An introduction to lattice gauge theory and spin systems*

John B. Kogut

Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

This article is an interdisciplinary review of lattice gauge theory and spin systems. It discusses the fundamentals, both physics and formalism, of these related subjects. Spin systems are models of magnetism and phase transitions. Lattice gauge theories are cutoff formulations of gauge theories of strongly interacting particles. Statistical mechanics and field theory are closely related subjects, and the connections between them are developed here by using the transfer matrix. Phase diagrams and critical points of continuous transitions are stressed as the keys to understanding the character and continuum limits of lattice theories. Concepts such as duality, kink condensation, and the existence of a local, relativistic field theory at a critical point of a lattice theory are illustrated in a thorough discussion of the two-dimensional Ising model. Theories with exact local (gauge) symmetries are introduced following Wegner's Ising lattice gauge theory. Its gauge-invariant "loop" correlation function is discussed in detail. Three – dimensional Ising gauge theory is studied thoroughly. The renormalization group of the two dimensional planar model is presented as an illustration of a phase transition driven by the condensation of topological excitations. Parallels are drawn to Abelian lattice gauge theory in four dimensions. Non-Abelian gauge theories are introduced and the possibility of quark confinement is discussed. Asymptotic freedom of O(n) Heisenberg spin systems in two dimensions is verified for $n \geq$ and is explained in . simple terms. The direction of present-day research is briefly reviewed.

CONTENTS

* Supported in part by the National Science Foundation under grant number NSF PHY77-25279.

I. INTRODUCTION—AN OVERVIEW OF THIS ARTICLE

This article consists of a series of introductory lectures on lattice gauge theory and spin systems. It is intended to explain some of the essentials of these subjects to students interested in the field and research physicists whose expertise lies in other domains. The expert in lattice gauge theory will find little new in the following pages aside from the author's personal perspective and overview. The style of this presentation is informal. The article grew out of a half-semester graduate course on lattice physics presented at the University of Illinois during the fall semester of 1978.

Lattice spin systems are familiar to most physicists because they model solids that are studied in the laboratory. These systems are of considerable interest in these lectures, but we shall also be interested in more abstract questions. In particular, we shall be using space–time lattices as a technical device to define cutoff field theories. The eventual goal of these studies is to construct solutions of cutoff theories so that field theories defined in real continuum Minkowski space–time can be understood. The lattice is mere scaffolding—an intermediate step used to analyze a difficult nonlinear system of an infinite number of degrees of freedom. Different lattice formulations of the same

field theory can be contemplated, just as different discrete versions of differential equations can be written down. Space–time symmetric lattices in four Euclidean dimensions will be discussed in detail because they are quite elegant. Another approach which leaves the "time" axis continuous and replaces continuum spatial axes by a three-dimensional lattice will also be discussed.

Once a lattice field theory has been formulated, the original field theory problem becomes one of statistical mechanics. This point will be developed in detail in this article through both general analyses and specific examples. The first step in understanding the theory is then to map out the phase diagram of the equivalent statistical mechanics system. The ground state of the theory changes qualitatively from one phase to another. In this review we shall be particularly interested in lattice gauge theories which model the strong interactions of particle physics. In some of these theories there is a range of parameters for which the ground state cannot tolerate the presence of an isolated quark. This is the quark-confining phase of the cutoff theory. This phase may be separated by a critical surface from another phase in which quarks can be isolated. After the phases of the theory have been established, the behavior of the theory in the critical region must be determined. One must determine, for example, the order of the transition occurring between the two phases. Only if the transition is continuous can one obtain a continuum, relativistic field theory from the lattice system. If the transition is continuous, the system's mass gap vanishes as a critical point is approached. One can then define a renormalized mass of the field theory, which can be held fixed as the lattice spacing is taken to zero. At the critical point the theory loses memory of the lattice and the continuous space–time symmetries of the field theory are reestablished. These points will be discussed at length throughout this article. In short, to use lattice formulations of field theory to construct continuum theories, one maps out the lattice theory's phase diagram, locates its critical points (lines or surfaces) of continuous phase transitions, and approaches the critical points in a well-defined, delicate fashion.

The first step in this program is the least difficult. The phases of a statistical mechanics system can usually be established using several methods of varying reliability: high- or low-temperature expansions, duality transformations, spin-wave analyses, mean field theory, etc. In particular cases one must use these techniques judiciously. If a system has complicated symmetries, it can be difficult indeed to establish its phase diagram with confidence. Frequently thorough numerical calculations using expansion methods or renormalization groups are required. These same methods can then be used to determine the nature of the system's critical points. In particular, at a continuous phase transition various thermodynamic functions of the system become singular and the degree of their singularities is recorded in their critical exponents. We shall illustrate the calculation of some of these exponents for model lattice spin systems in later chapters. The two-dimensional Ising model will also be solved in closed form to illustrate the fact that at its critical point it defines a relativistic, scale-invariant field theory.

Lattice gauge theories pose special problems to this traditional approach to studying spin systems. Ising gauge theories were invented in a remarkable article by F. Wegner in 1971 (Wegner, 1971). He elevated the global up $\rightleftharpoons$ down symmetry of the ordinary Ising model to a *local* symmetry of his new theory's action. He realized that such a model could undergo phase transitions, but they could not be accompanied by a spontaneous magnetization. The absence of a local magnetization challenged him to find a sensible way to label the phases of the theory. Wegner stressed the importance of correlation functions here. In fact, he was forced to construct a correlation function which respected the local up $\rightleftharpoons$ down symmetry of the model. This led him to the "loop correlation function" and the "area" and "perimeter" laws which will be discussed in Sec. V.

There are two phenomena which play an especially important role in the physics of lattice gauge theories. The first is the occurrence of topological excitations, which can label the phases of some of these models and determine the character of their ground states. The second is asymptotic freedom and the possibility that non-Abelian gauge theories exist only in a single, quark-confining phase. The importance of topological excitations will be illustrated in the two-dimensional Ising model, the planar model, and four-dimensional Abelian lattice gauge theory. Topological excitations are stationary configurations of the theory's action. They affect the local variables of the theory over an infinite domain of space and tend to disorder the system. These models' high-temperature phases can be described as condensates of such excitations. These facts will be illustrated in a renormalization group calculation of the planar model's phase diagram (Kosterlitz, 1974). It is believed that topological excitations which resemble dynamical magnetic monopoles play an important role in determining the character of the ground state of non-Abelian gauge theories in four dimensions ('t Hooft, 1978).

The discovery that non-Abelian gauge theories in four dimensions are asymptotically free ('t Hooft, 1972; Politzer, 1973; Gross and Wilczek, 1973) has profoundly affected the modern study of field theories. Roughly speaking, asymptotic freedom means that lower- frequency fluctuations in a theory are more strongly coupled than higher- frequency fluctuations. If the theory were formulated with a momentum space cutoff Λ and a coupling constant $g(\Lambda)$, then another formulation using a smaller cutoff Λ' would need a larger coupling $g(\Lambda')$ to describe the same physics at a given physical momentum scale. As $\Lambda \to \infty$, the coupling $g(\Lambda)$ vanishes and the theory is free at short distances. The fact that $g(\Lambda)$ vanishes as Λ grows allows one to make precise predictions for deep-inelastic scattering of electrons or neutrinos off strongly interacting particles such as protons and neutrons. These calculations compare extremely well with experiment and strongly support the conjecture that the underlying theory of strong interactions is based on an SU(3) "color" gauge group with quarks residing in the fundamental representation of the group (Fritzsch *et al.*, 1973). It is this experimental success that has made gauge theories of strong inter-

actions so attractive to the high-energy physics community.

Immediately after the discovery of asymptotic freedom, it was conjectured by innumerable theorists that quarks would be confined in such theories. Asymptotic freedom suggests the possibility that the coupling $g(\Lambda)$ grows large as Λ decreases, so that it would be energetically favorable for quarks to bind together into color-singlet bound states rather than to exist alone. This idea could then reconcile the great success of non-Abelian gauge theories in describing deep-inelastic phenomena with the absence of free quarks in the debris of these scattering events. It is easy to establish that lattice gauge theory confines quarks at strong coupling (Wilson, 1974). As in many spin systems, reliable calculations can be made in lattice gauge theories in the strong coupling (high-temperature) domain. From this point of view, however, the problem of obtaining a theory of strong interactions becomes one of establishing asymptotic freedom of the lattice theory formulated on a fine lattice with vanishingly small coupling constant and the absence of any critical points at finite coupling. Doing this would then constitute the first step toward making a computable theory of strong interactions in which asymptotic freedom and confined quarks could coexist. The lattice theory of colored quarks and gluons has not been analyzed in enough detail to decide whether this grand hope is realized. However, analogous behavior has been found in simpler systems. Later in this review we shall see that $O(n)$ Heisenberg spin systems are asymptotically free (Polyakov, 1975a). It is also easy to see that these spin systems are disordered at large coupling in the sense that their spin–spin correlation function is short ranged. This feature of the theory is analogous to the quark-confining property of lattice gauge theory at strong coupling. This point will be discussed at length in Sec. VIII. In addition, recent work strongly suggests that the $O(n)$ Heisenberg systems have well-behaved scattering matrices describing *massive* particles (Zamolodchikov and Zamolodchikov, 1978). This indicates that such models exist only in a disordered phase and that they possess a well-defined continuum limit given a relativistic, massive, interacting field theory. This model calculation constitutes an important, optimistic step in the lattice gauge theory program.

This article is organized as follows. First we review some phenomenology concerning phase transitions. Then the transfer matrix is discussed in detail and the connection between field-theoretic and statistical mechanics language is established. Statistical mechanics systems formulated on symmetric space–time lattices and very anisotropic lattices in which one axis (time) is left continuous are discussed. These general remarks are illustrated in a detailed discussion of the Ising model. Strong coupling expansions for the theory's critical behavior are illustrated. The self-duality of the model is established using both the partition function for the theory formulated on a symmetric lattice and the operator transfer matrix formulated on a lattice with one axis continuous. Kink condensation is discussed and the time-continuum version of the model is solved exactly to establish that the theory at the critical point becomes a scale-invariant, relativistic field theory—a free, massless fermion. Next, global and local symmetries and spontaneous symmetry breaking are discussed in preparation for the introduction of Ising lattice gauge theory. F. Wegner's Ising lattice gauge theory is introduced and the fact that its mean magnetization must vanish at all coupling is proved. The loop correlation function is introduced and the perimeter law at small coupling and the area law at strong coupling are established using reliable expansion methods. The two-dimensional gauge model is shown to be equivalent to the ordinary one-dimensional Ising model which exists only in a disordered phase. The time-continuum formulation of the theory is developed. The three-dimensional gauge theory is shown to be dual to the three-dimensional Ising model. This exercise sheds light on the phases of the gauge theory and kink condensation. Abelian lattice gauge theory is introduced next. The physical interpretation of the loop correlation function in terms of quark confinement is derived. The energetics and phases of the model are discussed by evaluating the loop correlation function at weak coupling where Coulomb's law emerges and at strong coupling where a linear potential confines quarks. The time-continuum formulation of the theory and the notion of flux tubes are discussed. We then turn to a detailed renormalization group analysis of the two-dimensional planar model to see an example of a phase transition without a local order parameter. The Kosterlitz–Thouless physical picture of vortices and the planar model phase transition is developed in detail (Kosterlitz and Thouless, 1973), and the renormalization group trajectories are found from the sine–Gordon representation of the theory's action. The vortex structure of the two-dimensional planar model is compared with the vortex-loop structures of the four-dimensional Abelian lattice gauge theory (Banks *et al.*, 1977). Non-Abelian lattice gauge theories are discussed next. The SU(2) theory is formulated and its special properties are discussed. Analogies to the two-dimensional $O(4)$ Heisenberg model are drawn, and it is shown that $O(n)$ Heisenberg models are asymptotically free for $n \geq 3$ (Polyakov, 1975a). Results from the Migdal recursion relation (Migdal, 1975) are reviewed and the possibility of deep relations between two-dimensional spin systems and four-dimensional gauge theories is noted. In a short section of concluding remarks some topics of current research interest are discussed.

The reader may wish to consult other reviews of lattice gauge theory while studying this article. The review by Kadanoff (1977), which discusses fermions and real space renormalization more thoroughly than this article, is recommended.

II. PHENOMENOLOGY AND PHYSICS OF PHASE TRANSITIONS

A. Facts about critical behavior

Let us begin by collecting some definitions and facts about statistical mechanics and phase transitions. We shall use the Ising model as an illustration. Consider a square lattice in two dimensions with sites labeled by a vector of integers,

$$n=(n_1,n_2)\,. \tag{2.1}$$

Place a "spin" variable $s(n)$ at each site and suppose that s can only be "up" ($s=+1$) or "down" ($s=-1$). Then the energy or "Action" of the model is

$$S=-J\sum_{n,\mu} s(n)s(n+\mu)\,, \tag{2.2}$$

where μ denotes one of the two unit vectors of the lattice as depicted in Fig. 1, and J is positive so the Action favors aligned spins. Two important properties of the model are (1) only nearest-neighbor spins are coupled. The Action is as local as possible. (2) The model has a global symmetry. If all the spins are flipped, S is left unchanged.

Placing the system of spins into an external magnetic field B changes its Action to

$$S=-J\sum_{n,\mu} s(n)s(n+\mu)-B\sum_{n} s(n)\,. \tag{2.3}$$

The external field breaks the system's global symmetry.

The statistical properties of the model follow from the hypothesis that the probability for a particular spin configuration is proportional to

$$P=\exp(-\beta S)\,, \tag{2.4}$$

where $\beta=1/kT$. The statistical physics can then be obtained from the partition function

$$Z=\sum_{\{\text{configs}\}}\exp(-\beta S)\,, \tag{2.5}$$

where the sum runs over all possible spin configurations. (On a lattice of N sites, there are 2^N such configurations.) For example, the free energy is

$$F=-kT\ln Z\,. \tag{2.6}$$

The mean magnetization per site can be expressed either as

$$M=\left\langle (1/N)\sum_n s(n)\right\rangle=\sum_{\{\text{configs}\}}\left[(1/N)\sum_n s(n)\right]e^{-\beta S}/Z \tag{2.7a}$$

or, using Eq. (2.6), as

$$M=\frac{1}{N}\frac{\partial}{\partial B}F\,. \tag{2.7b}$$

A measure of the response of the spins to an external infinitesimal magnetic field is given by the susceptibility per site,

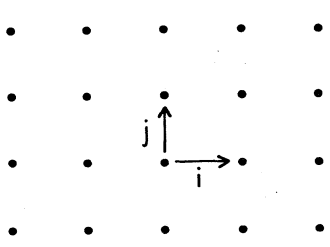


FIG. 1. The unit vectors of the square lattice in two dimensions.

$$\chi=\left.\frac{\partial M}{\partial B}\right|_{B=0}\,. \tag{2.8}$$

Using Eq. (2.7a) χ can be written in terms of a configurational average

$$\begin{aligned}\chi&=(1/NkT)[\langle s_{\text{tot}}^2\rangle-\langle s_{\text{tot}}\rangle^2]\\&=(1/NkT)[\langle(s_{\text{tot}}-\langle s_{\text{tot}}\rangle)^2\rangle]\,,\end{aligned} \tag{2.9}$$

where $s_{\text{tot}}=\sum_n s(n)$. This formula shows that the zero-field susceptibility is a measure of the *fluctuations* in the spins. Equation (2.9) is an example of the "fluctuation–dissipation" theorem (Stanley, 1971). We learn that χ will be large at those values of the temperature T where the spins are fluctuating considerably. χ can also be written in the form

$$\chi=(1/NkT)\left[\sum_{n,m}\langle s(n)s(m)\rangle-\left\langle\sum_n s(n)\right\rangle^2\right], \tag{2.10}$$

which shows that χ is also a measure of the *correlations* among the spins. Define the spin–spin correlation function

$$\Gamma(n)=\langle s(n)s(0)\rangle\,, \tag{2.11}$$

and suppose that the system has no net magnetization, $\langle s(0)\rangle=0$. Then, using the translational invariance of the system, Eq. (2.10) becomes

$$\chi=(1/kT)\sum_n\Gamma(n)\,. \tag{2.12}$$

Note that χ can diverge if the system has sufficiently long-range correlations.

The connection between long-range correlations and singular behavior in the system's thermodynamic properties is important. It will be discussed later in this section and will reappear several times throughout this article. But first we should discuss the phenomenology of the critical region.

Consider the magnetization M as a function of T and B. At fixed T, let B tend to zero. If M remains nonzero, the system is said to experience "spontaneous magnetization." This occurs only for T below a critical value T_c in our example, the Ising model. Otherwise M tends to zero as the magnetic field is removed. Note that a spontaneous magnetization indicates that the equilibrium state of the system does not possess the global up $\gtrless$ down symmetry of the system's action. Below T_c that symmetry is "spontaneously broken." M can serve as a "local order parameter" to label the phases of the system. As T approaches T_c from below, M vanishes. In many physical systems including our model it vanishes as a power

$$M\sim(T_c-T)^{\beta}\,, \tag{2.13}$$

where β is the magnetization critical exponent (Fisher, 1967). It takes the value 1/8 in the two-dimensional Ising model as noted in Table I, where other critical indices are collected. A good theory of critical phenomena should yield an estimate of β and the other indices.

Next consider the spin–spin correlation function $\Gamma(n)$ for $T>T_c$. At truly high temperature we expect thermal fluctuations to dominate the tendency of distant spins to align and correlate. One finds that, in fact, $\Gamma(n)$ falls

TABLE I. Exact critical indices of the two-dimensional Ising model.

Critical index	Relation	Value in 2-D Ising model
β	$M \sim (T_c - T)^{\beta}$	1/8
η	$\Gamma(\mathbf{n}) \sim \lvert\mathbf{n}\rvert^{-(d-2+\eta)}, (T = T_c)$	1/4
ν	$\xi \sim (T - T_c)^{-\nu}$	1
γ	$\chi \sim (T - T_c)^{-\gamma}$	1.75
α	$C \sim (T - T_c)^{-\alpha}$	0
δ	$M \sim B^{1/\delta}, (T = T_c)$	15

off exponentially with the distance between spins for any T above T_c,

$$\Gamma(n) \sim \exp[-|n|/\xi(T)], \quad |n| \gg 1, \tag{2.14}$$

where $\xi(T)$ is the system's "correlation length." $\xi(T)$ gives a measure of the size of patches of correlated spins in the system. For high T, $\xi(T)$, measured in units of the lattice spacing, is of order unity. For temperatures below the critical point,

$$\Gamma(n) \sim \langle s(0)\rangle^2 \neq 0, \quad |n| \gg 1. \tag{2.15}$$

There are no long-range correlations here either. The system is simply magnetized. In fact, the connected correlation function

$$\Gamma_{\text{con}}(n) = \langle s(0)s(n)\rangle - \langle s(0)\rangle^2, \tag{2.16}$$

is exponentially small in the Ising model. Precisely at the critical point $\Gamma(n)$ falls off as a power of $|n|$:

$$\Gamma(n) \sim |n|^{-(d-2+\eta)}, \quad T = T_c, \tag{2.17}$$

where η is another critical index recorded in Table I. So, the system has long-range correlations only at its critical temperature. In order that Eqs. (2.14) and (2.17) be compatible, it is necessary that the correlation length $\xi(T)$ diverge as T approaches T_c from above,

$$\xi(T) \sim (T - T_c)^{-\nu}, \tag{2.18}$$

where ν is another standard critical exponent. Also, our exercise with the "fluctuation–dissipation" theorem suggests that the susceptibility χ diverges as T approaches T_c,

$$\chi \sim (T - T_c)^{-\gamma}, \tag{2.19}$$

where γ is recorded in Table I. Another quantity of interest is the specific heat

$$C = -T\frac{\partial^2 F}{\partial T^2}. \tag{2.20}$$

It also may diverge as the temperature is reduced to T_c,

$$C \sim (T - T_c)^{-\alpha}, \tag{2.21}$$

where α is the specific heat critical exponent. Finally, if we adjust T to be precisely T_c and apply a small external magnetic field, we should expect the magnetization to respond sharply. A critical index δ characterizes this effect,

$$M \sim B^{1/\delta} \quad (T = T_c). \tag{2.22}$$

The essential point of this summary is the fact that at the critical temperature various thermodynamic functions develop singular behavior, and that this singular behavior is related to long-range correlations and large fluctuations. Although the underlying action of the Ising model has only short-range forces, correlations can appear in the model over an infinite range. One of the aims of the first several sections of this review is to provide an understanding of this point.

B. Correlation length scaling and the droplet picture

Experiment has shown that the critical indices of a wide variety of physical systems are identical. Such striking regularities have challenged physicists to pinpoint the essential physics of the critical region ($T \approx T_c$). One approach suggests that it is the divergence of $\xi(T)$ that is responsible for the singular dependence of *all* physical quantities on $T - T_c$ (Fisher, 1967). This is the "correlation length scaling hypothesis." It claims that in the critical region $\xi(T)$ is the only relevant length in the system. In some sense the system is no longer sensitive to the lattice and its small spacing. The fact that spins are correlated over distances of order $\xi(T)$ and that $\xi(T) \gg 1$ dominates the properties of the system.

It is interesting to accept this hypothesis and produce a physical picture of the Ising model for $T \approx T_c$ (Kadanoff, 1976a). As shown in Fig. 2 we expect regions of correlated spin of sizes ranging up to $\xi(T)$. These are depicted as droplets of overturned spins. But some thought indicates that this picture is not quite right. The point is that each droplet of size $\sim\xi(T)$ is itself a huge physical system near criticality. Therefore, to be consistent, it must consist of droplets of overturned spins whose sizes range from zero to $\xi(T)$. A better picture is shown in Fig. 3. According to this final view, there are fluctuations in the system on *all* length scales from zero to $\xi(T)$. If we choose $T = T_c$ where $\xi(T)$ diverges, then the system should appear identical on all length scales. It would be scale invariant.

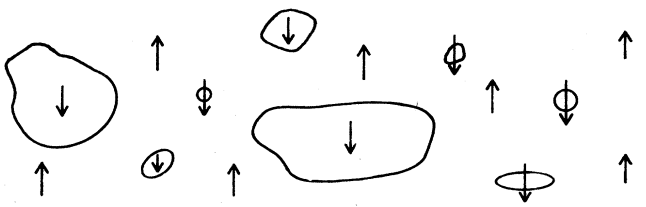


FIG. 2. Droplets in the Ising model near the critical point.

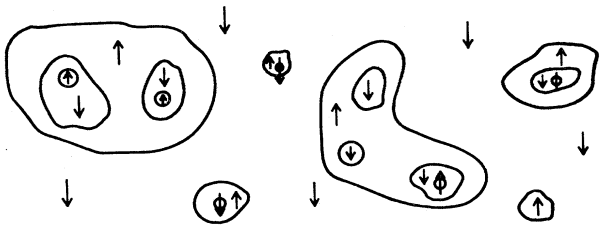


FIG. 3. Scale-invariant droplets.

The correlation length scaling hypothesis leads to relations among the various critical indices we have introduced. They read

$$\beta = \tfrac{1}{2}\nu(d-2+\eta),$$
$$\gamma = \nu(2-\eta),$$
$$\alpha = 2-\nu d,$$
$$\delta = (d+2-\eta)/(d-2+\eta). \tag{2.23}$$

The essential ingredient leading to these relations is the assumption that $\xi(T)$ is the only relevant length in the problem. General relations among the thermodynamic functions and dimensional analysis then lead to Eq. (2.23) (Fisher, 1967). More sophisticated arguments based on the operator product expansion (Kadanoff, 1976a) can also be made.

It is interesting to observe that the scaling relations are true for the critical indices of the two-dimensional Ising model collected in Table I. It is also interesting that other approaches to critical behavior produce values for the indices which violate these scaling laws and the exact results of the Ising model. For example, mean field theory gives $\beta = 1/2$, $\nu = 1/2$, and $\eta = 0$ independent of the dimensionality of the system. These results violate Eq. (2.23) except in four dimensions. This limitation of mean field theory is well known and is easily understood by recalling that it ignores fluctuations. Therefore, although it can be a very useful scheme, it misses the essence of the physics except in dimension four and above. This failure is another hint of the fact that fluctuations are an essential ingredient in models of critical phenomena in low-dimensional spaces.

Since many physical systems have identical critical indices, it is important to determine those features of the Action which are relevant to the critical behavior of the system. It appears that physical systems can be arranged into "universality classes," and that within each class each substance has common critical behavior (Kadanoff, 1976a). These classes are labled by (1) the dimension of the local variables in the action, (2) the symmetries of the coupling between the local variables, (3) the dimensionality of the system. The value of this classification is that it tells us what we can ignore. For example, the detailed lattice features are unimportant—they may affect the critical temperature T_c but they do not affect the singular parts of the thermodynamic functions in the critical region. For example, the Ising model can be formulated on anisotropic lattices and the same critical indices result.

Rules 1–3 are, in fact, not a complete list. We shall see that universality can be discussed accurately within the context of the renormalization group (Wilson and Kogut, 1974). Then the origin of rules 1–3 and the additional relevant features of the physical system which control its critical exponents can be understood. For example, according to rules 1–3, critical indices should not depend continuously on any parameter in the Action. However, the low-temperature phase of the planar Heisenberg spin system in two dimensions does not have this property. Later we shall review the renormalization group analysis for that model and shall then be able to formulate the idea of universality more precisely.

