, in which case $\tau_K \sim K^{z/\nu}$. In the mean field case, we found that $\tau_0 \sim 1/\tau$ as $\tau \to 0$ and that $\tau_K \sim 1/K^2$, so that $z = -1$ and $z/\nu = -2$, which leads to the result $\nu = 1/2$, in agreement with what we found earlier. Therefore, in addition to the thermodynamic critical indices, a new critical index z appears in the description of dynamic phenomena.

4.4 Conclusion

The problem of calculating the critical indices which define the behavior of thermodynamic functions near the critical points is very much simplified by the use of the scaling hypothesis. The basic idea is as follows. Close to the critical point the correlation length ξ becomes much larger than the distance a between particles, and there are many "block" lattices of size L containing L^d spins, such that $a < L < \xi$. It is natural to assume that in each block almost all spins would be or up or down, just as in the original site lattice. Therefore, the critical phenomena will be similar in all block lattices, provided that the external parameters, say temperature and magnetic field, are scaled, as compared with the original site lattice, by a simple power-law dependence on L. Although the two indices which define such a change in the external parameters remain unknown, all other critical indices can be easily expressed in terms of these two indices. The scaling relations deduced from this hypothesis are satisfied by the results of the exact solution of the two-dimensional Ising model and by those of the mean field theory. The simplified examination of dynamic phenomena near the critical points requires the knowledge of an additional (dynamic) critical index, and results in slowing down of the approach to equilibrium.

This page intentionally left blank

Chapter 5

The Renormalization Group

The aim of renormalization group theory (RG) is to find the critical indices, including the indices x and y, which were still unknown after using the scaling hypothesis in the previous chapter. RG is a group of transformations (or more rigorously a semi-group, since in general there is no inverse transformation) from a site lattice with lattice constant unity and the Hamiltonian H to a block lattice with lattice constant L and the Hamiltonian H' without changing the form of the partition function,

$$Z(H, N) = Z(H', N/L^d).$$

In fact, this idea was already used in Eq. (4.15), where we compared the free energies of site and block lattices. In this chapter, we consider the connection between Hamiltonians rather than between free energies.

5.1 Fixed Points of a Map

The key idea in the use of the renormalization group is that the transition from H to H' can be regarded as a rule, $K' = f(K)$, for obtaining the parameters K' of the Hamiltonian H' of a block lattice from those K of the Hamiltonian H of the site lattice. This process can be repeated, with the lattice of small blocks being treated as a site lattice for a lattice of larger blocks, so that for the latter $K'' = f(K') = f[f(K)]$, and so on indefinitely for larger and larger

blocks until their size reaches the correlation length. This is an example of a map, i.e., a uniquely defined rule relating two successive members of a sequence, x_n and x_{n+1} by means of $x_{n+1} = f(x_n)$. In such a map, it can happen that as $n \to \infty$ this process may stop at some stage, when $x^* = f(x^*)$, in which case x^* is called a fixed point of the map. For instance, one of the simplest examples of a non-linear map is the logistic map [19]

$$x_{n+1} = f(x_n) = \mu x_n(1 - x_n), \ \ 0 < x < 1, \ \ \ \mu > 1 \quad\quad (5.1)$$

so that the fixed points are the roots of the equation $x^* = \mu x^*(1-x^*)$. Since this is a quadratic map, it has two fixed points, namely $x^* = 0$ and $x^* = 1 - 1/\mu$. This situation is shown in Fig. 5.1 where the fixed points are the intersection of the graph of the parabola, $f(x_n)$, and of the diagonal of the unit box. Starting from some initial point x_0 we make a vertical shift to the point $x_1 = f(x_0)$ lying on the parabola, which (after the horizontal shift to the diagonal) will be the starting point x_1 for the next approximation, and so on, reaching finally the stable fixed point $x^* = 1 - 1/\mu$. In order to examine the stability of the fixed points, we introduce a perturbation ξ_n from x^* in the form

$$x_n = x^* + \xi_n \quad\quad (5.2)$$
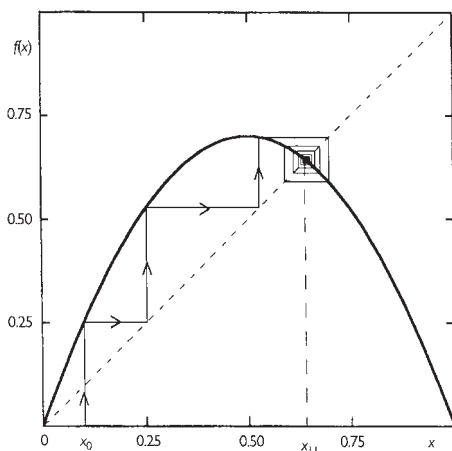
Fig. 5.1   Approach to the fixed point for the logistic map.

and substitute it in Eq. (5.1). On expanding $f(x_n)$ in a Taylor series around x^* and keeping only the terms linear in the small parameters ξ_n, we then find that

$$\xi_{n+1} = \left(\frac{df}{dx}\right)_{x=x^*} \xi_n. \qquad (5.3)$$

The small distance ξ_n from the fixed point will decrease at each stage if $|df/dx|_{x=x^*} < 1$, and in this case the fixed point will attract the set of points x_n so that the fixed point will be stable. On the other hand, if $|df/dx|_{x=x^*} > 1$, the fixed point is unstable. For the logistic map of Eq. (5.1), $(df/dx) = \mu(1 - 2x)$, and since $\mu > 1$ it is larger than unity for $x^* = 0$ and smaller than unity for $x^* = 1 - 1/\mu$, so that the latter fixed point is the stable one.

5.2   Basic Idea of the Renormalization Group

The above procedure is the basic idea of the renormalization group. Assuming that a fixed point is reached at some step of the iteration procedure for

$$K'' = f(K') = f[f(K)] = \cdots,$$

namely $K^* = f(K^*)$, where $K = J/(kT)$, one finds the value of $K^* = J/(kT_c)$ which defines the critical temperature, as discussed below. This estimate will be more precise the closer is the size of the block lattice to the correlation length, i.e., the more iteration steps that are used.

A very important property of the RG procedure is that it enables one to find not only the critical temperature but also the critical indices. Indeed, on using the RG relation $K' = f(K)$ near a fixed point, one obtains the equation

$$K' - K^* = f(K) - f(K^*) \sim \left(\frac{df}{dK}\right)_{K=K^*} (K - K^*) + \cdots. \qquad (5.4)$$

Instead of trying to solve the problem for a single lattice, which is usually impossible, we compare the solution of the problem for lattices of different sizes. Since the free energy per particle $g(K)$ must be

unaltered, for a block of L^d sites in d dimensions $L^d g(K_1) = g(K_2)$, or for the partition function $Z(K_1, N) = Z(K_2, N/L^d)$. As noted previously, we can only expect the properties of these lattices to be similar if the size of the blocks is less than the correlation length of the system. However, at a fixed point this similarity applies for blocks of arbitrary size. Therefore, we identify the fixed point with the critical point of a phase transition, since only there is the correlation length infinite. The existence of such a fixed point shows that a phase transition takes place. If a value K' of K is close to the fixed point K^*, we can rewrite Eq. (5.4) for $T = J/kK$ in the form

$$T' - T^* \sim \left(\frac{df}{dK} \right)_{K=K^*} (T - T^*). \tag{5.5}$$

On comparing Eq. (5.5) with the main scaling relation $\tau' = L^x \tau$, where $\tau = |T - T_c|/T_c$, one finds that

$$x = \ln \left[\left(\frac{df}{dK} \right)_{K=K^*} \right] / \ln(L). \tag{5.6}$$

If there is more than one external parameter, say K and H, then the function f will depend on more than one parameter, $f = f(K, H)$, and so we will have two RG equations

$$K' = f_1(K, H), \qquad H' = f_2(K, H) \tag{5.7}$$

and the fixed pointsare defined by the solutions of equations

$$K^* = f_1(K^*, H^*), \qquad H^* = f_2(K^*, H^*). \tag{5.8}$$

On repeating all the procedure leading to (5.5), one finds that

$$\begin{aligned}
T' - T^* &\sim \left(\frac{df_1}{dK} \right)_{K=K^*} (T - T^*) + \left(\frac{df_1}{dH} \right)_{K=K^*} (H - H^*), \\
H' - H^* &\sim \left(\frac{df_2}{dK} \right)_{K=K^*} (T - T^*) + \left(\frac{df_2}{dH} \right)_{K=K^*} (H - H^*).
\end{aligned} \tag{5.9}$$

Finally, on diagonalizing the 2×2 matrix and writing the diagonal elements in the form L^x and L^y, one finds the critical indices x and y.

It should be noted that the procedure described here is the simplest real space form of the renormalization techniques. In many

applications, the renormalization group technique is used in momentum space rather than in real space.

5.3   RG: 1D Ising Model

We start our examination of the application of the RG method by considering the simplest possible system, namely the 1D Ising model, even though we know that it does not exhibit a phase transition. For the original site lattice, the partition function (2.1) for the one-dimensional Ising model (2.3), with $K = J/(kT)$, has the form

$$Z(K, N) = \sum \exp[K(S_1 S_2 + S_2 S_3 + \cdots + S_{N-1} S_N)] \qquad (5.10)$$

where the summation is performed over all neighboring pairs among the N spins. In the block lattice, we omit all the spins associated with even-numbered sites and replace two spins by a single one, so that there are $N/2$ spins (which we referred to previously as μ_j, but in this case we write as S_{2j-1}) for which we have to find a partition function $Z(K', N/2)$ which involves summing only over the odd spins. The first two terms in the exponents of Eq. (5.10), after summation over $S_2 = \pm 1$, can be written in the following form

$$\exp[K(S_1 + S_3)] + \exp[-K(S_1 + S_3)] = F \exp[K'(S_1 S_3)] \quad (5.11)$$

and the same procedure can be performed for each two subsequent terms. In this way, we exclude from the sum all the spins with even indices. Thus, we can compare the partition functions for the site and block lattices, and write

$$Z(K, N) = F^{\frac{N}{2}} Z(K', N/2). \qquad (5.12)$$

Equation (5.11) gives for $S_1 = S_3$ and $S_1 = -S_3$

$$\begin{aligned} 2\cosh(2K) &= F \exp(K'), \\ 2 &= F \exp(-K'), \end{aligned} \qquad (5.13)$$

so that

$$K' = \frac{1}{2} \ln[\cosh(2K)], \qquad F = 2\sqrt{\cosh(2K)} \qquad (5.14)$$

Let us now compare the free energy per site, $f_1(K)$, for the site lattice and that, $f_2(K')$, for the block lattice. Since

$$f_1(K) = (1/N) \ln[Z(K, N)] \quad \text{and} \quad f_2(K') = (2/N) \ln[Z(K', N/2)],$$

it follows from Eq. (5.12)–(5.14) that

$$f(K') = 2f(K) - \ln\left(2\sqrt{\cosh(2K)}\right). \tag{5.15}$$

Moreover,

$$\begin{aligned}
K' &= \frac{1}{2} \ln\left[\{\exp(2K) + \exp(-2K)\}/2\right] \\
&< \frac{1}{2} \ln\left[\exp(2K)/2\right] = K - \frac{1}{2} \ln 2,
\end{aligned}$$

i.e., K' is smaller than K, and so by repeatedly doubling the size of the blocks one can proceed to a very small value of interactions between spins far away from each other. Therefore, on going in the reverse direction from small to large interactions, one can start with a very small value of the interaction, say $K' = 0.01$, so that $Z(K') \approx 2^N$ and $f \approx \ln(2)$. Using the above equations, one can find the renormalized interaction and the appropriate free energy. It turns out [18] that it takes eight steps to proceed from $K' = 0.01$ to the exact result.

The one-dimensional Ising model does not show a phase transition at a finite temperature, and our analysis was only designed to show by means of a simple example how the RG method works. We now turn to the two-dimensional Ising model, for which we have seen in Chapter 2 that a phase transition does occur.

5.4 RG: 2D Ising Model for the Square Lattice (1)

For the 2D Ising model on a square lattice, the partition function contains the interaction between a spin and its four nearest neighbors, so that the analysis is more complicated. However, we start by applying the same procedure as for the 1D Ising lattice, namely we form the block lattice by omitting the nearest neighbors of half the sites, as shown in Fig. 5.2, and summing over the interactions
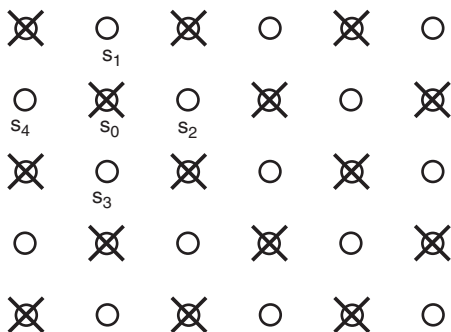
Fig. 5.2   Ising model on a square lattice with alternate sites removed.

with the spins of the sites that we omitted. Since the spin S_0 located at the site $(0,0)$ interacts with the spins S_1, S_2, S_3, S_4, located at the sites $(0,1), (1,0), (0,-1), (-1,0)$ respectively, we must consider the terms in the partition function arising from $S_0 = \pm 1$ with all the 16 possible values of (S_1, S_2, S_3, S_4). As we will see immediately, these 16 configurations are divided into four different groups which will give four different equations instead of the two equations (5.13) for the one-dimensional lattice. In order to satisfy these equations we are forced to introduce three different interactions in the partition function of a block lattice, i.e., to replace the cell of the site lattice shown in Fig. 5.2 by the following interactions in a block lattice

$$\exp[K(S_1 + S_2 + S_3 + S_4)] + \exp[-K(S_1 + S_2 + S_3 + S_4)]$$
$$= F\{\exp[(K_1/2)(S_1 S_2 + S_2 S_3 + S_3 S_4 + S_4 S_1)$$
$$+ K_2(S_1 S_3 + S_2 S_4) + K_3 S_1 S_2 S_3 S_4]\}. \tag{5.16}$$

(The reason for using $K_1/2$ rather than K_1 in the above equation is that the products of spins such as $S_1 S_2$ appear again in some other term of the partition function for the block lattice.) In order to determine the values of the four unknowns f, K_1, K_2, K_3 we equate the two sides of this equation for the four distinct values of (S_1, S_2, S_3, S_4), and take the logarithms of each side. For $(S_1, S_2, S_3, S_4) = (1,1,1,1)$, we find from Eq. (5.16) that

$$\ln[2\cosh(4K)] = \ln F + 2K_1 + 2K_2 + K_3. \tag{5.17}$$

Similarly, for $(S_1, S_2, S_3, S_4) = (1, 1, 1, -1)$ one obtains

$$\ln[2\cosh(2K)] = \ln F - K_3 \tag{5.18}$$

while $(S_1, S_2, S_3, S_4) = (1, 1, -1, -1)$ leads to

$$\ln 2 = \ln F - 2K_2 + K_3 \tag{5.19}$$

and $(S_1, S_2, S_3, S_4) = (1, -1, 1, -1)$ to

$$\ln 2 = \ln F - 2K_1 + 2K_2 + K_3. \tag{5.20}$$

This set of four linear equations for the four unknowns has the unique solution

$$
\begin{aligned}
K_2 &= K_1/2 = \frac{1}{8}\ln[\cosh(4K)], \\
K_3 &= K_2 - \frac{1}{2}\ln[\cosh(2K)], \\
F &= 2[\cosh(2K)]^{\frac{1}{2}}[\cosh(4K)]^{\frac{1}{8}}.
\end{aligned}
\tag{5.21}
$$

In contrast to the one-dimensional Ising model, for the two-dimensional Ising model, as well as in some others cases, the resulting interactions in a block lattice are more complicated than those in the original lattice, and so they are not suitable for an exact RG analysis. To proceed, one must use some approximation. The simplest approximation is to neglect K_2 and K_3, but then we come back to the equations considered in the one-dimensional case, where there is no phase transition. To obtain a physically meaningful result, we assume that the four-spin interaction is so weak that it can be ignored. The terms in K_1 and K_2 correspond, respectively, to interactions between nearest neighbors and next-nearest neighbors in the block lattice, and both favor the spins being parallel, and so we write $K_1 + K_2 = K'$. Thus,

$$K' = (3/8)\ln[\cosh(4K)]. \tag{5.22}$$

Finally, as discussed above, a phase transition occurs when K is a fixed point of the transformation, i.e., $K = K' = K_c$. For the 2D Ising model, this leads to $K_c = 0.50698$, while Onsager's exact result was $K_c = 0.44069$ [4], so that the error is about 13%. This is not

too bad for such an approximate theory, and suggests that its basic physical ideas are correct.

5.5 RG: 2D Ising Model for the Square Lattice (2)

Let us now choose a different form of the block lattice, in which we combine five neighboring spins into a block as shown in Fig. 5.3. The spin μ_a of the block lattice is determined by the majority rule

$$\mu_a = \text{sgn} \sum_{i=1}^{5} S_i^a. \tag{5.23}$$

As can be seen, the block lattice is also a square lattice with a distance between blocks of $L = \sqrt{5}$. For $\mu_a = 1$, there are sixteen possible values of S^a, which can be grouped into the six different configurations shown in Fig. 5.4, having characteristic energies 0, $\pm 2K$, and $\pm 4K$. For $\mu_a = -1$ one obtains the same energies as for $\mu_a = 1$ with signs reversed.