III. THE TRANSFER MATRIX–FIELD THEORY AND STATISTICAL MECHANICS

A. General remarks

The goal of this section is to establish the connection between statistical mechanics in four dimensions and field theory in three spatial and one time dimension. This will be done using simple examples and some general analysis. The key to this connection lies in the transfer matrix (Schultz, 1964), as will be discussed below. In this approach field theories are regulated by the space–time cutoff provided by the lattice itself. Even within this framework different cutoff procedures can be contemplated. We shall consider both space–time symmetric cutoffs (cubic lattices in four dimensions) and very anisotropic systems in which one axis is a continuum and the remaining three employ a lattice. The reason we consider these cases is that each has certain advantages. The space–time symmetric formulation is frequently more elegant and rigorous, while the second approach resembles Hamiltonian quantum mechanics and exposes some of the physical properties of the theory quite simply.

If one's ultimate interest is field theory formulated in a continuum space–time, then the lattice is of secondary interest. It is necessary to learn how to retrieve the physics of a continuum field theory after using the lattice as a scaffolding on which to formulate it precisely. We shall see that continuum limits of lattice theories exist at the critical points of the lattice theory (Wilson and Kogut, 1974). Therefore our discussions center upon the critical regions of the lattice theories. In many cases we shall be interested in finding the lattice theory's phase diagram as a first step toward understanding its various possible continuum limits.

In many of our discussions we shall be using different lattice versions of a single field theory. It should be kept in mind that these different lattice formulations describe the same physics. For example, different cutoff procedures in perturbation theory produce identical renormalized Feynman amplitudes. It will be one of our engineering problems to ensure that different lattice formulations of a field theory lead to the same physics.

B. The path integral and transfer matrix of the simple harmonic oscillator

Much of the formalism we wish to develop for field theories can be illustrated in the context of nonrelativ-

istic quantum mechanics. We shall discuss the relationship between the Feynman path integral and the Schrödinger equation for a one-dimensional potential problem (Creutz, 1977). This will be worked out in detail so that later discussions of field theories will only emphasize those new features encountered when dealing with many degrees of freedom in higher dimensions.

We begin with the Lagrangian for a one-dimensional simple harmonic oscillator,

$$\mathcal{L}=\tfrac{1}{2}(\dot{x}^2-\omega^2x^2). \tag{3.1}$$

In Feynman's formulation of quantum mechanics (Feynman, 1948) one considers the amplitude that the particle will be initially at (x_a, t_a) and finally at (x_b, t_b). The amplitude for this transition is then postulated to be

$$Z=\sum_{\text{paths}}\exp[(i/\hbar)S_m], \tag{3.2a}$$

where the sum is over all world lines between the initial and final points and S_m is the Minkowski space Action for a particular path,

$$S_m=\int_{t_a}^{t_b}\mathcal{L}dt. \tag{3.2b}$$

So, Eq. (3.2a) states that each path contributes to the transition through a weight $\exp[(i/\hbar)\int\mathcal{L}dt]$. A sensible definition of the "sum over all paths" must be provided before Eq. (3.2a) becomes useful. One way to proceed is to introduce a space–time lattice so that various paths can be labeled simply. For example, make the time axis discrete and call the spacing between slices ε

$$t_{i+1}-t_i=\varepsilon, \tag{3.3}$$

as shown in Fig. 4. It is also best to modify an expression such as Eq. (3.2a), which has rapidly oscillating phases, by continuing it to imaginary time

$$t=-i\tau. \tag{3.4}$$

Then each path will be weighted by an ordinary damped exponential factor and it becomes easier to distinguish important from unimportant paths. Substituting Eq. (3.4) into (3.2b) gives

$$\frac{i}{\hbar}S_m=-\frac{1}{2\hbar}\int\left[\left(\frac{dx}{d\tau}\right)^2+\omega^2x^2\right]d\tau. \tag{3.5}$$

So, if we define the Euclidean Action,

$$S=\frac{1}{2}\int\left[\left(\frac{dx}{d\tau}\right)^2+\omega^2x^2\right]d\tau, \tag{3.6}$$

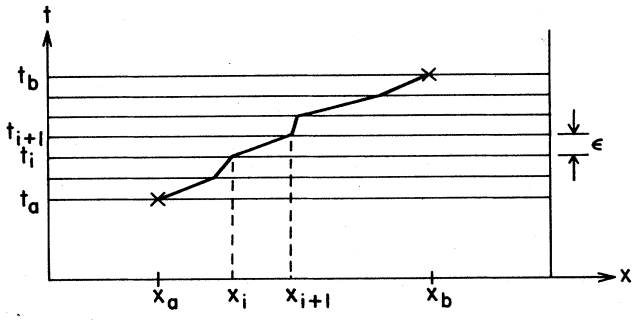


FIG. 4. A discrete time axis and illustrative path in quantum mechanics.

then the path integral becomes

$$Z=\int_{-\infty}^{\infty}\prod_i dx_i\exp\left(-\frac{1}{\hbar}S\right). \tag{3.7}$$

For discrete time slices,

$$S=\tfrac{1}{2}\varepsilon\sum_i\{[(x_{i+1}-x_i)/\varepsilon]^2+\omega^2x_i^2\}. \tag{3.8}$$

Before continuing to the transfer matrix and the Schrödinger description of this system, note that Eqs. (3.7) and (3.8) constitute a one-dimensional statistical mechanics problem. We have a one-dimensional lattice whose sites are labeled with the index i. On each site there is a variable x_i which takes on values between $-\infty$ and $+\infty$. The Action couples nearest-neighbor variables x_i and x_{i+1} together. The integral $\int\Pi dx_i$ is a sum over configurations, each weighted by an exponential $\exp[-(1/\hbar)S]$. It is important to note that $\hbar$ plays the role of the temperature in the statistical mechanics formulation. We usually think of the temperature as a measure of the fluctuations in a statistical mechanics problem and $\hbar$ as a measure of the fluctuations (through the uncertainty principle) of a quantum problem. Recall from quantum mechanics that the $\hbar\to 0$ limit picks out classical physics. In particular, as $\hbar\to 0$ the classical trajectory of the harmonic oscillator becomes the only path that contributes to the amplitude Z, and fluctuations are completely suppressed. These points make the correspondence

$$\hbar\longleftrightarrow T, \tag{3.9}$$

very appealing, since one visualizes the $T=0$ point of a statistical mechanics problem as frozen and without fluctuations.

Now let us return to our analysis and organize the evaluation of the integrals in Eq. (3.7) in an enlightening fashion. Since the Action only couples nearest-neighbor lattice variables x_i, we can write the partition function as

$$Z=\int\prod_i[dx_iT(x_{i+1},x_i)], \tag{3.10}$$

where

$$T(x_{i+1},x_i)=\exp\left\{-\frac{1}{2\hbar}\left[\frac{1}{\varepsilon}(x_{i+1}-x_i)^2+\frac{1}{2}\omega^2\varepsilon x_{i+1}^2+\frac{1}{2}\omega^2\varepsilon x_i^2\right]\right\}. \tag{3.11}$$

We can think of T as the matrix element of an operator—the transfer matrix—and establish the equivalence of the path integral with the Hamiltonian approach to quantum mechanics. First we must set up a space of states. Let there be operators $\hat{x}$ and $\hat{p}$. Eigenstates of $\hat{x}$ are states in which the particle is localized at the position x,

$$\hat{x}|x\rangle=x|x\rangle. \tag{3.12}$$

The states $|x\rangle$ have the continuum normalization

$$\langle x'|x\rangle=\delta(x'-x). \tag{3.13a}$$

We assume that they are complete, so

$$1=\int dx\,|x\rangle\langle x|\,. \tag{3.13b}$$

Finally, let there be a momentum operator $\hat{p}$ canonically conjugate to $\hat{x}$,

$$[\hat{p},\hat{x}]=-i\hbar\,. \tag{3.14}$$

Using the operators $\hat{p}$ and $\hat{x}$ we shall construct an operator $\hat{T}$ with the property

$$\langle x'|\,\hat{T}\,|x\rangle=T(x',x)\,, \tag{3.15}$$

as given by Eq. (3.11). Clearly $\hat{T}$ is the usual time evolution operator of quantum mechanics (imaginary time) evaluated over the interval ε. Using $\hat{T}$ we can obtain a useful expression for Z,

$$\begin{aligned}Z&=\int\prod_i dx_i\langle x_{i+1}|\,\hat{T}\,|x_i\rangle\\&=\int\langle x_b|\,\hat{T}\,|x_{N-1}\rangle dx_{N-1}\langle x_{N-1}|\,\hat{T}\,|x_{N-2}\rangle dx_{N-2}\langle x_{N-2}|\,\hat{T}\cdots\\&\quad\cdots\langle x_2|\,\hat{T}\,|x_1\rangle dx_1\langle x_1|\,\hat{T}\,|x_a\rangle\\&=\langle x_b|\,\hat{T}^N|x_a\rangle\,,\end{aligned} \tag{3.16}$$

where $N-1$ is the number of τ slices between τ_a and τ_b and we used completeness in arriving at Eq. (3.16). If we impose periodic boundary conditions and sum over all possible initial positions of the particle, Eq. (3.16) is replaced by the more familiar result,

$$Z=\mathrm{tr}\,\hat{T}^N\,. \tag{3.17}$$

So, if we can find an expression for $\hat{T}$ we are done. Note that

$$\langle x'|\,\exp(-\tfrac{1}{2}\varepsilon\hat{p}^2)\,|x\rangle=\text{const.}\exp[-\tfrac{1}{2}\varepsilon\hbar^2(x'-x)^2]\,, \tag{3.18a}$$

which is easily obtained using the complete set of momentum eigenstates $\{|p\rangle\}$,

$$\begin{aligned}\langle x'|\exp(-\tfrac{1}{2}\varepsilon\hat{p}^2)\,|x\rangle&=\int dpdp'\langle x'|\,p'\rangle\langle p'|\\&\quad\times\exp(-\tfrac{1}{2}\varepsilon\hat{p}^2\,|\,p\rangle\langle p\,|\,x\rangle\\&=\frac{1}{(2\pi)}\int dpdp'\exp(ip'x'/\hbar)\\&\quad\times\exp(-\tfrac{1}{2}\varepsilon p^2)\delta(p'-p)\exp(-ipx/\hbar)\\&=\frac{1}{(2\pi)}\int dp=\exp[ip(x'-x)/\hbar]\\&\quad\times\exp(-\tfrac{1}{2}\varepsilon p^2)\\&=\text{const.}\exp[-(\tfrac{1}{2}\varepsilon\hbar^2)(x'-x)^2]\,.\end{aligned} \tag{3.18b}$$

Using this result and inspecting Eq. (3.11), we have

$$\hat{T}=\exp[-(1/4\hbar^2)\varepsilon\omega^2\hat{x}^2]\exp[-(1/2\hbar)\varepsilon\hat{p}^2]\times\exp[-(1/4\hbar^2)\varepsilon\omega^2\hat{x}^2]\,, \tag{3.19}$$

and this is true for all lattice spacings ε. The operator ordering in Eq. (3.19) is important.

The transfer matrix is closely related to the theory's quantum Hamiltonian. Formally one can consider the operator

$$\hat{H}_s=-(\hbar/\varepsilon)\ln\hat{T}\,, \tag{3.20}$$

but it is rather complicated and not very enlightening. (The subscript s anticipates useful notation for later discussions.) However, if we let $\varepsilon\to 0$ so that the τ axis becomes a continuum and there are an infinite number of slices between τ_a and τ_b, then

$$\hat{H}_s\to\hat{H}\,, \tag{3.21a}$$

where

$$\hat{H}=\tfrac{1}{2}(\hat{p}^2+\omega^2\hat{x}^2)\,. \tag{3.21b}$$

$\hat{H}$ is the familiar Hamiltonian for the simple harmonic oscillator. The reason $\hat{H}$ is simple while $\hat{H}_s$ is not is the following. To write Eq. (3.19) as a simple exponential we apply the identity

$$\exp(\hat{A})\exp(\hat{B})=\exp(\hat{A}+\hat{B}+\tfrac{1}{2}[\hat{A},\hat{B}]+\cdots)\,, \tag{3.22}$$

with $\hat{A}=-\varepsilon\hat{p}^2/2\hbar$ and $B=-\varepsilon\omega^2\hat{x}^2/4\hbar^2$. But $[\hat{A},\hat{B}]=O(\varepsilon^2)$, so as $\varepsilon\to 0$ it is negligible compared to $\hat{A}$ and $\hat{B}$. Therefore

$$\hat{T}=\exp\{-\varepsilon/\hbar\,[\hat{H}+O(\varepsilon)]\}\,, \tag{3.23}$$

which we identify as the familiar definition of the Hamiltonian in terms of the evolution operator for an infinitesimal step in time (Messiah, 1962). The Schrödinger equation follows by observing that the Feynman interpretation of Z means that if the wave function of the particle is $\psi(x,\tau)$, then

$$\psi(x',\tau')=\int Z(x',\tau';x,\tau)\psi(x,\tau)\,dx\,. \tag{3.24}$$

Choosing $\tau'-\tau=\varepsilon$ to be infinitesimal so that this interval consists of just one slice,

$$\begin{aligned}Z=\langle x'|\,\hat{T}\,|x\rangle&=\langle x'|\exp[-(\varepsilon/\hbar)\hat{H}]\,|x\rangle\\&\approx\langle x'|1-(\varepsilon/\hbar)\hat{H}\,|x\rangle\\&=\delta(x'-x)-(\varepsilon/\hbar)H\delta(x'-x)\,.\end{aligned} \tag{3.25}$$

Then Eq. (3.24) becomes

$$\psi(x',\tau')=\psi(x,\tau)-(\varepsilon/\hbar)H\psi(x',\tau) \tag{3.26}$$

or

$$\hbar\frac{\partial}{\partial\tau}\psi(x,\tau)=-H\psi(x,\tau)\,, \tag{3.27}$$

which is the Euclidean version of Schrödinger's equation. H in Eq. (3.27) is just the Hamiltonian in the x-space realization. In summary, we have the usual operator form of quantum mechanics in the Schrödinger picture—operators are τ independent and state functions carry their τ dependence.

Since the path integral and Hamiltonian formulations of quantum mechanics are equivalent, convenience dictates which one uses in a certain application. Generally, the path integral is better for scattering processes, while the Hamiltonian is better for bound-state problems.

C. The transfer matrix for field theories

Now we shall extend the considerations of the previous section to field theories. Most of our efforts will be

aimed at establishing precise correspondences between properties of a statistical mechanics system and those of a field theory.

Consider a self-coupled scalar field in d space–time dimensions. The Lagrangian will be

$$L=\int d^{(d-1)}x\left[\frac{1}{2}(\partial\phi/\partial t)^2-\frac{1}{2}(\nabla\phi)^2-\frac{1}{2}\mu_0\phi^2-\lambda_0\phi^4\right], \tag{3.28}$$

and a formal expression for the path integral can be written down. As in the previous section it is best to analytically continue these expressions to imaginary time. We assume that this can always be done. It can be done order by order in perturbation theory for all the Green's functions of the theory (Schwinger, 1958). Granting this, we formulate the theory on an anisotropic space–time lattice by replacing integrals by sums and derivatives with discrete differences. Denote the lattice spacing in a "spatial" direction by "a" and that in the "temporal" direction by τ. Then the theory's Euclidean Action, the temporal integral of the Lagrangian, becomes

$$S=\sum_n\Big\{(1/2\tau)a^{d-1}[\Delta_0\phi(n)]^2+\tfrac{1}{2}\tau a^{d-3}\sum_k[\Delta_k\phi(n)]^2 +\tfrac{1}{2}\tau a^{d-1}\mu_0^2\phi^2(n)+\tau a^{d-1}\lambda_0\phi^4(n)\Big\}, \tag{3.29}$$

where $n=(n_0,n_1,n_2,n_3)$ labels the four-dimensional lattice, Δ_μ is a discrete difference operator

$$\Delta_\mu f(n)=f(n+\mu)-f(n), \tag{3.30}$$

$\mu=0$ is the temporal direction, and $\mu=k$ ($k=1,2,3$) labels the spatial directions. The notation $n=(n_0,\mathbf{n})$ will also be used from time to time. It is best to name the coefficients in Eq. (3.29)

$$K_\tau=(1/2\tau)a^{d-1},\quad K=\tfrac{1}{2}\tau a^{d-3},$$
$$b_0=\tfrac{1}{2}\tau a^{d-1}\mu_0^2,\quad u_0=\tau a^{d-1}\lambda_0. \tag{3.31}$$

Then K_τ is the strength of the nearest-neighbor coupling in the temporal direction and K is the coupling in the spatial direction. Note that they are equal only if $\tau=a$. Equation (3.29) is now

$$S=\sum_n\Big\{K_\tau(\Delta_0\phi)^2+K\sum_k(\Delta_k\phi)^2+b_0\phi^2+u_0\phi^4\Big\}. \tag{3.32}$$

The path integral for the lattice theory is

$$Z=\prod_n\int_{-\infty}^{\infty}d\phi(n)e^{-S}, \tag{3.33}$$

which can be thought of as the partition function of a four-dimensional statistical mechanics problem. The boundary conditions of Eq. (3.33) might consist of specifying ϕ on an "initial" temporal slice and a "final" temporal slice in direct correspondence with the example of the previous section. Again we can introduce a transfer matrix to "propagate" the field $\phi(n)$ in the temporal direction. We label the field ϕ on one time slice and ϕ' on the next time slice, and define

$$\langle\phi'|\hat{T}|\phi\rangle=\exp\Big[-\sum_n\Big(K_\tau[\phi'(\mathbf{n})-\phi(\mathbf{n})]^2+\tfrac{1}{2}K\sum_k\{[\phi'(\mathbf{n}+k)-\phi'(\mathbf{n})]^2+[\phi(\mathbf{n}+k)-\phi(\mathbf{n})]^2\} +\tfrac{1}{2}b_0[\phi'^2(\mathbf{n})+\phi^2(\mathbf{n})]+\tfrac{1}{2}u_0[\phi'^4(\mathbf{n})+\phi^4(\mathbf{n})]\Big)\Big]. \tag{3.34}$$

To obtain an operator expression for $\hat{T}$ we introduce second-quantized fields $\hat{\phi}(\mathbf{n})$ and $\hat{\pi}(\mathbf{n})$, which are conjugate variables,

$$[\hat{\pi}(\mathbf{n}'),\hat{\phi}(\mathbf{n})]=-i\delta_{\mathbf{n}',\mathbf{n}}. \tag{3.35}$$

Then manipulations generalizing those of our previous section give the partition function

$$Z=\mathrm{tr}\,\hat{T}^{N+1}, \tag{3.36}$$

where N is the number of time slices (spacing τ) between the initial and final times. The trace appears here because we have identified the initial and final fields and have summed over it. The operator expression for $\hat{T}$ is

$$\hat{T}=\exp\Big[-\tfrac{1}{2}\sum_{\mathbf{n}}\Big\{K\sum_k[\hat{\phi}'(n+k)-\hat{\phi}'(n)]^2+b_0\hat{\phi}'^2(n)+u_0\hat{\phi}'^4(n)\Big\}\Big] \times\exp\Big[-\sum_{\mathbf{n}}(1/4K_\tau)\hat{\pi}^2(n)\Big]\exp\Big[-\tfrac{1}{2}\sum_n\Big\{K\sum_k[\hat{\phi}(n+k)-\hat{\phi}(n)]^2+b_0\hat{\phi}^2(n)+u_0\hat{\phi}^4(n)\Big\}\Big]. \tag{3.37}$$

In general $\hat{T}$ is not the exponential of a simple, familiar operator. If we choose $\tau=a$ so that the lattice is symmetric, we can still write

$$\hat{T}=e^{-\hat{H}_s\tau}, \tag{3.38}$$

if we are willing to consider a complicated "Hamiltonian" $\hat{H}_s$. The notion of infinitesimal time translations does not exist here, but the statistical mechanics problem is quite elegant because $K=K_\tau$. Alternatively we can take the τ-continuum limit and compute the same partition function by subdividing the overall time interval into an infinite number of steps. In that case $\tau\to 0$

while a is held fixed and

$$\hat{T}=e^{-\hat{H}\tau}\approx 1-\tau\hat{H}+\cdots, \tag{3.39}$$

where $\hat{H}$ is the familiar canonical Hamiltonian of the original field theory [Eq. (3.28)] formulated on a spatial lattice. From Eq. (3.31) we note that, as $\tau\to 0$ for fixed a,

$$K_\tau\to\infty \text{ and } K\to 0 \tag{3.40a}$$

in such a way that their product remains fixed,

$$K_\tau K=\tfrac{1}{4}a^{2(d-2)}. \tag{3.40b}$$

This result is not surprising: in order that the physics be the same in various lattice formulations of a theory, the couplings must be adjusted appropriately. We see that taking a τ-continuum limit forces us to consider very anisotropic statistical mechanics systems.

Now we want to establish detailed connections between the partition function and the operator formulations of a field theory. We shall do this within the context of the symmetric formulation $\tau=a$ for ease of presentation. The τ-continuum formulation could also be used; the details are left to the reader and later sections of this review. So, we have a statistical mechanics problem on a symmetric lattice and field-theoretic formulation using operators $\hat{\phi}(n)$, $\hat{\pi}(n)$, and $\hat{H}_s$ defined on a symmetric lattice which provides the theory with an ultraviolet cutoff. The correspondences we shall find are:

Statistical Mechanics	*Field Theory*
Free energy density	Vacuum energy density
Correlation function	Propagator
Reciprocal of the correlation length	Mass gap

These results will give us additional insight into the nature of phase transitions and the continuum limits of lattice field theories.

Since the transfer matrix is Hermitian ($\hat{T}=\hat{T}^\dagger$), we know that its spectrum can be arranged into an orthonormal set $\{|1\rangle\}$ with real eigenvalues $[\exp(-E_i\tau)]$. A spectral decomposition can be written

$$\hat{T}=\sum_i |i\rangle e^{-E_i\tau}\langle i|. \tag{3.41}$$

To obtain the partition function we need T raised to a large power $N+1$,

$$\hat{T}^{N+1}=\sum_i |i\rangle \exp[-(N+1)E_i\tau]\langle i|. \tag{3.42}$$

Now a significant simplification occurs as $N\to\infty$. Suppose that the lower eigenvalue of $\hat{H}_s$ is unique, E_0. Then the first term in Eq. (3.42) dominates and the relative error in dropping all other terms goes to zero exponentially as $N\to\infty$,

$$\hat{T}^{N+1}\to |0\rangle e^{-E_0 T}\langle 0|, \tag{3.43}$$

where T is the difference between the initial and final times. So, the partition function becomes

$$Z=e^{-E_0 T}. \tag{3.44}$$

But the free energy of a statistical mechanics problem is

$$Z=e^{-F}, \tag{3.45}$$

where we have absorbed a factor of kT into our definition of F. But F is an extensive quantity in four dimensions, so

$$F=fVT, \tag{3.46}$$

where V is the volume of space and f is a free energy density. Similarly E_0 is an extensive quantity in three dimensions, so

$$E_0=\omega_0 V, \tag{3.47}$$

where ω_0 is the energy density of the quantum vacuum. Collecting these results,

$$f=\omega_0 \tag{3.48}$$

as claimed above. The reader should note that the free energy—not the internal energy—of statistical mechanics enters this correspondence. F contains entropy effects,

$$F=U-TS \quad (T=\text{temperature}), \tag{3.49}$$

where U is "internal" energy.

The assumption that the largest eigenvalue of the transfer matrix be unique is an important one. We shall see that it means that the system is in its high-temperature phase.

Next consider the field-theoretic propagator in Minkowski space

$$\Delta(t,\mathbf{x})=\langle 0|T\hat{\phi}(t,\mathbf{x})\hat{\phi}(0,\mathbf{x})|0\rangle, \tag{3.50}$$

where $|0\rangle$ is the exact ground state of $\hat{H}_s$ and $\hat{\phi}(\mathbf{x},t)$ is the field operator expressed in the Heisenberg picture. $\hat{\phi}(x,t)$ is related to the Schrödinger (time-independent) fields we have been using by

$$\hat{\phi}(t,\mathbf{x})=e^{i\hat{H}_s t}\phi(\mathbf{x})e^{-i\hat{H}_s t}. \tag{3.51}$$

Choosing $t>0$, Eq. (3.50) becomes

$$\Delta(t,\mathbf{x})=\langle 0|\hat{\phi}(\mathbf{x})e^{-i\hat{H}_s t}\hat{\phi}(0)|0\rangle e^{iE_0 t}, \tag{3.52}$$

where $\mathbf{x}$ and t label points on the space–time lattice.

Now consider the correlation function,

$$\Gamma(n_0,\mathbf{n})=Z^{-1}\int\prod_{n_0',\mathbf{n}'} d\phi(n_0',\mathbf{n}')\phi(n_0,\mathbf{n})\phi(0,0)e^{-S}. \tag{3.53}$$

Organize the configurational sums so that for each time slice n_0' the integrals are done over all spatial sites $\mathbf{n}$. Then, only for the slices $n_0'=0$ and $n_0'=n_0$ will the calculations be different from those for Z itself. So, we find a factor of $\langle\phi'|\hat{T}|\phi\rangle$ appearing between adjacent slices. Following the steps which lead to Eq. (3.36), it is easy to obtain an operator expression for the correlation function,

$$\Gamma(n_0,\mathbf{n})=\mathrm{tr}\{\hat{T}^P\hat{\phi}(\mathbf{n})\hat{T}^{n_0}\hat{\phi}(0)\hat{T}^L\}/\mathrm{tr}\,\hat{T}^{N+1}, \tag{3.54}$$

where $P+n_0+L=N+1$. Now let the number of slices P, n_0, and L go to infinity so that Eq. (3.43) can be used for the transfer matrix. We then find

$$\Gamma(n_0,\mathbf{n})=\langle 0|\phi(\mathbf{n})e^{-n_0\hat{H}_s\tau}\phi(0)|0\rangle e^{n_0 E_0\tau}. \tag{3.55}$$

Comparing this result with Eq. (3.52) we have the identity

$$\Gamma(n_0, \mathbf{n}) = \Delta(-in_0\tau, \mathbf{n}), \tag{3.56}$$

which gives the second correspondence of interest. So, by computing the statistical mechanics correlation function, we learn the propagator of the field theory at imaginary time. The analytic continuation to real time must then be done. The general postulates of field theory (Streater and Wightman, 1964) guarantee that propagators have sufficiently simple analytic structure that this continuation is possible.

From these results it is easy to relate the mass gap m of the field theory to the correlation length ξ of the statistical mechanics. This important result will be

$$m = 1/(\xi a). \tag{3.57}$$

This formula implies that to make a field theory with particles of small masses, the underlying statistical mechanics must be nearly critical. In particular, taking the continuum limit $a \to 0$ we shall reach an interesting theory, i.e., one in which there is an excitation spectrum of finite masses, only if $\xi \to \infty$. One of our main interests will be, therefore, the character of lattice theory's phase diagrams and the nature of their critical regions.

To establish Eq. (3.57) we recall that the correlation function of a system not at a critical point falls exponentially,

$$\Gamma(n_0, 0) \sim \exp(-|n_0|/\xi), \tag{3.58}$$

for $|n_0| \gg \xi$. Now consider the propagator $\Delta(-in\tau, 0)$. Inserting complete sets of states $\{|l\rangle\}$ of the Hamiltonian $\hat{H}_s$ into Eq. (3.52), we find

$$\Delta(-in\tau, 0) = \sum_l \exp[-(E_l - E_0)n\tau]\,|\langle 0|\hat{\phi}(0)|l\rangle|^2. \tag{3.59}$$

For $n\tau \gg 1$, the right-hand side of this equation will be dominated by that state having the smallest value of $E_l - E_0$. But this is just the lightest particle state at zero momentum. The matrix element $\langle 0|\hat{\phi}(0)|l\rangle$ is almost certainly nonzero in this ϕ^4 field theory, and the smallest value of $E_l - E_0$ is m, the physical mass of that particle. So,

$$\Delta(-in\tau, 0) \sim \exp(-mn\tau). \tag{3.60}$$

Using the general relation Eq. (3.56) and comparing Eq. (3.58) with (3.60), we obtain our desired result, Eq. (3.57). Note that $\tau = a$ here because we are using a symmetric space–time lattice.

Now we can return to our assumption that the largest eigenvalue of the transfer matrix is unique. This means that m must be different from zero, which in turn implies that ξ must be finite. In other words, the parameters in the Action must be chosen so that the system is not critical. In addition, the temperature of the underlying statistical mechanics must be above T_c. To understand this consider the Ising model. Below T_c the system is magnetized and the two spin configurations, one with magnetization M and the other with magnetization $-M$, have identical free energies. The largest eigenvalue of the transfer matrix is then doubly degenerate for $T < T_c$.

IV. THE TWO-DIMENSIONAL ISING MODEL

A. Transfer matrix and τ-continuum formulation

In this section we shall begin to illustrate the general remarks of the previous section. We shall emphasize the τ-continuum Hamiltonian approach to the two-dimensional Ising model (Fradkin and Susskind, 1978) and shall discuss partition function analyses of the model formulated on a symmetric lattice in a later section.

Consider a two-dimensional lattice and place variables $\sigma_3(n) = \pm 1$ on sites. Denote the unit lattice vector in the temporal direction by $\hat{\tau}$ and that in the spatial direction by $\hat{x}$. Then the Action is

$$S = -\sum_n [\beta_\tau \sigma_3(n+\hat{\tau})\sigma_3(n) + \beta\sigma_3(n+\hat{x})\sigma_3(n)], \tag{4.1}$$

where the temporal coupling β_τ and the spatial coupling β are free parameters. Let us construct the transfer matrix and find the τ-continuum Hamiltonian of this model. It is then better to write the Action as

$$S = \tfrac{1}{2}\beta_\tau \sum [\sigma_3(n+\hat{\tau}) - \sigma_3(n)]^2 - \beta \sum_n \sigma_3(n+\hat{x})\sigma_3(n), \tag{4.2}$$

which differs from Eq. (4.1) by an unimportant constant. Consider two neighboring spatial rows and label the spin variables in one row $\sigma_3(m)$ and those in the next row $s_3(m)$. The argument m runs over the integers labeling the sites of a spatial row as shown in Fig. 5. The Action can now be written as a sum over these rows

$$S = \sum_{n_0} L(n_0 + 1, n_0), \tag{4.3a}$$

where

$$L = \tfrac{1}{2}\beta_\tau \sum_m [s_3(m) - \sigma_3(m)]^2 - \tfrac{1}{2}\beta \sum_m [\sigma_3(m+1)\sigma_3(m) + s_3(m+1)s_3(m)]. \tag{4.3b}$$

If there are M sites on each spatial row, there will be 2^M spin configurations on each row. Therefore the transfer matrix will be a $2^M \times 2^M$ matrix. Consider a diagonal element of the matrix. Then $\sigma_3(m) = s_3(m)$ for all m, and Eq. (4.36) reduces to

$$L(0 \text{ flips}) = -\beta \sum_m \sigma_3(m+1)\sigma_3(m). \tag{4.4}$$

If there is one spin flipped between the two rows, then

$$L(1 \text{ flip}) = 2\beta_\tau - \tfrac{1}{2}\beta \sum_m [\sigma_3(m+1)\sigma_3(m) + s_3(m+1)s_3(m)], \tag{4.5}$$

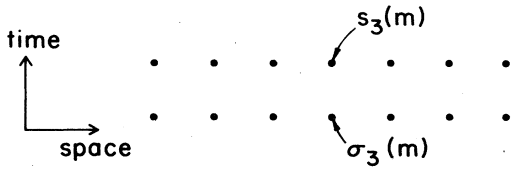


FIG. 5. Spin variables on adjacent spatial rows of the two-dimensional Ising model.

and if there were n spin flips,

$$L(n \text{ flips}) = 2n\beta_\tau - \tfrac{1}{2}\beta \sum_n [\sigma_3(m+1)\sigma_3(m) + s_3(m+1)s_3(m)] . \quad (4.6)$$

Next we must determine how β_τ and β must be adjusted so that the transfer matrix has the form

$$\hat{T} = e^{-\tau\hat{H}} \approx 1 - \tau\hat{H} , \quad (4.7)$$

as τ, the lattice spacing in the time direction, approaches zero. Consider various matrix elements of $\hat{T}$,

$$\hat{T}(0 \text{ flips}) = \exp\left\{\beta \sum_m \sigma_3(m+1)\sigma_3(m)\right\} \approx 1 - \tau\hat{H}\big|_{0 \text{ flips}} \quad (4.8a)$$

$$\hat{T}(1 \text{ flip}) = \exp(-2\beta_\tau) \times \exp\left\{\tfrac{1}{2}\beta \sum_m [\sigma_3(m+1)\sigma_3(m) + s_3(m+1)s_3(m)]\right\} \approx \tau\hat{H}\big|_{1 \text{ flip}} \quad (4.8b)$$

$$\vdots$$

$$\hat{T}(n \text{ flips}) = \exp(-2n\beta_\tau) \times \exp\left\{\tfrac{1}{2}\beta \sum [\sigma_3(m+1)\sigma_3(m) + s_3(m+1)s_3(m)]\right\} \approx -\tau\hat{H}\big|_{n \text{ flips}} . \quad (4.8c)$$

These equations will determine the τ dependence of β_τ and β. From Eq. (4.8a) we learn that

$$\beta \sim \tau , \quad (4.9a)$$

and from Eq. (4.8b),

$$\exp(-2\beta_\tau) \sim \tau . \quad (4.9b)$$

Therefore β and $\exp(-2\beta_\tau)$ must be proportional. Define the proportionality constant λ

$$\beta = \lambda \exp(-2\beta_\tau) . \quad (4.10)$$

We can identify the temporal lattice spacing

$$\tau = \exp(-2\beta_\tau) , \quad (4.11a)$$

so that the coupling between nearest-neighbor spins within a spatial row is

$$\beta = \lambda\tau . \quad (4.11b)$$

In summary, these results show that in order to define a smooth τ-continuum theory the couplings must be adjusted so that the temporal coupling grows large while the spatial coupling becomes weak. This general feature was noticed earlier for the ϕ^4 field theory, but the detailed scaling relations are different here. The physical interpretation of the parameter λ and the scaling relations will become clear as we continue.

Using the relations Eq. (4.11), we can identify the Hamiltonian H. Equations (4.8) become

$$\hat{T}(0 \text{ flips}) \approx 1 + \tau\left[\lambda \sum_m \sigma_3(m+1)\sigma_3(m)\right] = 1 - \tau\hat{H}\big|_{0 \text{ flips}} , \quad (4.12a)$$

$$\hat{T}(1 \text{ flip}) \approx \tau \approx -\tau\hat{H}\big|_{1 \text{ flip}} , \quad (4.12b)$$

$$\hat{T}(n \text{ flips}) \approx \tau^n \approx -\tau\hat{H}\big|_{n \text{ flips}} . \quad (4.12c)$$

Equation (4.12c) implies that the only matrix elements of $\hat{H}$ which survive as $\tau \to 0$ are the zero and one-spin-flip cases. This is a nice simplification which only occurs in the limit. To write $\hat{H}$ in operator form we need spin-flip operators. Place a Pauli matrix $\hat{\sigma}_3$ on each site m,

$$\hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (4.13a)$$

Let spin up (at site m) be represented by the vector $\binom{1}{0}$ and spin down by $\binom{0}{1}$. Then $\hat{\sigma}_1(m)$ is a spin-flip operator for site m. To check this recall that in the representation where Eq. (4.13a) holds one also has

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad (4.13b)$$

so

$$\hat{\sigma}_1\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}, \quad \hat{\sigma}_1\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} \quad (4.13c)$$

as claimed. To produce Eq. (4.12) the Hamiltonian must be

$$\hat{H} = -\sum_m \hat{\sigma}_1(m) - \lambda \sum_m \hat{\sigma}_3(m+1)\hat{\sigma}_3(m) . \quad (4.14)$$

So, we have a one-dimensional quantum Hamiltonian which can be interpreted as an Ising model in a transverse magnetic field (Pfeuty, 1970). We shall use it to study the phase diagram and critical region of the "classical" two-dimensional Ising model. According to our general considerations it must have the same phase diagram and critical indices as that model.

It is interesting to understand Eqs. (4.9)–(4.11) in more detail. Consider the formulation of the Ising model, Eq. (4.1), which began this discussion. That Action is parametrized by two variables (β_τ, β). It is known that for a certain range of these variables the system is magnetized (ordered, ferromagnetic), while for other values it is not magnetized (disordered, paramagnetic). There is a critical curve separating these two phases in the (β_τ, β) plane. In a later section, in which we shall study Eq. (4.1) and its partition function directly, we shall discuss calculational methods which easily yield the critical curve

$$\sinh(2\beta_\tau)\sinh(2\beta) = 1 . \quad (4.15)$$

This result is drawn in Fig. 6 and the phases are labeled appropriately. Choosing (β_τ, β) on the critical curve produces a theory with an infinite correlation length. Consider the form of the critical curve in the limit $\beta_\tau \to \infty$. Then

$$\sinh 2\beta_\tau \to \tfrac{1}{2}\exp(2\beta_\tau) \quad (4.16a)$$

and

$$\sinh 2\beta \to 2\beta . \quad (4.16b)$$

So, to maintain criticality in this limit, the parameters must satisfy

$$\beta = \exp(-2\beta_\tau) . \quad (4.17)$$

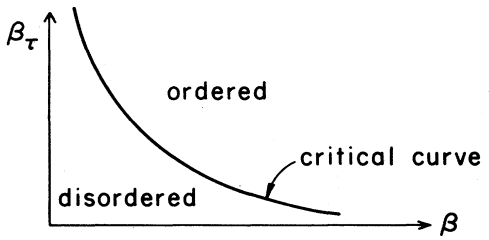


FIG. 6. Critical curve and phase diagram of the anisotropic two-dimensional Ising model.

Note that this is the same scaling relation we found in Eq. (4.10) except that λ has the specific value of unity. This shows that we can view the τ-continuum version of the theory as a natural limiting case of the general model, and that the parameter λ can be used to label its phases. If $\lambda > 1$, the model lies in the disordered phase. So λ is a temperaturelike variable with the correspondence

$$1/\lambda \sim \text{temperature}. \tag{4.18}$$

We shall often refer to $1/\lambda$ simply as temperature. The critical temperature is $\lambda = 1$.

It is interesting to understand the scaling relation Eq. (4.10) more intuitively (Fradkin and Susskind, 1978). To do this consider the spin–spin correlation function for the general model Eq. (4.1). If $\beta_\tau = \beta$, the lattice is symmetric under rotations through 90°, and the correlation function $\Gamma(n) = \langle \sigma_3(n)\sigma_3(0)\rangle$ shares this symmetry. Curves of constant $\Gamma(n)$ are shown in Fig. 7 for large $|n|$. They are approximately circular. If β_τ is now increased, correlations in the τ direction will become stronger, so curves of constant Γ will be distorted into ellipses with their major axes in the τ direction. An example is shown in Fig. 8. However, suppose we demand the same physics of the two lattice formulations. Then we must compensate the inequality $\beta_\tau > \beta$ by distorting the lattice so that $\tau < a$. If the lattice is adjusted appropriately, the contours of constant Γ can be left invariant. Clearly, as $\beta_\tau \to \infty$, the lattice spacing in the time direction must be squeezed to zero. The result $\tau = \exp(-2\beta_\tau)$ is the scaling relation which insures that the long-distance physics of the various formulations is kept essentially unchanged.

B. Self-duality of the Ising model

Our first step towards developing a detailed understanding of the Ising model is to derive a mapping between its high- and low-temperature behaviors. As a by-product of this construction we shall understand the special significance of the point $\lambda = 1$. This mapping is called a "duality" transformation (Kramers and Wannier, 1941).

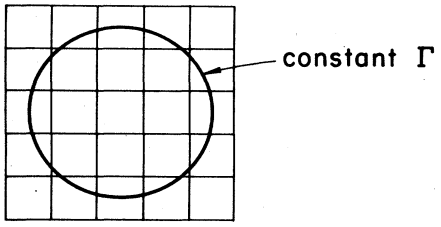


FIG. 7. A curve of constant $\Gamma(n)$ for the symmetric model.

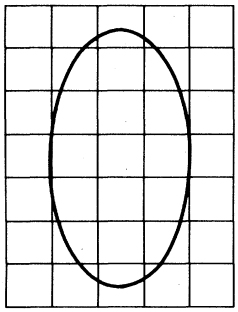


FIG. 8. A curve of constant $\Gamma(n)$ if $\beta_\tau > \beta$.

Consider the one-dimensional quantum-mechanical formulation of the model Eq. (4.14). From here on we shall omit the "hats" from operators. The first element of the duality transformation associates a new ("dual") lattice with our original spatial lattice. Sites of the original lattice will be associated with links of the "dual" lattice and vice versa. The transformation is visualized in Fig. 9. Operators are placed on the dual lattice so a spin system complete with a Hamiltonian will be generated from the original system. Define the operators on the dual lattice

$$\mu_1(n) = \sigma_3(n+1)\sigma_3(n)\,,$$

$$\mu_3(n) = \prod_{m<n} \sigma_1(m)\,. \tag{4.19}$$

So, $\mu_1(n)$ senses whether the spins on adjacent sites are aligned or not. $\mu_3(n)$ flips all the spins to the left of n. The dual operators have several important properties: (1) They satisfy the same Pauli spin algebra as σ_1 and σ_3. (2) The Hamiltonian can be rewritten simply using the dual spin operators. To check the first point, recall the algebra that defines the Pauli matrices. On a given site,

$$\sigma_1(n)\sigma_3(n) = -\sigma_3(n)\sigma_1(n)\,,$$

$$\sigma_1^2(n) = \sigma_3^2(n) = 1\,, \tag{4.20a}$$

and on different sites they commute,

$$[\sigma_1(n), \sigma_3(m)] = 0\,, \text{ etc.}\,, \quad \text{if } n \neq m\,. \tag{4.20b}$$

One checks that $\mu_1(n)$ and $\mu_3(m)$ satisfy the same algebra by using Eqs. (4.19) and (4.20) directly. For example, to check that

$$\mu_1(n)\mu_3(n) = -\mu_3(n)\mu_1(n)\,, \tag{4.21}$$

note that when Eq. (4.21) is written out only one factor of σ_1 and σ_3 are on the same site. Then the first rela-

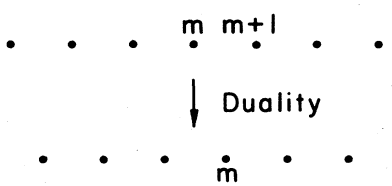


FIG. 9. Dual transformation for a spatial lattice.

tion in Eq. (4.20a) implies Eq. (4.21). One proves that

$$\mu_1(n)\mu_3(m) = \mu_3(m)\mu_1(n)\,, \quad n \neq m\,, \tag{4.22}$$

by observing that an even number of σ_1's and σ_3's are interchanged when passing from one side of Eq. (4.22) to the other. An even number of interchanges always produces a positive sign.

The second point is verified just as directly. Since

$$\sigma_1(m) = \mu_3(m+1)\mu_3(m)\,, \tag{4.23}$$

the Hamiltonian is

$$H = -\sum_n \mu_3(n)\mu_3(n+1) - \lambda \sum_n \mu_1(n)$$

$$= \lambda\left[-\sum_n \mu_1(n) - \lambda^{-1}\sum_n \mu_3(n)\mu_3(n+1)\right], \tag{4.24}$$

which has the same form as the Hamiltonian written in terms of the σ's,

$$H(\sigma:\lambda) = \lambda H(\mu;\lambda^{-1})\,. \tag{4.25}$$

But since the σ's and μ's have the same algebra, this is really an expression of symmetry for the original model alone—it states that the high-temperature and low-temperature properties of the model map onto one another! It is called "self-duality." In particular, Eq. (4.25) implies that each eigenvalue of H satisfies the relation

$$E(\lambda) = \lambda E(\lambda^{-1})\,. \tag{4.26}$$

This has a very important implication. Consider the mass gap $G(\lambda)$ of the Hamiltonian as a function of λ. Suppose it vanishes at some particular point. This would be a critical point of the theory because the correlation length would diverge there. But Eq. (4.26) states that if the gap vanishes at a certain λ, it must also vanish at λ^{-1}! Therefore, if we assume that the critical point is unique, the self-duality of the model implies that it occurs at

$$\lambda_c = 1\,. \tag{4.27}$$

This result agrees with our earlier observations. In the next section we shall compute the mass gap and find

$$G(\lambda) = 2|1-\lambda|\,, \tag{4.28}$$

in agreement with these considerations.

Note that the self-duality of the model yields the critical point only if one *assumes* that the point is unique. This is a reasonable (and true) assumption for the simple Ising model, but it is not true for more intricate, self-dual theories (Elitzur *et al.*, 1979).

C. Strong coupling expansions for the mass gap, weak coupling expansions for the magnetization

How does one determine the phase diagram and compute critical indices for a statistical mechanics system? There are various methods:

(1) Solve the model exactly.

(2) Renormalization group calculations.

(3) Perturbation expansions.

All of these approaches will be used within this review. But first we shall consider method 3 and discuss strong and weak coupling expansions for the quantum Hamiltonian formulation of the Ising model. From these calculations we shall determine the critical temperature λ, the mass gap critical index ν, and the magnetization index β. Because of the simplicity of the model, exact answers will be found in all cases. It is probably worthwhile to understand these expansion methods since they apply equally simply to more complicated theories where exact solutions and renormalization groups are not available.

We begin with a calculation of the mass gap of the Hamiltonian,

$$H = \sum_n [1-\sigma_1(n)] - \lambda \sum_n \sigma_3(n)\sigma_3(n+1)\,. \tag{4.29}$$

Consider the theory at high temperature, i.e., λ very small. If we write

$$H = H_0 + \lambda V\,, \tag{4.30a}$$

with

$$H_0 = \sum_n [1-\sigma_1(n)]\,,$$

$$V = -\sum_n \sigma_3(n)\sigma_3(n+1)\,, \tag{4.30b}$$

then we are ready to do perturbation expansions in λ. First we must determine the $\lambda = 0$ ground state. To minimize H_0 we must choose the spins at every site such that

$$\sigma_1(n)|0\rangle = |0\rangle \quad \text{(for all } n\text{)}\,. \tag{4.31}$$

It is convenient to refer to $|0\rangle$ as having all spins "up." This terminology is different from that used earlier in connection with Eq. (4.13). In that basis the eigenstates of σ_1 are $\binom{1}{1}$ with eigenvalue +1 and $\binom{1}{-1}$ with eigenvalue −1. Then σ_3 acts as a spin-flip operator within this basis. Therefore we can describe the perturbation V, Eq. (4.30b), as an operator that flips the spins on adjacent sites.

To calculate the mass gap we must obtain the expansion for the ground-state energy and the first excited state. In the $\lambda = 0$ limit, the first excited state above $|0\rangle$ consists of just one flipped spin. This state is N-fold degenerate (N is the number of sites of the lattice) because the flipped spin could occur anywhere on the spatial lattice. However, this degeneracy is resolved by constructing states which have definite momentum. Consider the zero-momentum state

$$|-1\rangle = (1/\sqrt{N})\sum_n \sigma_3(n)|0\rangle\,, \tag{4.32}$$

whose energy is the mass gap. The notation $|-1\rangle$ indicates that σ_3 has flipped one spin down. The state equation (4.32) is properly normalized, i.e., $\langle -1|-1\rangle = 1$.

We want to calculate the λ expansion for the mass gap to sufficiently high orders that we can sensibly discuss

values of λ near unity where the phase transition is expected. So, we need the formulas of Raleigh–Schrödinger perturbation theory to high order. They are easily obtained from the elegant Wigner–Brillouin formulas (Baym, 1969). The expansion for the energy of a state $|a\rangle$ whose zeroth-order energy is $\varepsilon_0 (H_0|a\rangle = \varepsilon_0|a\rangle)$ is

$$E_a = \varepsilon_0 + \varepsilon_1\lambda + \varepsilon_2\lambda^2 + \varepsilon_3\lambda^3 + \varepsilon_4\lambda^4 + \cdots , \tag{4.33a}$$

where

$$\varepsilon_1 = \langle a|V|a\rangle ,$$

$$\varepsilon_2 = \langle a|VgV|a\rangle ,$$

$$\varepsilon_3 = \langle a|VgVgV|a\rangle - \langle a|V|a\rangle\langle a|Vg^2V|a\rangle ,$$

$$\varepsilon_4 = \langle a|VgVgVgV|a\rangle - \langle a|VgV|a\rangle\langle a|Vg^2V|a\rangle + \langle a|V|a\rangle\langle a|V|a\rangle\langle a|Vg^3V|a\rangle - \langle a|V|a\rangle\langle a|VgVg^2V + Vg^2VgV|a\rangle \tag{4.33b}$$

and g is the resolvent

$$g = (1 - |a\rangle\langle a|)/(\varepsilon_0 - H_0) . \tag{4.33c}$$

It is easy to apply these formulas and determine the mass gap. It is helpful to represent the terms of the perturbation expansion pictorially. Let a single vertical line (Fig. 10) represent a flipped spin at a certain site. The potential V can act on the flipped spin and move it one lattice spacing to the left or right. Alternatively, V can act on the $\lambda = 0$ vacuum and flip two nearest-neighbor spins over. These effects are shown in Fig. 11. These figures should be read vertically as a guide to Eq. (4.33b). For example, Fig. 11(a) shows a contribution to $\langle a|V|a\rangle$: the flipped spin is initially on site n, then V, which is represented by a horizontal link, acts and flips that spin down while raising a nearest neighbor. The spin-up nearest neighbor is present in the final state. Figure 11(c) shows a spin configuration which contributes to the state $V|a\rangle$.

Let us compute the gap through λ^3. The zeroth-order energy is just $\varepsilon_0 = 2$. This is so because when one spin is flipped, one term in H_0 is increased by two units. The first-order coefficient ε_1 is -2. It is minus because of the minus sign in the expression for V, and the 2 records the presence of two graphs, Fig. 11(a) and 11(b). To obtain this result mechanically, evaluate the matrix element

$$\langle -1|V|-1\rangle = -(1/N)\sum_{n,n'}\langle 0|\sigma_3(n) \times \sum_m \sigma_3(m)\sigma_3(m+1)\sigma_3(n')|0\rangle . \tag{4.34}$$

There are two contributions: $n = m$ and $m+1 = n'$ or $n' = m$ and $m+1 = n$. Since $\sigma_3^2 = 1$, each term gives a factor N, canceling the normalization N^{-1}. Other terms in the sums contribute nothing, since in those cases the state $\sigma_3(n)\sigma_3(m)\sigma_3(m+1)\sigma_3(n')|0\rangle$ contains some flipped spins, so its projection onto $|0\rangle$ vanishes.

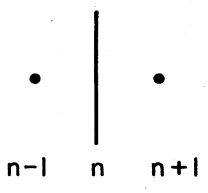


FIG. 10. A flipped spin on site n.

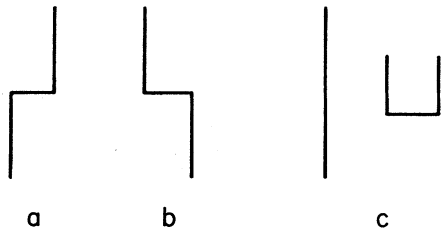


FIG. 11. Two possible actions of the potential V on a state initially containing a flipped spin.