In order to calculate the partition function Z_μ for the Hamiltonian H_μ of the block lattice it is convenient to write $H_\mu = H_0 + V$, where H_0 contains the interactions between the spins within a block and V
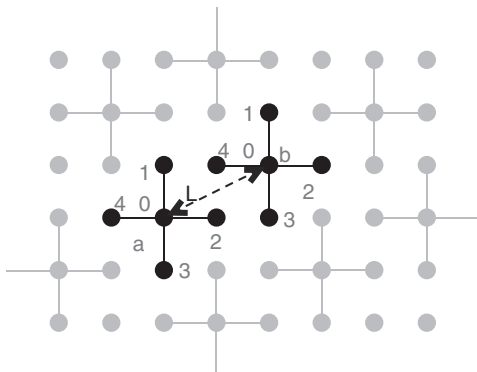
Fig. 5.3   Ising model on a square lattice with five sites in a block.
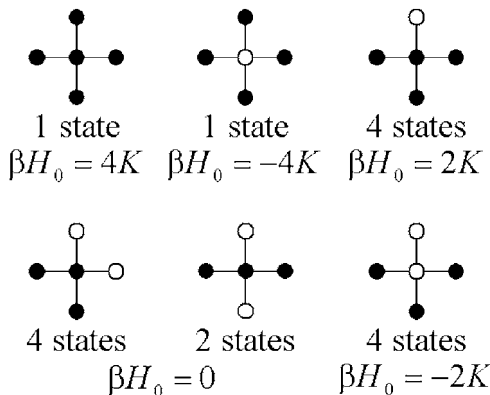
Fig. 5.4   The different spin configurations for the five sites in a block on the
square lattice, where black and white circles represent spins of opposite signs.
The number of states in each configuration and their energy are shown.

those between spins on adjacent blocks, so that

$$
\begin{aligned}
Z_\mu &= \sum \exp[-\beta H_\mu] \\
&= \sum \exp(-\beta H_o) \exp(-\beta V) \\
&= \sum \exp(-\beta H_o) \left[\frac{\sum \exp(-\beta H_o) \exp(-\beta V)}{\sum \exp(-\beta H_o)} \right]
\end{aligned}
\tag{5.24}
$$

where the first factor in the final expression in Eq. (5.24) is related
to the independent blocks, and therefore reduces to a multiple of the
partition function $Z_0(K)$ of a single block

$$
\sum \exp(-\beta H_o) = [Z_0(K)]^N.
\tag{5.25}
$$

The sixteen configurations that contribute to $Z_0(K)$, which are
shown in Fig. 5.4 for $\mu_a = 1$, lead to

$$
Z_0(K) = 2\cosh(4K) + 8\cosh(2K) + 6.
\tag{5.26}
$$

The second factor in Eq. (5.24) is the weighted average of
$\exp(-\beta V)$ with weight function $\exp(-\beta H_o)$, which will be denoted
by $\langle \exp(-\beta V) \rangle$.

   The formula (5.24) is still exact. However, we only evaluate
$\langle \exp(-\beta V) \rangle$ approximately, by the so-called cumulant expansion. In

this one writes

$$\langle e^W \rangle = \left\langle 1 + W + \frac{W^2}{2} + \frac{W^3}{6} + \cdots \right\rangle = 1 + \langle W \rangle + \left\langle \frac{W^2}{2} \right\rangle + \cdots. \tag{5.27}$$

For $|x| < 1$, $\ln(1 + x) = x - x^2/2 + \cdots$, and we only retain terms up to order W. Then, $\ln[\langle e^W \rangle] \approx \langle W \rangle$, so that

$$\langle e^W \rangle \approx \exp\langle W \rangle. \tag{5.28}$$

On substituting Eqs. (5.25) and (5.28) in Eq. (5.24), we obtain

$$\beta H_\mu = N \ln[Z_0(K)] - \langle \beta V \rangle. \tag{5.29}$$

The function $\langle \beta V \rangle$ connects adjacent spins of different blocks. As one can see from Fig. 5.3, the interaction between nearest neighbor spins on two adjacent blocks a and b contains only three terms,

$$\beta V_{ab} = K \left(S_2^a S_4^b + S_1^a S_4^b + S_2^a S_3^b \right). \tag{5.30}$$

Since the spins S^a and S^b are on different blocks, while the Hamiltonian H_0 neglects the interactions between blocks, on averaging with respect to H_0 we find that

$$\langle \beta V_{ab} \rangle = 3K \langle S_2^a \rangle \langle S_4^b \rangle \tag{5.31}$$

In evaluating $\langle S_2^a \rangle$ the configurations with βH_0 which contain $S_2^a = +1$ and $S_2^a = -1$ cancel out, and so one finds for $\mu_a = 1$ that

$$\langle S_2^a \rangle = \mu_a \left[\frac{2 \cosh(4K) + 4 \cosh(2K)}{2 \cosh(4K) + 8 \cosh(2K) + 6} \right]. \tag{5.32}$$

An analogous treatment for $\mu_a = -1$ shows that equation (5.32) applies also for this case. From Eqs. (5.31) and (5.32), we find that

$$\langle \beta V_{ab} \rangle = 3K \left(\frac{\cosh(4K) + 2 \cosh(2K)}{\cosh(4K) + 4 \cosh(2K) + 3} \right)^2 \mu_a \mu_b. \tag{5.33}$$

Table 5.1   The critical value of K, $K_c = kT_c/J$, and
the critical index x for different forms of renormal-
ization group theory applied to the Ising model for
a square lattice. The column Half is for the blocks of
RG-2D(1) with half the spins removed, and the column
5-spin is for the blocks of 5 spins of RG-2d(2), while
the column Exact presents the exact results obtained
by Onsager [4].

	Half	5-spin	Exact
K_C	0.5070	0.5931	0.4406
x	1.07	0.901	1

Hence, to first order in V, $\beta H_\mu = \text{Const} + K' \sum_{a,b} \mu_a \mu_b$, where

$$K' = 3K \left(\frac{\cosh(4K) + 2\cosh(2K)}{\cosh(4K) + 4\cosh(2K) + 3} \right)^2. \tag{5.34}$$

The fixed point corresponds to $K' = K = K^*$ in this equation. On
solving for it, we find for $kT_c = J/K^*$ that

$$kT_c = J \cosh^{-1} \left(\frac{2 - \sqrt{3} + \sqrt{9 - 2\sqrt{3}}}{2\left(\sqrt{3} - 1\right)} \right). \tag{5.35}$$

In Table 5.1 we show the results of calculations of the critical
value of K, $K_c = kT_c/J$, and of the critical index x defined by
$(dK'/dK)_{K^*} = L^x$, for the square lattice, obtained by the RG
method with the two different blocks considered above, and also
the exact results of Onsager [4]. By using cumulant expansions, one
can also perform the renormalization group calculations for the two-
dimensional triangular lattice [20], and the results obtained are close
to the exact results of Onsager.

5.6   Conclusion

The same idea of a comparison of site and block lattices which
was used in the previous chapter for the formulation of the scal-
ing hypothesis can be used also as a basis for the renormalization

group technique. The RG transformations, which were originally used in quantum field theory, are suitable for the phenomenon of phase transitions, where the main feature is the long-range correlations. Therefore, on going from small to larger block lattices, we gradually exclude the small-scale degrees of freedom. For each new block lattice one has to construct effective interactions, and find their connection with the interactions of the previous lattice. These steps are carried out repeatedly. The fixed points of RG transformations define the critical parameters, and series expansions near the critical point provide the values of the critical indices. We note that the fixed points are a property of transformations which are not particularly sensitive to the original Hamiltonian, and this is the basis of the idea of universality which will be considered in Chapter 7.

This page intentionally left blank

Chapter 6

Phase Transitions in Quantum Systems

Our treatment of phase transitions so far has considered only phenomenological properties and model systems, without any microscopic analysis of real physical systems. In this chapter, we extend our treatment by considering many-body quantum systems, in which the two best-known phase transitions are superfluidity and superconductivity. Since these phenomena are associated with the symmetry of the wave function and quantum statistical physics, we start with a brief review of these two topics.

6.1 Symmetry of the Wave Function

According to Heisenberg's uncertainty principle, the product of the uncertainty in the position, Δr, and that in its momentum, Δp, can never be smaller than Planck's constant, $\Delta r \Delta p \geq h$. As a result, if we could identify all the particles of the same type, for instance electrons, in a system at some given moment by the exact position of each one, we would have no idea what are their momenta, and so where each one would be at any subsequent time. Hence, we could no longer distinguish between them. As a result of this indistinguishability of identical particles, the amplitude of the wave function describing them must be left unchanged if the labels of the particles are permuted. Thus, for a pair of identical particles with coordinates (position and spin) ξ_1 and ξ_2, the wave function $\psi(\xi_1, \xi_2)$ describing their state can only be multiplied by a phase factor when the

particles are interchanged,

$$\psi(\xi_2, \xi_1) = e^{i\alpha}\psi(\xi_1, \xi_2).$$

On interchanging the particles once again, one finds that

$$\psi(\xi_1, \xi_2) = e^{i\alpha}\psi(\xi_2, \xi_1) = e^{2i\alpha}\psi(\xi_1, \xi_2),$$

and so, since one has now returned to the original system, $e^{2i\alpha}$ must equal unity, so that $e^{i\alpha} = \pm 1$. Hence $\psi(\xi_2, \xi_1) = \pm\psi(\xi_1, \xi_2)$, i.e., the wave function is either symmetric or anti-symmetric with respect to the interchange of particles. It is known from relativistic quantum theory that particles with integer spin, which are called bosons, have symmetric wave functions, and those with half-integer spin, which are called fermions, have anti-symmetric wave functions. In the latter case, two particles cannot have the same coordinates, since the requirement that $\psi(\xi_1, \xi_1) = -\psi(\xi_1, \xi_1)$ can only be satisfied if $\psi(\xi_1, \xi_1) = 0$, and this property is known as the Pauli exclusion principle.

If the wave function of a pair of particles $\psi(\xi_1, \xi_2)$ is expressed in terms of the product of single particle wave functions $\phi_1(\xi)$ and $\phi_2(\xi)$, then the normalized wave function is

$$\psi(\xi_1, \xi_2) = \frac{1}{\sqrt{2}}[\phi_1(\xi_1)\phi_2(\xi_2) \pm \phi_1(\xi_2)\phi_2(\xi_1)] \tag{6.1}$$

where the signs plus and minus correspond to bosons and fermions, respectively. For a system of N bosons, the symmetric wave function $\psi_s(\xi_1, \xi_2, \ldots, \xi_N)$ is given by

$$\psi_s(\xi_1, \xi_2, \ldots, \xi_N) = C_N \sum P\phi_1(\xi_1)\phi_2(\xi_2)\cdots\phi_N(\xi_N) \tag{6.2}$$

where the sum is over all possible permutations P of the coordinates ξ_j and C_N is a normalization constant. For N fermions the anti-symmetric wave function $\psi_a(\xi_1, \xi_2, \ldots, \xi_N)$ can be expresses as

a Slater determinant

$$\psi_a(\xi_1, \xi_2, \ldots, \xi_N) = A_N \left\| \begin{array}{cccc} \phi_1(\xi_1) & \cdots & \cdots & \phi_1(\xi_N) \\ \vdots & & & \vdots \\ \phi_N(\xi_1) & \cdots & \cdots & \phi_N(\xi_N) \end{array} \right\| \tag{6.3}$$

where A_N is a normalization constant. For this wave function, if two particles are in identical states then two rows of the determinant are identical, so that the wave function vanishes, in accordance with the exclusion principle.

6.2 Exchange Interactions of Fermions

Let us now consider a pair of fermions and distinguish between their spatial coordinates $\mathbf{r}_j$, and their spin coordinates σ_j. In the absence of spin-orbit interaction, we can express the anti-symmetric function $\psi_a(\xi_1, \xi_2)$ as a product of a function $\Phi(\mathbf{r}_1, \mathbf{r}_2)$ of the space coordinates and one $K(\sigma_1, \sigma_2)$ of the spin coordinates, i.e.,

$$\psi_a(\xi_1, \xi_2) = \Phi(\mathbf{r}_1, \mathbf{r}_2) K(\sigma_1, \sigma_2). \tag{6.4}$$

The condition that ψ_a be antisymmetric with respect to an interchange of the coordinates ξ_j then requires that one of these functions be symmetric and one anti-symmetric. If one expresses the function $K(\sigma_1, \sigma_2)$ in terms of the product of single particle spin functions, $\chi_1(s_1)$ and $\chi_2(s_2)$, then the symmetric state (triplet), with total spin $S = 1$, can be formed in three different ways,

$$\chi_1(s_1)\chi_1(s_2), \quad \chi_2(s_1)\chi_2(s_2) \quad \text{and} \quad \chi_1(s_1)\chi_2(s_2) + \chi_1(s_2)\chi_2(s_1),$$

while the anti-symmetric state (singlet), with total spin $S = 0$, has the form

$$\chi_1(s_1)\chi_2(s_2) - \chi_1(s_2)\chi_2(s_1).$$

For the triplet state the spatial wave function, $\Phi_-(\mathbf{r}_1, \mathbf{r}_2)$, is anti-symmetric, where for the singlet state, $\Phi_+(\mathbf{r}_1, \mathbf{r}_2)$, it is symmetric,

where the wave functions are

$$\Phi_\pm(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}[\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \pm \phi_1(\mathbf{r}_2)\phi_2(\mathbf{r}_1)]. \qquad (6.5)$$

According to first order perturbation theory, an interaction potential of the form $V(\mathbf{r}_1 - \mathbf{r}_2)$, such as the Coulomb potential, changes the energy of the states $\Phi_\pm$ by

$$\Delta E = \int |\Phi_\pm(\mathbf{r}_1, \mathbf{r}_2)|^2 V(\mathbf{r}_1 - \mathbf{r}_2) d\tau_1 d\tau_2. \qquad (6.6)$$

Substituting Eq. (6.5) into Eq. (6.6), one can rewrite Eq. (6.6) in the form

$$\Delta E = A \pm J_1 \qquad (6.7)$$

where A is the Coulomb interaction between the particles and J_1 is the exchange interaction

$$\begin{aligned} A &= \int d\tau_1 d\tau_2 |\phi_1(\mathbf{r}_1)|^2 |\phi_2(\mathbf{r}_2)|^2 V(\mathbf{r}_1 - \mathbf{r}_2), \\ J_1 &= \int d\tau_1 d\tau_2 \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\phi_1^*(\mathbf{r}_2)\phi_2^*(\mathbf{r}_1) V(\mathbf{r}_1 - \mathbf{r}_2). \end{aligned} \qquad (6.8)$$

It follows that $\Delta E = A + J_1$ for the singlet state, $S = 0$, while $\Delta E = A - J_1$ for the triplet state, $S = 1$, so that the physically important difference between these two energies is just $2J_1$. This difference can be expressed in terms of the spin operators $\mathbf{s}_1$, $\mathbf{s}_2$ and $\mathbf{S} = \mathbf{s}_1 + \mathbf{s}_2$ or $S^2 = s_1^2 + s_2^2 + 2\mathbf{s}_1 \cdot \mathbf{s}_2$. Since $s_1^2 = s_2^2 = \frac{1}{2}\left(\frac{1}{2} + 1\right) = \frac{3}{4}$ and $S^2 = S(S + 1)$ equals zero for the singlet state and two for the triplet states, the eigenvalues of the operator $\mathbf{s}_1 \cdot \mathbf{s}_2$ are equal to $-\frac{3}{4}$ and $\frac{1}{4}$ respectively for these two states. Hence, the eigenvalues of the operator $-J_1/2 - 2J_1\mathbf{s}_1 \cdot \mathbf{s}_2$, are just $\pm J_1$. Omitting the unimportant constants, we conclude that the two energies (6.7) are the eigenvalues of the operator $-2J_1\mathbf{s}_1 \cdot \mathbf{s}_2$. If one has many pairs of electrons associated with the sites of a lattice, one can describe them by the

so-called Heisenberg quantum operator

$$H = -2J_1 \sum_{(i,j)nn} \mathbf{s}_i \cdot \mathbf{s}_j \qquad (6.9)$$

Thus, the antisymmetry of the wave function, a purely quantum property, allows us to describe the effective interaction between the particles in the form of the spin operators, even though the original Hamiltonian was independent of spin. For the two electrons in the helium atom, the exchange integral (6.8) is negative, and, therefore, according to Eq. (6.9), the minimum energy corresponds to the antiparallel configuration of the electron spins. On the other hand, for ferromagnetic systems it is positive, and the parallel configuration of spins is preferable. We see that the Ising model (2.1) was a simplified version of the Heisenberg model (6.9), with scalar spins instead of vector operators.

One has to distinguish between the above "exchange interaction" (6.9) between spins and the interaction of the elementary magnetic moments μ connected with the spin $\hbar/2$ through the gyromagnetic ratio $e/(mc)$, $\mu = e\hbar/(2mc)$. While the former interaction is strong, being of order of $1000\,\mathrm{K}$, the latter is of order of $1\,\mathrm{K}$ and it is able to explain only the so-called dipole ferromagnetism. This makes clear what were the problems which P. Weiss had in 1907 (before the discovery of quantum mechanics) with his mean field theory described in Chapter 1.

6.3   Quantum Statistical Physics

Let us now consider some aspects of the statistical physics of quantum systems. We consider N non-interacting particles inside a vessel which is in contact with a thermal reservoir and a reservoir of particles at fixed volume. Such a system is described by the partition function of a grand canonical ensemble which has the form

$$Z = \sum_N \exp\left(\frac{N\mu}{kT}\right) \sum_n \exp\left(-\frac{E_{n,N}}{kT}\right) \qquad (6.10)$$

where $E_{n,N}$ are the energies of the different states n of a system containing N particles, and μ is the chemical potential. For such a system of non-interacting particles, the total number of particles present is $N = \sum_K n_K$ and the total energy $E = \sum_K n_K \epsilon_K$, where n_K is the number of particles with energy ε_K. Then the free energy $\Omega = -kT \ln Z$ can be written in the form

$$\Omega = -kT \ln \left\{ \sum_{K,n_K} \exp[\beta n_K(\mu - \epsilon_K)] \right\} \equiv \sum_K \Omega_K \qquad (6.11)$$

where as previously $\beta = 1/kT$.

For fermions, in accordance with the Pauli exclusion principle $n_K = 0$ or 1, and so

$$\Omega_K = -kT \ln\{1 + \exp[\beta(\mu - \epsilon_K)]\}.$$

The situation is more complicated for bosons, where there are no restrictions on n_K. On summing the geometric series in the exponent of (6.11), one finds that

$$\Omega_K = -kT \ln \left\{ \frac{1}{[1 - \exp[\beta(\mu - \epsilon_K)]]} \right\}$$

provided that $\exp[\beta(\mu - \epsilon_K)] < 1$, i.e., that $\mu < \varepsilon_K$. Since for the state of lowest energy $\epsilon_K = 0$, this condition requires μ to be negative. The mean number of particles $\langle n_K \rangle$ is defined from the relation $\langle n_K \rangle = -\partial\Omega_K/\partial\mu$, so that

$$\langle n_K \rangle = \frac{1}{\exp[\beta(\epsilon_K - \mu)] \pm 1} \qquad (6.12)$$

where the signs plus and minus are related to fermions and bosons, respectively.

We now consider the consequence of the requirement of negative μ for bosons. In order to do this, we need to derive the equation that determines μ. The volume of phase space available for a single state of a particle is $(2\pi\hbar)^{-1}$ [1], so that the number of states in phase

space with momentum between p and $p + dp$ is

$$N(p)dp = \frac{4\pi g V p^2 dp}{(2\pi\hbar)^3} \tag{6.13}$$

where V is the volume occupied by the system and $g = 2S + 1$ accounts for the different possible values of the spin. The chemical potential μ is defined from the conservation law of the number of particles,

$$N = \int N(p)\langle n(p)\rangle dp. \tag{6.14}$$

On transforming to the variable of integration $\epsilon = p^2/2m$, substituting Eqs. (6.12) and (6.13) into Eq. (6.14) and making some simple transformations, one finds that

$$\frac{N}{V} = A \int \frac{\sqrt{\epsilon}d\epsilon}{\exp[\beta(\epsilon - \mu)] - 1} \tag{6.15}$$

where A is some constant. With the change of variable $z = \beta\varepsilon$, this equation can be written in the form

$$\frac{N}{V} = A(kT)^{3/2} \int \frac{\sqrt{z}dz}{\exp(z - \beta\mu) - 1}. \tag{6.16}$$

Since N/V is constant, as the temperature decreases the integral must increase, and so the positive quantity $-\mu$ must decrease. However, μ vanishes not at $T = 0$, but at some finite temperature $T = T_c$ which is defined by the condition

$$\frac{N}{V} = A(kT_c)^{3/2} \int \frac{\sqrt{z}dz}{\exp(z) - 1}. \tag{6.17}$$

Up to now we have considered only particles in states with $\epsilon > 0$, and not allowed for the states with $\epsilon = 0$. In order to adjust Eqs. (6.16) and (6.17) so that they can be satisfied for $T < T_c$, one has to assume that Eq. (6.17) represents the number of particles in states with $\epsilon > 0$, and that all the remaining N_0 particles must be in the

state with energy $\epsilon = 0$. Hence,

$$N_0 = N[1 - (T/T_c)^{3/2}] \tag{6.18}$$

and a phase transition occurs at $T = T_c$, where the particles start filling up the state with $\epsilon = 0$. This quantum effect, which is called Bose–Einstein condensation, is the only known example of a phase transition in a system without interactions.

The question arises of how to observe Bose–Einstein condensation experimentally. In the above treatment it was assumed that the system considered was an ideal gas of bosons, with no interactions between the particles. However, because of inter-particle attractive interactions most gases solidify before the temperature becomes low enough for such quantum effects to be observable. A necessary condition for quantum effects to manifest themselves is that the thermal de Broglie wavelength λ of the particles be larger than the average distance r_0 between them. Since

$$\lambda = \frac{h}{p} = \frac{h}{Mv} = \frac{h}{M\sqrt{kT/M}},$$

where M is the mass of the particles, this condition can be written approximately as

$$\frac{h}{\sqrt{MkT}} > r_0. \tag{6.19}$$

This requirement is quite difficult to satisfy, since a small value of r_0 implies a high concentration of particles, which favors solidification. In order to satisfy the inequality (6.19) one needs either very light particles or very low temperatures. For instance, the lightest simple boson is the He[4] atom, and for this Kapitza and Allen observed in 1937 superfluid behavior at atmospheric pressure and temperatures of around $2\,\mathrm{K}$. Another possibility for satisfying the inequality (6.19) is to use very low temperatures for bosonic alkali metals, in which the inter-atomic interactions are comparatively weak, even though the mass of the atoms is not so small. Indeed, Bose–Einstein condensation has been observed recently in such systems at the incredibly low temperature of $10^{-7}\,\mathrm{K}$. We will now consider first Bose–Einstein

condensation in helium and the related phenomenon of superfluidity, and after that Bose–Einstein condensation in alkali metals.

6.4 Superfluidity

In this book, we will consider the simplest argument about superfluidity, which was presented by Landau [1]. Let us imagine a fluid flowing with velocity **u** along a tube at zero temperature. In the rest frame of the fluid, the walls are moving with velocity $-\mathbf{u}$, and if there is friction the fluid near the wall must move at some velocity **v** between 0 and $-\mathbf{u}$, and so the energy of the fluid increases. If this is impossible, then the fluid will flow without friction, which is just the phenomenon of superfluidity. An increase of the fluid's energy can only occur if the walls transfer some energy to the liquid. According to quantum mechanics, the energy must be transferred in the form of quasi-particles with energy ϵ and momentum **p**. In the original frame of reference, the energy obtained from the walls will be $\epsilon + \mathbf{p} \cdot \mathbf{v}$, and such an energy transfer is energetically favorable if this quantity is negative, $\epsilon + \mathbf{p} \cdot \mathbf{v} < \mathbf{0}$. The minimum value of the left hand side of this expression occurs when **p** is anti-parallel to **v**, in which case the condition becomes $v > \epsilon/p$. Since only under this condition is an energy transfer possible, the crucial question that arises is the minimum value of the ratio ϵ/p. Two possibilities are shown in Fig. 6.1. For a free particle $\epsilon = p^2/(2m)$, and the minimum value of ϵ/p (which occurs at $p = 0$) is zero, so that such an energy transfer is always possible and the fluid flow will always be accompanied

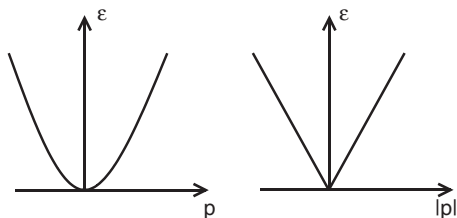Fig. 6.1 The energy ϵ as a function of the momentum p (a) for free particles; (b) for particles with $\epsilon \sim |p|$.

by friction. However, in systems in which $\epsilon \sim |p|$ (as for phonons in solids), the minimum value of $d\epsilon/dp$ is c_0. It is found [21] that in liquid helium, the excitations have a dispersion law of the form $\epsilon \sim c_0|p|$, so that for velocities less than c_0 the fluid must flow without friction, i.e., it behaves as a superfluid. The above considerations apply to a system at zero temperature. For finite temperatures T that are less than T_c, some thermal excitations already exist, but new ones will not appear for velocities smaller than c_0. Therefore, for temperatures $0 < T < T_c$, He^4 behaves as a mixture of a normal fluid and a superfluid. Such a behavior was first observed in 1937 in liquid He^4, for which the value of c_0 is $60\,cm/s$, and was explained by Landau in the 1940's.

6.5 Bose–Einstein Condensation of Atoms

For over fifty years, liquid helium was the only "quantum" fluid showing superfluidity. Over the years, it became clear that the same effect exists, in principle, for atoms heavier than helium, as soon as, according to criterion (6.19), one reaches a temperature much lower than $2.2\,K$ which is the phase transition temperature for Bose–Einstein condensation in helium. The real breakthrough in low-temperature experimental physics occurred in the 1980s and 1990s, when researchers managed to attain the unbelievably low temperatures of a few hundred nano-kelvin, which is a few hundred billionths of a degree above absolute zero! (Note that the lowest temperature found in Nature, which came from the Big-Bang residual radiation, is "only" $3\,K$.) At these ultra-low temperatures atoms are practically "stopped", having a velocity of a few centimeters per second. This achievement is especially impressive since it is impossible to control the behavior of neutral atoms by the use of the electric and magnetic fields commonly used for charged particles. The slowing down of atoms was achieved with the help of magneto-optical traps, using laser cooling by three mutually orthogonal laser beams and evaporative cooling by an inhomogeneous magnetic field [22]. It is no coincidence that after the 1997 Nobel prize "for the development of

methods to cool and trap atoms with laser light" the turn came of the 2001 Nobel prize "for the achievement of Bose–Einstein condensation (BEC) in dilute gasses of alkali atoms . . .". A Bose–Einstein condensate is a collection of atoms in a single macroscopic quantum state described by one single wave function. This wave function of a "super-atom" can now be calculated or observed as well as photographed! Many experimental groups have worked on this subject, and BEC has been observed in rubidium, potassium, lithium, cesium and spin-polarized hydrogen. Already in the first experiments on rubidium, the existence of a condensate was proved by measurements of the velocity distributions of the atoms, which after reaching the transition temperature changed from the Gaussian distribution appropriate to a classical gas to a narrow peak of velocities centred at zero velocity. Additional experiments include the resonance effect of two separate condensates, and the observation of a vortex lattice in a rotating condensate. The Bose–Einstein condensate is a new state of matter, and its study is very important for a full theoretical understanding of the macroscopic manifestation of quantum laws, which also appear in mesoscopic physics, a popular topic nowadays. Possible applications of BEC may include "atomic lasers" and different uses in nano-technology. This field is developing and changing rapidly, and we suggest that those reader interested in the subject should follow developments in the scientific periodicals. A simple presentation of the subject can be found in the Scientific American articles [23], and in Cornell's and Wieman's Nobel lectures [24].

6.6 Superconductivity

The experiments on and theories of superconductivity, and in particular of high temperature superconductivity, are a vast subject, which we will not even attempt to summarize in this book. However, since the phase transition from a normal to a superconducting state is of great theoretical and practical importance, it should not be omitted from a book on phase transitions. Accordingly, we present here a brief description of the main experimental facts and of the BCS (Bardeen–Cooper–Schrieffer) theory of type I superconductors,

which is sufficient for appreciating the special features of the super-conductivity phase transition. After this, we will consider the main features of the high temperature superconductors.

Superconductivity consists of the flow of electric currents in a material without any resistance, and so is a special case of super-fluidity, in which the electrons which are fermions create a bosonic "fluid". It was first observed by Kamerling-Ohnes in 1911, soon after he succeeded in liquefying helium and so obtaining very low temperatures. He originally expected that his experiments would support Lord Kelvin's idea (the dominant one at that time) that at very low temperatures the electrons will be localized around ions, rather than be free, so that the conductivity will vanish [25]. However, in complete contrast to this, he found that as the temperatures T is lowered through a critical temperature T_c the electrical resistivity of mercury suddenly drops to zero (and stays there as the temperature is lowered more). Some other important properties of superconductor materials, discovered subsequently, include the following:

(1) Magnetic fields are expelled from type I superconductors. This absence of a magnetic field inside a superconductor, which does not follow from the infinite conductivity, is known as the Meissner effect. Incidentally, this effect is not found in type II superconductors (the category to which belong all the high temperature superconductors discovered so far), which consist of coexisting superconducting and ordinary states.

(2) Although superconductivity is associated with the motion of electrons there is an isotope effect, so that for different isotopes of the same atom the critical temperature T_c is proportional to $1\sqrt{M}$, where M is the mass of the isotope.

(3) Superconductivity can be destroyed by a sufficiently high magnetic field or electrical current.

The standard theoretical explanation of superconductivity is the BCS theory. The main idea of this is that there is an attractive force between pairs of electrons, in spite of the Coulomb repulsion between them. Such a force is associated with the motion of the atoms in the solid, just as a ball rolling on the surface of a drum will depress the

surface, and surrounding balls will be attracted into this depression. The most important property of the resulting Cooper pairs of electrons is that they have integer spin and so are bosons, since (as we saw above) Bose–Einstein condensation and superfluidity are only possible for such particles. In a metal, the electrons of interest are those originally in states close to the Fermi surface, i.e., with energies close to the Fermi energy ϵ_F, which are the ones that carry the electric current in normal metals. Since for electrons at the Fermi surface the minimum value of $d\epsilon/dp$ is finite, such a "bosonic" gas of electrons will exhibit superfluidity which — for charged particles — means superconductivity.

The only property which still has to be explained is the origin of these electron pairs. Here we follow the two-page paper of Cooper (from which originates the name Cooper pairs) published in 1956 [26]. We consider the wave function $\psi(\mathbf{r}_1, \mathbf{r}_2)$ of a pair of electrons, and assume that it depends only on the distance between them, $\psi = \psi(\mathbf{r}_1 - \mathbf{r}_2)$, while the potential also depends only on this distance. Schrodinger's equation for $\psi(\mathbf{r}_1, \mathbf{r}_2)$ can be written in the form

$$(-\hbar^2/2m)(\nabla_1^2 + \nabla_2^2)\psi + V(\mathbf{r}_1 - \mathbf{r}_2)\psi = (E + 2\epsilon_F)\psi \qquad (6.20)$$

where the energy E is that of the pair relative to the energy at the Fermi surface, and must be negative for a pair to be formed. We take the Fourier transform of this equation, and write

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\mathbf{K}} g_{\mathbf{K}} \exp[i\mathbf{K}\cdot(\mathbf{r}_1 - \mathbf{r}_2)] \qquad (6.21)$$

which corresponds to one electron of the pair having momentum $\hbar\mathbf{K}$ and the other momentum $-\hbar\mathbf{K}$. On substituting Eq. (6.21) into Eq. (6.20) we find that

$$\frac{\hbar^2 K^2}{m} g_{\mathbf{K}} + \sum_{\mathbf{K}'} V_{\mathbf{K}-\mathbf{K}'} g_{\mathbf{K}'} = (E + 2\epsilon_F)g_{\mathbf{K}}. \qquad (6.22)$$

Let us now assume a potential such that

$$V_{\mathbf{K}-\mathbf{K}'} = -V_0 \text{ for } \epsilon_F < \epsilon_K, \epsilon_{K'} < \epsilon_F + \hbar\omega_D \qquad (6.23)$$

and $V_{\mathbf{K}-\mathbf{K}'}$ is zero otherwise, where ω_D is the Debye frequency of the phonons in the metal, which is proportional to $1/\sqrt{M}$. Such a potential is suitable for describing the phonon induced coupling between pairs of electrons in a spherical shell just above the Fermi surface, and its microscopic derivation is not important for understanding the superconducting phase transition. It then follows from Eqs. (6.22) and (6.23) that

$$g_K = \frac{V_0 \sum_{\mathbf{K}'} g_{\mathbf{K}'}}{\hbar^2 K^2/m - E - 2\epsilon_F}. \tag{6.24}$$

The summation of both sides of Eq. (6.24) over K leads to

$$1 = \sum_K \frac{V_0}{\hbar^2 K^2/m - E - 2\epsilon_F}. \tag{6.25}$$

On going from a sum over K to an integral over the spherical shell in which V_0 is non-zero, and converting the variable of integration from K to

$$\xi = \hbar^2 K^2/(2m) - \epsilon_F,$$

we then find that

$$V_0 \int_0^{\hbar\omega_D} \frac{n(\xi)d\xi}{2\xi - E} = 1 \tag{6.26}$$

where $n(\xi)$ is the density of states in the spherical shell, which to a good approximation is the density of states $n(\epsilon_F)$ at the Fermi level. Hence,

$$\ln \frac{2\hbar\omega_D - E}{-E} = \frac{2}{V_0 n(\epsilon_F)},$$

or

$$E \exp\left(\frac{2}{V_0 n(\epsilon_F)}\right) = E - 2\hbar\omega_D. \tag{6.27}$$

Since V_0 is assumed to be small and $n(\epsilon_F)$ finite, it follows that

$$\exp\left(\frac{2}{V_0 n(\epsilon_F)}\right) \gg 1.$$

Thus, finally,

$$E \approx -2\hbar\omega_D \exp\left(-\frac{2}{V_0 n(\epsilon_F)}\right) < 0 \qquad (6.28)$$

and so it it is energetically favorable for the two electrons to be bound together in a pair.

We now consider a number of other consequences of the Cooper formula (6.28):

(a) Firstly, we note that the binding energy E of a pair is proportional to ω_D and so to $1/\sqrt{M}$, so that, since one expects that $kT_c \sim |E|$, the critical temperature T_c will also be proportional to $1/\sqrt{M}$, which is just the isotope effect.

(b) Secondly, we note that V_0 enters the Cooper formula in the form

$$\exp\{-2/[V_0 n(\epsilon_F)]\},$$

which has an essential singularity at $V_0 = 0$. Hence, although V_0 is very small, one cannot obtain Cooper pairs by applying any sort of perturbation theory which uses expansions in powers of V_0.

(c) Thirdly, in order to destroy the coupling (so that the electrons become independent and contribute to normal electric conductivity rather than to superconductivity), one has to supply them with an energy of order of $|E|$ by increasing the temperature above T_c, or by an electric current.

(d) Fourthly, an essential element of the theory is that the electron pairs appear above the Fermi level, so that the relevant density of states is $n(\epsilon_F)$, and not some positive power of ϵ which would become zero at $\epsilon = 0$.

(e) Finally, the natural question arises as to why the Coulomb repulsion between the electrons does not destroy the Cooper pair. Leaving the full answer to the microscopic theory [21], the qualitative answer lies in the fact that the size of the Cooper pair is very large so that the Coulomb repulsion is small. In order to estimate the size a of the Cooper pair, we use the Heisenberg

uncertainty principle, and write

$$a \approx \frac{h}{\delta p_F} \approx \frac{h p_F}{m \delta \epsilon_F} \approx \frac{h p_F}{m \hbar \omega_D} \approx \frac{h}{p_F} \frac{\epsilon_F}{\hbar \omega_D} \gg \frac{1}{k_F}. \qquad (6.29)$$

Since the mean distance between electrons is of order $1/k_F$, it follows that the electrons in a pair are very far apart, with many other electrons between them, which explains why their mutual Coulomb repulsion is so small.

6.7 High Temperature (High-T_c) Superconductors

More than ten thousand scientific articles on high-T_c superconductivity have appeared since its discovery. It is impossible, therefore, to describe here even the main lines of investigations, which also change rapidly, and so we restrict ourselves to a very general picture, and some recent (at the time of writing this book) results.

For more than seventy years after the discovery of superconductivity, it was found in hundreds of pure materials, binary alloys and more complicated materials. In some of these materials it was quite unexpected, for instance in MoC where neither component is a superconductor. However, the applications of these materials were quite restricted due to the low critical currents (so that large currents, which produce a large enough magnetic field, destroy the superconductivity) and to the low critical temperatures. The highest critical temperature, 23.1 K, which was achieved in Nb_3Ge, is only slightly above the boiling point of liquid hydrogen (20.3 K at a pressure of one atmosphere). Among the restricted number of applications of low-temperature superconductors, one of quite general interest is Magnetic Resonance Imaging, which uses superconductors as the source of strong magnetic fields and is used to obtain images of soft issues in the human body, similar to those of bones obtained by X-rays.