Second-order graphs are shown in Fig. 12. In these cases there is a sum over intermediate states. The energy denominators are $(2-6)^{-1}$, since 2 is the energy of the initial state and 6 is the energy of the intermediate state (three spin flips, each producing two units of energy). The intermediate state is labeled by the dashed horizontal line in the figure. The second-order results organized by figure number are

$$\text{Fig. 12(a)} = 2\left(\frac{1}{2-6}\right)\lambda^2 ,$$

$$\text{Fig. 12(b)} = (N-2)\left(\frac{1}{2-6}\right)\lambda^2 . \tag{4.35}$$

The third-order contributions are obtained similarly. There are contributions from the $\langle a|VgVgV|a\rangle$ piece of Eq. (4.33b) shown in Fig. 13, as well as the $-\langle a|V|a\rangle\langle a|Vg^2V|a\rangle$ piece which is easily obtained from the lower-order calculations. Collecting these results

$$\text{Fig. 13(a)} = -2\left(\frac{1}{4\cdot 4}\right)\lambda^3 ,$$

$$\text{Fig. 13(b)} = -2\left(\frac{1}{4\cdot 4}\right)\lambda^3 ,$$

$$\text{Fig. 13(c)} = -2\left(\frac{1}{4\cdot 4}\right)\lambda^3 ,$$

$$\text{Fig. 13(d)} = -2(N-3)\left(\frac{1}{4\cdot 4}\right)\lambda^3 ,$$

$$-\langle a|V|a\rangle\langle a|Vg^2V|a\rangle = -(-2\lambda)\left[\frac{2}{16} + \frac{(N-2)}{16}\right]\lambda^2 .$$

To finally obtain the mass gap we must calculate the shift in the vacuum energy through this order in λ and take the difference of the two series. The zeroth-order

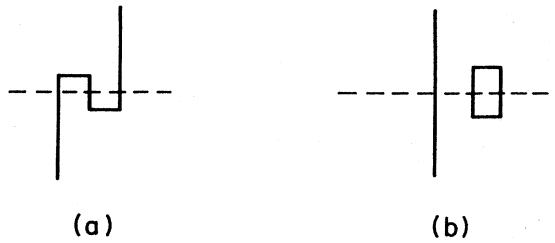


FIG. 12. Second-order contributions to the energy of $|-1\rangle$. The dashed line denotes an intermediate state.

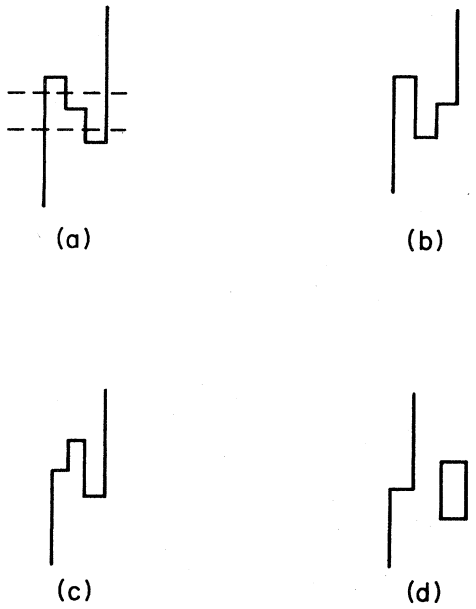


FIG. 13. Third-order contributions to the energy of $|-1>$. The dashed lines labeling intermediate states are shown only for the first graph.

energy is zero—the constant term $\sum_n \cdot 1$ was added to the Hamiltonian to achieve this. The first-order correction is also zero, since $V|0\rangle$ does not project onto $|0\rangle$. Second-order effects are shown in Fig. 14,

$$\text{Fig. (14)} = -(N/4)\lambda^2 . \tag{4.37}$$

Third-order effects vanish identically.

Collecting these results we find the series for the gap

$$G(\lambda) = (2-2)\lambda + (0)\lambda^2 + (0)\lambda^3 + \cdots , \tag{4.38}$$

where the second- and third-order coefficients have vanished identically! Higher-order calculations confirm the obvious suspicion—the series truncates after the first term. So, we find an *exact* result

$$G(\lambda) = 2(1-\lambda) , \tag{4.39}$$

for $\lambda < 1$. Note that $G(\lambda)$ vanishes at $\lambda_c = 1$. This proves that the Ising model has just two phases separated by a continuous phase transition. We also learn from Eq. (4.39) that the critical exponent ν is unity

$$\nu = 1 , \tag{4.40}$$

which is another exact result. The fact that we have obtained the correct answer (see Table I) constitutes an example of universality—the singularities in the critical region of the model are independent of its detailed lattice structure.

Since the gap must satisfy the duality condition Eq. (4.26)

$$G(\lambda) = \lambda G(\lambda^{-1}) , \tag{4.41}$$

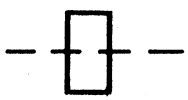


FIG. 14. Second-order contribution to the vacuum energy.

the resulting equation (4.39), which was derived only for $\lambda < 1$, can be extended to all λ,

$$G(\lambda) = 2|1-\lambda| , \tag{4.42}$$

as was recorded earlier.

Now we turn to the magnetization. In this case we organize the calculation around zero temperature, $1/\lambda = 0$, where the system is definitely magnetized, and develop a perturbation series in powers of $1/\lambda$. To begin recall a simple fact about perturbation theory. Suppose we want to calculate the matrix element of an operator θ perturbatively. Then if we modify the Hamiltonian

$$H \to H + h\theta , \tag{4.43a}$$

where h is a parameter, the matrix element of θ in the state $|\psi\rangle$, an eigenstate of H with energy E, is

$$\langle\psi|\theta|\psi\rangle = \frac{\partial}{\partial h} E(h)\Big|_{h=0} . \tag{4.43b}$$

This equation is just the familiar statement of perturbation theory in the term $h\theta$ about the Hamiltonian H. The useful feature of the formula is that in some applications it is convenient to calculate $E(h)$ directly and take its first derivative at $h = 0$ afterwards.

To develop perturbation series about $1/\lambda = 0$, consider the operator

$$(1/\lambda)H = -\sum_n \sigma_3(n)\sigma_3(n+1) - (1/\lambda)\sum_n \sigma_1(n) . \tag{4.44}$$

It is more convenient to add constants and use the operator W

$$W = \sum_n [1 - \sigma_3(n)\sigma_3(n+1)] - (1/\lambda)\sum_n \sigma_1(n) . \tag{4.45}$$

Finally, choose the operator

$$\theta = \sum_n \sigma_3(n) , \tag{4.46}$$

and incorporate it into W,

$$W \to W + h\sum_n \sigma_3(n) . \tag{4.47}$$

To do perturbation theory define the operators W_0 and V

$$W = W_0 + (1/\lambda)V , \tag{4.48a}$$

where

$$W_0 = \sum_n [1 - \sigma_3(n)\sigma_3(n+1)] + h\sum_n \sigma_3(n) ,$$

$$V = \sum_n \sigma_1(n) . \tag{4.48b}$$

To begin we must determine the ground state of W_0. It would be doubly degenerate, having either all spins "up,"

$$\sigma_3(n)|0\rangle = |0\rangle \quad (\text{all } n) , \tag{4.49a}$$

or all spins "down,"

$$\sigma_3(n)|0\rangle = -|0\rangle \quad (\text{all } n) , \tag{4.49b}$$

if h were set to zero. (The terminology for "up" and

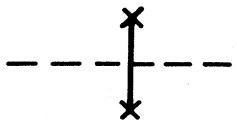


FIG. 15. Second-order contribution to the $1/\lambda$ expansion of the ground-state energy.

"down" now refers to the eigenstates of σ_3. The operator σ_1 acts as a spin-flip operator, as in the discussion of Eq. (4.13).) However, the operator $h\sum_n \sigma_3(n)$ breaks the global up $\rightleftharpoons$ down symmetry and picks out one of the two degenerate ground states. Choosing h to be small and negative singles out the all-spins-"up" state as the true ground state around which we shall do perturbation theory. The perturbation V now flips spins in this state. As we calculate to higher and higher orders in $1/\lambda$ we shall encounter intermediate states with more and more spins turned over.

Since we want the ground-state expectation value of the magnetization, we must calculate $E_0(h)$, the energy of that state. In zeroth order we have hN, where N is the number of lattice sites. There is no first-order effect because $\langle 0|V|0\rangle$ vanishes identically. However, in second order we find contributions in which V acts on $|0\rangle$ flipping a spin over. If the second application of V in the matrix element $\langle 0|VgV|0\rangle$ flips that spin back to its original position, a nonzero contribution results. It is shown pictorially in Fig. 15, and gives a contribution

$$\text{Fig. (15)} = [N/(-4+2h)](1/\lambda)^2 . \tag{4.50}$$

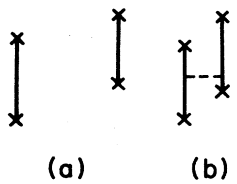
The crosses in the figure denote the action of the perturbation V and the vertical line labels the flipped spin. The energy of the intermediate state is

$$4 + (N-1)h - h . \tag{4.51a}$$

The four accounts for the fact that flipping one spin breaks two bonds in W_0 and each broken bond costs two units of energy. The $(N-2)h$ is just the energy from the external magnetic field. The energy denominator is then

$$Nh - [4 + (N-2)h] = -4 + 2h , \tag{4.51b}$$

as stated in Eq. (4.50). There are no third-order contributions. In fourth order there are two distinct types of contributions. The two flipped spins may or may not be nearest neighbors. If they are not nearest neighbors (Fig. 16a) four bonds are broken, otherwise only two bonds are broken (Fig. 16b). Thus the energy denominators must be calculated separately. One must be sure to sum over the different orderings of the vertices. These simple computations give



(a) (b)

FIG. 16. Fourth-order contributions to $E_0(h)$. The horizontal dotted line means that the two flipped spins are nearest neighbors. The dashed lines labeling intermediate states have been omitted.

$$\text{Fig. 16(a)} = \frac{2N(N-3)}{(-4+2h)^2(-8+4h)}(1/\lambda)^4 , \tag{4.52a}$$

$$\text{Fig. 16(b)} = \frac{4N}{(-4+2h)^2(-4+4h)}(1/\lambda)^4 . \tag{4.52b}$$

And finally there is the term

$$-\langle a|VgV|a\rangle\langle a|Vg^2V|a\rangle = -\frac{N^2}{(-4+2h)^3}(1/\lambda)^4 \tag{4.53}$$

which contributes to the fourth-order coefficient ε_4 of Eq. (4.33b). The only effect of this last term is to subtract off the N^2 term of Eq. (4.52) and render the vacuum energy extensive, i.e., proportional to N.

Collecting these results, we have

$$E_0(h)/N = h + \frac{1}{(-4+2h)}(1/\lambda)^2 + \frac{1}{(-4+2h)^2}\left[\frac{6}{(-8+4h)} + \frac{4}{(-4+4h)}\right](1/\lambda)^4 + \cdots \tag{4.54}$$

and the magnetization series becomes

$$M = \frac{1}{N}\frac{\partial E_0}{\partial h}\bigg|_{h=0} = 1 - \frac{1}{8}(1/\lambda)^2 - \frac{7}{128}(1/\lambda)^4 - \cdots . \tag{4.55}$$

Since this series does not terminate, we need a procedure to test whether it is "tending to vanish" at a finite value of $1/\lambda$. According to earlier discussions, we expect that in the critical region the magnetization might vanish as a power of $(\lambda - \lambda_c)$. We can test the compatibility of this hypothesis with our series for M by applying the ratio test (Stanley, 1971). The general idea of the method is the following. Suppose we have a series

$$f(x) = \sum_n a_n x^n , \tag{4.56a}$$

for a function $f(x)$ which develops a power singularity in the vicinity of a critical point x_c.

$$f(x) \underset{x\to x_c}{\sim} b(x_c - x)^{-\gamma} = b x_c^{-\gamma}(1 - x/x_c)^{-\gamma}, \quad x < x_c . \tag{4.56b}$$

Using the binomial theorem, the Taylor series for this power-behaved function is

$$f(x) \sim b x_c^{-\gamma}\left\{1 + \gamma\left(\frac{x}{x_c}\right) + \frac{\gamma(\gamma+1)}{2!}\left(\frac{x}{x_c}\right)^2 + \cdots \frac{\gamma(\gamma+1)\cdots(\gamma+l-1)}{l!}\left(\frac{x}{x_c}\right)^l + \cdots\right\} . \tag{4.56c}$$

The power behavior shows up clearly in the ratio of successive coefficients

$$R_l = a_l/a_{l-1} = (1/x_c)[1 + (\gamma - 1/l)] . \tag{4.57}$$

Therefore a plot of R_l vs $1/l$ would produce a straight line with an intercept x_c^{-1} and a slope $\gamma - 1$. In this way an estimate of the critical point and the critical index γ can be made. One can also consider the "linear extrapolant"

$$E_l = lR_l - (l-1)R_{l-1} , \tag{4.58}$$

which equals x_c^{-1} identically for a pure power. In addition, the slopes

$$S_l = (R_l - R_{l-1})/[1/l - 1(l-1)] , \tag{4.59}$$

equal $x_c^{-1}(\gamma-1)$ for a pure power.

Table II shows an application of the ratio test to the magnetization series. Six terms in its series [three beyond our illustrative calculation Eq. (4.55)] are recorded there (Hamer and Kogut, 1979). We see that the power-law hypothesis is exact! E_l are all exactly unity, giving $\lambda_c=1$ as always. In addition,

$$S_l=-1.125 \quad (\text{all } l), \tag{4.60}$$

so, if we use standard terminology

$$M \underset{x\to x_c}{\sim} (1-\lambda^{-2})^{\beta}, \tag{4.61}$$

we have

$$-\beta-1=-1.125, \tag{4.62a}$$

which implies

$$\beta=1/8. \tag{4.62b}$$

So, we have the exact result

$$M=(1-\lambda^{-2})^{1/8} \tag{4.63}$$

for all temperatures below T_c (Pfeuty, 1970).

Of course, expansion methods do not usually yield exact results. It just happens that in these examples the exact results are simple functions of the expansion parameters λ and λ^{-1}, so we could reconstruct precise answers. Applying the same calculational methods to the susceptibility and using the ratio test, one obtains an estimate of the critical index γ (Hamer and Kogut, 1979),

$$\gamma=1.76\pm.03 \tag{4.64}$$

to be compared with the exact result 1.75. Similarly the specific heat exponent α is found to be

$$\alpha=0.01\pm.02, \tag{4.65}$$

which compares favorably with the exact result of zero. It should be remarked that the quantities E_l and S_l are not very useful in these approximate calculations, since subdominant power-behaved singularities make them rather irregular functions of l (Fisher, 1967).

It is hoped that the "dirty" calculations presented here familiarize the reader with the "nuts and bolts" of spin systems.

D. Kink condensation and disorder

The magnetization M serves as a local order parameter for the Ising model. It is nonzero at low temperature, vanishes continuously as T passes through T_c, and remains identically zero for all higher temperatures. The magnetization indicates that the low-temperature phase spontaneously breaks the up $\rightleftharpoons$ down global symmetry of the system's Action.

TABLE II. Series analysis for the magnetization of the Ising model.

l	a_l	R_l	E_l	S_l
0	1			
1	−0.125	−0.125		
2	−0.054 687 5	0.437 5	1.0 (exact)	−1.125 (exact)
3	−0.034 179 69	0.625	1.0	−1.125
4	−0.024 566 65	0.718 75	1.0	−1.125
5	−0.019 039 15	0.775	1.0	−1.125

The self-duality of the model allows us to view the magnetization and order from another interesting perspective (Fradkin and Susskind, 1978). Consider the Hamiltonian

$$(1/\lambda)H+h\sum_n \sigma_3(n), \tag{4.66}$$

that we used in the computation of the magnetization. Applying the duality transformation equation (4.19), we obtain

$$(1/\lambda)H(\sigma;\lambda)+h\sum_n \sigma_3(n)=H(\mu;\lambda^{-1})+h\sum_n\prod_{m<n}\mu_1(m), \tag{4.67}$$

where we used the fact that $\mu_1(n)=\sigma_3(n+1)\sigma_3(n)$ to find that

$$\sigma_3(n)=\prod_{m<n}\mu_1(n). \tag{4.68}$$

But first-order perturbation theory in h now implies that

$$\langle 0|\sum_n \sigma_3(n)|0\rangle_\lambda=\langle 0|\sum_n\prod_{m<n}\mu_1(m)|0\rangle_{\lambda^{-1}}, \tag{4.69}$$

where $|0\rangle_\lambda$ is the vacuum of $H(\sigma;\lambda)$ and $|0\rangle_{\lambda^{-1}}$ is the vacuum of $H(\mu;\lambda^{-1})$. Since σ_1 and σ_3 are equivalent to μ_1 and μ_3, Eq. (4.69) tells us that the operator

$$\prod_{m<n}\sigma_1(n), \tag{4.70}$$

has a nonvanishing expectation value in the high-temperature phase of the original model. It is conventional to call this operator a "disorder" parameter (Kadanoff and Ceva, 1971). It has an interesting physical interpretation. Suppose it acts on a completely ordered state having all spins "up," $\sigma_3(n)|\text{state}\rangle=|\text{state}\rangle$, for all sites. Then it flips all the spins to the left of site n. It creates a "kink." This observation exposes why the expectation value of the disorder operator vanishes in the low-temperature phase.

Since the energy of the kink state is localized in the vicinity of site n, one can think of it as a particle excitation. A zero-momentum kink state would be

$$|1 \text{ kink}\rangle=\frac{1}{\sqrt{N}}\sum_n\prod_{m<n}\sigma_1 n)|0\rangle_{\lambda^{-1}=0}. \tag{4.71}$$

It follows from the duality relation [Eq. (4.67)] that the computation of the kink's mass in powers of λ^{-1} is identical to that of the flipped spin states' mass in powers of λ. Therefore the mass gap formula $G(\lambda)=2|1-\lambda|$ applies to the kink in the low-temperature phase.

Since the vacuum expectation value of the disorder parameter is nonzero in the high-temperature phase, the ground state of that phase is a "kink condensate." The short-range character of the spin–spin correlation function can be understood intuitively in these terms. Since the disorder operator flips spins over an infinite region of space, a condensate of kinks random-

izes the spin variables and leads to a finite correlation length. It is also clear that kinks have a *topological* significance. Applying the kink operator to the low-temperature ground state produces a state which interpolates between the two degenerate ground states of the low-temperature phase. Kinks alter the boundary conditions of the system—by inspecting the spins at the extremities of the spatial lattice one can determine whether there are an even or odd number of kinks in the system.

The concepts of topological disorder and kink condensation will become more significant as we continue.

E. Self-duality of the isotropic Ising model

It is interesting to use the partition function of the Ising model to derive its self-duality. We shall do it here for a square lattice and obtain the exact critical temperature. The manipulations we shall make also constitute good preparation for analyses we shall do later on lattice gauge theories.

Our strategy will consist of developing both high-temperature and low-temperature expansions of the partition function Z. These calculations will expose an exact mapping between the high-T and low-T properties of the model.

Consider the partition function

$$Z(K)=\sum_{\sigma_N=\pm1}\cdots\sum_{\sigma_1=\pm1}\exp\left\{K\sum_{\langle ij\rangle}\sigma_i\sigma_j\right\}, \tag{4.72}$$

where $K=J/kT$, the notation $\langle ij\rangle$ indicates a sum over nearest-neighbor sites i and j, and N is the number of lattice sites. If K is very small, $Z(K)$ can be expanded in a power series. First observe that if σ is a generic Ising variable ($\sigma=\pm1$), then

$$\exp(K\sigma)=\cosh K+\sigma\sinh K\,. \tag{4.73}$$

This is a very useful identity. Applying it to Eq. (4.72) gives

$$\begin{aligned}Z(K)&=\sum_{\sigma_N=\pm1}\cdots\sum_{\sigma_1=\pm1}\prod_{\langle ij\rangle}(\cosh K+\sigma_i\sigma_j\sinh K)\\&=(\cosh K)^{2N}\sum_{\sigma_N}\cdots\sum_{\sigma_1}\prod_{\langle ij\rangle}(1+\sigma_i\sigma_j\tanh K)\,.\end{aligned} \tag{4.74}$$

The high-temperature expansion now consists of using Eq. (4.74) to develop a power series for $Z(K)(\cosh K)^{-2N}$ in the variable $\tanh K$. To identify nonzero contributions, note the trivial identities

$$\sum_{\sigma=\pm1}\sigma=0\,,\quad\sum_{\sigma=\pm1}1=2\,. \tag{4.75}$$

Therefore, as we expand $\Pi(1+\sigma_i\sigma_j\tanh K)$, only those terms without a free spin variable will contribute. The product $\Pi_{\langle ij\rangle}$ indicates that we place "bonds," $\tanh K\sigma_i\sigma_j$, on the links of a square lattice. Only if the bonds form closed paths will the term contribute (only then will all the σ_i's occur squared). In addition, no link can be covered more than once. Systematic rules can be developed to allow very-high-order calculations (Wortis, 1972). The first several terms are shown pictorially in Fig. 17, and the expression for $Z(K)$ is

$$\begin{aligned}Z(K)(\cosh K)^{-2N}\cdot2^{-N}&=1+N(\tanh K)^4+2N(\tanh K)^6\\&\quad+\tfrac{1}{2}N(N-5)(\tanh K)^8+\cdots\,.\end{aligned} \tag{4.76}$$

The factors of N record the number of ways the graph can occur on the lattice.

Next consider the low-temperature expansion of $Z(K)$. Since all the spins are aligned at zero temperature, this will be an expansion in the number of flipped spins. Choose the $T=0$ ground-state spin configuration to be all "up." When one spin is flipped, four bonds are broken. If two spins are flipped, then six bonds are broken if those spins are nearest neighbors and eight bonds are broken otherwise. The first several terms in the low-temperature expansion are visualized in Fig. 18, and the expression for $Z(K)$ is

$$Z(K)e^{-2NK}=1+Ne^{-8K}+2Ne^{-12K}+\tfrac{1}{2}N(N-5)e^{-16K}+\cdots\,. \tag{4.77}$$

The correspondence between the high- and low-T expansions should be clear. If the flipped spins of the low-temperature expansion are surrounded by a square, the graphical rules of the low-T expansion map onto those of the high-T expansion! To make the relationship precise, define a coupling K^* such that

$$\tanh K=\exp(-2K^*)\,. \tag{4.78}$$

Then comparing Eqs. (4.77) and (4.76) gives

$$\frac{Z(K^*)}{(e^{2K^*})^N}=\frac{Z(K)}{2^N(\cosh^2K)^N}\,. \tag{4.79}$$

Note that Eq. (4.78) defines a mapping which takes large K into small K^* and vice versa. There is more symmetry in these expressions than meets the eye. Manipulating hyperbolic functions shows that Eq. (4.78) can be written

$$\sinh(2K)\sinh(2K^*)=1\,, \tag{4.80}$$

and Eq. (4.79) can be written

$$\frac{Z(K^*)}{\sinh^{N/2}(2K^*)}=\frac{Z(K)}{\sinh^{N/2}(2K)}\,. \tag{4.81}$$

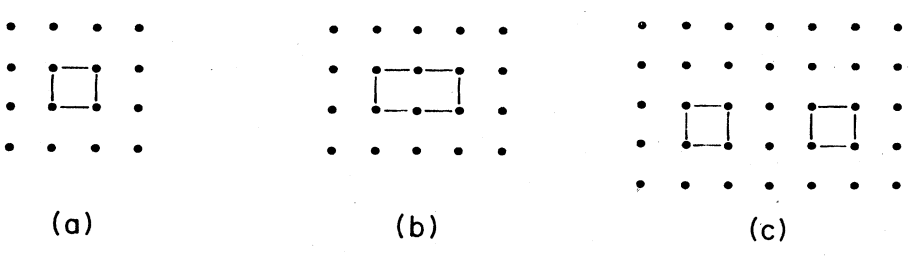


FIG. 17. Low-order contributions to the high-temperature expansion of $Z(K)$.

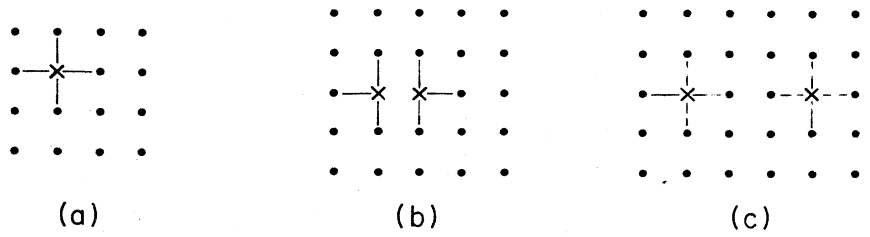


FIG. 18. Low-temperature expansion graphs for $Z(K)$. A cross indicates a flipped spin.

So, these equations define a duality mapping which exposes a special symmetry between the model's high- and low-temperature properties. We can argue again that if the model has a unique critical point it must be at a temperature such that $K = K^*$. Equation (4.80) then yields the critical value

$$\sinh^2(2K_c) = 1 \tag{4.82a}$$

or

$$e^{2K_c} = \sqrt{2} + 1 . \tag{4.82b}$$

This critical point was found long before the Ising model was solved (Kramers and Wannier, 1941). Note that Eq. (4.82a) is a special case of the critical curve [Eq. (4.15)] cited earlier. Using the same methods employed here and doing a bit more bookkeeping, the reader can obtain that result.