The new era which marked the birth of high-T_c superconductivity started in 1986 when Bednorz and Muller [27] found a material with a critical temperature of around 35 K. In the following year (!) they received a Nobel prize for this discovery, and following their

discovery a large number of solid states laboratories started both an intensive study of different aspects of this phenomenon and a search for materials with even higher transition temperatures. Nowadays, a critical temperature of order $90\,K$ is obtained quite routinely. One important advantage of such high critical temperatures is that they are easily achieved by the liquefaction of nitrogen at $77\,K$. This is a much cheaper process than the liquefaction of helium which had to be used previously to obtain superconductivity. In addition to their high transition temperatures, the high-T_c superconductors also have high critical currents, which is very important for applications. The typical high-T_c superconductors (called in professional jargon "bisko" for BSCCO and "ibco" for YBCO) have the so-called perovskite-type structure with copper Cu, and oxygen O, located in well separated parallel planes of cuprite oxide, and atoms such as yttrium Y, bismuth B, strontium S, lanthanum La, or calcium Ca located between these planes. However, these materials have the disadvantage of being ceramics, which are difficult to manufacture in the form of wires as is required for most applications. This is one reason for the continuing intensive search for new superconducting materials which are easier to process mechanically and/or cheaper to produce, and not just for materials with higher critical temperatures or fields. One recent (April 2003) promising material in this class is magnesium diboride MgB_2, whose transition temperature of $39\,K$ is unexpectedly high for this type of material. Although this is well below the liquefaction temperature of nitrogen, it could lead to a family of similar materials with higher critical temperatures. This is why an extensive article in the March, 2003 issue of Physics Today was devoted to MgB_2 and also a special issue of a scientific journal [28] and the special April, 2003 Cambridge conference. Although the fragile ceramic structure of the standard high-T_c materials poses major problems for their practical applications, the Washington Post reported that American Superconductor Co. set a 1200-foot long electrical cable in the tunnel near Detroit which supplies electrical power for $14,000$ customers. The same company supplied 18 miles of superconducting wires to Pirelli Co. The era of magnetically levitated high-speed trains and loss-free electrical lines is approaching . . . .

The high-T_c superconductors raise many theoretical questions as well, the most important of which is whether they are described by the BCS theory which was so successful for describing the low temperature superconductors. Some additional ideas involve describing the Cooper pair as having the *d*-wave symmetry rather than spherical symmetric *s*-waves, as having a double-energy gap (for MgB_2, for example) rather than the simple gap described in the previous section, as having special "strips" connecting between atoms, etc. In order to understand the avalanche of scientific publications we suggest studying first the simply written article in Contemporary Physics [29], as well as different monographs specially dedicated to high-T_c superconductors.

6.8   Conclusion

The exchange interactions of fermions, which follow from their symmetry properties, is responsible, among other things, for the ordering of spins in ferromagnets described by the Heisenberg Hamiltonian. The statistical properties of bosons give rise to the phenomenon of Bose–Einstein condensation (BEC), in which particles accumulate in the state of zero energy. Superfluidity (the non-viscous motion of a system) is a macroscopic manifestation of BEC. A related phenomenon is superconductivity, in which electrons ordered in charged bosonic pairs also move without friction, or in other words without resistance. New developments in this area are connected with the discovery of BEC of atoms and superconductivity at high temperatures.

Chapter 7

Universality

In the previous chapters, we have seen that different models exhibiting phase transitions, such as the mean field model and the Ising model in two and three dimensions, have different sets of critical indices associated with the transition. Phase transitions are usually classified according to their sets of critical indices, such that systems with phase transitions having the same set of critical indices belong to the same universality class. This name is used because, as we show in this chapter, there are often a range of different systems belonging to the same universality class.

7.1 Heisenberg Ferromagnet and Related Models

As we saw in the last chapter, the exchange interaction between electrons leads to the Heisenberg quantum Hamiltonian

$$\hat{H} = -J \sum_{(i,j)nn} \mathbf{s}_i \cdot \mathbf{s}_j \tag{7.1}$$

(with $2J_1 = J$) and the associated quantum partition function

$$Z = Tr\{\exp[-\hat{H}/(kT)]\}.$$

Since near a phase transition the correlation length ξ is very large, quantum effects are unimportant here, and so we can replace the quantum operator $\hat{H}$ by the Hamiltonian of the classical Heisenberg

model,

$$H = -J \sum_{(i,j)nn} \mathbf{S}_i \cdot \mathbf{S}_j. \qquad (7.2)$$

This model is isotropic, but in real systems crystal symmetry in conjunction with spin-orbit coupling can lead to anisotropy, with different values of J for the different principal symmetry axes of the crystal. Accordingly, we now consider the generalized Heisenberg Hamiltonian, for which the x-axis and y-axis are equivalent but differ from the z-axis, so that

$$H = -J' \sum_{(i,j)nn} S_i^z S_j^z - J \sum_{(i,j)nn} (S_i^x S_j^x + S_i^y S_j^y). \qquad (7.3)$$

The Ising model, for which $J = 0$, and the ordinary Heisenberg model, for which $J' = J$, are special cases of this generalized Hamiltonian. It is known that the critical indices for the Ising and Heisenberg models are different. For the Hamiltonian (7.3), the critical indices of the phase transition are found to be the same as for the Ising model whenever $J < J'$. Only for $J' = J$ does one obtains the same critical indices as for the Heisenberg model. A further special case, $J' = 0$, the so-called x–y model, will be considered separately later on in this chapter. Thus, the critical indices do not depend continuously on the interaction energies J and J', but rather jump from one "universality class", that of the Ising model for all $J < J'$, to the Heisenberg universality class for $J = J'$. In addition, for spatial dimensions $d \geq 4$ the critical indices belong to the mean field universality class which is different from both the Ising and the Heisenberg universality classes.

In general, the universality class of a system depends only on its spatial dimensions d, the number of components (dimensionality) D of the order parameter and on whether the range r_c of the interactions is finite or infinite. With regard to the dependence on d and D, we have already seen that for the Ising and Heisenberg models there is a difference between one, two and three dimensions, while the difference between the Ising and Heisenberg models is in the value of D, as the dimension of the order parameter for the Ising model is unity and for the Heisenberg model is three.

In order to obtain a better understanding of the role of the range of the interaction, let us consider the infinite-range Ising model of N spins described by the Hamiltonian

$$H = -\frac{J}{N} \sum_{(i,j)} S_i S_j \tag{7.4}$$

where the summation is performed over all pairs of sites and not only over the nearest neighbors as it was in all previous formulae. The factor $1/N$ is needed since the energy has to be proportional to N, while the number of pairs in the sum is proportional to N^2. The double sum in Eq. (7.4) can be written as

$$\sum_{(i,j)} S_i S_j = \left(\sum_j S_j \right)^2 - N \;\; = \;\; S^2 - N \tag{7.5}$$

where $S = \sum_j S_j$ is the total spin of the system. The substitution of Eq. (7.5) into Eq. (7.4) leads to the partition function (2.1) becoming

$$Z_N = Tr \left[\exp \left(-\beta J + \frac{\beta J}{N} S^2 \right) \right] = \exp \left(-\beta J \right) \sum_{S_j = \pm 1} \exp \left(\frac{\beta J}{N} S_j^2 \right). \tag{7.6}$$

The last exponent in Eq. (7.6) can be written as

$$\exp \left(a S_j^2 \right) = \frac{1}{\sqrt{4\pi a}} \int_{-\infty}^{\infty} du \exp \left(-\frac{u^2}{4a} + u S_j \right) \tag{7.7}$$

so that, on inserting Eq. (7.7) into Eq. (7.6), we find that

$$\mathbf{Z}_N = \sqrt{\frac{N}{4\pi\beta J}} \exp \left(-\beta J \right) \sum_{S_j = \pm 1} \int_{-\infty}^{\infty} du \exp \left(-\frac{N u^2}{4\beta J} + u S_j \right)$$

$$= \sqrt{\frac{N}{4\pi\beta J}} \exp \left(-\beta J \right) \int_{-\infty}^{\infty} du \exp \left(-\frac{N u^2}{4\beta J} \right) [2 \cosh(u)]^N. \tag{7.8}$$

As $N \to \infty$ the Gaussian in Eq. (7.8) has a maximum that resembles a delta-function, and one can use the well-known method of steepest descents to show that the peak of the integrand is centered at

$$u^* = 2\beta J \tanh(u^*). \tag{7.9}$$

This equation is identical to Eq. (1.2), with u^* in place of M/M_0. It follows that the infinite-range Ising model (7.4) belongs to the mean field universality class. The above approximate calculation is appropriate for any dimension as long as each spin interacts with each other spin. The same mean field result follows from the exact solution of the one-dimensional model [30], in which the interaction $f(r)$ is very weak and decays exponentially with distance,

$$f(r) = -\gamma \exp\left[\frac{-r}{1/\gamma}\right] \quad \text{with } \gamma \to 0.$$

The reason that both systems with infinite-range interactions and systems in more than four space dimensions belong to the same universality class, that of mean field systems, is that such a large number of interactions affect each spin that their contribution can be well approximated by their mean value. Hence, systems with identical parameters d, D and the same range of interaction (i.e., finite or infinite) belong to the same universality class while other parameters do not influence the critical behavior. We now bring some examples.

(1) The Potts model [5], which has the same Hamiltonian as the Ising model but a spin S_j that can have the $(2S + 1)$ possible values $S, S-1, \ldots, -S$ for arbitrary integer or half-integer values of S, belongs to the same universality class as the Ising model, for which $S_j = \pm 1$.

(2) The critical behavior does not depend on the interaction energy and on the local structure of a system. For instance, it does not make any difference whether the set of spins considered are located on a simple cubic lattice, a body-centered one, a face-centered one, or a diamond-type lattice, even though the number of nearest neighbors of a site vary from four for the diamond lattice to twelve for the face-centered cubic lattice.

7.2 Many-Spin Interactions

We now describe briefly some more model systems. So far we have considered different forms of pair interactions between spins. Another group of models takes into account many-spin interactions described by a Hamiltonian of the form

$$H = -J_2 \sum_{(i,j)} S_i S_j - J_3 \sum_{(i,j,k)} S_i S_j S_k - J_4 \sum_{(i,j,k,l)} S_i S_j S_k S_j. \quad (7.10)$$

If $J_3 = J_4 = 0$ and sites i and j are nearest neighbors, this is just the Ising model. However, new phenomena appear when one takes into account four-spin interactions. For $J_3 = 0$ with J_2 and J_4 non-zero, which is known as the eight-vortex model, exact results were obtained for two dimensional systems by Baxter [31], who found that the critical indices depend continuously on the ratio J_2/J_4, so that there is no universality. Such non-universal behavior turns out to be an unpleasant surprise in the orderly picture of the theory of phase transitions.

There are several other cases in which non-universality can appear. For instance, let us consider the Union Jack lattice shown in Fig. 7.1, which is described by the following Hamiltonian

$$H = -J_1 S_0 (S_1 + S_2 + S_3 + S_4)$$
$$- \frac{J_2}{2}(S_1 S_2 + S_2 S_3 + S_3 S_4 + S_4 S_1) - J_3(S_1 S_3 + S_2 S_4). \quad (7.11)$$
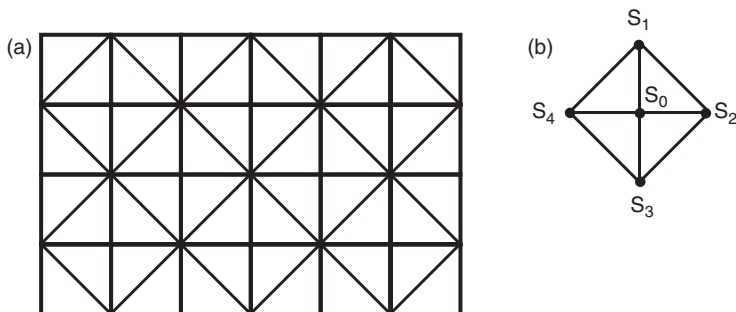

Fig. 7.1   (a) The Union Jack lattice; (b) elementary cell.

The factor $\frac{1}{2}$ in the second term of Eq. (7.11), like that in Eq. (5.16), takes into account the fact that each of the corresponding interactions appears twice, in two adjacent cells. The particular cases, $J_1 = J_3 = 0$ and $J_2 = J_3 = 0$, correspond to the nearest-neighbor Ising model. Moreover, when $J_3 = 0$ and both J_1 and J_2 are non-zero, the exact solution [32] again exhibits the Ising type critical behavior. However, it has been shown [33] that the inclusion of an additional next-nearest-neighbor interaction J_3, the value of which is determined by J_1 and J_2, leads to non-universal eight-vortex type behavior.

Although the last example seems to be quite artificial, new types of phase transitions may appear for models with competing two-spin interactions. So far we have considered only one type of interaction in any given system, ferromagnetic for $J > 0$, and antiferromagnetic for $J < 0$. However, the possibility exists that in some systems the interactions are competing. For instance the interaction between nearest neighbors can be ferromagnetic and that between next nearest neighbors antiferromagnetic. In the next chapter we will consider the so-called spin glasses, which are characterized by such competing interactions.

7.3 Gaussian and Spherical Models

The discrete Ising models with two-spin interactions can be generalized by other means. Two such models, the Gaussian model and the spherical model, were introduced by Mark Kac, to join "the small and select company of models in statistical mechanics which can be rigorously discussed" [34]. In both of these models instead of the spins S_j having a finite number of discrete values (± 1 in Ising model), they are continuous variables that can have any real value ($-\infty < S_j < \infty$). The similarity of these models to the Ising model is as follows. For the Gaussian model, it lies in the additional requirements that for each spin

$$\langle S_j^2 \rangle = 1. \tag{7.12}$$

The condition (7.12) is a "relaxed" version of the "strong" condition $S_j^2 = 1$ in the Ising model. The spherical model is closer to the Ising

model than the Gaussian model in that not only must Eq. (7.12) be satisfied, but there is also the overall restriction

$$\sum_{j=1}^{N} S_j = N. \tag{7.13}$$

Both conditions (7.12) and (7.13) are obviously fulfilled in the Ising model.

Let us start with the Gaussian model. A Gaussian distribution of the spin variables is assumed, with $\langle S_j \rangle = 0$ and $\langle S_j^2 \rangle = 1$, so that the probability $P(S_j)\, dS_j$ of finding a value S_j of the spin between S_j and $S_j + dS_j$ is given by

$$P(S_j)\, dS_j = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} S_j^2\right) dS_j. \tag{7.14}$$

Then, the partition function Z has the following form

$$Z = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dS_1 \cdots dS_N$$
$$\times \exp\left(-\frac{1}{2}\sum_j S_j^2 + \frac{J}{kT}\sum_{(i,j)nn} S_i S_j\right). \tag{7.15}$$

The integration in Eq. (7.15) can be easily performed using Eq. (7.5), and the result leads to the conclusion that the Gaussian model does not show a phase transition in any number of spatial dimensions. This result is not so surprising, since after transforming the integral in Eq. (7.15) to "normal coordinates" the partition function does not contain any interaction between these coordinates, and in such a system without interactions there are no phase transitions in any number of spatial dimensions.

For the spherical model, the integration in Eq. (7.15) has to be performed under the additional constraint (7.13) which can be written in the form of the delta function, $\delta\left(\sum_{j=1}^{N} S_j^2 - N\right)$. If the integral

representation of the delta function

$$\delta\left(\sum_j S_j^2 - N\right) = (2\pi)^{-1} \int_{-\infty}^{\infty} \exp\left[iu\left(\sum_j S_j^2 - N\right)\right] du \quad (7.16)$$

is substituted in Eq. (7.15), the following form of the partition function is obtained for the spherical model:

$$Z = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}+1} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dS_1 \cdots dS_N$$
$$\times \exp\left[iu\left(\sum_i S_i^2 - N\right) - \sum_j \frac{S_j^2}{2} + \frac{J}{\kappa T}\sum_{(i,j)nn} S_i S_j\right]. \quad (7.17)$$

Once again the transformation (7.5) can be used to perform integrations over S_j, just at was done for the Gaussian model, and the remaining integral over u, for the large values of N in which we are interested, can be carried out by the method of steepest descent. It turns out that a phase transition in the spherical model occurs for three and higher dimensional systems with critical indices different from those of the Ising model [35].

A geometric picture helps to clarify the relation between the Ising, Gaussian and spherical models [36]. If one considers each S_j as a component of an N-dimensional vector, then the Gaussian model corresponds to integration throughout space, the spherical model to integration over a spherical shell of radius $\sqrt{N}$, and the Ising model to summation over the corners of the unit hypercube inscribed within the unit sphere.

7.4 The x–y Model

We now turn to a model with properties very different from the above ones, namely the x–y model, for which the Hamiltonian is

$$H = -J \sum_{(i,j)nn} (S_i^x S_j^x + S_i^y S_j^y) \quad (7.18)$$

which corresponds to the generalized Heisenberg Hamiltonian of equation (7.3) with $J' = 0$. This model, with a two-component order

parameter, is equivalent to one with $\psi = \psi_0 e^{i\theta}$, since one can write $S_i^x = \psi_0 \cos(\theta_i)$ and $S_i^y = \psi_0 \sin(\theta_i)$, where θ_i is the polar angle. Due to the common multiplier J in Eq. (7.18), we can put $\psi_0 = 1$. Then

$$
\begin{aligned}
H &= -J \sum_{(i,j)nn} (\cos\theta_i \cos\theta_j + \sin\theta_i \sin\theta_j) \\
&= -J \sum_{(i,j)nn} \cos(\theta_i - \theta_j).
\end{aligned}
\tag{7.19}
$$

For low temperatures the argument of the cosine is small, and so we can expand it in terms of $\theta_i - \theta_j$. Then, apart from a constant term, the Hamiltonian (7.19) takes the following form

$$
H = \frac{J}{2} \sum_{(i,j)nn} (\theta_i - \theta_j)^2.
\tag{7.20}
$$

Since the Hamiltonian (7.20) is a function of the polar angle θ, it is invariant under the transformation

$$
\theta_i \to \theta_i + 2\pi.
\tag{7.21}
$$

We now go from the lattice model, in which the angle θ_j is defined on the sites j of a crystal lattice, to the continuous argument $\mathbf{r}$, $\theta = \theta(\mathbf{r})$. Then, the Hamiltonian (7.20) becomes

$$
H = \frac{J}{2} \int |\nabla\theta(\mathbf{r})|^2 \, d^d r.
\tag{7.22}
$$

Let us consider the correlation function $g(r_1 - r_2)$ of the phase of the order parameter

$$
g(\mathbf{r}_1 - \mathbf{r}_2) = \int \exp\{i[\theta(\mathbf{r}_1) - \theta(\mathbf{r}_2)]\} \exp\left[-\frac{J \int |\nabla\theta(\mathbf{r})|^2 \, d^d r}{2kT}\right] d\theta.
\tag{7.23}
$$

Using the (real) Fourier transform of $\theta(\mathbf{r})$:

$$
\theta(\mathbf{r}) = \sum_{\mathbf{K}} [\phi_{\mathbf{K}} \cos(\mathbf{K} \cdot \mathbf{r}) + \zeta_{\mathbf{K}} \sin(\mathbf{K} \cdot \mathbf{r})]
\tag{7.24}
$$

one can rewrite Eq. (7.23) as

$$g(R) = \iint d\phi_{\mathbf{K}} d\zeta_{\mathbf{K}} \exp \sum_{\mathbf{K}} \left[A_{\mathbf{K}} \phi_K + B_{\mathbf{K}} \zeta_{\mathbf{K}} - \frac{JK^2}{2kT} \left(\phi_{\mathbf{K}}^2 + \zeta_{\mathbf{K}}^2 \right) \right] \tag{7.25}$$

where $\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$ and

$$A_{\mathbf{K}} = \cos(\mathbf{K} \cdot \mathbf{r}_1) - \cos(\mathbf{K} \cdot \mathbf{r}_2), \qquad B_{\mathbf{K}} = \sin(\mathbf{K} \cdot \mathbf{r}_1) - \sin(\mathbf{K} \cdot \mathbf{r}_2). \tag{7.26}$$

On going from a sum over $\mathbf{K}$ in the exponent of Eq. (7.25) to an integral in d dimensions, and completing the squares in ϕ_K and $\zeta_{\mathbf{K}}$, one finds finally that

$$g(R) = \exp \left\{ -\frac{kT}{2J} \int \frac{d^d K}{K^2} \sin^2 \left(\frac{KR}{2} \right) \right\}. \tag{7.27}$$

It follows that for the different spatial dimensions d as $R \to \infty$:

$$
\begin{aligned}
g(R) &\sim \exp\left(-\frac{R}{\text{Const}} \right), & d &= 1; \\
g(R) &\sim \text{Const} \cdot R^{\frac{-kT}{2\pi J}}, & d &= 2; \\
g(R) &\sim \text{Const} \cdot \exp\left(\frac{\text{Const}}{R} \right), & d &= 3.
\end{aligned} \tag{7.28}
$$

In general, when $R \to \infty$ the correlation function $g(R)$ for a disordered system will decay exponentially, while in a completely ordered state it will be constant, showing the presence of long range correlations. It follows from Eq. (7.28) that there is no phase transition in a one-dimensional system while in three dimensions the system has long-range correlations. However, for $d = 2$ the correlation function has a power law decay, which corresponds neither to order nor to disorder, but rather to what is called quasi-order. Our calculation represents a special case of the general Mermin–Wagner theorem about the absence of a phase transition in a one-dimensional and a two-dimensional Heisenberg model [37]. In the last chapter we will consider such a power-law dependence as a fingerprint of self-organized criticality in non-equilibrium systems. For the present, we

examine in more detail the two-dimensional x–y model, and compare it with the Landau mean field theory.

The order parameter η in the Landau theory can be written as

$$\eta = \Psi_0 \exp(i\theta_0).$$

In the standard (with no fluctuations) Landau theory, Eq. (3.6), a phase transition occurs at a temperature T_L, and the amplitude of the order parameter Ψ_0 in the ordered phase is equal to $\sqrt{-A/(2B)}$, while the phase θ_0 is arbitrary. This theory ignores fluctuations, and so gives too low a value for the entropy S. According to Eq. (2.8), a decrease of entropy leads to an increase of the phase transition temperature. Thus, the critical temperature T_L predicted by the Landau theory will be higher than that predicted by a theory that takes fluctuations into account. In order to include fluctuations in the Landau theory, we proceed as in Chapter 3, using the amplitude and the phase of the order parameter. We replace M in Eq. (3.19) by

$$\psi(\mathbf{r}) = \psi_0 + \psi_1(\mathbf{r})e^{i\theta(\mathbf{r})},$$

where $|\psi_0|^2 = -\frac{A}{2B}$ as predicted by the Landau theory. Then

$$G = G_0 - \int \left[2A\left|\psi_1\left(r\right)\right|^2 - \left|\nabla\psi_1\left(r\right)\right|^2\right] d^d r$$
$$+ \psi_0^2 \int |\nabla\theta\left(r\right)|^2) d^d\mathbf{r} + \cdots. \tag{7.29}$$

Just as in the calculations of Chapter 3, we use the Gaussian approximation for fluctuations, and leave only quadratic terms in Eq. (7.29). The angular part of the Hamiltonian (7.29) is identical to the x–y model (7.22). For high temperatures, $T > T_L$, the system is in a disordered state with exponentially decaying correlations. At $T = T_L$, a non-zero amplitude of the order parameter appears, but there are still fluctuations of the phase of the order parameter. Another phase transition, the so-called Kosterlitz–Thouless phase transition occurs at some temperature T_{KT} smaller than T_L, $T_{KT} < T_L$. This phase transition is connected with the unbinding of vortices, which are bound in pairs for $T < T_{KT}$, and become free for $T_{KT} < T < T_L$, as discussed below.

7.5   Vortices

The equilibrium states corresponding to the x–y Hamiltonian (7.22) can be obtained from the Lagrange equation

$$\frac{\delta H}{\delta \theta} - \nabla \frac{\delta H}{\delta (\nabla \theta)} = 0 \qquad (7.30)$$

which for the x–y model reduces to

$$\nabla^2 \theta = 0. \qquad (7.31)$$

For the two-dimensional case in which we are interested, the solution of Eq. (7.31) has the form

$$\nabla \theta = \frac{1}{r}, \qquad \theta = \ln(r). \qquad (7.32)$$

The non-physical singularity of the function $\theta(r)$ at $r = 0$ can be removed by cutting out a small circle of radius a near the origin. This means that one includes a special type of fluctuation with zero order parameter in the small region surrounding $r = 0$. Such excitations around the equilibrium state are called topological charges or vortices, and the region $0 < r < a$ is called the core of the vortex. Vortices are characterized by their winding numbers, which are defined as follows. Consider the integral $\oint \nabla \theta \cdot d\mathbf{r}$ around any closed path enclosing the origin $r = 0$. Each circuit around the origin will, according to the symmetry property (7.21), bring us back to the original system with an increase of $2n\pi$ in the value of θ,

$$\oint \nabla \theta \cdot d\mathbf{r} = 2\pi n. \qquad (7.33)$$

Here n, which is called the winding number of the vortex, is a positive (for vortices) or negative (for anti-vortices) integer, depending on whether the rotation is clockwise or anti-clockwise.

  To answer the question whether vortices will appear as result of thermal fluctuations at temperature T, one has to calculate the change in the free energy $F = E - TS$ associated with the appearance of vortices. It follows from Eq. (7.22) and (7.32) that the energy

required to produce a vortex is

$$E = \frac{J}{2} \int_a^L |\nabla\theta|^2 \, d^2r = \pi J \ln\left(\frac{L}{a}\right) \qquad (7.34)$$

where L is a linear dimension of the system. On the other hand, since a vortex can appear everywhere in the system, the entropy of a vortex is

$$S = k \ln\left(\frac{\pi L^2}{\pi a^2}\right) = 2k \ln\left(\frac{L}{a}\right). \qquad (7.35)$$

Therefore if $T > T_{KT} = \pi J/(2k)$, the free energy $F = E - TS$ is negative, so that the appearance of a single vortex is energetically favorable. As we will see in the next section, for temperatures $T < T_{KT}$, the interaction between vortices results in the binding of vortex–antivortex pairs.

Incidentally, there is another example of a topological charge which is quite well known, though not necessarily by this name. This is the case of domain walls in ferromagnetic materials, which separate regions having opposite magnetizations. The number of domain walls increases as the critical temperature is approached from below, and the ferromagnetic–paramagnetic phase transition can be regarded as a topological one, originating from the proliferation of domain walls.

7.6   Interactions Between Vortices

The final point that we want to consider in connection with the x–y model is the interaction between vortices, and how they annihilate each other when the temperature is lowered below the Kosterlitz–Thouless temperature T_{KT}. Using Stokes theorem, one can rewrite Eq. (7.33) as

$$\oint \nabla\theta \cdot d\mathbf{r} = \int (\nabla \times \nabla\theta) \cdot d\mathbf{S} = 2\pi n. \qquad (7.36)$$

The second integral in Eq. (7.36) is identically zero unless there are singularities at isolated points $r = r_i$ within the contour,

in which case

$$(\nabla \times \nabla\theta) = 2\pi\mathbf{e}_z \sum_i n_i \delta(\mathbf{r} - \mathbf{r}_i) \equiv 2\pi\mathbf{n}(\mathbf{r}) \qquad (7.37)$$

where $\mathbf{e}_z$ is the unit vector perpendicular to the xy-plane. From these two equations one finds that

$$\int \mathbf{n}(\mathbf{r}) \cdot d\mathbf{S} = n. \qquad (7.38)$$

Let us define the velocity $\mathbf{v}$ of the phase θ to be its spatial gradient $\nabla\theta$, $\mathbf{v} \equiv \nabla\theta$. Since any vector $\mathbf{v}$ can be expressed as the sum of a longitudinal part $\mathbf{v}_l$ and a transverse part $\mathbf{v}_t$,

$$\mathbf{v} = \mathbf{v}_l + \mathbf{v}_t,$$

where $\nabla \cdot \mathbf{v}_t = 0$ and $\nabla \times \mathbf{v}_l = 0$, we assume such a division for $\mathbf{v}$, and so consider only $\mathbf{v}_t$, for which $\nabla \times \mathbf{v}_t = 2\pi\mathbf{n}(\mathbf{r})$ and $\nabla \cdot \mathbf{v}_t = 0$, so that

$$\nabla \times (\nabla \times \mathbf{v}_t) = -\nabla^2\mathbf{v}_t = 2\pi\nabla \times \mathbf{n}(\mathbf{r}). \qquad (7.39)$$

We solve this equation using the Green's function $G(\mathbf{r} - \mathbf{r}')$ for the operator ∇^2, i.e., the solution of

$$\nabla^2 G(\mathbf{r} - \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \qquad (7.40)$$

which for a two dimensional system is given by

$$G(\mathbf{r} - \mathbf{r}') = \frac{1}{2\pi} \ln\left(\frac{|\mathbf{r} - \mathbf{r}'|}{a}\right). \qquad (7.41)$$

Then, the solution of Eq. (7.39) can be written as

$$\begin{aligned} v_t(r) &= -2\pi \int d^2r' G(\mathbf{r} - \mathbf{r}') \left[\nabla \times \mathbf{n}(\mathbf{r}')\right] \\ &= -2\pi \int d^2r' \mathbf{n}(\mathbf{r}') \times \nabla G(\mathbf{r} - \mathbf{r}'). \end{aligned} \qquad (7.42)$$

by using the standard formula for $\nabla \times (G\mathbf{n})$ and Stokes' theorem. The energy of the system of vortices, U, is obtained by the substitution

of Eq. (7.42) into Eq. (7.22), which gives

$$U = 2\pi^2 J \int d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' [\nabla G(\mathbf{r} - \mathbf{r}') \times \mathbf{n}(\mathbf{r}')] \cdot [\nabla G(\mathbf{r} - \mathbf{r}'') \times \mathbf{n}(\mathbf{r}'')]. \tag{7.43}$$

On performing the partial integration over $\mathbf{r}$ and using Eq. (7.40) and (7.41), one finds that

$$U = \pi J \int d\mathbf{r}' d\mathbf{r}'' \ln\left(\frac{|\mathbf{r}' - \mathbf{r}''|}{a}\right) n\left(\mathbf{r}'\right) n(\mathbf{r}''). \tag{7.44}$$

This equation, which describes the energy U of interaction between two vortices, can be compared to the energy E of an individual vortex given by Eq. (7.34). While E is positive and diverges logarithmically with the size L of the system, the interaction energy U is negative for the vortex–antivortex pair. Thus, if single vortices and antivortices were to appear at low temperatures, $T < T_{KT}$, they would immediately bind together. As we have seen earlier, only for temperatures $T > T_{KT}$ does the appearance of individual vortex becomes energetically favorable. Without going into any further details, we only note [38] that for the topological Kosterlitz–Thouless phase transition the correlation length ξ is always infinite below T_{KT}, while above T_{KT} it diverges according to the non-trivial law,

$$\xi \sim \exp(\text{Const}/\sqrt{T - T_{KT}}).$$

Note the clear analogy between the behavior of vortices and that of the two-dimensional Coulomb gas with charges n and their interaction described by Eq. (7.44). In electrostatic language, the Kosterlitz–Thouless transition is analogous to unbinding the dipoles into single charges (a sort of insulator–metal transition).

7.7 Vortices in Superfluids and Superconductors

The x–y model is isomorphic to the superfluid and superconducting transitions, where the two-dimensional order parameter describes the wave functions of the Bose–Einstein condensate and of the Cooper pairs, respectively, while the interaction energy J has to be replaced

by the superfluid density or by the electrical resistance of the two-dimensional films. In a three-dimensional superconductor, the presence of an external magnetic field will make the system effectively two dimensional in the plane perpendicular to the field. The discovery of the high temperature superconductors stimulated a huge interest in the theory of vortices, since these materials are type II superconductors where the magnetic fields enters the superconducting materials in the form of vortices. We refer the reader to the current scientific literature for further details. As a starting point, one can use the comprehensive reviews of the theory of vortices in superfluids and superconductors [39, 40].

7.8 Conclusion

The Ising model can be extended to include many-spin interactions, which can lead to drastic changes of its properties, and, in particular, to a dependence of the critical indices on the interaction energies (non-universality). The Ising model, as such, is a special case of the more general Heisenberg model, which for the three-dimensional case gives a phase transition with critical indices different from those of the Ising model. Another particular case of the Heisenberg model is the x–y model, with vortices as an essential part of it. The latter play a crucial role in the modern theory of superfluidity and superconductivity. Another pair of soluble models are the Gaussian and spherical models, with the critical indices in the spherical model different from those in other models. Nevertheless, all these models can be arranged by dividing them into a few universality classes. The university class to which a system belongs is specified by the following three factors:

(1) The spatial dimension d. In this respect, the lower critical dimensions, i.e., the minimum dimensions for the occurrence of phase transitions, are two for the Ising and three for the Heisenberg model. The upper critical dimension for these models is four. Indeed, for four and higher dimensions the phase transitions are described by mean field theory, with its own universality class.

(2) The dimension (number of components) D of the order parameter. This dimension is unity for the Ising model and three for the Heisenberg model, and so these two models lead to different critical behavior in three-dimensional systems.

(3) The finite or infinite radius of the interaction. In the latter case, each object interacts with all the others in the system, and one can neglect fluctuations (a situation similar to that in high-dimensional systems) which immediately leads to the mean field universality class.

All other factors, which define the short-range behavior (number of states at each site, type of the crystal lattice, etc.) of the system, do not influence the critical behavior.

This page intentionally left blank

Chapter 8

Random and Small World Systems

Most of the systems that we have considered so far in this book have been ordered ones, with definite well-defined spatial arrangements and interactions. In this chapter, we consider the effects of disorder or randomness. We consider first a simple example of a disordered or random system, namely that of a lattice in which there is a given probability that a bond exists between a pair of adjacent sites. Although the presence or absence of a bond is a purely geometrical property, with no thermal effects involved, this system is found to have many features similar to those of a phase transition, such as a critical fraction of bonds for percolation to occur, critical indices, and the applicability of renormalization group techniques, and so we consider it first. In fact, it is formally possible (although not very instructive) to map the bond percolation problem on to the Potts model [36] mentioned in Chapter 7. After this we consider the Ising model with random interactions or sites removed at random, and spin glasses, in which there are competing interactions. Finally, we consider small world systems, in which there are a limited number of long-range interactions between randomly chosen sites.

8.1   Percolation

The conceptually simplest example of a random system is a lattice of sites linked by bonds between adjacent sites in which a fraction x of the bonds is removed at random. In such a system, the disorder or randomness is uniquely specified by a single parameter, the fraction

x of bonds or sites that have been removed or the fraction $p = 1 - x$ of bonds that are present. For $x = 0$, which corresponds to $p = 1$, it is obvious that a continuous path of bonds exists spanning the system from one side to the other, and this will still be true if x increases and p decreases slightly. On the other hand, if $p = 0$ there is obviously no such path, and this will also be true for sufficiently small values of p. Hence it is reasonable to expect that there is a critical value p_c of p such that if $p < p_c$ then no continuous path exists that spans the system, and only isolated clusters of connected sites exist, while if $p > p_c$ an infinite cluster exists that spans the whole system. The critical value p_c is known as the bond percolation threshold, and the phenomenon is known as percolation, since if the bonds correspond to open pores in a block of material then once this fraction of pores is opened a fluid will percolate from one side of the block to the other side. It is found empirically that for a three-dimensional lattice on which the coordination number is z [41, 42],

$$p_c \approx 1.5/z,$$

while for a square lattice it can be proved that $p_c = 0.5$. A related problem is that of site percolation, in which sites rather than bonds are either occupied or empty.

The percolation transition is a geometrical phase transition, in which the critical concentration p_c separates a phase of finite clusters for $p < p_c$ from a phase with $p > p_c$ in which an infinite cluster is present. This is similar to the thermal phase transitions considered in previous chapters, with the concentration of bonds p playing the role of the temperature. Just as there, the transition is characterized by the properties of clusters close to the percolation threshold. For instance, a quantity that behaves like an order parameter is the probability P_∞ that a bond belongs to an infinite cluster. By definition, $P_\infty = 0$ for $p < p_c$, while for $p > p_c$ it is found that it is related to $p - p_c$ by a power law

$$P_\infty \sim (p - p_c)^\beta \tag{8.1}$$

similar to that for the order parameter in Table 4.1. Another quantity of interest is the correlation length ξ, which is the mean distance

between two sites in a finite cluster. For p close to p_c, this is given by

$$\xi \sim |p - p_c|^{-\nu} \tag{8.2}$$

which can be compared to the behavior of the correlation length in Table 4.1. Other quantities of importance having similar types of behavior include the mean number of bonds in a finite cluster, the distribution of finite clusters of different sizes, and for $p > p_c$ the probability that a bond belongs to the infinite cluster. The problem of percolation, and especially in a continuum, is of great practical importance in many fields, such as the conductivity of a mixture of conductors and insulators and the extraction of oil from an oil-field. A comprehensive review of the percolation transition and its relationship to thermal phase transitions was presented by Essam [43].