The self-duality of the Ising model can also be obtained without the aid of expansion methods. But the expansions are of considerable interest for practical calculations as well as rigorous analyses. One can prove, for example, that the high- and low-temperature expansions of the Ising model in more that one dimension have finite radii of convergence (Domb, 1972). Therefore the expansion alone teach us that at sufficiently low but nonzero T the system magnetizes in more than one dimension, while at sufficiently high but finite T the system is disordered and the magnetization vanishes. Proofs that high-temperature expansions have finite radii of convergence can be given for a wide class of models. Analysis of this sort is usually the first step one takes in trying to determine the phase diagram of an unfamiliar physical system. We shall do such analyses for lattice gauge theories later in this review.

F. Exact solution of the Ising model in two dimensions

Finally, we want to solve the τ-continuum formulation of the Ising model exactly. The point in doing this, besides the illustration of interesting technical topics, is to show that at the critical point the lattice theory becomes a relativistic field theory: the spatial lattice becomes irrelevant and the desired continuous space–time symmetries are restored. We shall see that the model becomes that of a free, massless, self-charge conjugate fermion (Pfeuty, 1970). The explicit solution illustrates the general strategy of using lattice formulations of quantum field theory: one first determines the theory's phase diagram and critical points, and then by approaching the critical points one constructs appropriately relativistic field theories.

Consider the Hamiltonian

$$H = -\sum_n \sigma_3(n) - \lambda \sum_n \sigma_1(n)\sigma_1(n+1) . \tag{4.83}$$

which is equivalent to our previous expressions such as Eq. (4.14). There is a standard construction to expose the fermion hiding here. Recall the Jordan–Wigner transformation (Jordan and Wigner, 1928) which accomplishes this. First define raising and lowering operators

$$\sigma^+(n) = \tfrac{1}{2}[\sigma_1(n) + i\sigma_2(n)] , \qquad \sigma^-(n) = \tfrac{1}{2}[\sigma_1(n) - i\sigma_2(n)] . \tag{4.84}$$

Label the spatial lattice sites $n = -N, -N+1, \ldots, N$. Then fermion operators $c(n)$ can be constructed from the Pauli matrices

$$c(n) = \prod_{j=-N}^{n-1} \exp[i\pi\sigma^+(j)\sigma^-(j)]\sigma^-(n)$$
$$c^\dagger(n) = \sigma^+(n) \prod_{j=-N}^{n-1} \exp[-i\pi\sigma^+(j)\sigma^-(j)] . \tag{4.85}$$

The idea behind this construction is the following: The Pauli matrices σ^+ and σ^- anticommute on the same site. Their squares are zero. In this sense they resemble fermions. However, fermion operators anticommute even when they have different spatial arguments. The "strings" of operators in Eq. (4.85) are inserted to insure this.

To check the claim that $c(n)$ and $c^\dagger(n)$ are fermion operators, it is best first to simplify the Pauli spin algebra in Eq. (4.85). Observe that

$$\sigma^-(n)\sigma^+(n) = \tfrac{1}{2}[1 - \sigma_3(n)] ,$$
$$\sigma^+(n)\sigma^-(n) = \tfrac{1}{2}[1 + \sigma_3(n)] \tag{4.86}$$
$$\exp[i(\pi/2)\sigma_3] = i\sigma_3 .$$

These identities allow us to write

$$c(n) = \prod_{j=-N}^{n-1} [-\sigma_3(j)]\sigma^-(n) ,$$
$$c^\dagger(n) = \sigma^+(n) \prod_{j=-N}^{n-1} [-\sigma_3(j)] . \tag{4.87}$$

It is now easy to verify that the operators $c(n)$ and $c^\dagger(n)$ satisfy fermion anticommutation relations,

$$\{c(n), c^\dagger(m)\} = \delta_{nm} \quad \{c(n), c(m)\} = 0 . \tag{4.88}$$

For example, using Eqs. (4.86) and (4.87), we see that

$$c(n)c^\dagger(n) + c^\dagger(n)c(n) = \sigma^-(n)\sigma^+(n) + \sigma^+(n)\sigma^-(n) = 1 . \tag{4.89}$$

In addition, choose $m < n$ and compute

$$c(m)c^\dagger(n)+c^\dagger(n)c(m)=\sigma^-(m)\prod_{j=m}^{n-1}[-\sigma_3(j)]\sigma^+(n)+\sigma^+(n)\prod_{j=m}^{n-1}[-\sigma_3(j)]\sigma^-(m). \tag{4.90}$$

Using the fact that

$$\sigma^-(m)\sigma_3(m)=-\sigma_3(m)\sigma^-(m), \tag{4.91}$$

we see that the right-hand side of Eq. (4.90) vanishes identically. The remaining parts of Eq. (4.88) are verified similarly.

Our next task is to write H in terms of the fermion operators. Using Eq. (4.86) we have

$$\sigma_3(n)=2\sigma^+(n)\sigma^-(n)-1=2c^\dagger(n)c(n)-1. \tag{4.92}$$

The coupling terms require a bit more work. Begin with

$$\sigma_1(n)\sigma_1(n+1)=[\sigma^+(n)+\sigma^-(n)][\sigma^+(n+1)+\sigma^-(n+1)]. \tag{4.93}$$

Consider the product

$$c^\dagger(n)c(n+1)=\sigma^+(n)[-\sigma_3(n)]\sigma^-(n+1). \tag{4.94}$$

But

$$\sigma^+(n)\sigma_3(n)=-\sigma^+(n), \tag{4.95}$$

so

$$c^\dagger(n)c(n+1)=\sigma^+(n)\sigma^-(n+1). \tag{4.96}$$

It is clear now that the coupling term can be written in terms of fermion operators which are also only coupled if they are nearest neighbors. The fact that H can be expressed in terms of *local* products of the fermion operators makes this change of variables useful. Note that the Ising character of the original degrees of freedom, $\sigma^2(n)=1$, was essential in leading to this result. Other products of fermion operators are computed as easily:

$$\begin{aligned} c(n)c^\dagger(n+1)&=-\sigma^-(n)\sigma^+(n+1),\\ c^\dagger(n)c^\dagger(n+1)&=\sigma^+(n)\sigma^+(n+1),\\ c(n)c(n+1)&=-\sigma^-(n)\sigma^-(n+1). \end{aligned} \tag{4.97}$$

Collecting all this we have

$$\sigma_1(n)\sigma_1(n+1)=[c^\dagger(n)-c(n)][c^\dagger(n+1)+c(n+1)]. \tag{4.98}$$

The Hamiltonian becomes

$$H=-2\sum_n c^\dagger(n)c(n)-\lambda\sum_n[c^\dagger(n)-c(n)][c^\dagger(n+1)+c(n+1)]. \tag{4.99}$$

The solubility of the model is now clear, since H is only a quadratic form. It is convenient to write H in momentum space, so define operators

$$a_k=\left(\frac{1}{2N+1}\right)^{1/2}\sum_{n=-N}^{N}e^{ikn}c(n), \tag{4.100}$$

where the complete set of wave vectors is

$$k=0,\pm\frac{2\pi}{2N+1},\pm\frac{4\pi}{2N+1},\ldots,\pm\frac{2\pi N}{2N+1}. \tag{4.101}$$

This choice implies a certain treatment of boundary conditions (Schultz *et. al.*, 1964). Since we shall only be interested in the energy–momentum relation and the mass gap of the model, we need not be careful with such points. We shall frequently discard constants as we simplify H. The boundary conditions are not without interest, however, and the reader is referred to the literature for a discussion (Schultz *et. al.*, 1964). Anyway, it is easy to verify that the operators a_k and $a_k^\dagger$ have fermion anticommutation relations

$$\begin{aligned} \{a_{k'}^\dagger, a_k\}&=\delta_{k',k},\\ \{a_{k'}, a_k\}&=\{a_{k'}^\dagger, a_k^\dagger\}=0. \end{aligned} \tag{4.102}$$

Simply use the completeness relations

$$\begin{aligned} \frac{1}{2N+1}\sum_k e^{-ik(n-m)}&=\delta_{n,m},\\ \frac{1}{2N+1}\sum_n e^{in(k-l)}&=\delta_{k,l}, \end{aligned} \tag{4.103}$$

and verify by explicit calculation that the fermion anticommutation relations for $c(n)$ and $c^\dagger(n)$ imply Eq. (4.102).

Next we must write H in terms of the a_k and $a_k^\dagger$. Since Eq. (4.100) can be inverted,

$$c(n)=\left(\frac{1}{2N+1}\right)^{1/2}\sum_k e^{-ikn}a_k, \tag{4.104}$$

the computations are straightforward. For example,

$$\begin{aligned} \sum_n c^\dagger(n)c^\dagger(n+1)&=\frac{1}{2N+1}\sum_n\sum_k\sum_{k'}e^{ikn}e^{ik'(n+1)}a_k^\dagger a_{k'}^\dagger,\\ &=\sum_k\sum_{k'}e^{ik'}\delta_{k,-k'}a_k^\dagger a_{k'}^\dagger,\\ &=\sum_k e^{-ik}a_k^\dagger a_{-k}^\dagger. \end{aligned} \tag{4.105}$$

Collecting similar results for the other terms in Eq. (4.99) gives

$$\begin{aligned} H&=-2\sum_k a_k^\dagger a_k-\lambda\sum_k(e^{-ik}a_k^\dagger a_{-k}^\dagger+e^{-ik}a_k^\dagger a_k+e^{ik}a_k^\dagger a_k-e^{ik}a_k a_{-k})\\ &=-2\sum_k(1+\lambda\cos k)a_k^\dagger a_k-\lambda\sum_k(e^{-ik}a_k^\dagger a_{-k}^\dagger-e^{ik}a_k a_{-k}). \end{aligned} \tag{4.106}$$

It will prove more convenient to write the sum over modes just over those with $k>0$ and include the others by simply writing them out. Then Eq. (4.106) becomes

$$H=-2\sum_{k>0}(1+\lambda\cos k)(a_k^\dagger a_k+a_{-k}^\dagger a_{-k})+2i\lambda\sum_{k>0}\sin k(a_k^\dagger a_{-k}^\dagger+a_k a_{-k}). \tag{4.107}$$

Note that the vacuum of H is not the vacuum of the operator a_k because of the presence of $(a_k^\dagger a_{-k}^\dagger + a_k a_{-k})$ in Eq. (4.107). We want to write H in the form

$$H = \sum_k \Lambda_k \eta_k^\dagger \eta_k + \text{constant}, \tag{4.108}$$

so we can identify the single-particle excitations above the vacuum $(\eta_k|0\rangle = 0)$. This can be done by making a canonical transformation from the operators $a_k, a_k^\dagger$ to a new set $\eta_k, \eta_k^\dagger$. Define

$$\begin{aligned} \eta_k &= u_k a_k + i v_k a_{-k}^\dagger, \quad \eta_{-k} = u_k a_{-k} - i v_k a_k^\dagger, \\ \eta_k^\dagger &= u_k a_k^\dagger - i v_k a_{-k}, \quad \eta_{-k}^\dagger = u_k a_{-k} + i v_k a_k, \end{aligned} \tag{4.109}$$

where $k > 0$ everywhere. The functions u_k and v_k will be determined by two criteria: (1). The $\eta_k, \eta_k^\dagger$ should be fermion operators. (2). The Hamiltonian should be diagonalized when written in terms of the $\eta_k, \eta_k^\dagger$. The choice of the precise form of Eq. (4.109) is made with some hindsight, i.e., the two criteria can be met if u_k and v_k are real, even functions of k. The appearance of i in Eq. (4.109) also leads to more elegant algebra later. Anyway, it is easy to verify that the first requirement

$$\begin{aligned} &\{\eta_{k'}, \eta_k^\dagger\} = \delta_{k',k}, \\ &\{\eta_{k'}, \eta_k\} = \{\eta_{k'}^\dagger, \eta_k^\dagger\} = 0, \end{aligned} \tag{4.110}$$

leads to a relation between u_k and v_k

$$u_k^2 + v_k^2 = 1. \tag{4.111}$$

To write H in terms of η_k and $\eta_k^\dagger$ first invert Eq. (4.109)

$$\begin{aligned} a_k &= u_k \eta_k - i v_k \eta_{-k}^\dagger, \quad a_{-k} = u_k \eta_{-k} + i v_k \eta_k^\dagger, \\ a_k^\dagger &= u_k \eta_k^\dagger + i v_k \eta_{-k}, \quad a_{-k}^\dagger = u_k \eta_{-k}^\dagger - i v_k \eta_k. \end{aligned} \tag{4.112}$$

Substituting into the Hamiltonian and collecting terms gives

$$\begin{aligned} H = &\sum_{k>0} [-2(1+\lambda\cos k)(u_k^2 - v_k^2) + 4\lambda \sin k u_k v_k] \\ &\times (\eta_k^\dagger \eta_k + \eta_{-k}^\dagger \eta_{-k}) \\ &+ \sum_{k>0} [4i(1+\lambda\cos k) u_k v_k + 2i\lambda \sin k (u_k^2 - v_k^2)] \\ &\times (\eta_k^\dagger \eta_{-k}^\dagger + \eta_k \eta_{-k}). \end{aligned} \tag{4.113}$$

Demanding that H have the form of Eq. (4.108) implies

$$4(1+\lambda\cos k) u_k v_k + 2\lambda \sin k (u_k^2 - v_k^2) = 0. \tag{4.114}$$

In light of Eq. (4.111) it is sensible to parametrize the coefficients with an angle θ_k,

$$u_k = \cos\theta_k, \quad v_k = \sin\theta_k. \tag{4.115}$$

Then

$$2u_k v_k = \sin 2\theta_k, \quad u_k^2 - v_k^2 = \cos 2\theta_k \tag{4.116}$$

so Eq. (4.114) becomes

$$2(1+\lambda\cos k)\sin 2\theta_k + 2\lambda \sin k \cos 2\theta_k = 0 \tag{4.117a}$$

or

$$\tan 2\theta_k = -\frac{\lambda \sin k}{(1+\lambda\cos k)}. \tag{4.117b}$$

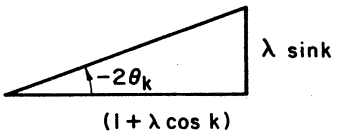


FIG. 19. The parameters of the canonical transformation which diagonalizes the Hamiltonian.

This relation can be represented with a right triangle, sketched in Fig. 19. The hypotenuse is

$$\sqrt{(1+\lambda\cos k)^2 + \lambda^2 \sin^2 k} = \sqrt{1 + 2\lambda\cos k + \lambda^2}, \tag{4.118}$$

and we choose sign conventions so that

$$\begin{aligned} \sin(2\theta_k) &= \frac{\lambda \sin k}{\sqrt{1+2\lambda\cos k + \lambda^2}}, \\ \cos(2\theta_k) &= -\frac{1+\lambda\cos k}{\sqrt{1+2\lambda\cos k + \lambda^2}}. \end{aligned} \tag{4.119}$$

Now we can return to Eq. (4.113) and evaluate the coefficient of the operator $(\eta_k^\dagger \eta_k + \eta_{-k}^\dagger \eta_{-k})$. Some algebra yields

$$H = 2\sum_k \sqrt{1+2\lambda\cos k + \lambda^2}\, \eta_k^\dagger \eta_k + \text{const}, \tag{4.120}$$

so

$$\Lambda_k = 2\sqrt{1+2\lambda\cos k + \lambda^2}, \tag{4.121}$$

which is plotted in Fig. 20. Note that the minimum value of Λ_k is obtained at $k = \pm\pi$ where

$$\Lambda_{\pm\pi} = 2|1-\lambda|, \tag{4.122}$$

in agreement with our earlier analysis! The critical index $\nu = 1$ as listed in Table I.

Finally, consider the theory in its critical region. To discuss the energy–momentum relation we must restore physical units. Since Λ_k is a minimum at $k = \pm\pi$, we measure momentum from π. Let

$$k = \pi + k'a, \tag{4.123a}$$

so that k' has dimensions of momentum. Also, define

$$E(k') = \Lambda_k / 2a \tag{4.123b}$$

so that the energy has correct dimensions. Since we want to consider finite values of k' as the lattice spacing $a \to 0$, k is forced to π. Then Λ_k simplifies

$$\tfrac{1}{2}\Lambda_k = \sqrt{1+2\lambda\cos(\pi + k'a) + \lambda^2} \approx (1-\lambda)^2 + \lambda(k'a)^2. \tag{4.124}$$

Therefore the energy–momentum relation becomes

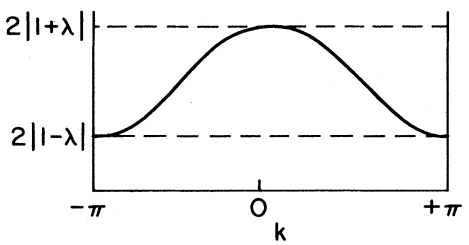


FIG. 20. The energy–momentum relation for the Ising model.

$$E(k') = \sqrt{\frac{(1-\lambda)^2}{a^2} + \lambda k'^2}. \tag{4.125}$$

So, if λ is not in the critical region, $E \sim 1/a$ as $a \to 0$ and we do not have a nontrivial continuum limit. However, if $\lambda = 1$,

$$E(k') = |k'|, \tag{4.126}$$

and we have the relativistic spectrum of a massless particle. We have seen that the particle is a fermion. In fact, it is a self-charge conjugate, Majorana field (Schultz *et. al.*, 1964).

V. WEGNER'S ISING LATTICE GAUGE THEORY

A. Global symmetries, local symmetries, and the energetics of spontaneous symmetry breaking

Since one of our first steps in understanding a spin or gauge system consists of mapping out its phase diagram, it is useful to develop an intuitive understanding of how symmetries can be realized in these systems. Our study of the two-dimensional Ising model has shown us how a global symmetry can be spontaneously broken. At low T that system's ground state is doubly degenerate. One state has positive magnetization, the other has negative magnetization. Since these two states do not mix through any finite order of perturbation theory, the spectrum of the theory must be based on only one of the two alternatives. In this way the space of states will not respect the up $\rightleftarrows$ down symmetry of the Action. Note that the symmetry which spontaneously breaks down here acts on an infinite number of degrees of freedom occupying an infinite volume. In fact, it appears that only symmetries of this kind can break down spontaneously. If a quantum system has a finite number of degrees of freedom, then its ground state is unique and symmetries of its Action are also symmetries of that state.

To appreciate these points consider some simple examples. The Hamiltonian of the nonrelativistic hydrogen atom is invariant under rotations, and its ground state is also spherically symmetric and unique. Next consider a particle in a one-dimensional, symmetric, double-well potential sketched in Fig. 21. The system's Hamiltonian is invariant under the operation $x \to -x$. Classically the ground state is doubly degenerate—the particle can sit in one of the two minima. The symmetry is not respected by either state. However, if the problem is treated quantum-mechanically, there is a finite probability that the particle will tunnel from one minimum to the other.

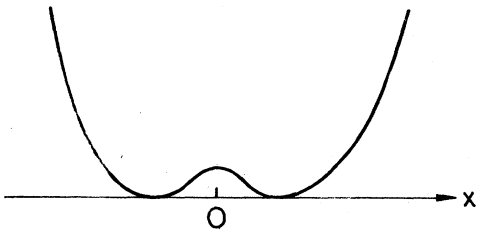


FIG. 21. Double-well potential.

This effect lifts the classical degeneracy and produces a symmetric ground state. The ground state becomes the symmetric combination of the alternative that the particle reside either in the left or right well. The antisymmetric combination has an energy which is higher by an amount proportional to the tunneling amplitude. The symmetric state has the lower energy because its wave function is smoother—it has no nodes.

It is interesting to think of the two-dimensional Ising model as consisting of a double-well potential at each site. Nearest-neighbor sites are then coupled together in the usual way. Now the symmetry can be spontaneously broken. Let us ignore the quantitative analysis we have done earlier which exposed this fact, and try to understand it in crude terms. Imagine the system with all spins up. One spin may flip by tunneling through a finite potential barrier (four bonds are broken). Such fluctuations will happen at low temperature, and they are responsible for a smooth decrease of the magnetization with increasing T. Now consider other fluctuations which tend to disorder the system more. Let a whole region of spins flip and suppose the region has a perimeter L, as shown in Fig. 22. The only broken bonds are on the *surface* of the block, so this configuration contributes a term

$$Z(\text{block}) = \exp(-2\beta L), \tag{5.1}$$

to the partition function. But such blocks can occur anywhere on the lattice, and for a given perimeter they can occur in many shapes. Summing over shapes gives

$$Z(\text{block of perimeter } L) \approx \mu^L \exp(-2\beta L) + \cdots, \tag{5.2}$$

where the factor μ^L counts blocks of various shapes and μ is a small number which can be estimated using properties of random walks. As L grows, blocks of perimeter L become less and less likely, and the magnetization is not destroyed at sufficiently low temperatures. However, if the temperature is large enough, the entropy factor μ^L in Eq. (5.2) grows faster than the temperature-dependent exponential falls. Then large blocks are not suppressed, fluctuations are important, and the magnetization vanishes.

The energetics of this example should be contrasted with the one-dimensional Ising model. That model is trivially soluble, and one knows that it is magnetized only at $T = 0$ (Stanley, 1971). In fact, one can prove

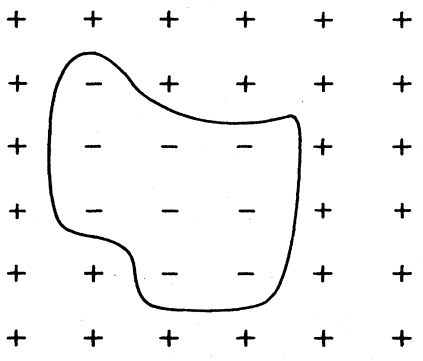


FIG. 22. A block of flipped spins. The pluses (minuses) denote spins pointing up (down).

that spin systems with nearest-neighbor couplings and discrete global symmetries can only experience spontaneous breaking at nonzero temperature in more than one dimension (Stanley, 1971). We can understand this by considering a block of length L of flipped spins as shown in Fig. 23. Since only two bonds are broken, such a block contributes a finite term to Z for *any* length L. If $L \to \infty$, so that we have a kink configuration, an infinite number of spins are flipped with only a finite cost in Action. This means that for $T \neq 0$ the ground-state expectation value of the magnetization in zero and the global symmetry is restored. Clearly the energetics of this model resembles that of a single double-well potential. This point can be made precise by considering the τ-continuum Hamiltonian formulation of the one-dimensional Ising model. Then the quantum Hamiltonian acts on only one spin variable. Following the discussion of Sec. IV.A we easily obtain the Hamiltonian

$$H = -\sigma_1 , \tag{5.3}$$

and solve for its spectrum

$$E_0 = -1 , \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \tag{5.4}$$

$$E_1 = +1 , \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} .$$

The up $\rightleftharpoons$ down symmetry of the model is realized through the spin-flip operator σ_1,

$$\sigma_1 H \sigma_1 = H . \tag{5.5}$$

This is clearly a symmetry of the spectrum Eq. (5.4). The ground state is unique and symmetric.

These examples illustrate the importance of dimensionality of the physical system to its phase diagram. Another essential ingredient is the dimensionality of the local degrees of freedom occurring in the Action. Later in this article we shall study theories with a continuous global symmetry

$$S = -J \sum_{n,\mu} \mathbf{s}(n) \cdot \mathbf{s}(n+\mu) , \tag{5.6}$$

where $\mathbf{s}(n)$ is a two-dimensional unit vector in the "planar" (or O_2) model and a three-dimensional unit vector in the O_3 Heisenberg model. There are rigorous theorems (Mermin and Wagner, 1966) which state that the global continuous symmetries of such models cannot break down spontaneously in two or fewer dimensions. We shall understand this result later after some detailed analysis. Roughly speaking, these models do not magnetize in two dimensions while the Ising model does, due to the presence of smoothly varying spin configurations. They tend to disorder the system and their multiplicity is so great that they succeed in two dimensions at any nonzero temperature. One says that the "critical dimension" of the O_n models is two because only above this number of dimensions will they have two phases. The critical dimension of the Ising model is one.

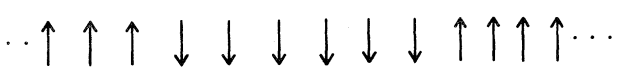


FIG. 23. A block of flipped spins in the one-dimensional Ising model.

B. Constructing an Ising model with a local symmetry

As a first step toward lattice gauge theory formulations of the strong interactions, we shall consider Ising lattice gauge theory. We want to take the degrees of freedom of the Ising model and couple them together in such a way that the global symmetry of the old model is elevated to a *local* symmetry. When this construction is generalized to systems with continuous symmetries, we shall recognize the models as lattice versions of theories with gauge symmetries.

F. Wegner invented Ising lattice gauge theory in 1971 (Wegner, 1971). His motivation was to obtain models which could not magnetize but would have nontrivial phase diagrams. His inspiration was the Planar model in two dimensions. It was known that such a system could not magnetize, but it was suspected that the theory had a phase transition because high-temperature expansions of the theory's susceptibility indicated a singularity at a reasonable temperature (Stanley and Kaplan, 1966). The idea of a "phase transition" without a local order parameter is quite novel, since it challenges us to find a symmetry of the system which can distinguish the two phases. We shall see that lattice gauge theories pose similar conceptual problems, and Wegner's work sheds much light here. He realized that endowing the lattice system with a local invariance group forbade the occurrence of a magnetization, since a local symmetry *cannot* break down spontaneously. We shall prove this fact (Elitzur, 1975) in the next section. Wegner also showed that his models had phase transitions, and he suggested how the various phases could be labeled and distinguished. Although his work is not generally cited, his 1971 paper ranks among the most significant in the field.