8.2 Ising Model with Random Interactions

So far, we have considered the Ising model

$$H = - \sum_{(i,j)nn} J_{ij} S_i S_j \tag{8.3}$$

with uniquely specified values for the interactions J_{ij} between the spins on different sites. When these interactions are all ferromagnetic, $J_{ij} = J > 0$, the equilibrium state at any given temperature is uniquely defined, and if a phase transition occurs, it is between a well-ordered state and a disordered one, with a change of the symmetry of the system at the phase transition point. The Ising model with an antiferromagnetic interaction, i.e., with $J_{ij} = J < 0$, usually has similar properties. However, for certain types of lattice, there is not a unique equilibrium state at low temperatures because of frustration. The simplest example of frustration is the antiferromagnetic triangular lattice in two dimensions. Let us consider three mutually adjacent atoms, A, B, and C. If the spin at site A is up, the interaction favors the spins at sites B and C to both be down, while the interaction between the spins at B and C wants them to be of opposite sign. We now examine the Ising model for random systems, in

which the values of J_{ij} are random variables with a probability distribution $P(J_{ij})$, rather than the random values of the spin considered in Section 7.3.

A simple type of random system is one with diluted ferromagnetic bonds,

$$P(J_{ij}) = x\delta(J_{ij}) + (1-x)\delta(J_{ij} - J)$$

for (i, j) nearest neighbors and zero otherwise, which is the usual Ising model with a fraction x of the bonds removed. A related system is the ferromagnetic system with fraction x of the sites removed,

$$H = -J \sum_{(i,j)nn} c_i c_j S_i S_j$$

with $P(c_i) = x\delta(c) + (1-x)\delta(c-1)$. In these cases, for sufficiently low values of the dilution parameter x the system exhibits a phase transition from a paramagnetic state to a ferromagnetic one as the temperature is reduced. However, for sufficiently large values of x, there is no ferromagnetic phase but only a transition from the paramagnetic phase to the Griffiths phase [44], in which there are clusters of ordered spins but not all the clusters are ordered in the same direction. The above results are connected with percolation theory. For sufficiently large values of x, which correspond to sufficiently small values of p in the percolation problem, there is no connected cluster of spins spanning the system but only isolated clusters of connected spins, and at sufficiently low temperatures the spins in each cluster will be ordered, but those for different clusters will be ordered in arbitrary directions.

Another type of system is one in which the interactions are not only of random strength but also of random sign, for instance with $P(J_{ij})$ having a Gaussian distribution,

$$P(J_{ij}) = (2\pi\Delta)^{-\frac{1}{2}} \exp[-J_{ij}^2/(2\Delta)]$$

or a bimodal distribution

$$P(J_{ij}) = x\delta(J_{ij} + A) + (1-x)\delta(J_{ij} - A),$$

frequently studied for the symmetric case $x = \frac{1}{2}$. In systems with such combinations of ferromagnetic and anti-ferromagnetic interactions, frustration of the spins occurs, similar to that mentioned above for the antiferromagnetic Ising model on a triangular lattice. As a result, these systems, which are known as spin glasses, have no unique ground state.

8.3 Spin Glasses

There are a variety of magnetic systems in nature that behave as spin glasses. One type is that of magnetic ions such as iron or cobalt inserted as substitutional impurities at random sites in a lattice of non-magnetic metal ions such as copper. These systems have frozen disorder, similar to that of a glass at low temperatures, and the interactions between the spins on the magnetic ions are indirect, i.e., they are mediated by the spins of the electrons attached to the host atoms or by free electrons. In the latter case, there is an effective magnetic interaction of the Ruderman–Kittel form [45]

$$\phi(r) = \phi_0 \sin(k_F r)/(k_F r)^3,$$

where k_F is the wave number of electrons on the Fermi surface, so that the sign of this interaction varies with the distance r between the magnetic ions.

A major difference exists between ferromagnetic system and spin glasses with respect to their symmetry properties. In ferromagnetic materials, when the temperature is raised through T_c the equilibrium state of the system changes to the paramagnetic one, with a sudden change in the magnetic moment M and symmetry breaking. Moreover, in ferromagnetic systems such as the original Ising model, at temperatures $T < T_c$ the system is ferromagnetic, with all the spins pointing in the same direction, and there is degeneracy between the states with all spins up and all spins down. In the thermodynamic limit, these two states are separated by an infinite potential barrier, so that no transition between them can occur. As a result, in the ferromagnetic state not all states in phase space are allowed. Since

a system is ergodic only when it is free to sample all points in phase space, this means that the ergodicity is broken at T_c. Incidentally, the degeneracy of the ferromagnetic state can be removed (in order to produce a unique ground state) by applying a small magnetic field or by introducing suitable boundary conditions. Thus, for ferromagnetic systems, both symmetry and ergodicity are broken at T_c. In spin glasses, by contrast, while there is a critical temperature T_c, the mean magnetic moment $\langle M \rangle$ is zero both below and above it, and the transition is characterized by a cusp-like singularity of the susceptibility at T_c [46]. Below the critical temperature there are many competing ground states, separated by infinite barriers. This means that for spin glasses the ergodicity is broken at T_c (as a result of frustration), but this broken ergodicity is not accompanied by broken symmetry.

In all random systems, the macroscopic properties are those of a configurational average of the system, which we denote by $\langle \cdots \rangle$, i.e., the property has to be calculated separately for each possible accessible configuration of the system and an average taken (with the appropriate weights of the probability of that configuration being found) over all the possible configurations. In general, there are two types of random system, annealed or quenched, depending on the rate of change of the disorder, for instance as a result of the mobility of magnetic ions in spin glasses. In annealed systems this rate is fast, so that one requires the average of the positions of the impurities which enter the partition function, and one has to average the partition function $\langle Z \rangle$. On the other hand, in quenched systems this rate is slow and the disorder is essentially fixed, so that one has to minimize, the free energy $F = -kT \ln Z$ individually for each possible configuration, and so the important quantity is $\langle \ln Z \rangle$. In spin glasses the disorder is quenched, so that for the macroscopic system one has to average over the free energy for different configurations rather than over the partition function. This is the main problem in the theoretical analysis of spin-glasses.

There are two well-known models for spin glasses, the "short-range" Edwards–Anderson and the "infinite-range" Sherrington–Kirkpatrick models, which overcome the difficulty stemming from

the nonlinearity of the logarithm function appearing in the expression for the quenched free energy by the so-called replica method [47] based on the self-evident identity,

$$\ln Z = \lim_{n \to 0} \frac{Z^n - 1}{n}. \tag{8.4}$$

Equation (8.4) means that one can calculate the quenched average free energy by creating multiple copies of the system, averaging them, and then letting the number of copies tend to zero. The calculation of Z^n is similar to that performed in Chapter 7 for the spherical model, but with many delicate points, such as the meaning of the limit $n \to 0$, its interchange with the thermodynamic limit, etc.

There are a number of problems from different fields that can be related to the spin glass problem, and we will mention two of them. The theory of spin glasses can be used to derive computer algorithms for the solution of optimization problems, such as that of the travelling salesman [48] (who has a large number of possible routes, and if he spends all his time trying to work out the best one he will never sell anything). Another problem related to spin glasses is that of neurons in the brain, which can be in one of two states ("firing" or "non-firing"), just like the spins in the Ising model. An appropriate model to describe the possible patterns of neural connections was introduced by Hopfield [49], and has led to an enormous amount of work on neural networks, with applications to understanding such functions of the brain as learning, and to attempts to mimic them so as to obtain self-correcting systems.

8.4   Small World Systems

We are living in a large world, and already in 1999 the population of the Earth exceeded six billion people. On the other hand, if we suddenly meet our neighbor somewhere in the Amazon jungle, we exclaim: "This is a small world!". The simple "large-small" games can be illustrated by the Bacon and Erdos numbers. The famous actor Kevin Bacon played in many movies with different actors, who can thereby be characterized by their different Bacon numbers, where

the Bacon number of an actor is one if he appeared in the same film as Kevin Bacon (1,472 people), two if he appeared in the same film as an actor with Bacon number one (110,315 people), etc. The number of people with Bacon number three is 260,123, and so on. In the world of science, a similar process is connected with the name of a legendary mathematician Paul Erdos, who published more than 1,500 scientific papers with 458 collaborators, each of whom has Erdos number one. Each of the additional 4,500 scientists who collaborated with the latter have the Erdos number two, and so on. It turns out that majority of scientists publishing articles in the last sixty years have quite small Erdos numbers. One of the authors of this book (M.G.) has the Erdos number three, having worked with G. Weiss who worked with J. Gillis who worked with P. Erdos. Correspondingly, the second author (V.H.) has at least Erdos number four, as a result of working with the first author.

However the first real, though quite bizarre, experiment was carried out by the psychologist Milgram [50], who tried to find the "distance" between two randomly chosen American citizens, which is defined as the number of acquaintances needed to connect these two individuals. The results that one expects from such an experiment are not at all clear. On the one hand, each person knows (calling them by the first name) about one hundred individuals, each one of those knows another hundred, etc., so 280 million Americans will be covered already in five steps $(280,000,000 \ll (100)^5)$. This simple argument is obviously wrong, since generally friends of my friends are likely also to be my friends. The opposite point of view is that an infinite number of steps is needed to find a connection between, say, a single homeless person and the president of the USA. The surprising result of Milgram's experiment was that everyone in the world is connected to everyone else through a chain of at most six mutual acquaintances. The legendary phrase "six degrees of separation" became an integral part of our "small world".

However, this subject was given a firm scientific basis, and the concept of "small world" became a normal scientific term, only thirty years later in the outstanding article of Strogatz and Watts [51]. They considered the transition from a regular to a random graph. In

general, a graph is defined as a set of sites or vortices with connections between them which are called bonds or links. A regular graph is one in which each vortex is linked to its k closest neighbors (which can include nearest neighbors, nearest and next nearest neighbors, etc.), while in a random graph each vortex is linked to k arbitrary vortices. There are two important parameters which characterize the long-range and short-range properties of a graph containing $N = L^d$ sites and $\frac{1}{2}Nk$ links, where L is the linear size of the graph and d is the spatial dimension. The long-range parameter is the average minimal number of links $h(l)$ required to go from one site to another one a distance l from it, and this is called the shortest path distance between the sites. The l-dependence of this shortest path distance is quite different for regular and random graphs. While for a regular graph, it is obvious that $h(l) \sim l$, for a random graph $h(l) \sim \ln(l)$. For a simple argument to support this result, we note that the each site has on the average k sites linked to it, and each of these is also linked to k sites, so that a given site has k^2 sites with $h = 2$, etc. Thus after h steps one can reach a site located a distance l away, where $k^h = l$ or $h = \ln(l)/\ln(k)$. One can also define a clustering coefficient c which is the average fraction of neighboring sites linked to a given site. This parameter defines a short-range order, and is obviously close to unity for a regular graph and very small for random graphs. Hence, for a regular graph h and c are both large, while for a random graph they are both much smaller.

A new type of graph, intermediate between regular and random, as shown in Fig. 8.1, was introduced by Strogatz and Watts in the following way [51]. They started from a regular one-dimensional graph, where each site is connected to its k nearest neighbors. They then replaced some of these connections by links between randomly chosen pairs of sites ("shortcuts"), with probability p per link, which we will call the randomizing number. This leads to an average of $pkN/2$ such shortcuts. The two limit cases are $p = 0$ (regular graph) and $p = 1$ (random graph). For all intermediate values, $0 < p < 1$, the graphs are partially random (disordered). Strogatz and Watts founded that already for a small value of p, when the graph is still locally ordered having only a small number of "shortcuts", the graph exhibits a
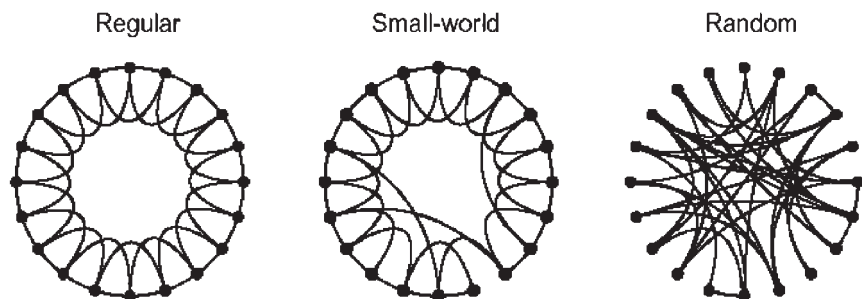
Fig. 8.1   Three types of graphs: regular, random and regular with a small number of "shortcuts" (small world). Reprinted by permission from Nature [51] copyright 2003, Macmillan Publishers Ltd.

different behavior, namely $h(l) \sim \ln(l)$, as in random graphs, but c remains close to unity, similar to regular graphs. Such a behavior is known as "small world" [51]. The drop in h from $h \sim l$ to $h \sim \ln(l)$ is connected with the small number of shortcuts which bring together previously distant sites.

Now that we have three types of graph, regular ("large world"), disordered ("small world") and random, we have to decide which type is most common in real life. To do this, it is convenient to change our terminology, and assume that the sites in the graphs are occupied by some objects, and that links correspond to interactions between these objects. These new systems will be called networks. The concept of networks is very general, and incorporates such systems as interacting atoms and galaxies, different groups of people, sites on the Internet, etc. In their original article Strogatz and Watts considered three different networks (a sociological system of film actors, a biological system of the worm's neural network, and the man-made power grid in the western US), and found that all of them exhibit the small world phenomena. A more recent development is that many other networks belong to the small world category (see Watts' book [52] and some recent reviews [53,54]). In addition to the topological characteristics h and c of the network, some new parameters (efficiency and cost, geographic distances [55], etc.) have been introduced for the description of real networks.

8.5 Evolving Graphs

So far we have examined graphs with a fixed number of sites and links. However, most real systems are "living", i.e., their size L changes with time. Accordingly, let us consider graphs which are characterized by three parameters, the number k of closest neighbors, the randomizing number p, and the total number of sites L^d. In the previous section we considered the p-dependence of two characteristic parameters, the path length $h(l)$ and the cluster coefficient c, provided that k and L remained constant. As we saw, the shortest path distance $h(l)$ between two sites distance l apart is quite different for regular and random graphs, being proportional to l and $\ln(l)$, respectively. The question arises as to the conditions for a transition from one regime to another to take place, or in other words how large has the disorder to be for a transition to take place from a regular lattice to a random one. Roughly speaking, a graph is regular when the fraction of random links N_{rd} is small, $N_{rd} \ll 1$, and it is random when this fraction is large, $1 - N_{rd} \ll 1$. The total number of links is $\frac{1}{2}kL^d$, where d is the spatial dimension, and the probability of a random link is p, so that the average fraction of random links N_{rd} is given by $N_{rd} = \frac{1}{2}kL^dp$. Since N_{rd} is small for regular graphs and close to unity for random ones, it seems natural to assume that the transition occurs at some "crossover" size L^* such that

$$N_{rd} = \frac{1}{2}kL^{*d}p = \frac{1}{2}, \qquad (8.5)$$

i.e., the graph with given k and p is regular when $L < L^*$ and random when $L > L^*$. The heuristic arguments which lead to Eq. (8.5) are supported by the renormalization group analysis [56]. For a one-dimensional system, it follows that

$$L^* \sim \frac{1}{p}. \qquad (8.6)$$

We note that the transition from a regular to a random graph occurs at some value L^* of L no matter how small is the randomizing number p. Moreover, an analogy exists between this "geometric" phase transition and the "physical" phase transitions considered in previous

chapters, with L and p playing the role of the correlation length and the dimensionless temperature, respectively. Later on we will compare the characteristic length L^* with the correlation length ξ.

8.6 Phase Transitions in Small World Systems

In the previous section we considered the "geometric" phase transition controlling the transition from regular to random graphs. There are, however, also "physical" phase transitions, if the small world defines the underlying topology for physical objects located on sites of these graphs, with the links defining the interactions between these objects. In such a way we obtain physical objects with random interactions. A simple example is provided by the Ising model applied to linear polymers, where the short-range interactions between neighboring monomers in the chain are supplemented by the random interactions between monomers that are close in space (and not along the chain). This example demands a slight change of the original model [51] in which the shortcuts do not replace the short-range interactions but rather are added to them, i.e., no links are removed from the regular lattice but extra links are introduced.

Let us recall the main results of the regular one-dimensional Ising model considered in Chapter 2. Such a system with N sites has no phase transition at finite temperatures in the thermodynamic limit $N \to \infty$. However, a mean field type phase transition is found in the Kac model [30] where, in addition to the short-range interactions, one adds a large (of order of N^2) number of weak interactions between all pairs of spins. The ferromagnetic one-dimensional Ising model has been analyzed analytically and numerically in small world systems. It is found from analytical calculations [57, 58] that a mean field transition occurs at a finite temperature $T_c(p)$ for any $p > 0$ provided that the number of sites is large enough, and numerical simulations show that the critical indices have their mean field values [59]. Close to $p = 0$ the transition temperature T_c goes to zero as $|\ln(p)|^{-1}$. Hence, in order to obtain the mean field phase transition in a one-dimensional Ising system there is no need for long-range interactions

between all the spins, and it is enough to have a comparatively small number of randomly chosen long-range interactions. We note that a mean field type behavior has also been found by Monte Carlo simulations for the $x - y$ model based on two-dimensional small world graphs [60].

The problem becomes more interesting in two and three dimensions where phase transitions already occur in the ordered ($p = 0$) systems, and one expects that the appearance of a small amount of disorder ($p \ll 1$) will result in a change of the universality class of this transition from that of the Ising model for $p = 0$ to that of the mean field for $p > 0$. If this change occurs at finite p, the question arises as to the dependence of the critical indices on p in the transition regime. It turns out, however, that the transition to the mean-field type behavior probably occurs for all $p > 0$, provided that the number of sites is large enough, although numerical simulations [60, 61] have been restricted to $p \geq 10^{-3}$. If this last statement is correct, the question arises as to the dependence on p of the transition temperature T_c for small p. An elegant answer to this question has been found by Herrero [61]. Let us compare the two characteristic lengths appearing in our problem, the characteristic length L^* which defines the transition from regular (large world) to random (small world) behavior, and the correlation length ξ of the "physical" phase transition. When ξ is smaller than L^*, the system behaves as a regular lattice. However, as the temperature decreases ξ becomes larger than L^*, and so the random interactions become important and give rise to the mean field behavior. Thus the transition between the regular lattice (Ising type behavior) and the small world (mean field behavior) occurs for $\xi \approx L^*$. For the one-dimensional Ising model the correlation length at low temperature is given by $\xi \sim \exp[2J/(kT)]$ [36], and for $L^* \sim p^{-1}$, the condition $\xi \sim L^*$ results in $T_c \sim |\ln(p)|^{-1}$ in agreement with [58]. For higher dimensions, $\xi \sim |T - T_{c,p=0}|^{-\nu}$ and $L^* \sim p^{-1/d}$. Thus, the change in the temperature of the phase transition for a non-zero randomizing parameter p is given, for small p, as

$$T_c - T_{c,p=0} \sim p^{\frac{1}{\nu d}}. \tag{8.7}$$

8.7   Conclusion

Although systems with random interactions have been studied for many years, especially in the framework of spin glass theory, a fresh point of view was introduced a few years ago with the discovery of small world phenomena. It turns out that most real "many-body" natural and human-made systems exhibit small world properties, having a short-range structure like that of ordered systems and a long-range structure similar to that of random systems. These unusual properties lead to many important applications in science and sociology [53, 54].

Chapter 9

Self-Organized Criticality

Many-particle systems can be divided into two classes. On the one hand, one can start with a superficially simple system such as an atom, and then consider more and more complex systems as one probes further and further into its structure, electrons and a nucleus, the nucleus containing protons and neutrons, these being composed of quarks, and so on. On the other hand, one can consider large systems that are inherently complex because they contain many objects, with very many degrees of freedom and a large number of connections between them. Just as in the previous chapter, these systems can be described in terms of networks of sites and bonds. Such a network can refer not only to the atoms or molecules in a solid or a fluid but also to groups of people and their interactions, competing companies on the stock market, the sites and links of the Internet, and numerous other systems. In our analysis we consider those properties which are independent of the type of objects located at the sites and connected by the bonds, but rather depend on the topological properties of networks. Such properties are very appropriate to phase transitions, where as we saw in Chapter 7 the universality classes do not depend on the detailed properties of the objects being associated with the infinite correlation length.

In the previous chapter we classified networks by the parameters h and c. Another way of doing so is to introduce the probability $P(l) \sim l^{-\delta}$ that two sites at a distance l are connected. If $\delta = 0$, $P(l)$ is constant, so that links of all sizes are equally probable, as in the random graphs discussed in the previous chapter. On the

other hand, if δ is very large then long-range links are restricted, and one effectively obtain a regular graph with only short-range links. For intermediate values of δ, one gets the original Strogarz–Watts model of a small world [51]. It turns out [62] that for $\delta < \delta_1$ the system behaves as a random network, for $\delta_1 < \delta < \delta_2$ the behavior is of the small world type, and for $\delta > \delta_2$ it behaves as a regular network, while for a one-dimensional system $\delta_1 = 1$ and $\delta_2 = 2$.

For the probability of a connection between two sites, we introduced the power law distribution $P(l) \sim l^{-\delta}$. In general, for independent random events with a finite mean a and variance σ, the central limit theorem states that a property x will have the Gaussian distribution,

$$P(x) = (2\pi\sigma)^{-\frac{1}{2}} \exp[-(x-a)^2/2\sigma].$$

This bell-shape curve distribution is probably the most common type of figure in the scientific literature. As we have seen in the previous chapters, in some special cases the Gaussian distribution is replaced by a power-law one. The two examples that we examined previously are the spatial dependence of the correlation function at critical points and of the correlation length in the low temperature region of the x–y model. In contrast to these special cases of systems in equilibrium, which only apply to some special temperature region, there is a wide class of non-equilibrium phenomena which are of the power law type. Since these systems evolve spontaneously into critical states because of their internal dynamics, with no dependence on the value of an external parameter, their behavior is called self-organized criticality (SOC). An alternative name for such systems is scale-free systems, since the power law distribution is scale-invariant and does not involve a characteristic length scale. This situation is similar to that which occurs close to the critical points in equilibrium systems, where the short-range lengths are of no importance and the only characteristic length is the correlation length. This diverges at the critical point, so that there is no characteristic length at this point.

9.1 Power Law Distributions

In considering power-law distributions, we start with some old empirical laws. One of the oldest examples (1897) is the Pareto distribution [63] describing the distribution of large wealth w in different societies. It turns out that the number of people $N(w)$ having wealth w, has a form $N(w) \sim w^{-\alpha}$ with α between 2 and 3. Very recently [64], the effects have been considered of using fiscal policy to change the value of α, and some surprising results have beeen found with regard to the probability of an individual becoming very wealthy (wealth condensation). A similar power-law distribution was found by Guttenberg and Richter [65] in 1954 for the number $N(E)$ of earthquakes which release a certain amount of energy E. This dependence has the form $N(E) \sim E^{-\alpha}$, with $\alpha = 1.5$ independent of the geographic region examined. Finally, in the 1930's, Zipf found [66] the power-law distributions of towns according to their populations and of the frequency of appearance of different words in the English language. There are many more systems with power-law distributions, such as climate fluctuations, cloud formation, rainfall, volcanic eruptions, forest fires, commodity prices, scientific collaborations, citations, the structure of the Internet web and traffic jams. One of the most recently found and exotic examples is the power-law distribution of the wealth of the inhabitants of the ancient Egyptian city Akhetaten, based on the excavations and on the assumption that wealth was proportional to the house area [67].

The power-law distribution, unlike the Gaussian distribution, has the form of a straight line in a logarithmic plot and has no distinctive value. This feature has some interesting consequences, of which two are of special importance. First of all, it is meaningless to define a typical income or the average strength of an earthquake, which would correspond to the peak of a Gaussian distribution. Secondly, and of greater importance, the straight line logarithmic distribution suggests that the same mechanism is responsible for earthquakes of all sizes, including the largest ones. Such large values are often regarded as exceptions, or catastrophic, but the power law distribution shows that although they are rare in occurrence, they are expected to occur

occasionally, and do not require any special explanation. This may be counter to our physical intuition, in which we are used to systems in thermal equilibrium within deep potential minima, where the response is proportional to the disturbance. In that case, small disturbances lead to small consequences and some special mechanism is required to produce large effects. However, there are occasional exceptions even for systems in equilibrium. One example is the divergent susceptibility at the critical point, which leads to a small disturbance producing a large response near this point. In this manner, the gravitational field, which in general is negligible for all laboratory experiments, becomes important near the liquid-gas critical point. Because of the infinite compressibility $\partial\rho/\partial p$ at the critical point, even the small change in pressure δp along a vertical tube 10 cm long leads to a large difference $\delta\rho \sim (\partial\rho/\partial p)\delta p$ between the fluid's density at the top and bottom of the tube ("gravity effect" [68]), which can be as large as 10%. Similarly, in systems driven at a frequency close to the system's resonance frequency a small force can produce a catastrophic response, as happened in the collapse of the Takoma Bridge on the 7th November 1940.

A basic difference between the above systems and those that we are considering now is that the latter are not in thermodynamic equilibrium. One can imagine, for example, a shallow valley in the potential landscape, in which case only a small stimulus is required to lift the system out of this valley, and it can then fall into another potential valley of arbitrary depth. A simple example of this is provided by a man standing on a mountain. If he is at the bottom of the mountain, a small push will have very little effect. If he is standing on a ridge which has a wider ridge slightly below it, then a small push will at most cause him to fall a short distance. However, if he is standing at the top of the mountain with a sheer drop to down below, then a small push away from the edge will have little effect, but a small push over the edge will have catastrophic results. Let us bring two illustrative examples of such catastrophic effects. While one can easily find the obvious reason for the crash of New York stock exchange after 11th September 2001, it is very hard to find a special reasons for the "black Monday" of 19th October 1987, the

greatest one-day crash in history, when that market lost 20% of its value in one day. The second example concerns the disappearance of dinosaurs. It is possible that this was due to the collision of Earth with an asteroid, or due to a drastic change of the climate, or due to a huge volcanic eruption. However, there may be no need for a special explanation, as this phenomena — the disappearance of dinosaurs — could have happened due to the internal properties of the ecosystem.

9.2   Sand Piles

The paradigm for power-law distributions, and the one which was the subject of the first serious theoretical studies, is provided by sand piles. As grains of sand are dropped one by one, slowly and uniformly, onto a table, at first they stay where they land, forming an initial thin layer, and then grains will accumulate on top of one another creating a pile. However, this pile will not grow taller forever. At some stage, the sand pile will enter a state of self-organized criticality (SOC), with a slope of $34°$–$37°$, and then a qualitative change will occur. A new falling grain has the following three possibilities:

(1) It can fall to the bottom of the pile, and so enlarge the pile's base.
(2) It can stick to the side of the pile at some point, thereby increasing slightly the angle of slope of the pile.
(3) It may increase the angle of slope above a critical value and so trigger an avalanche in which S grains fall to the bottom of the pile, some of which may have been located nowhere near the place where the grain which started the avalanche fell. This happens because in the SOC state the whole pile has become connected, so that even a single grain (small disturbance) in one part of the pile may trigger a large avalanches (catastrophic event) anywhere else. In the SOC state the global dynamics controls the behavior of the sand pile as a whole rather than that of the individual grains. In this SOC state the number $N(S)$ of avalanches involving S grains is found to have the power law distribution $N(S) \sim S^{-\alpha}$ with $\alpha \approx 1$. It is intuitively clear that there is no

typical size of an avalanche, and there are no obvious precursors of a huge, catastrophic avalanche. Since each grain "knows" only its short-range environment, it does not expect to be involved in an avalanche, and no predictions of the catastrophic events can be made.

For a mathematical model of such a process, one describes the system as a box which is divided into cells, and assumes that there is an upper limit to the number of particles that can accumulate in a single cell. When a cell contains this maximum permissible number of particles, the arrival of an additional particle will cause some particles to leave it for other cells. If one or more of these particles move to cells which are located near the boundaries of the box and already contain the maximum number of particles, then more and more particles will leave the system creating thereby an avalanche.

A similar model can be used for study of forest fires. For the model of a forest, the boxes consist of cells with trees inside them. Then, one of the trees is set on fire, and one studies how many trees will be consumed before the fire extinguishes itself. A study of this system permits the planning of how to plant forests in such a way as to minimize the risks of large fires. It may be useful to initiate occasionally small fires so as to reduce the connections between different groups of trees (cells), and kept the whole system below the critical SOC level when fires of arbitrary size (including huge avalanches) become possible. It is possible that the wrong policy of keeping the ecosystem in its virgin form, and so suppressing the small fires, was responsible for not preventing the huge 1988 fire in Yellowstone National Park in USA. The forest fire professionals now recognize that the best way to prevent the largest fires is to allow small and medium size fires to burn.

9.3 Distribution of Links in Networks

As in the previous chapter, let us consider a network consisting of a large number N of sites each having a number k links to other sites in the network. Different sites may have different number of links, and

so we introduce the probability distribution of the number k of links, $P(k)$, where k_i is the number of links (also called connectivities) of site i. For a scale free systems $P(k)$ is the power law function, $P(k) \sim k^{-\gamma}$.

In order to show how a power-law distribution can appear, we associate with each site i a random number $x_i[x_i \in (0, \infty)]$, which we call the rank of site i and use it to measure the importance of this site [69]. Let $f(x)$ be the probability distribution of the ranks, which for scale free systems we take to be a power law distribution similar to those considered previously in this chapter. In addition, we assume that the probability that a link exists between a pair of sites i and j is $\psi(x_i, x_j)$. Incidentally, for random graphs, this probability is the same for all sites coupled, and $\psi(x_i, x_j)$ is a constant, $\psi(x_i, x_j) = b$.

The mean connectivity of a site of rank x is

$$\langle k(x) \rangle = \int_0^\infty \psi(x, y) f(y) dy \equiv N F(x) \tag{9.1}$$

so that $x = F^{-1} \left(\frac{\langle k \rangle}{N} \right)$. On the other hand, by definition

$$\langle k \rangle = \int_0^\infty k P(k) dk. \tag{9.2}$$

On combining Eqs. (9.1) and (9.2), one obtains, using $P(k) = f(x) dx / dk$,

$$P(k) = f \left[F^{-1} \left(\frac{\langle k \rangle}{N} \right) \right] \frac{d}{dk} \left[F^{-1} \left(\frac{\langle k \rangle}{N} \right) \right]. \tag{9.3}$$

As a special example, we consider $\psi(x_i, x_j) = x_i x_j / x_{\max}^2$, where $x_{\max}$ is the largest rank in the network. Then,

$$\langle k(x) \rangle = \frac{Nx}{x_{\max}^2} \int_0^\infty y f(y) dy = N \frac{x \langle x \rangle}{x_{\max}^2} \tag{9.4}$$

and

$$P(k) = \frac{x_{\max}^2}{N \langle x \rangle} f \left(\frac{x_{\max}^2}{N \langle x \rangle} k \right). \tag{9.5}$$

One sees from Eq. (9.5), that the distribution of connectivities $P(k)$ is completely determined by the distribution of ranks $f(x)$, so that if $f(x)$ is a power law distribution, then so is the distribution $P(k)$ of connectivities.

9.4   Dynamics of Networks

So far we have considered only static networks. However, most networks are not static, as the number of sites can change with time, new sites and links can appear, and some of the old ones may disappear. For instance, in the world wide web of the Internet, new sites are continually being created, as are new links between sites, while some sites may also disappear. Similarly, in the financial world, new businesses are continually being created and others liquidated, while assets are also transferred from one firm to another.

The main assumption which leads to a power law distribution of links is that of preferential attachment [70], which means that new links prefer to be connected to those old sites that already have many connections (which corresponds to "rich becomes richer"). Let us consider a model system that initially contains m_0 bonds and n_0 sites. At each of the time steps $t_1, t_2, \ldots, t_N$, where we choose $t_r = r$ ($r = 1, 2, 3, \ldots$), one site with m bonds is added to the system so that after t time steps the network has $N = n_0 + t$ sites and $m_0 + mt$ bonds. According to the assumption of preferential attachment, the increase of the connectivity k_i on site i is chosen to be proportional to k_i itself,

$$\frac{\partial k_i}{\partial t} = A \frac{k_i}{\sum_j k_j} \tag{9.6}$$

where the normalization parameter $\sum_j k_j$ determines the total number of bonds, and at long times $\sum_j k_j \simeq 2mt$. The factor two in this formula appears because each bond links two sites. On summing Eq. (9.6) over all i, we find that the coefficient A determines the change of the total number of connectivities per unit time, i.e., $A = m$. The substitution of this result in Eq. (9.6) and solution of

the resultant equation leads to

$$k_i = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}} \quad \text{or} \quad \frac{dt_i}{dk_i} = -\frac{2m^2 t}{k_i^3}. \tag{9.7}$$

The distribution of links $P(k)$ is related to the (homogeneous) distribution of the times t_i when the sites were created, $P(t_i) = 1/t_i$ by

$$P(k)dk = -P(t_i)dt_i. \tag{9.8}$$

Here, the minus sign is connected with the fact that the older sites (with small t_i) have larger connectivity ("rich become richer"). On combining Eqs. (9.7) with (9.8), we finally obtain

$$P(k) = \frac{2m}{k^3}. \tag{9.9}$$

According to Eq. (9.6), the increase of the connectivity of a given site is fully determined by the present value of its connectivity, and this gives an obvious advantage to the "old" sites as compared to the "new" ones which only recently jointed the network. However, in addition to its "age", the "quality" of a site should influence its future development. To extend the model of equation (9.6) in this respect, we now introduce a fitness parameter γ_i for site i, with $0 \leq \gamma_i \leq 1$, which describes how capable it is of receiving links. The increase of the connectivity depends now not only on the present value of the connectivity k_i of the site i, as in Eq. (9.6), but also on its fitness. Accordingly we assume that

$$\frac{dk_i}{dt} = m \frac{\gamma_i k_i}{\sum (\gamma_j k_j)}. \tag{9.10}$$

An interesting analogy exists [71] between networks with the dynamics described by (9.10) and the equilibrium Bose gas considered in Chapter 6. In order to establish this, we associate with the fitness γ_i an energy ε_i through the relation

$$\varepsilon_i = -T \ln(\gamma_i) \tag{9.11}$$

so that $0 < \gamma_i < 1$ corresponds to $\infty > \varepsilon_i > 0$, where in this section we have set the Boltzmann constant equal to unity. Let a link

between the two sites i and j (which have fitnesses γ_i and γ_j or energies ε_i and ε_j) be replaced by two non-interacting particles. Then the addition of a new site l with m links to the network means the addition of $2m$ particles, half of which have the energies ε_l, and the other half of which are distributed between other energy levels corresponding to the endpoints of the new links. Instead of the connectivity of a site depending only on time, as in Eq. (9.6), the connectivity k_i on site i in the generalized theory will also be a function of the number of links (particles) which the site has at time t, $k_i = k_i(\varepsilon_i, t, t_i)$. Equation (9.10) will take the following form

$$\frac{\partial k_i(\varepsilon_i, t, t_i)}{\partial t} = m \frac{\exp(-\varepsilon_i/T) k_i(\varepsilon_i, t, t_i)}{\sum_j \exp(-\varepsilon_j/T) k_j(\varepsilon_j, t, t_j)}. \tag{9.12}$$

By analogy with Eq. (9.7), we look for a solution of Eq. (9.12) of the form

$$k_i = m \left(\frac{t}{t_i}\right)^{f(\epsilon_i)}. \tag{9.13}$$

One can rewrite the "partition function" $Z(t)$ in the denominator of Eq. (9.12) as

$$Z(t) = \int_0^\infty d\varepsilon g(\varepsilon) \int_1^t dt' \exp\left(-\frac{\varepsilon}{T}\right) k(\varepsilon, t, t') \tag{9.14}$$

where $g(\varepsilon)$ is the distribution of energies obtained by Eq. (9.11) from the distribution of fitnesses $\phi(\gamma_i)$. On substituting Eq. (9.13) into Eq. (9.14) one can perform the integration over t_i, to obtain

$$Z(t) = mt \int_0^\infty \frac{d\varepsilon \, g(\varepsilon) \exp(-\varepsilon/T)}{1 - f(\varepsilon)}. \tag{9.15}$$

We now introduce the "chemical potential" μ as

$$\exp\left(-\frac{\mu}{T}\right) \equiv \lim_{t\to\infty} \frac{Z(t)}{mt}. \tag{9.16}$$

From Eqs. (9.16), (9.12) and (9.13), one then finds that

$$f(\varepsilon) = \exp\left(-\frac{\varepsilon - \mu}{T}\right) \tag{9.17}$$

and obtains the following equation for μ

$$\int_0^\infty \frac{d\varepsilon \, g(\varepsilon)}{\exp[(\varepsilon - \mu)/T] - 1} = 1. \tag{9.18}$$

Equation (9.18) is nothing else than Eq. (6.15) for the chemical potential of a Bose gas. The mapping of our problem to that of a Bose gas means that the typical property of a Bose gas, Bose–Einstein condensation, must also occur in an evolving network. Therefore, one predicts, in addition to the usual self-organized criticality with preferential attachment ("rich become richer"), the appearance of a new phase, analogous to the Bose condensate, where all the new sites will correspond to $\varepsilon = 0$, which according to Eq. (9.11) corresponds to the site with the highest fitness $\gamma = 1$. It is just this site that will have links to all new sites ("winner takes all").