Consider a cubic lattice in d-dimensional Euclidean space–time. Label links of the lattice by a site n and a unit lattice vector μ. The same link can be labeled as (n, μ) or $(n+\mu, -\mu)$. Place Ising spins ($\sigma_3 = \pm 1$) on links. Define a local gauge transformation at the site n as the operation $G(n)$ of flipping all the spins on links connected to that site. An example is shown in Fig. 24. The Action has a huge invariance group because $G(n)$ can be applied anywhere. A nontrivial Action having this symmetry consists of the product of spins around primitive squares, or "plaquettes," of the lattice,

$$S = -J \sum_{n,\mu\nu} \sigma_3(n,\mu)\sigma_3(n+\mu,\nu) \times \sigma_3(n+\mu+\nu,-\mu)\sigma_3(n+\nu,-\nu) . \tag{5.7}$$

The arguments of the spin variables label the relevant link. We shall sometimes use the generic notation

$$S = -J \sum \sigma_3\sigma_3\sigma_3\sigma_3 \tag{5.8}$$

for easy presentation.

We must first check that S is invariant under arbitrary local gauge transformations. If the operation G is

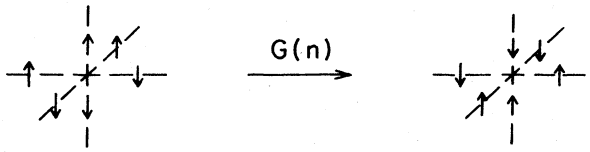


FIG. 24. A local symmetry operation in Ising lattice gauge theory in three dimensions.

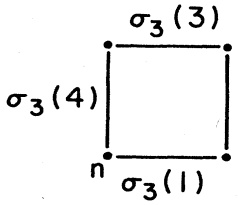


FIG. 25. A plaquette of spins.

applied at the site n, then both spins $\sigma_3(n,\mu)$ and $\sigma_3(n+\nu,-\nu)=\sigma_3(n,\nu)$ change sign. Therefore S is unchanged. A little thought shows that the essential ingredient in guaranteeing the gauge invariance of S is that it be constructed from the products of σ_3 taken around *closed* paths of links. In constructing S one chooses primitive squares of four links to keep it as local as possible.

If we want to discuss the energetics of Ising lattice gauge theory, we must use a language which respects the local gauge invariance of the Action. Recall that when dealing with the ordinary Ising model we could speak of broken or unbroken bonds—concepts which are invariant to the global symmetry of that model's Action. However, the concept of a broken bond is not locally gauge invariant. For example, consider the plaquette of spins shown in Fig. 25. Applying a gauge transformation at n changes the character of the bonds $\sigma_3(1)\sigma_3(2)$ and $\sigma_3(3)\sigma_3(4)$. So, it is only the relative orientation of four spins around a plaquette that is gauge invariant. There are two possibilities:

$$\prod_i \sigma_3(i)=1 \tag{5.9a}$$

or

$$\prod_i \sigma_3(i)=-1\,. \tag{5.9b}$$

One says that in the second case the spins are "frustrated" (Anderson, 1976), while in the first case they are not.

C. Elitzur's theorem—the impossibility of spontaneously breaking a local symmetry

We want to know if Ising lattice gauge theories can have interesting phase diagrams and, if so, how one can label their phases. The first point to make is that transitions with local order parameters cannot occur. This is a consequence of Elitzur's theorem, which states that the spontaneous magnetization of this model vanishes identically at all T as a consequence of the local gauge symmetry (Elitzur, 1975). In light of our intuitive discussion of spontaneous symmetry breaking, this result is certainly plausible. The proof is not difficult.

To test whether a system sustains a spontaneous magnetization, one places it in an external field h, which couples through a term $h\sum_{n,\mu}\sigma_3(n,\mu)$, computes $\langle\sigma_3(m,\nu)\rangle$, and then takes the limit $h\to 0$. If the expectation value of σ_3 is nonzero in the limit, the system is magnetized. For lattice gauge theory, a nonzero expectation value would imply the spontaneous breaking of not just the global up $\rightleftharpoons$ down symmetry, but the local symmetry as well. Consider the expectation value

$$\langle\sigma_3(n,\nu)\rangle_h=\frac{\sum_{\text{spin config}}\sigma_3(n,\nu)\exp\left\{\beta\sum\sigma_3\sigma_3\sigma_3\sigma_3+h\sum\sigma_3\right\}}{\sum_{\text{spin config}}\exp\left\{\beta\sum\sigma_3\sigma_3\sigma_3\sigma_3+h\sum\sigma_3\right\}}\,. \tag{5.10}$$

We shall use the *local* gauge invariance of the Action in the absence of h to show that the right-hand side of Eq. (5.10) vanishes smoothly as $h\to 0$. Consider a local gauge transformation at the site n. Denote the set of links emanating from site n by $\{l_n\}$. The four-spin term in the Action is invariant to this gauge transformation, but the external field term becomes

$$h\sum\sigma_3=h\sum\sigma_3'-h\sum\delta\sigma_3\,, \tag{5.11}$$

where σ_3' is a transformed spin and

$$\begin{aligned}&\delta\sigma_3(l_n)\equiv\sigma_3'(l_n)-\sigma_3(l_n)=-2\sigma_3(l_n)\,,\\&\delta\sigma_3(l)=0 \text{ if, } l\notin\{l_n\}\,.\end{aligned} \tag{5.12}$$

Making this change of variables $\sigma_3\to\sigma_3'$ in the expression Eq. (5.10) gives

$$\begin{aligned}\langle\sigma_3(n,\nu)\rangle_h&=-\sum\sigma_3'(n,\nu)\\&\quad\times\exp\left\{\beta\sum\sigma_3'\sigma_3'\sigma_3'\sigma_3'+h\sum\sigma_3'+h\sum\delta\sigma_3\right\}/Z\,,\\&=\left\langle-\sigma_3(n,\nu)\exp\left\{-h\sum_{\{l_n\}}\delta\sigma_3\right\}\right\rangle_h\,.\end{aligned} \tag{5.13}$$

Now we can produce a bound,

$$|\langle\sigma_3(n,\nu)\rangle_h-\langle-\sigma_3(n,\nu)\rangle_h|=\left|\left\langle-\sigma_3(n,\nu)\left[\exp\left\{-h\sum_{\{l_n\}}\delta\sigma_3\right\}-1\right]\right\rangle_h\right|\le|e^{4dh}-1||\langle\sigma_3(n,\nu)\rangle_h|\,, \tag{5.14}$$

where d is the dimension of space-time. But the right-hand side of Eq. (5.14) approaches zero as $h\to 0$, so

$$\langle\sigma_3(n,\nu)\rangle_{h\to 0}=\langle-\sigma_3(n,\nu)\rangle_{h\to 0}\,, \tag{5.15}$$

which implies

$$\langle\sigma_3(n,\nu)\rangle=0 \tag{5.16}$$

as claimed.

What was the essential ingredient in this proof? The point is that even in the presence of h, spin configura-

tions connected by the operation $G(n)$ differ in Action only by a *finite* amount. Therefore only a finite energy barrier is set up between two spin configurations, one of which has $\sigma_3(n,\nu)$ positive and the other $\sigma_3(n,\nu)$ negative. Thus the energetics resemble a quantum mechanics problem involving only a few local degrees of freedom, and we are familiar with the fact that in this setting symmetries do not break down spontaneously.

Elitzur's theorem applies to a wide class of theories. Theories with a continuous local symmetry group also cannot have a spontaneous magnetization. The proof given here easily generalizes to these cases, which will be studied in later sections of this review. The theorem can also be applied to the expectation values of more complicated operators. The expectation value of any operator which is not invariant under all local gauge transformations must vanish identically. Accordingly we shall restrict our attention to gauge-invariant operators from here on.

D. Gauge-invariant correlation functions

Now we can return to our original question—how to label the phases of a theory having a local symmetry group. Wegner suggested that we consider the spatial dependence of correlation functions. His inspiration was the two-dimensional planar model, which undergoes a phase transition without the appearance of a spontaneous magnetization. The two-phase character of the model shows up in the system's spin-spin correlation function. At low temperature spin-wave analyses show that the correlation function falls off as a power,

$$\langle \mathbf{s}(0)\cdot\mathbf{s}(n)\rangle \sim |n|^{-kT/2\pi J}, \tag{5.17}$$

while high-temperature expansions show that the behavior is exponential for T sufficiently large,

$$\langle \mathbf{s}(0)\cdot\mathbf{s}(n)\rangle \sim \exp(-|n|/\xi(T)). \tag{5.18}$$

These estimates mean that the correlation length is finite at high temperatures and is infinite at low temperatures. A phase transition must occur in the intermediate temperature region. More powerful analysis is necessary to probe the details of that region of the phase diagram.

Wegner invented and studied a gauge-invariant correlation function for Ising gauge theories. Since one must consider the product of spin variables around a closed path of links, it is natural to consider

$$\prod_{l\in C}\sigma_3(l), \tag{5.19}$$

where the arguments of the spin variables denote links and C is a closed contour. The expectation value of such an operator will depend on the characteristics of the contour C. In particular, the loop has a certain perimeter P, a certain minimal enclosed surface area A. Wegner showed (Wegner, 1971) that at high T

$$\left\langle \prod_{l\in C}\sigma_3(l)\right\rangle \sim \exp(-A), \tag{5.20a}$$

while at low T

$$\left\langle \prod_{l\in C}\sigma_3(l)\right\rangle \sim \exp(-P). \tag{5.20b}$$

So, at high T the correlation function falls very quickly as the loop is taken larger and larger, while at low T it falls off at a qualitatively slower rate. This result proves that the system has distinct high- and low-temperature phases.

It is easy and instructive to obtain these results. The methods employed are simple generalizations of the high- and low-temperature expansions of the Ising model reviewed in Sec. IV.E. First consider high T. We use the identity

$$\begin{aligned}\exp(\beta\sigma_3\sigma_3\sigma_3\sigma_3) &= \cosh\beta + \sigma_3\sigma_3\sigma_3\sigma_3\sinh\beta \\ &= (1+\sigma_3\sigma_3\sigma_3\sigma_3\tanh\beta)\cosh\beta,\end{aligned} \tag{5.21}$$

so that

$$\left\langle \prod_{l\in C}\sigma_3(l)\right\rangle = \sum_{\substack{\text{spin}\\ \text{config}}}\prod_{\text{plaquettes}}(1+\sigma_3\sigma_3\sigma_3\sigma_3\tanh\beta) \times \prod_C\sigma_3 \Big/ \sum\prod(1+\sigma_3\sigma_3\sigma_3\sigma_3\tanh\beta). \tag{5.22}$$

If β is small compared to unity, it is sensible to expand Eq. (5.22) in powers of $\tanh\beta$ as in Sec. IV.E. The fact that

$$\sum_{\sigma_3=\pm1}\sigma_3 = 0, \quad \sum_{\sigma_3=\pm1}\sigma_3^2 = \sum_{\sigma_3=\pm1}1 = 2, \tag{5.23}$$

means that enough powers of $\sigma_3\sigma_3\sigma_3\sigma_3\tanh\beta$ must be collected from the expansion of the Action so that no isolated factors of σ_3 belonging to $\Pi_C\sigma_3$ are left unmatched. A little thought indicates that the first nonvanishing contribution in the numerator of Eq. (5.22) occurs when there is a plaquette of operators $\sigma_3\sigma_3\sigma_3\sigma_3$ for each square of the minimal surface bounded by C,

$$\left\langle \prod_C\sigma_3\right\rangle = (\tanh\beta)^{N_c}+\cdots, \tag{5.24}$$

where N_c is the number of squares on that surface. But Eq. (5.24) confirms Eq. (5.20a) because it can be written

$$\left\langle \prod_C\sigma_3\right\rangle = \exp\{\ln\tanh\beta A\}+\cdots,\ \beta\ll 1. \tag{5.25}$$

If we calculated to higher orders, the "area law" found here would remain, but the coefficient $\ln\tanh\beta$ would become a more complicated expression,

$$\left\langle \prod_C\sigma_3\right\rangle = \exp[-f(\beta)A], \tag{5.26}$$

where the leading term in the expansion of $f(\beta)$ is $\ln\tanh\beta$. The point is that the "area law" holds for finite but large temperatures. The methods used to prove that high-temperature expansions of simpler Ising systems have finite radii of convergence apply also to Ising gauge theory.

Now consider low temperatures and suppose that $d>2$. To develop a low-temperature expansion for the correlation function it is convenient to organize the sum over spin configurations in a particular way.

A configuration of spins $\{\sigma_3(n,\nu)\}$ may be transformed to another configuration $\{\sigma_3'(n,\nu)\}$ by applying gauge transformations at various sites. We have seen that the physics of the model consists only of its gauge-invariant content, so the configurations $\{\sigma_3(n,\nu)\}$ and $\{\sigma_3'(n,\nu)\}$ are said to be "gauge equivalent." A gauge-invariant configuration of the system consists of a spin configuration $\{\sigma_3(n,\nu)\}$ and all of its gauge-equivalent copies. It is often convenient, however, to label a gauge-invariant configuration by one of its representative configurations of definite spins. When computing gauge-invariant expectation values we can do this with no loss of generality or rigor. In such a computation the sum over all spin configurations is replaced by a sum over representative spin configurations labeling different gauge-invariant configurations. When this is done, a common multiplicative counting factor is removed from the numerator and denominator of an expression such as Eq. (5.22). In some of our illustrative calculations which follow, we shall often use the language of "flipped spins" and "representative spin configurations" rather than the explicitly gauge-invariant concepts of "frustration" and "gauge-invariant configurations." In particular, it is helpful when discussing the low-temperature expansion to choose the representative spin configuration at $T=0$ to have "all spins up" so that the expansion proceeds in the "number of flipped spins." So consider the expectation value

$$\left\langle \prod_C \sigma_3 \right\rangle = \sum \prod \sigma_3 \exp\left\{\beta \sum \sigma_3\sigma_3\sigma_3\sigma_3\right\}/Z . \tag{5.27}$$

The first term in the expansion is unity because $\Pi_C\sigma_3 = 1$ when all spins are up. Now let there be one spin flipped. Then $2(d-1)$ plaquettes are frustrated, so the Action of this configuration measured relative to the Action of the "all spins up" configuration is $4(d-1)\beta$. Note that only if the flipped spin occurs on the contour C will the numerator be different from unity. In those cases it is -1. If N is the number of links of the lattice and L is the number of links on the contour, we have

$$\left\langle \prod_C \sigma_3 \right\rangle = \{1 + (N-2L)\exp[-4(d-1)\beta] + \cdots\}/ \{1 + N\exp[4(d-1)\beta] + \cdots\} . \tag{5.28}$$

Now let's estimate the contribution to the numerator due to n spin flips. Treating them as completely independent, we have

$$\frac{1}{n!}(N-2L)^n \exp[-4n(d-1)\beta] . \tag{5.29a}$$

The n-spin-flip contribution to the partition function itself is

$$\frac{1}{n!}N^n \exp[-4n(d-1)\beta] . \tag{5.29b}$$

Summing over n,

$$\left\langle \prod_C \sigma_3 \right\rangle \cong \{1 + (N-2L)\exp[-4(d-1)\beta] + \tfrac{1}{2}(N-2L)^2 \exp[-8(d-1\beta)] + \cdots\}/Z , \tag{5.30a}$$

where

$$Z \cong 1 + N\exp[-4(d-1)\beta] + \tfrac{1}{2}N^2\exp[-8(d-1)\beta] + \cdots . \tag{5.30b}$$

So, both numerator and denominator exponentiate, the dependence on N cancels as expected, and

$$\left\langle \prod_C \sigma_3 \right\rangle \approx \exp[-2e^{-4(d-1)\beta}L] , \tag{5.31}$$

and we have verified the "perimeter law," Eq. (5.20b). If more graphs were summed we would find that

$$\left\langle \prod_C \sigma_3 \right\rangle \approx \exp[-h(\beta)L] , \tag{5.32}$$

where the leading term in a low-temperature series for $h(\beta)$ is $2\exp[-4(d-1)\beta]$. To calculate $h(\beta)$ systematically one should develop a connected graph formalism (Wortis, 1972). Then one could take account of the excluded volume effects which we ignored when discussing the n spin-flipped configurations in Eq. (5.29). All this can be done systematically, and it does not affect the leading term in the series for $h(\beta)$. Since this low-temperature expansion has a finite radius of convergence, we have learned that the system has a low-temperature phase which is distinct from its high-temperature phase.

Our argument for the "perimeter law" does not apply in two dimensions. In that case the low-temperature expansion has a vanishing radius of convergence. There are several ways to see this, and we shall discuss two of them which are instructive. First observe that the energetics of the two-dimensional model is very special. If a single spin is flipped, then two plaquettes are frustrated. However, if a line of spins are flipped, then again only two plaquettes are frustrated—those at the ends of the line, as shown in Fig. 26. So, there is a special degeneracy problem in two dimensions which "disorders" the system at any nonzero temperature and leads to an "area law" as we shall now see. This degeneracy problem does not occur in higher dimensions because then plaquettes which come out of the plane of Fig. (26) are frustrated by the line of flipped spins. Let us calculate the gauge-invariant correlation function by taking into account only configurations of flipped spins in which one of the ends of the lines extends to infinity. Such configurations are very efficient at disordering the system, and we shall see by comparison with an exact calculation that they are a good guide to the qualitative character of the

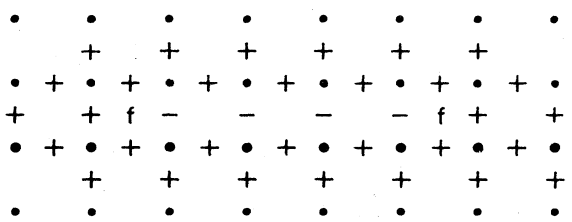


FIG. 26. A line of flipped spins and the accompanying frustrated plaquettes f in two dimensions.

system. In gauge-invariant language, we shall obtain an expansion in terms of free, isolated frustrations. Consider the case in which there is only one frustration f in the neighborhood of the contour C. If f lies outside C, its string of flipped spins affect an even number of spins on C, producing a positive contribution to $\langle \Pi_C \sigma_3 \rangle$. However, if f is inside C, an odd number of links on C are affected and a -1 is obtained. If N_c is the number of links enclosed within C, then

$$\left\langle \prod_C \sigma_3 \right\rangle = \{1 + (N - 2N_c)\exp(-2\beta) + \cdots\} / \{1 + N\exp(-2\beta) + \cdots\}. \tag{5.33}$$

If we treat the configuration of n frustrations ignoring correlations, then numerator and denominator exponentiate as in Eq. (5.30), and we find

$$\left\langle \prod_C \sigma_3 \right\rangle \approx \exp(-2e^{-2\beta}N_c), \approx \exp(-4e^{-2\beta}A), \tag{5.34}$$

where A is the area enclosed by C. Since the high-temperature expansion is well-behaved in any number of dimensions, Eq. (5.20a) labels the high-temperature region of the model. Therefore the gauge-invariant correlation function shows no evidence for a phase transition—it appears that the system only resides in a disordered phase. This is, in fact, the case as we shall see now.

Let's return to the partition of two-dimensional Ising lattice gauge theory. Some simple analysis will yield the following equivalence:

$$\begin{pmatrix}\text{Two-dimensional}\\ \text{Ising gauge theory}\end{pmatrix} \equiv \begin{pmatrix}\text{One-dimensional}\\ \text{Ising model}\end{pmatrix}. \tag{5.35}$$

This equivalence is the first example in this review of some deep relations between gauge systems and spin systems. The equivalence is most easily exposed by "choosing a convenient gauge." In other words, we shall label gauge-invariant spin configurations with a well-chosen representative spin state. Consider a two-dimensional lattice such as Fig. 26 and label vertical links "temporal" and horizontal links "spatial." Choose representative configurations such that, on each temporal link,

$$\sigma_3(n, \hat{\tau}) = 1. \tag{5.36}$$

It should be clear that given any configuration of $\sigma_3(n, \mu)$ one can apply local gauge transformations appropriately so that Eq. (5.36) is satisfied. As long as only gauge-invariant operators are considered, one can work with configurations satisfying this condition and no error is made. The choice Eq. (5.36) is known as the "temporal gauge." The Action now becomes

$$S = -J \sum \sigma_3(n, \hat{x})\sigma_3(n + \hat{\tau}, \hat{x}). \tag{5.37}$$

So, there is no coupling between spins in the horizontal direction and the coupling in the $\hat{\tau}$ direction is that of a one-dimensional Ising model. Thus Eq. (5.35) is established. Since the one-dimensional model is disordered for all nonzero T, we learn that the gauge system has only one phase. It is characterized by the area law. The correspondence between the disordered character of the spin system and the area law of the gauge system comes about as follows. Consider the correlation function $\Pi_C \sigma_3$ in the temporal gauge. Let C be a rectangular contour whose temporal extent is T and whose spatial extent is R, as shown in Fig. 27. In the temporal gauge,

$$\prod_C \sigma_3 = \sigma_3(0, 0; \hat{x})\sigma_3(0, 1; \hat{x}) \ldots \sigma_3(0, R; \hat{x}) \times \sigma_3(T, 0; \hat{x})\sigma_3(T, 1; \hat{x}) \ldots \sigma_3(T, R; \hat{x}), \tag{5.38}$$

since σ_3 on each temporal link is unity. With this result and Eq. (5.37) it becomes clear that the computation of $\langle \Pi_c \sigma_3 \rangle$ reduces to the product of R spin–spin correlation functions of the one-dimensional Ising model. These correlation functions are all short ranged (Stanley, 1971);

$$\langle \sigma_3(T, 0; \hat{x})\sigma_3(0, 0; \hat{x}) \rangle \big|_{\substack{1-d \\ \text{Ising}}} \underset{T\to\infty}{\sim} \exp(-T/\xi), \tag{5.39}$$

where ξ is the correlation length of that model. Therefore the gauge theory correlation function is

$$\left\langle \prod_C \sigma_3 \right\rangle \sim [\exp(-T/\xi)]^R = \exp(-T/\xi), \tag{5.40}$$

which gives the area law

$$\left\langle \prod_C \sigma_3 \right\rangle \sim \exp(-A/\xi), \tag{5.41}$$

again.

E. Quantum Hamiltonian and phases of the three-dimensional Ising gauge theory

We want to gain more insight into the character of phase transitions which occur in gauge systems. Since these systems cannot have local order parameters, it is not clear what symmetry operator distinguishes their phases. We shall see that these theories possess topological singularities which accomplish this. To illustrate these points we shall study three-dimensional Ising gauge theory and see that it is dual to the ordinary three-dimensional Ising model. Since that model has an ordinary well-understood phase transition, we can learn about the gauge system by pursuing the duality mapping.

To obtain the τ-continuum Hamiltonian it is convenient to work within the temporal gauge, Eq. (5.36).

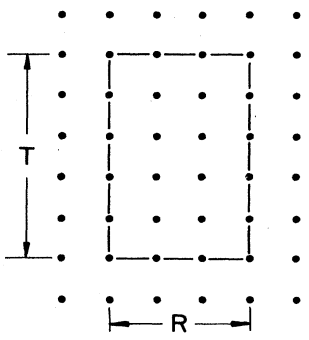


FIG. 27. A rectangular contour.

This condition does not label the representative spin configuration of a gauge-invariant set uniquely. In fact, within the temporal gauge the theory is invariant to a huge set of local symmetry operations. These consist of all τ-independent gauge transformations. Consider the application of such a transformation on a particular column of sites of definite spatial position. The column runs from $\tau=-\infty$ to $\tau+\infty$ as sketched in Fig. 28. The point is that since each spin on a temporal link is either not affected by the transformation or is flipped twice, the condition Eq. (5.36) is maintained.

To obtain the τ-continuum Hamiltonian, begin with the Action written with anisotropic couplings. Let β_τ be the coupling for plaquettes containing some temporal links and let β be the coupling for the remaining cases,

$$S=-\beta_\tau\sum_{\{P_\tau\}}\sigma_3\sigma_3\sigma_3\sigma_3-\beta\sum_{\{P_s\}}\sigma_3\sigma_3\sigma_3\sigma_3\,, \tag{5.42}$$

where P_τ and P_s denote the two different types of plaquettes. The first term simplifies because of the choice of gauge

$$-\beta_\tau\sum_{\{P_\tau\}}\sigma_3(n,\hat{x})\sigma_3(n+\hat{\tau},\hat{x})\,. \tag{5.43}$$

Aside from an irrelevant additive constant, this term can be written

$$\tfrac{1}{2}\beta_\tau\sum_{\{P_\tau\}}[\sigma_3(n+\hat{\tau},\hat{x})-\sigma_3(n,\hat{x})]^2\,, \tag{5.44}$$

so

$$S=\tfrac{1}{2}\beta_\tau\sum_{\{P_\tau\}}[\sigma_3(n+\hat{\tau},\hat{x})-\sigma_3(n,\hat{x})]^2-\beta\sum_{\{P_\tau\}}\sigma_3\sigma_3\sigma_3\sigma_3\,. \tag{5.45}$$

This expression should be compared to Eq. (4.2) of the two-dimensional Ising model. The construction of the transfer matrix, a τ-continuum limit, and a quantum Hamiltonian parallels that discussion. The same limiting procedure

$$\beta_\tau\to\infty\,,\quad \beta\to\lambda\exp(-2\beta_\tau)\,, \tag{5.46}$$

applies and leads to the quantum Hamiltonian

$$H=-\sum_{\mathbf{n},i}\sigma_1(\mathbf{n},i)-\lambda\sum\sigma_3\sigma_3\sigma_3\sigma_3\,, \tag{5.47}$$

where the σ_1 and σ_3 are now operators. The lattice underlying Eq. (5.47) is now purely spatial, of course. The vector $\mathbf{n}$ labels sites and i labels the unit vectors of the lattice.