In addition to fitness, there are some other generalizations of the main dynamic equation (9.6).

(a) If one assumes in the phenomenological expression (9.6), $\partial k_i / \partial t \sim k^\beta$ with $\beta \neq 1$, one obtains [72] an exponential distribution for $P(k)$ if $\beta < 1$, the limit situation of "winner takes all" if $\beta > 1$, and only for $\beta = 1$ does $P(k)$ exhibit power law behavior.

(b) The power law distribution function was obtained from the assumption of preferential attachment (9.6). On the other hand, in random graphs and small world systems, a new site randomly creates new links with each one of the existing sites. In contrast to the preferential attachment, this assumption leads to an exponential form of the distribution function. It can be assumed [73] that a realistic network grows in time according to an attachment rule that is neither completely preferential nor completely random, i.e.,

$$\frac{\partial k_i}{\partial t} = \frac{(1-p)k_i + p}{\sum[(1-p)k_i + p]} \tag{9.19}$$

where $0 \leq p \leq 1$ is a parameter characterizing the relative weights of the deterministic and random attachments. An

analysis similar to Eqs. (9.6)–(9.9) leads to the following result

$$P(k) \sim \left(\frac{k/m+b}{1+b}\right)^{-\gamma}, \qquad \gamma = 3+b, \qquad b = \frac{p}{m(1-p)} \quad (9.20)$$

where, as previously, m is the number of new links added at each time step. The power law behavior for scale free networks is recovered from Eq. (9.20) if $p \to 0$, while the exponential distribution dominates for $p \to 1$.

(c) The preferential attachment (9.6) gives a clear advantage to the "old" sides, and the fitness described above was designed to correct this "injustice". Another way of obtaining the same effect is to give an advantage to new sites by introducing in the dynamic equation (9.6) an "aging" factor τ_i for each site i, and assuming that the probability of connecting a new site with some old one is proportional not only to the connectivity of the old site but also to a power of its age, $\tau_i^{-\alpha}$. It turns out [74] that the power law distribution is still obtained for $\alpha < 1$, while for $\alpha > 1$ the exponential law applies, with the intermediate case $P(k) = \text{const}$ for $\alpha = 1$. The "winner takes all" behavior occurs for large α, such as $\alpha > 10$.

9.5   Mean Field Analysis of Networks

In the previous section, we applied to the analysis of networks some ideas from the thermodynamic theory of equilibrium processes, such as a mean field type dynamic equation (9.6), and the "chemical potential" approach which led to Bose condensation. We now present an elegant general approach [75] based on the Landau mean field theory of phase transitions considered in Chapter 3. In order to do this, we need to characterize a network by the thermodynamic potential G which is a function of the order parameter η and the field H conjugate to it, $G = G(\eta, H)$, just as for the magnetic Ising system with order parameter M the free energy G is a function of M and the magnetic field H conjugate to M, $G = G(M, H)$. As usual, we require that the order parameter η vanishes in the disordered phase.

A site with connectivity k feels, on the average, k of its neighbors, i.e., it is subject an average field $k\eta$ in addition to the external field H. Hence, the Landau expansion of $G(\eta, H)$ contains both of the fields H and $k\eta$, and can be written in the following form

$$G(\eta, H) = -\eta H + \sum_k P(k)\phi(\eta, k\eta), \qquad (9.21)$$

i.e., it depends on the distribution function $P(k)$. The following analysis [75] is based on the expansion of the function $\phi(x, y)$ in a Taylor series for small x and y,

$$\varphi(x, y) = \sum_m \sum_l \varphi_{ml} x^m y^l,$$

and analyzing the first terms of this expansion. The coefficients in this expansion are of the form

$$\frac{d^n G(\eta, H)}{d\eta^n}\bigg|_{\eta=0} = n! \sum_{l=0}^n \varphi_{n-l,l} \langle k^l \rangle. \qquad (9.22)$$

The average values $\langle k^l \rangle$ entering Eq. (9.22) are completely determined by the distribution function $P(k)$. For the power law distribution, $P(k) \sim k^{-\gamma}$, we find that

$$\langle k^l \rangle \equiv \int_1^\infty k^l P(k) dk \sim k^{l-\gamma+1}|_{k\to\infty} \qquad (9.23)$$

which means that all moments with $l \geq \gamma$ diverge. For these divergent moments, the function (9.21) contains singular coefficients (9.22), and, therefore, the Landau mean field approach breaks down. Just as for equilibrium phenomena, such a breakdown is caused by strong fluctuations. The most connected sites in the networks, which correspond to strong fluctuations in the local connectivity, are the source of the inapplicability of mean field theory. As the power γ in $P(k) \sim k^{-\gamma}$ is increased, the relative number of highly connected sites decreases, and the phase transition becomes of the mean field type. Calculations show [75] that this happens for $\gamma > 5$, while there is a logarithmic corrections for $\gamma = 5$. (This latter feature can be compared with the analogous behavior near the upper critical

dimension $d = 4$ of the thermodynamic critical phenomena). Other results are that for $3 < \gamma < 5$, the critical indices are functions of γ ($\eta \sim (|T - T_c|/T_c)^{\frac{1}{\gamma-3}}$ etc.), while no phase transition occurs at any finite temperature for $2 < \gamma \leq 3$. These results have been obtained in the absence of the third order term in η in the expansion of G. If this term is included, G contains singular terms of the form:

(a) $\eta^3 \ln \eta$ for $\gamma = 4$,
(b) $\eta^{\gamma-1}$ for $2 < \gamma < 3$ and $3 < \gamma < 4$, and
(c) $\eta^2 \ln(\eta)$ for $\gamma = 3$.

One further comment should be made. We saw in the previous chapter that for small world phenomena, a small fraction p of long-range bonds brings a system into the mean field universality class. However, for scale-free systems, in spite of the existence of the long-range bonds, the situation is quite different. As we have seen, the mean field behavior only takes place for large enough values of the power γ in the distribution function of k, $\gamma > 5$, while for smaller values of γ the critical behavior is model independent but non-universal, and the critical indices depend on γ.

9.6   Hubs in Scale-Free Networks

Scale-free networks with a power law distribution function have a topological structure very different from that of usual ("exponential") networks where the distribution function has a Gaussian form. In the exponential networks nearly all sites have more or less the same number of links, close to the average one, and the system is homogeneous. The situation is quite different for scale-free networks. As we have seen, the dynamics of these networks ensures that "rich becomes richer" or even that the "winner takes all". This means that scale free networks are very inhomogeneous, with a few sites ("hubs") having a very large number of links and most of the sites having only very few links, as shown in Fig. 9.1. By the way, such a structure is typical for the routes of large airlines which fly between many sites using the hubs as intermediate stations for connecting flights. The
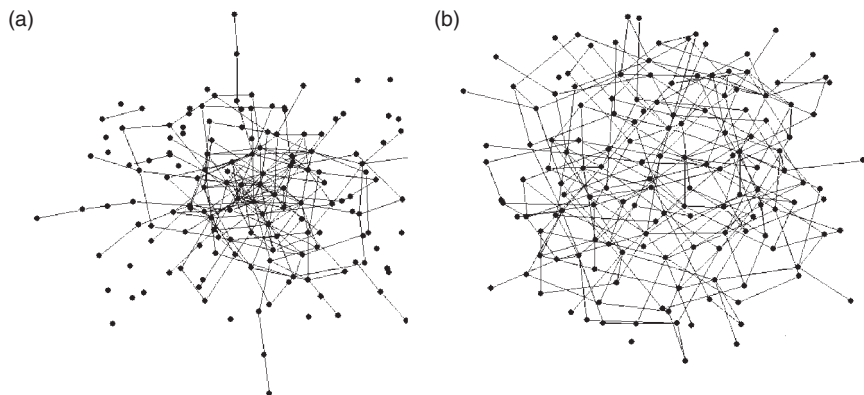
Fig. 9.1 Two networks containing 150 sites and 4950 links which are distributed among the sites according to (a) a power-law and (b) an exponential law. The former have a few sites with many links ("hubs") such as Chicago and Atlanta for air flights, or Google and Yahoo for Internet.

existence of hubs is very important for phase transitions, such as the ordering of spins located on sites of scale free system. Since the hubs have many links, they keep the majority of spins in an ordered state. To destroy the order or reverse the direction of the spins, it is enough to flip the spins located at a small numbers of hubs, some $10^{-3} - 10^{-4}$ of all the sites [76].

Another interesting consequence of the distinction between exponential and scale free networks is their different response to damage [77]. Damage means the removal of some sites with all their links. A distinction must be made between random (failure) and intentional (attack) damage. While for the exponential networks there is no difference between these two types of damage since all sites are more or less equivalent, a striking difference exists for scale free networks [78, 79]. It is intuitively clear that scale free networks are extremely resilient to random damage but are very sensitive to intentional damage focused on the hubs, which destroys many links, and so interrupts the connection between different parts of the network.

A specific example of the above process, which is of great practical interest, concerns the stability of the Internet, which can be studied on the basis of percolation theory. The main question is

how many sites need to be destroyed (or incapacitated) before the Internet becomes unstable, i.e., the connectivity of the world wide web is broken and it disintegrates into small disconnected clusters. It was found [78,79] that the Internet network would never disintegrate if the system were infinite. While the Internet is a finite rather than an infinite network, it remains connected even if more than 90% of the sites are removed at random. However, if an attack is planned to eliminate preferentially the most connected sites (the hubs), even if less than 10% of these sites are removed the Internet will break up into disconnected sections.

9.7   Conclusion

Many natural and man-made systems belong to the class of scale-free networks, which are characterized by power law distributions, $P(k) \sim k^{-\gamma}$, of the number k of connections between one object and others in these systems. The value of the exponent γ turns out to be sensitive to the details of the network structure, and it is usually in the range $2 < \gamma < 3$. There is no conclusive theory which either predicts the value of γ or classifies networks according to the magnitude of γ. In contrast to small world systems, where the mean field type transition occurs for all values of the characteristic parameter p, scale-free systems show a whole range of different types of phase transitions for different values of γ. This subject is developing very rapidly, and several reviews [80–82] have been published recently.

Bibliography

[1] L. Landau and E. Lifshitz, *Statistical Physics* (Pergamon, 1980).

[2] P. Weiss, *J. Phys.* **6**, 661 (1907).

[3] H. E. Stanley, *Rev. Mod. Phys.* **71**, S358 (1999).

[4] L. Onsager, *Phys. Rev.* **65**, 117 (1944).

[5] R. B. Potts, *Proc. Cambridge Philos. Soc.* **48**, 106 (1952).

[6] S. G. Brush, *Rev. Mod. Phys.* **39**, 883 (1967).

[7] J. M. Ziman, *Properties of the Theory of Solids* (Cambridge, 1964).

[8] N. M. Svrakic, *Phys. Lett. A* **80**, 43 (1980).

[9] B. Liu and M. Gitterman, *Am. J. Phys.* **71**, 806 (2003).

[10] K. Binder and E. Luijten, *Phys. Rep.* **344**, 179 (2001).

[11] V. L. Ginzburg, *Sov. Phys. Solid State* **2**, 1824 (1960).

[12] K. Wilson, *Rev. Mod. Phys.* **47**, 773 (1975).

[13] K. Wilson, *Physica* **73**, 119 (1974).

[14] A. Hankey and H. E. Stanley, *Phys. Rev. B* **6**, 3515 (1972).

[15] G. S. Rushbrooke, *J. Chem. Phys.* **39**, 8942 (1963).

[16] R. B. Griffiths, *Phys. Rev. Lett.* **14**, 623 (1965).

[17] L. Kadanoff, *Physics* **2**, 263 (1966).

[18] H. J. Maris and L. P. Kadanoff, *Amer. J. Phys.* **46**, 652 (1978).

[19] G. Nicolis, *Introduction to Nonlinear Science* (Cambridge, 1995).

[20] Th. Niemeijer and J. M. J. van Leeuwen, *Physica* **71**, 17 (1974).

[21] E. M. Lifshitz and L. P. Pitaevski, *Statistical Physics*, Part 2 (Pergamon, 1980).

[22] B. G. Levi, *Physics Today*, December 1997, p.17.

[23] W. Ketterle, *Scientific American*, December 1999, p.30: K. Burnett, M. Edwards, and C. W. Clark, *ibid.* p.37.

[24] E. A. Cornell and C. E. Wieman, *Rev. Mod. Phys.* **74**, 875 (2002).

[25] P. Meijer, *Amer. J. Phys.* **62**, 1105 (1994).

[26] L. N. Cooper, *Phys. Rev.* **104**, 1189 (1956).

[27] J. G. Bednorz and K. A. Muller, *Z. Phys. B* **64**, 189 (1986).

[28] *Physica C* **385**, issue 1-2 (2003).

[29] P. J. Ford and G. A. Sanders, *Contemporary Physics* **38**, 63 (1997).

[30] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, *J. Math. Phys.* **4**, 216, 229 (1963).

[31] R. J. Baxter, *Exact Soluble Models in Statistical Mechanics* (Academic, London, 1984).

[32] V. G. Vaks, A. I. Larkin, and Y. N. Ovchinnikov, *Sov. Phys. JETP* **22**, 820 (1965).

[33] M. Gitterman and P. Hemmer, *J. Phys. C* **13**, L329 (1980).

[34] M. Kac, *Physics Today*, October 1964. p. 40.

[35] T. H. Berlin and M. Kac, *Phys. Rev.* **86**, 821 (1952).

[36] J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, *The Theory of Critical Phenomena* (Clarendon, Oxford, 1992).

[37] M. D. Mermin and H. Wagner, *Phys. Rev. Lett.* **17**, 133 (1966).

[38] J. M. Kosterlitz, *J. Phys. C* **7**, 1046 (1974).

[39] G. Blatter, M. V. Feigelman, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, *Rev. Mod. Phys.* **66**, 1125 (1994).

[40] N. B. Kopnin, *Rep. Prog. Phys.* **65**, 1633 (2002).

[41] V. A. Vyssotsky, S. B. Gordon, H. L. Frisch and J. M. Hammersley, *Phys. Rev.* **123**, 1566 (1961).

[42] J. M. Ziman, *J. Phys. C* **1**, 1532 (1968).

[43] J. W. Esssam, *Rep. Prog. Phys.* **43**, 833 (1980).

[44] R. B. Griffiths, *Phys. Rev. Lett.* **23**, 17 (1969).

[45] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996).

[46] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glasses Theory and Beyond*, (World Scientific, Singapore, 1987).

[47] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*, (World Scientific Lecture Notes in Physics, Vol. 54, 2000).

[48] P. Larranaga, C. M. H. Kuijpers, R. H. Murga, I. Inza and S. Dizdarevic, *Artificial Intelligence Review* **13**, 129 (1999).

[49] J. J. Hopfield, *Proc. Nat. Acad. Sci.* (USA) **79**, 2554 (1982) and **81**, 3088 (1984).

[50] S. Milgram, *Psychol. Today* **2**, 60 (1967).

[51] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).

[52] D. J. Watts, *Small World* (Princeton, New Jersey, 1999).

[53] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).

[54] M. E. J. Newman, *J. Stat. Phys.* **101**, 819 (2000).

[55] V. Latora and M. Marchiori, *Physica A* **314**, 109 (2002); *Eur. Phys. J. B* **32**, 249 (2003).

[56] M. E. J. Newman and D. J. Watts, *Phys. Lett. A* **263**, 341 (1999).

[57] A. Barrat and M. Weigt, *Eur. Phys. J. B* **13**, 547 (2000).

[58] M. Gitterman, *J. Phys. A* **33**, 8373 (2000).

[59] H. Hong, B. J. Kim, and M. Y. Choi, *Phys. Rev. E* **66**, 018101 (2002).

[60] B. J. Kim, H. Hong, G. S. Jeon, P. Minnhagen, and M. Y. Choi, *Phys. Rev. E* **64**, 056135 (2001).

[61] C. P. Herrero, *Phys. Rev. E* **65**, 066110 (2002).

[62] P. Sen, K. Banerjee, and T. Biswas, *Phys. Rev. E* **66**, 037102 (2002).

[63] V. Pareto, *Le Cours dE'conomie Politique* (MacMillan, London, 1897).

[64] Z. Burda, D. Johnston, J. Jurkiewicz, M. Kaminski, M. A. Nowak, G. Papp, and I. Zahed, *Phys. Rev. E* **65**, 026102 (2002).

[65] B. Gutenberg and C. F. Richter, *Seismity of the Earth and Associated Phenomena*, 2nd ed. (Princeton, NJ, Princeton University Press, 1954)

[66] G. K. Zipf, *The Psycho-biology of Language* (Houghton Mifflin Co., Boston, 1935).

[67] A. Y. Abul-Magd, *Phys. Rev. E* **66**, 057104 (2002).

[68] M. Gitterman, *Rev. Mod. Phys.* **50**, 85 (1978).

[69] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, *Phys. Rev. Lett.* **89**, 258702 (2002).

[70] A.-L. Barabasi and R Albert, *Science* **286**, 509 (1999).

[71] G. Bianconi and A.-L. Barabasi, *Phys. Rev. Lett.* **86**, 5632 (2001).

[72] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).

[73] Z. Liu, Y.-C. Lai, N. Ye, and P. Dasgupta, *Phys. Lett. A* **303**, 337 (2000).

[74] S. N. Dorogovstev and J. F. F. Mendes, *Phys. Rev. E* **62**, 1842 (2000).

[75] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **67**, 026123 (2003).

[76] A. Aleksieyuk, J. A. Holyst, and D. Stauffer, *Physica A* **310**, 260 (2002).

[77] R. Albert, H. Jeong, and A.-L. Barabasi, *Nature*, **406**, 378 (2000).

[78] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).

[79] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2000).

[80] A.-L. Barabasi, and E. Bonabeau, *Sci. Am.* **288**, 60 (2003).

[81] R. Albert and A.-L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).

[82] S. N. Dorogotsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).

This page intentionally left blank

Index