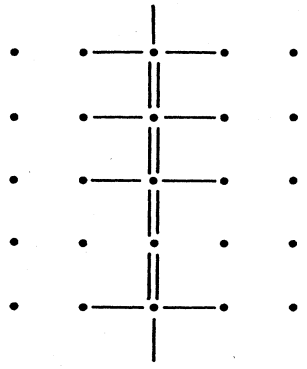


FIG. 28. A τ-independent gauge transformation.

Our first task is to understand the local symmetries of the theory using the operator formulation. It is easy to construct a τ-independent operator $G(\mathbf{n})$ which flips the spins $\sigma_3(\mathbf{n},i)$ on all the links emanating from $\mathbf{n}$,

$$G(\mathbf{n})=\prod_{\pm i}\sigma_1(\mathbf{n},i)\,. \tag{5.48}$$

For example on a three-dimensional spatial lattice $\pm i$ runs over six values, as shown in Fig. 29. Since σ_1 is a spin-flip operator for σ_3, we have

$$\begin{aligned}&G^{-1}(\mathbf{n})\sigma_1(\mathbf{m},i)G(\mathbf{n})=\sigma_1(\mathbf{m},i)\,,\\ &G^{-1}(\mathbf{n})\sigma_3(\mathbf{n},i)G(\mathbf{n})=-\sigma_3(\mathbf{n},i)\,,\\ &G^{-1}(\mathbf{n})\sigma_3(\mathbf{m},i)G(\mathbf{n})=\sigma_3(\mathbf{m},i)\,,\end{aligned} \tag{5.49}$$

where the last formula applies only if the link $(\mathbf{m},i)$ does not coincide with any of those originating from site $\mathbf{n}$. It follows that

$$G^{-1}(\mathbf{n})HG(\mathbf{n})=H\,, \tag{5.50}$$

so the Hamiltonian has the desired local invariance. Elitzur's theorem implies that the space of states $\{|\psi\rangle\}$ is invariant to these local operations,

$$G(\mathbf{n})|\psi\rangle=|\psi\rangle\,. \tag{5.51}$$

This statement subsumes the fact that the system cannot magnetize. In particular, consider the matrix element

$$\begin{aligned}\langle\psi|\sigma_3(\mathbf{n},i)|\psi\rangle&=\langle\psi|G(\mathbf{n})G^{-1}(\mathbf{n})\sigma_3(\mathbf{n},i)G(\mathbf{n})G^{-1}(\mathbf{n})|\psi\rangle\\&=-\langle\psi|\sigma_3(\mathbf{n},i)|\psi\rangle\,,\end{aligned} \tag{5.52}$$

which implies that

$$\langle\psi|\sigma_3(\mathbf{n},i)|\psi\rangle=0\,. \tag{5.53}$$

Now we can study the three-dimensional Ising gauge theory in detail. It is easy to see that it is dual to the ordinary three-dimensional Ising model. To do this we follow the same strategy used to see that the two-dimensional gauge theory was trivial; write the theory in terms of a minimal number of degrees of freedom by eliminating all the residual gauge freedom. If we always work within the gauge-invariant subspace Eq. (5.51), we have an operator identity

$$G(\mathbf{n})=\prod_{\pm i}\sigma_1(\mathbf{n},i)=1\,. \tag{5.54}$$

So

$$\sigma_1(\mathbf{n},\hat{y})\sigma_1(\mathbf{n},-\hat{y})\sigma_1(\mathbf{n},\hat{x})\sigma_1(\mathbf{n},-\hat{x})=1\,, \tag{5.55a}$$

which allows us to solve for $\sigma_1(\mathbf{n},\hat{y})$,

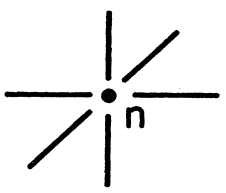


FIG. 29. The links affected by the gauge transformation $G(n)$.

$$\sigma_1(\mathbf{n},\hat{y}) = \sigma_1(\mathbf{n},\hat{x})\sigma_1(\mathbf{n},-\hat{x})\sigma_1(\mathbf{n},-\hat{y}) . \tag{5.55b}$$

But $\sigma_1(\mathbf{n},-\hat{y})$ can be treated similarly by considering a gauge transformation at the site $\mathbf{n}-\hat{y}$. Iterating this procedure, one can write σ_1 on any $\hat{y}$ link just in terms of those on $\hat{x}$ links,

$$\sigma_1(\mathbf{n},\hat{y}) = \sigma_1(\mathbf{n},\hat{x})\sigma_1(\mathbf{n},-\hat{x})\sigma_1(\mathbf{n}-\hat{y},\hat{x})\sigma_1(\mathbf{n}-\hat{y},-\hat{x}) \times \sigma_1(\mathbf{n}-2\hat{y},\hat{x})\sigma_1(\mathbf{n}-2\hat{y},-\hat{x}) \ldots . \tag{5.56}$$

This procedure means that we are treating σ_1 on links which point in the $\hat{y}$ direction as *dependent* variables. To be consistent, the spin variables $\sigma_3(\mathbf{n},\pm\hat{y})$ must also be eliminated from the collection of independent degrees of freedom of the theory. In particular, H will have no operators which do not commute with $\sigma_3(\mathbf{n},\hat{y})$. Therefore, choose it to be a constant,

$$\sigma_3(\mathbf{n},\hat{y}) = 1 . \tag{5.57}$$

Now $\sigma_1(\mathbf{n},\hat{x})$ and $\sigma_3(\mathbf{n},\hat{x})$ are the only variables in H and they are independent.

The duality mapping can now be defined. Associate a site $\mathbf{n}^*$ of a "dual lattice" with each plaquette of the original lattice. Define a "dual spin-flip" operator on this site,

$$\mu_1(\mathbf{n}^*) = \sigma_3\sigma_3\sigma_3\sigma_3 , \tag{5.58}$$

where the four σ_3's are those of the plaquette associated with $\mathbf{n}^*$, as shown in Fig. 30. "Dual spin" variables are defined by

$$\mu_3(\mathbf{n}^*) = \prod_{n'\geq 0} \sigma_1(\mathbf{n}-n'\hat{y},\hat{x}) . \tag{5.59}$$

As in the case of our discussion of the two-dimensional Ising model, the dual variables satisfy the Pauli algebra, and the Hamiltonian can be written simply in terms of them. It is clear that

$$\mu_1^2(\mathbf{n}^*) = \mu_3^2(\mathbf{n}^*) = 1 . \tag{5.60}$$

Also,

$$\mu_1(\mathbf{n}^*)\mu_3(\mathbf{n}^*) = -\mu_3(\mathbf{n}^*)\mu_1(\mathbf{n}^*) , \tag{5.61}$$

since these operators, when written in terms of σ_1 and σ_3, have one link in common where $\sigma_1\sigma_3 = -\sigma_3\sigma_1$ holds. Finally,

$$\mu_1(\mathbf{n}^*)\mu_3(\mathbf{m}^*) = \mu_3(\mathbf{m}^*)\mu_1(\mathbf{n}^*) \quad (\mathbf{n}^* \neq \mathbf{m}^*) , \tag{5.62}$$

since the identity $\sigma_1\sigma_3 = -\sigma_3\sigma_1$ must be applied an even number of times to interchange the two dual operators. To write the Hamiltonian in terms of the dual operators, note that

$$\mu_3(\mathbf{n}^*)\mu_3(\mathbf{n}^*-\hat{y}) = \sigma_1(\mathbf{n},\hat{x}) , \tag{5.63}$$

follows trivially from Eq. (5.59). In addition

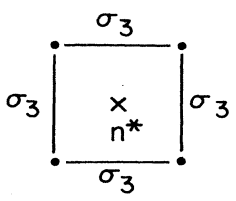


FIG. 30. The dual lattice sites in two spatial dimensions.

$$\mu_3(\mathbf{n}^*)\mu_3(\mathbf{n}^*-\hat{x}) = \sigma_1(\mathbf{n},\hat{y}) , \tag{5.64}$$

can be recognized as a consequence of Eq. (5.56). Therefore the Hamiltonian is

$$H = -\sum_{\mathbf{n}^*,i} \mu_3(\mathbf{n}^*)\mu_3(\mathbf{n}^*+i) - \lambda \sum_{n^*} \mu_1(\mathbf{n}^*)$$

$$= \lambda\left\{ -\sum_{\mathbf{n}^*} \mu_1(\mathbf{n}^*) - (1/\lambda) \sum_{n^*,i} \mu_3(\mathbf{n}^*)\mu_3(\mathbf{n}^*+i) \right\}, \tag{5.65}$$

which we recognize as the quantum Hamiltonian formulation of the three-dimensional Ising model at a temperature $\lambda^* = \lambda^{-1}$,

$$H(3\text{-}D \text{ Ising gauge};\lambda) = \lambda H(3\text{-}D \text{ Ising model};\lambda^{-1}) . \tag{5.66}$$

Therefore we have a mapping between the high- (low-) temperature properties of the gauge system and the low- (high-) temperature properties of the spin system. A great deal is known about the three-dimensional Ising model. It is a two-phase system with a spontaneous magnetization labeling the ordered phase. It undergoes a continuous phase transition at T_c with power-behaved singularities. We can now use the fact that this model has a local order parameter to learn how to label the phases of the gauge theory. Since $\langle 0|\mu_3(\mathbf{n}^*)|0\rangle$ is nonzero in the low-temperature phase of the Ising model, consider the Hamiltonian in an external field,

$$H(3\text{-}D \text{ Ising model};\lambda^*) + h \sum_{\mathbf{n}^*} \mu_3(\mathbf{n}^*) . \tag{5.67a}$$

This operator can be written in terms of the variables σ_1 and σ_3,

$$(1/\lambda)H(3\text{-}D \text{ Ising gauge};\lambda) + h \sum_{\mathbf{m}} \prod_{n\geq 0} \sigma_1(\mathbf{m}-n\hat{y},\hat{x}) . \tag{5.67b}$$

So, first-order perturbation theory implies the equality

$$\langle 0|\mu_3(\mathbf{m}^*)|0\rangle\big|_{\substack{\text{Ising}\\ \text{model}\\ \lambda^*=\lambda^{-1}}} = \langle 0|\prod_{n\geq 0} \sigma_1(\mathbf{m}-n\hat{y},\hat{x})|0\rangle\big|_{\substack{\text{gauge model}\\ \lambda}} . \tag{5.68}$$

Since $\mu_3(\mathbf{m}^*)$ is an order parameter for the Ising model, and since the duality mapping interchanges high and low temperatures, we can identify a *nonlocal disorder* parameter for the gauge theory,

$$\langle 0|\prod_{n\geq 0} \sigma_1(\mathbf{m}-n\hat{y},\hat{x})|0\rangle = 0 \quad (\lambda \text{ large})$$

$$\langle 0|\prod_{n\geq 0} \sigma_1(\mathbf{m}-n\hat{y},\hat{x})|0\rangle \neq 0 \quad (\lambda \text{ small}) . \tag{5.69}$$

The similarity of these statements to our discussion of the two-dimensional Ising model should be noted. We can think of $\Pi_{n\geq 0}\sigma_1(\mathbf{m}-n\hat{y},\hat{x})$ as a kink operator. Equation (5.69) states that the low-temperature ground state of the gauge theory is free of kinks, while for all T above a critical point there is a kink condensate. Note

also that the *gauge-invariant* characterization of the operator is stated in terms of *frustation*. If the operator is applied to the $T=0$ state of the gauge theory, which is free of frustrated plaquettes, then a frustration is made at the plaquette $\mathbf{m}^*$. The phase transition in the model can therefore be viewed as "frustration condensation." Since frustration and kinks have a topological significance, we have exposed a symmetry criterion to label the system's phases. Of course, the kink operator is a nonlocal object, so it can have a vacuum expectation value and not violate Elitzur's theorem. The kink condensation idea also exposes the physical mechanism which makes the gauge-invariant correlation function satisfy the area law at temperatures above T_c. If we consider a purely spatial contour C and compute $\langle 0|\Pi_C\sigma_3|0\rangle$ in the approximation that treats the high-temperature ground state as a kink condensate, then the area law is obtained in the same way that we found the area law for the two-dimensional Ising gauge theory in Eq. (5.34).

Before moving on to our next topic, let us review some facts about the *four*-dimensional Ising gauge theory. Using methods we have already discussed, one can show that the theory is self-dual (Wegner, 1971). The self-dual point is $\lambda=1$ in the τ-continuum Hamiltonian language. Expansions for the theory's free energy indicate that it is a two-phase system with a first-order phase transition at $\lambda=1$ (Balian *et al.*, 1975). This fact has been confirmed by computer simulations (Creutz, *et. al.*, 1979). It has been suggested that the theory is soluble (Polyakov, 1977) and can be diagonalized in terms of "fermionic string" variables. Research in this direction is being actively pursued by several groups.

VI. ABELIAN LATTICE GAUGE THEORY

A. General formulation

F. Wegner's Ising lattice gauge theories were generalized to continuous, Abelian gauge groups by K. G. Wilson (Wilson, 1974). Lattice gauge theories were independently invented by A. M. Polyakov (Polyakov, 1975). The quantum Hamiltonian approach was developed by the author and L. Susskind (Kogut and Susskind, 1975).

Let us begin the discussion of Abelian models by recalling the character of spin systems with a global Abelian symmetry group. Place a planar spin on each lattice site n,

$$\mathbf{s}(n)=\begin{bmatrix}\cos\theta(n)\\ \sin\theta(n)\end{bmatrix}. \tag{6.1}$$

Couple nearest neighbors together in the traditional fashion,

$$S=-J\sum_{n,\mu}\mathbf{s}(n)\cdot\mathbf{s}(n+\mu) \tag{6.2}$$

$$=-J\sum_{n,\mu}\cos[\theta(n)-\theta(n+\mu)].$$

This can be written more elegantly if we introduce a finite difference operator in the direction μ,

$$\Delta_\mu\theta(n)=\theta(n+\mu)-\theta(n), \tag{6.3}$$

so that Eq. (6.2) becomes

$$S=-J\sum_{n,\mu}\cos(\Delta_\mu\theta(n)). \tag{6.4}$$

This Action has a global, continuous symmetry, i.e., rotate *all* the spins through a common angle α. S remains unchanged since it is constructed from inner products.

Now construct a theory which elevates this continuous symmetry to a local continuous symmetry. We shall do this by generalizing Wegner's Ising gauge theories in a direct fashion. Experience with electrodynamics is also helpful here. Place planar spins on each link of a lattice. Suppose that there are local symmetry operators at every site n, which rotate all the spins on the links originating from site n by a common angle $\chi(n)$. This should be an exact symmetry of the theory's Action. To state this precisely let $\theta_\mu(n)$ be an angular variable on the link (n,μ). Since the link (n,μ) can also be labeled $(n+\mu,-\mu)$, we need a definition of the variable $\theta_{-\mu}(n+\mu)$. We choose

$$\theta_{-\mu}(n+\mu)=-\theta_\mu(n), \tag{6.5}$$

for reasons which will be explained shortly. With each plaquette of the lattice, associate the following "discrete curl":

$$\theta_{\mu\nu}(n)=\Delta_\mu\theta_\nu(n)-\Delta_\nu\theta_\mu(n)$$

$$=\theta_\nu(n+\mu)-\theta_\nu(n)-\theta_\mu(n+\nu)+\theta_\mu(n) \tag{6.6}$$

$$=\theta_\mu(n)+\theta_\nu(n+\mu)+\theta_{-\mu}(n+\mu+\nu)+\theta_{-\nu}(n+\nu),$$

where we have used Eq. (6.5) to express the curl as the sum of angular variables around the *directed* plaquette, as shown in Fig. 31. The interesting characteristic of the curl is that it is invariant under local gauge rotations. Consider a gauge transformation at the site n,

$$\theta_\mu(n)\to\theta_\mu(n)-\chi(n). \tag{6.7}$$

Equation (6.5) implies that under this gauge transformation

$$\theta_{-\mu}(n+\mu)\to\theta_{-\mu}(n+\mu)+\chi(n). \tag{6.8}$$

Therefore it is clear that $\theta_{\mu\nu}(n)$ is invariant to this operation. We can make this invariance principle appear more familiar by considering simultaneous gauge transformations one at site n involving the angle $\chi(n)$ and one at the site $n+\mu$ involving the angle $\chi(n+\mu)$. Then $\theta_\mu(n)$ transforms as

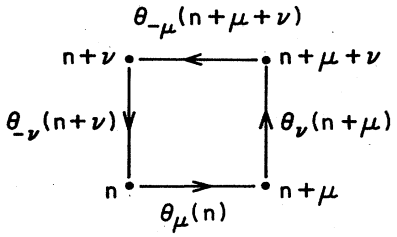


FIG. 31. A directed plaquette and its angular variables.

$$\theta_\mu(n) \to \theta_\mu(n) - \chi(n) + \chi(n+\mu) \to \theta_\mu(n) + \Delta_\mu\chi(n), \tag{6.9a}$$

while the curl is invariant,

$$\theta_{\mu\nu}(n) \to \theta_{\mu\nu}(n). \tag{6.9b}$$

These equations are recognized as a discrete form of the local gauge invariance of electrodynamics,

$$A_\mu \to A_\mu + \partial_\mu\chi, \quad F_{\mu\nu} \to F_{\mu\nu}. \tag{6.9c}$$

So, to make an Action which is locally gauge invariant, we use the variables $\theta_{\mu\nu}(n)$ and define

$$S = J\sum_{n,\mu\nu}[1 - \cos\theta_{\mu\nu}(n)]. \tag{6.10}$$

This Action has two important symmetries. It is locally gauge invariant because it is constructed from the lattice field $\theta_{\mu\nu}(n)$. It is also a *periodic* function of $\theta_{\mu\nu}(n)$, in the same way that the simpler spin model Eq. (6.4) is a periodic function of $\Delta_\mu\theta(n)$. We are interested in this type of Action for reasons which will be discussed when non-Abelian gauge theories are introduced.

Before studying the details of the Action equation (6.10) let us discuss its local invariance properties in a more geometric fashion. At this point our model consists of angular variables on links and symmetry operations on sites. To visualize the symmetry operations, imagine a planar "reference frame" at each site. They are not oriented relative to each other in any particular way. Interpret $\theta_\mu(n)$ as the relative orientation of the frame at $n+\mu$ to that at n. We see now that the definition equation (6.5) is necessary to achieve a consistent geometric interpretation. The local gauge invariance of the model is now seen as the requirement that the orientations of the local frames be unobservable. This is the discrete version of the interpretation of local gauge invariance first discussed by C. N. Yang and R. Mills (Yang and Mills, 1954). In particular, Eq. (6.9a) is the transformation implied by this geometric interpretation when simultaneous, independent rotations of the frames at sites n and $n+\mu$ are made.

The resemblance of this lattice theory to continuum electrodynamics has already been noted. Let us make this connection more suggestive by considering the Action equation (6.10) at weak coupling (low temperature), where $\theta_\mu(n)$ is expected to be a smooth, slowly varying field. We can then expand the cosine,

$$1 - \cos\theta_{\mu\nu} \approx 1 - [1 - \tfrac{1}{2}\theta_{\mu\nu}^2 + \dots] \approx \tfrac{1}{2}\theta_{\mu\nu}^2, \tag{6.11}$$

and replace the sum over lattice plaquettes by an integral

$$\sum \to \int \frac{d^4x}{a^4}, \tag{6.12}$$

in four dimensions. Then Eq. (6.10) becomes

$$S \approx J\int \frac{d^4x}{a^4}\,\frac{1}{2}\,\theta_{\mu\nu}^2. \tag{6.13}$$

Therefore, if we identify

$$\theta_{\mu\nu} = a^2 g F_{\mu\nu}, \quad J = 1/2g^2 \tag{6.14}$$

Eq. (6.13) becomes

$$S \approx \frac{1}{4}\int d^4x\, F_{\mu\nu}F_{\mu\nu}, \tag{6.15}$$

which is the Euclidean Action of electrodynamics. In doing this we have identified the lattice variable $\theta_\mu(n)$ and the electrodynamic potential $A_\mu(r)$ using Eq. (6.14),

$$\theta_\mu(n) = agA_\mu(r) \tag{6.16}$$

where g is a lattice theory's coupling constant.

The analysis leading from Eq. (6.10) to (6.15) is plausible but not foolproof even at low temperature. It is similar to spin-wave (or Gaussian) analyses of spin models, and these analyses are, as a matter of fact, usually quite good. However, one should really study the lattice theory thoroughly, explore its phase diagram, develop a renormalization group for it, and prove that the Actions [Eqs. (6.10) and (6.15)] lead to the same physics at large distance if T is sufficiently small. Later we shall carry out such a program for the planar spin model, Eq. (6.4), and see that the analogous statements do hold true. Such a thorough study of Eq. (6.10) has not been done, but several approximate discussions have appeared (Banks *et al.*, 1977), and they support our naive manipulations.

There is one last ingredient in this discussion we have not motivated, and that is the appearance of the coupling constant in Eq. (6.16). In fact, we have made this identification with an eye toward coupling charged matter fields to the gauge fields in a sensible, gauge-invariant fashion. To make this clear, recall some facts about local gauge invariance and relativistic field theories first stressed by J. Schwinger (Schwinger, 1959). In order to regulate the ultraviolet divergence of quantum electrodynamics, he considered point-split operators,

$$\bar\psi(r+\xi)\gamma_\mu\psi(r), \tag{6.17}$$

where ψ is a quantum field which creates a quantum of charge e. Under local gauge transformations the fields transform as

$$\begin{aligned}&\psi(r) \to \exp[ie\Lambda(r)]\psi(r),\\ &\bar\psi(r) \to \exp[-ie\Lambda(r)]\bar\psi(r),\\ &A_\mu(r) \to A_\mu(r) + \partial_\mu\Lambda(r),\end{aligned} \tag{6.18}$$

so it is easy to see that Eq. (6.17) is not gauge invariant so long as $\xi \neq 0$. Therefore that operator must be modified if it is to be used in dynamical calculations which are to be consistent. Schwinger suggested the operator

$$\bar\psi(r+\xi)\gamma_\mu \exp\left[ie\int_r^{r+\xi} A_\mu\, dx^\mu\right]\psi(r), \tag{6.19}$$

because the line integral restores the desired local gauge invariance. So, if matter fields ψ were placed on a space–time lattice, Eq. (6.19) would read

$$\bar\psi(r+\nu)\gamma_\mu \exp[iaeA_\nu(r)]\psi(r), \tag{6.20}$$

and the combination aeA_ν has emerged in a natural way.

B. Gauge-invariant correlation functions, physical interpretations, and phase diagrams

The partition function of Abelian lattice gauge theory is

$$Z=\prod_{r,\mu}\int_0^{2\pi} d\theta_\mu(r)\exp\left\{-\frac{1}{2g^2}\sum[1-\cos(\Delta_\mu\theta_\nu-\Delta_\nu\theta_\mu)]\right\}. \tag{6.21}$$

We want to understand the phase diagram of the theory. Many of the lessons we learned from Ising gauge theory carry over to this discussion. In particular, Elitzur's theorem can be proved here, and again states that the local continuous gauge symmetry cannot break down spontaneously. Therefore the ground-state expectation value of $\cos\theta_\mu(r)$ vanishes. So, we turn to the gauge-invariant correlation function to distinguish ordered from disordered phases of the theory. Consider a directed contour C and the sum of angular variables around C,

$$\exp\left\{i\sum_C\theta_\mu(r)\right\}. \tag{6.22}$$

The same argument that showed us that $\theta_{\mu\nu}(r)$ is gauge invariant generalizes to a closed loop of any size and proves that Eq. (6.22) is locally gauge invariant. Therefore we consider (Wilson, 1974)

$$\left\langle\exp i\sum_C\theta_\mu(r)\right\rangle=\prod_{r,\mu}\int_0^{2\pi}d\theta_\mu(r)\times\exp\left[i\sum_C\theta_\mu(r)\right]\exp(-S)/Z. \tag{6.23}$$

Suppose $g^2\gg 1$. Then we can estimate Eq. (6.23) using standard high-temperature expansion methods. For large loops C, low orders in the expansion give vanishing contributions because

$$\int d\theta_\mu(r)\,e^{i\theta_\mu(r)}=0. \tag{6.24a}$$

Thus any exposed phase factor $\exp[i\theta_\mu(r)]$ within the integrand Eq. (6.23) produces a zero. However, if all the phases are canceled then we meet integrals,

$$\int_0^{2\pi}d\theta_\mu(r)=2\pi, \tag{6.24b}$$

which give finite contribution. Consider the exponential of the Action

$$\begin{aligned}\exp\left\{\frac{1}{2g^2}\sum\cos\theta_{\mu\nu}\right\}&=\prod_{r,\mu\nu}\exp\left\{\frac{1}{2g^2}\cos\theta_{\mu\nu}(r)\right\},\\&=\prod_{r,\mu\nu}\exp\left\{\frac{1}{4g^2}[e^{i\theta_{\mu\nu}(r)}+e^{-i\theta_{\mu\nu}(r)}]\right\},\\&=\prod_{r,\mu\nu}\sum_n\frac{1}{n!}\left\{\frac{1}{4g^2}[e^{i\sum_p\theta_\mu}+e^{-i\sum_p\theta_\mu}]\right\}^n,\end{aligned} \tag{6.25}$$

where $\sum_p$ means a sum over the links of the plaquette $(r,\mu\nu)$. If we pick out the $n=1$ term for each plaquette making up the minimal area bordered by the large contour C, we pick up a nonvanishing contribution to Eq. (6.23) which has the fewest possible powers of $1/4g^2$. This is visualized in Fig. 32. We have the estimate

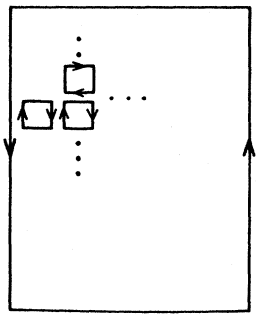


FIG. 32. The plaquettes which fill the interior of the contour C and produce a nonzero correlation function.

$$\left\langle\exp\left\{i\sum_C\theta_\mu(r)\right\}\right\rangle\approx\left(\frac{1}{4g^2}\right)^A=\exp[-\ln(4g^2)A], \tag{6.26}$$

where A is the area or number of plaquettes making up the minimal surface determined by C. This high-temperature expansion has a finite radius of convergence, so if we computed to higher order Eq. (6.26) would become

$$\left\langle\exp\left\{i\sum_C\theta_\mu(r)\right\}\right\rangle=\exp[-f(g^{-2})A], \tag{6.27}$$

where $f(g^{-2})$ is a finite function for g^2 large enough, and in an expansion its leading term is $\ln(4g^2)$. Clearly all these high-temperature calculations differ only in detail from the corresponding discussions for Ising lattice gauge theory.

Next consider weak coupling. If $g^2\ll 1$, we would guess that a naive continuum limit could be taken and the Action could be replaced by the Gaussian approximation, Eq. (6.15). Let us assume this and use suggestive continuum notation,

$$\left\langle\exp\left(ig\oint_C A_\mu dx\right)\right\rangle=\prod_{r,\mu}\int dA_\mu(r)\times\exp\left(-\frac{1}{4}\int d^4x F_{\mu\nu}^2+ig\oint_C A_\mu\,dx\right)\Big/Z, \tag{6.28}$$

where the vector potential A_μ ranges from $-\infty$ to $+\infty$. We must face two technical problems before we can evaluate Eq. (6.28). First, the expression is meaningful only if a particular gauge is chosen (Abers and Lee, 1973). Since the correlation function is manifestly gauge invariant, all gauge choices will lead to the same answer. We shall choose the Feynman gauge below. Second, the theory is formulated with a lattice cutoff to regulate potentially divergent self-energy effects, so four-dimensional *lattice* propagators will be encountered in the evaluation of Eq. (6.28). Lattice propagators have finite values at the origin and are typically well approximated by their naive continuum expressions elsewhere (Spitzer, 1964). Since we want only the dominate spatial dependence of the correlation function for large loops, we shall approximate the lattice propagator,

$$\langle A_\mu(x)A_\mu(0)\rangle = \delta_{\mu\nu}\Delta(x), \tag{6.29}$$

by

$$\Delta(x-y) = \Delta(0)\delta_{x,y} + \Delta'(x-y), \tag{6.30a}$$

where

$$\Delta'(x) = \frac{1}{2\pi^2}\,\frac{1}{|x|^2}, \text{ if } |x| > a$$
$$\Delta'(x) = 0, \text{ otherwise}. \tag{6.30b}$$

Finally, the functional integrals in Eq. (6.28) can be done by inspection since they are all Gaussians,

$$\left\langle \exp\left(ig\oint_C A_\mu dx^\mu\right)\right\rangle = \exp\left(-\frac{1}{2}g^2\oint_C\oint_C \langle A_\mu(x)A_\nu(y)\rangle dx_\mu dy_\nu\right) = \exp\left(-\frac{1}{2}g^2\oint_C\oint_C \Delta(x-y)dx_\mu dy_\mu\right). \tag{6.31}$$

Since the explicit evaluation of the integrals here will be rather tedious for electrodynamics, let us first understand the general features of the calculations and the types of answers expected. First observe that Eq. (6.31) can be neatly visualized. It states that the line element dx_μ interacts with the line element dy_μ through the exchange of a virtual particle described by the lattice propagator $\Delta(x)$, as shown in Fig. 33. Therefore the dependence of the correlation function on the dimensions of the loop reflects the long-distance character of the propagator. Suppose first that $\Delta(x)$ is short ranged—perhaps a Yukawa potential with a screening distance μ^{-1}. Then we obtain significant contributions to Eq. (6.31) only when dx_μ and dy_μ are within this distance. Therefore the correlation function satisfies the *perimeter* law. In electrodynamics $\Delta(x)$ falls off only as a power of $|x|$, so there will be additional dependence on the contour besides the simple perimeter effect. Later we shall understand this electrodynamic calculation in very simple terms.

To evaluate Eq. (6.31) explicitly, choose a rectangular contour as displayed already in Fig. 27. It is convenient to separate the calculation into two distinct parts as shown in Fig. 34. In Fig. 34(a) one of the four edges of the contour interacts with itself, while in Fig. 34(b) two different edges interact. Neighboring edges do not interact because in that case $dx_\mu dy_\mu = 0$. The first graph gives

$$\iint \Delta'(x-y)dx_\mu dy_\mu = \frac{2}{2\pi^2}\int_0^T dy\int_0^{y-a} dx\,\frac{1}{(y-x)^2}$$
$$= \frac{2}{2\pi^2}[T/a - \ln(T/a)], \tag{6.32}$$

and the second graph produces

$$\iint \Delta'(x-y)dx_\mu dy_\mu = -\frac{2}{2\pi^2}\int_0^T dx\int_0^T dy\,\frac{1}{R^2+(x-y)^2} = -\frac{2}{\pi^2}\left[\frac{T}{R}\tan^{-1}(T/R) - \frac{1}{2}\ln(1+T^2/R^2)\right]. \tag{6.33}$$

The results for the other edges are obtained from these calculations by inspection. We shall see that contours which are much longer in the temporal direction than the spatial direction have the simplest physical interpretation. Therefore we specialize to the case $T \gg R$, simplify Eqs. (6.32) and (6.33), and collect all the terms,

$$\oint\oint \Delta(x-y)dx_\mu dy_\mu \cong \left(\Delta(0) + \frac{1}{\pi a^2}\right)P - (1/\pi)T/R - 4/\pi^2\ln(R/a), \tag{6.34}$$

where $P = 2(T+R)$ is the perimeter of C. So, the final result is

$$\left\langle \exp\left(ig\oint_C A_\mu dx_\mu\right)\right\rangle \approx \exp\left\{-\frac{1}{2}g^2cP + \frac{g^2}{2\pi}\frac{T}{R} + \frac{2g^2}{\pi^2}\ln(R/a)\right\}, \tag{6.35}$$

where c is a constant depending on the lattice propagator at the origin. We see that the long-range character of the massless propagator $\Delta(x)$ has generated additional long-range effects in addition to the perimeter law. We shall see below that Eq. (6.35) is just Coulomb's law in disguise.

Compare Eqs. (6.35) and (6.26) and note that the correlation function has qualitatively different behavior at weak and strong coupling. Although the analysis of the weak coupling limit was rather naive, this result strongly suggests that Abelian lattice gauge theory is a two-phase

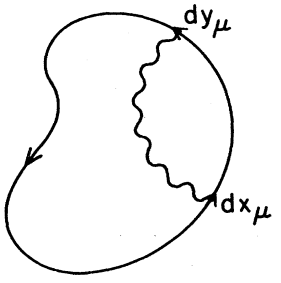


FIG. 33. The loop integral correlation function in weak coupling.

FIG. 34. Calculating the loop integral for a rectangular contour.

system.

The gauge-invariant correlation function plays a major role in our understanding of the qualitative features of many lattice gauge theories. We should relate it to some physical measurement. In fact, it measures the *force law* between static charged quanta which can be placed into the system. To see this, recall that one couples external currents J_μ to the electromagnetic field by adding a term,

$$e\int A_\mu(x)J_\mu(x)d^4s\,, \tag{6.36}$$

to the theory's Action. Since charge is conserved, we can take a *closed* current loop as a interesting, nontrivial external current. Denote the current loop C. J_μ is unity on the directed loop and zero elsewhere. Choose the contour to be a rectangle with $T\gg R$ as in Fig. 27 again. Then if we consider a fixed-τ slice of the system we see that at $x=0$, $J_0=-1$, while at $x=R$, $J_0=+1$. Therefore, from the transfer matrix point of view, we have a system with a static charge at $x=R$ and its antiparticle partner at $x=0$. The expectation value of the loop integral can then be interpreted as the ratio of the partition function for the system, including the external charges $Z(J)$, to that in which they are omitted, $Z(0)$,

$$\left\langle \exp\left(ie\oint_C A_\mu dx_\mu\right)\right\rangle = Z(J)/Z(0)\,. \tag{6.37}$$

If we call $\mathcal{F}(J)$ the free energy of the system with the charges, we have

$$\left\langle \exp\left(ie\oint_C A_\mu dx_\mu\right)\right\rangle = \exp\{-[\mathcal{F}(J)-\mathcal{F}(0)]\}, \tag{6.38}$$

where a factor of kT has been absorbed into $\mathcal{F}$ [cf. Eq. (2.6)]. Finally, recall from our discussion of the transfer matrix that free energy density can be identified with the ground-state energy density of a Hamiltonian description of the system, Eq. (3.48). In particular, taking $T\to\infty$ and appealing to the extensive character of the free energy, we have

$$\mathcal{F}(J)-\mathcal{F}(0)\propto T\,, \tag{6.39}$$

where the constant of proportionality is the energy difference between the ground state of the Hamiltonian with the charges included and with them omitted. Since the charges are static, this energy difference is pure potential,

$$\mathcal{F}(J)-\mathcal{F}(0)=V(R)\,T\,. \tag{6.40}$$

Therefore Eq. (6.38) becomes

$$\left\langle \exp\left(ie\oint_C A_\mu dx_\mu\right)\right\rangle = \exp[-V(R)T]\,, \tag{6.41}$$

so the potential energy required to hold the static charges a distance R apart is (Wilson, 1974)

$$V(R)=-\lim_{T\to\infty}\frac{1}{T}\ln\left\langle \exp\left(-ie\oint_C A_\mu dx_\mu\right)\right\rangle. \tag{6.42}$$

This is a beautiful connection between the partition function and transfer matrix approaches.

Now we can interpret our calculations of the loop integral correlation function. At strong coupling we have the "area law," which means

$$V(R)\sim|R|\,. \tag{6.43}$$

Thus it would require an infinite amount of energy to separate the charged sources! This result is called "quark confinement" for obvious reasons. Since confinement occurs so naturally in lattice gauge theories and since strong coupling problems are handled easily in this framework, lattice gauge theories give an unconventional perspective to field theory. Next, note that the "perimeter law" produces a quark potential,

$$V(R)\sim\text{const.}\,, \tag{6.44}$$

which indicates that charged quanta could easily be pulled free in such a theory. The forces of the underlying gauge theory are short ranged in this case. Finally, our free electrodynamic calculation Eq. (6.35) produces a potential,

$$V(R)\sim\text{const.}-(e^2/2)(1/R)\,, \tag{6.45}$$

which reproduces Coulomb's law! Thus the manipulations leading to Eq. (6.35) constitute "cracking a nut with a sledgehammer." It is amusing that lattice physics makes strong coupling problems and confinement appear natural, while weak coupling and free charges appear peculiar and difficult. The fact that both problems can be handled within a single formalism is, however, quite significant.

Before leaving this topic it is interesting to consider two-dimensional Abelian lattice gauge theory in some detail. In this case there will not be a phase transition separating weak and strong coupling. There are several ways to see this result. For example, if we repeated the weak coupling calculation Eq. (6.31), we would obtain an area law, since the two-dimensional massless propagator does not fall off with increasing x. Alternatively we could follow the same line of reasoning which showed that two-dimensional Ising gauge theory is equivalent to the one-dimensional Ising model. This would lead to the correspondence

$$\begin{bmatrix}\text{Two-dimensional}\\ \text{Abelian gauge}\end{bmatrix}\sim\begin{bmatrix}\text{One-dimensional}\\ \text{planar spin}\end{bmatrix}. \tag{6.46}$$

Since the one-dimensional planar spin model is disordered for all nonzero temperatures, the gauge theory inherits the area law for all nonzero couplings. Let us pursue this observation in detail and calculate the loop integral correlation function exactly (Balian *et al.*, 1975). Choose the temporal gauge so that $\exp[i\theta_\mu(r)]\equiv 1$ for μ in the τ direction. In other words,

$$\theta_0(n)=0\,, \tag{6.47}$$

is chosen to label representative spin configurations. The correlation function becomes

$$\left\langle \exp\left(i\sum_C\theta_\mu\right)\right\rangle = \int\prod d\theta_1(n) \times\exp\left\{\beta\sum_{n,\mu\nu}\cos\theta_{\mu\nu}+i\sum_C\theta_\mu\right\}\Big/Z\,, \tag{6.48}$$

where $\beta=1/2g^2$. If we choose the contour to be a rec-

tangle, then there are phase factors just on its horizontal (spatial) links. Equation (6.48) can be written in a manifestly gauge-invariant form by transforming to plaquette variables $\theta_{\mu\nu}(n)$. Note that in the temporal gauge

$$\theta_1(\tau,x) = -\sum_{\tau'<\tau} \theta_{\mu\nu}(\tau',x)\,. \tag{6.49}$$

Therefore the loop integral which appears in Eq. (6.48) can be written as

$$\sum_C \theta_\mu(n) = \sum_{\{P_c\}} \theta_{\mu\nu}(n)\,, \tag{6.50}$$

where $\{P_c\}$ is the set of plaquettes enclosed within C. Equation (6.50) is a lattice version of Stokes' law. Using Eq. (6.49) to change integration variables, we have

$$\left\langle \exp i \sum_C \theta_\mu \right\rangle = \frac{\int \prod_{\{P_c\}} d\theta_{\mu\nu}(n) \exp\left\{\beta \sum_{n,\mu\nu} \cos\theta_{\mu\nu} + i \sum_{P_s} \theta_{\mu\nu}\right\}}{\int \prod_{\{P_c\}} d\theta_{\mu\nu}(n) \exp\left\{\beta \sum_{n,\mu\nu} \cos\theta_{\mu\nu}\right\}}, \tag{6.51}$$

after cancelling some common factors from the numerator and denominator. Note that Eq. (6.51) is explicitly gauge invariant. In addition, it expresses the correlation function as the *product* of independent integrations over decoupled plaquettes,

$$\left\langle \exp\left(i \sum_C \theta_\mu\right)\right\rangle = \left[\frac{\int_0^{2\pi} d\theta_{\mu\nu} \exp(\beta\cos\theta_{\mu\nu} + i\theta_{\mu\nu})}{\int_0^{2\pi} d\theta_{\mu\nu} \exp(\beta\cos\theta_{\mu\nu})}\right], \tag{6.52}$$

where A is the number of enclosed plaquettes. The integrals in Eq. (6.62) are just Bessel functions of imaginary argument,

$$\left\langle \exp\left(i \sum_C \theta_\mu\right)\right\rangle = (I_1(\beta)/I_0(\beta))^A\,, \tag{6.53}$$

which gives us the expected area law for all coupling. Consider Eq. (6.53) in the limiting cases of strong and weak coupling. For $g^2 \gg 1$ where $\beta = 1/2g^2 \ll 1$,

$$I_1(\beta)/I_0(\beta) \approx \tfrac{1}{2}\beta = 1/4g^2\,, \tag{6.54}$$

so,

$$\left\langle \exp\left(i \sum_C \theta_\mu\right)\right\rangle \approx \exp[-\ln(4g^2)A]\,. \tag{6.55}$$

Choosing $g^2 \ll 1$,

$$I_1(\beta)/I_0(\beta) \approx 1 - 1/2\beta = 1 - g^2\,, \tag{6.56}$$

so

$$\left\langle \exp\left(i \sum_C \theta_\mu\right)\right\rangle \simeq \exp[\ln(1-g^2)A]\,, \approx \exp(-g^2 A)\,. \tag{6.57}$$

In summary, the two-dimensional model confines for all coupling g. The strength of the interquark potential is a smooth function of g which varies between $g^2|R|$ at weak coupling and $\ln(4g^2)|R|$ at strong coupling.

This example shows the utility of plaquette variables in computing the correlation function [Eq. (6.51)] in two dimensions. It is natural to ask why this elegant analysis does not generalize to higher dimensions. The point is that plaquette variables can only be chosen as independent integration variables in two dimensions. To see this, let us expose a nontrivial constraint among these variables. Consider a three-dimensional cube. Label the links comprising each face as shown in Fig. 35. There is a plaquette variable $\theta_{\mu\nu}$ associated with each of these oriented squares. The orientations have been chosen so that the sum of all six $\theta_{\mu\nu}$ corresponding to the six faces of cube sum to zero,

$$\sum_{\text{faces}} \theta_{\mu\nu} = 0\,. \tag{6.58}$$

Such constraints are manifestations of flux continuity (Gauss' law).

C. The quantum Hamiltonian formulation and quark confinement

Much of the physics of Abelian lattice gauge theory becomes more accessible using the τ-continuum Hamiltonian approach. The Hamiltonian is obtained by considering the Action with anisotropic couplings,

$$S = \beta_\tau \sum_{n,k} [1 - \cos\theta_{0k}(n)] - \beta \sum_{n,ik} \cos\theta_{ik}(n)\,, \tag{6.59}$$

where spatial links are labeled with Latin indices and the temporal direction τ with the index 0. As usual we choose the gauge $\theta_0(n) = 0$. Then τ-independent gauge transformations are still manifest local symmetries of the system, and they are constructed as in our parallel discussion of the Ising gauge theories. We shall obtain an operator formalism for these symmetry operations below.

To obtain the Hamiltonian, note that the first term of Eq. (6.59) simplifies because

$$\theta_{0k}(n) = \theta_k(n + \hat{\tau}) - \theta_k(n)\,. \tag{6.60}$$

We know from the general discussion in Sec. III.C that β_τ will be forced to infinity in the τ-continuum limit. This will force $\theta_{0k}(n)$ to be small and slowly varying. We can, therefore, make the approximation

$$1 - \cos\theta_{0k} \approx \tfrac{1}{2}\theta_{0k}^2\,, \approx \frac{1}{2} a_\tau^2 \left(\frac{\partial\theta_k}{\partial\tau}\right)^2\,, \tag{6.61}$$

where a_τ denotes the lattice spacing in the τ direction. Sums over the lattice can be replaced by integrals over the continuum τ axes and sums over the spatial lattice,

$$\sum_{n,k} \to \int \frac{1}{a_\tau} d\tau \sum_{\mathbf{n},k} \tag{6.62}$$

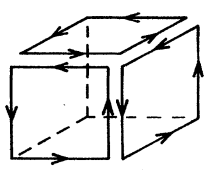


FIG. 35. Oriented links on a three-dimensional cube.

Now the Action becomes

$$S \to \int d\tau \left\{ \frac{1}{2}\beta_\tau a_\tau \sum_{\mathbf{n},k} [\dot{\theta}_k(\tau,\mathbf{n})]^2 - \frac{1}{a_\tau}\beta \sum_{\mathbf{n},ik} \cos\theta_{ik}(\tau,\mathbf{n}) \right\}. \tag{6.63}$$

So, to obtain a sensible limit, β_τ must scale as a_τ^{-1} and β must scale as a_τ. The constant of proportionality will be identified with the coupling constant g^2 which is held fixed,

$$\begin{aligned} &\beta_\tau = g^2/a_\tau \to \infty\,, \\ &\beta = a_\tau/g^2 \to 0\,. \end{aligned} \tag{6.64}$$

The identification of g^2 with the electrodynamic coupling constant will become clear later in this section. Finally, to set up a proper Hamiltonian, we must set up a Hilbert space for each fixed-τ surface. There is a quantum field $\theta_k(\mathbf{n})$ and its conjugate momentum density $L_k(\mathbf{n})$. Their commutation relations are postulated to be

$$[L_i(\mathbf{n}), \theta_k(\mathbf{n}')] = i\delta_{ik}\,\delta_{\mathbf{n},\mathbf{n}'}\,. \tag{6.65}$$

Inspecting Eq. (6.63) and using our experience gained from Sec. III. C, we identify the Hamiltonian

$$aH = \sum_{\mathbf{n},k} \tfrac{1}{2} g^2 L_k^2(\mathbf{n}) - (1/g^2) \sum_{\mathbf{n},ik} \cos\theta_{ik}(\mathbf{n})\,, \tag{6.66}$$

where "a" is the spatial lattice spacing. The factor of "a" is inserted into Eq. (6.66) so that H has the correct units.

We want to understand the formal and physical content of these last two equations. First consider τ-independent local gauge transformations. Such a transformation should add a common angle to all the angular variables associated with links originating on the site $\mathbf{n}$. An operator which does this is

$$G_\chi(\mathbf{n}) = \exp\left\{ i \sum_{\pm j} L_j(\mathbf{n})\chi \right\}. \tag{6.67}$$

To verify this, note that Eq. (6.65) allows us to interpret the $L_i(\mathbf{n})$ as planar angular momentum operators. Thus Eq. (6.67) is a rotation operator and induces the change $\theta_i(\mathbf{n}) \to \theta_i(\mathbf{n}) + \chi$ for all links originating at $\mathbf{n}$. A gauge transformation which acts at all the sites and induces spatially dependent rotations through an arbitrary angle $\chi(\mathbf{n})$ is

$$G(\chi) = \exp\left\{ i \sum_{\mathbf{n},j} L_j(\mathbf{n})\chi(\mathbf{n}) \right\}. \tag{6.68}$$

It is easy to check that

$$G(\chi)\theta_k(\mathbf{n})G^{-1}(\chi) = \theta_k(\mathbf{n}) - \chi(\mathbf{n}) + \chi(\mathbf{n}+k) = \theta_k + \Delta_k\chi\,, \tag{6.69}$$

which is an ordinary τ-independent gauge transformation. The Hamiltonian is, of course, locally gauge invariant,

$$G(\chi)HG^{-1}(\chi) = H\,. \tag{6.70}$$

In addition, Elitzur's theorem assures us that the physical space of states is also locally gauge invariant,

$$G(\chi)|\rangle = |\rangle\,. \tag{6.71}$$

Next we should like to understand the commutation relation Eq. (6.65) in more detail. In particular, we should associate the link angular momentum variable $L_i(\mathbf{n})$ with a conventional variable of electrodynamics. From earlier discussions we have the identification

$$\theta_k(\mathbf{n}) = agA_k(\mathbf{n})\,. \tag{6.72}$$

This allows Eq. (6.65) to be written in a suggestive form,

$$(g^2/a)[L_i(\mathbf{n}), A_j(\mathbf{n}')] = i\delta_{ik}(1/a^3)\delta_{\mathbf{n},\mathbf{n}'}\,. \tag{6.73}$$

If we identify $a^{-3}\delta_{\mathbf{n},\mathbf{n}'}$ as the discrete form of the Dirac delta function and recall the continuum theory's commutation relation,

$$[E_i(\mathbf{r}), A_j(\mathbf{r}')] = i\delta_{ij}\delta(\mathbf{r}-\mathbf{r}')\,, \tag{6.74}$$

where $\mathbf{E}(r)$ is the conventional electric field, we can identify

$$E_i(\mathbf{n}) = (g/a^2)L_i(\mathbf{n})\,. \tag{6.75}$$

This teaches us an important fact about the lattice theory. Since $L_i(\mathbf{n})$ is an angular momentum operator, i.e., it is conjugate to the angular variable $\theta_k(\mathbf{n})$, its spectrum is discrete,

$$L_i(\mathbf{n}) = 0, \pm 1, +2, \ldots\,. \tag{6.76}$$

Therefore Eq. (6.75) implies that the electric flux on a link, $a^2E_i(\mathbf{n})$, is quantized in units of the charge g. Electric flux cannot subdivide into arbitrarily small units on individual links! The origin of this aspect of the lattice theory is the fact that $agA_k(\mathbf{n})$ is an angular variable. If we had formulated the theory without this condition, the quantization of electric flux would not have occurred. We take this path, however, because it is the most natural one in non-Abelian lattice gauge theories where the underlying gauge group is compact. This point will be discussed again in a later section.

If we write the Hamiltonian using Eq. (6.75) we find

$$H = a^3 \sum_{\mathbf{n},k} \tfrac{1}{2} E_k^2(\mathbf{n}) - (1/g^2a) \sum_{\mathbf{n},ik} \cos\theta_{ik}(\mathbf{n})\,. \tag{6.77}$$

So, the first term is just the simplest lattice form of $\frac{1}{2}\int d^3x\mathbf{E}^2$. The second term should, therefore, be a lattice form of $\frac{1}{2}\int d^3x\mathbf{B}^2$. To see this, recall that in the continuum electrodynamics $B_i = \varepsilon_{ijk}\partial_j A_k$. Therefore we should make the identification

$$\theta_{jk} = a^2 gB_i \quad (ijk \text{ cyclic})\,. \tag{6.78}$$

Then in a smooth, classical continuum limit,

$$\begin{aligned} -\frac{1}{g^2 a}\sum \cos\theta_{ik} &\approx -(1/g^2a) \sum_{\mathbf{n},ij} (1 - \tfrac{1}{2}\theta_{ik}^2) \\ &\approx a^3 \sum_{\mathbf{n},k} \tfrac{1}{2} B_k^2(\mathbf{n}) + \text{const.}\,, \end{aligned} \tag{6.79}$$

and conventional electrodynamics is retrieved.

Now consider the phase diagram of Abelian lattice gauge theory using the quantum Hamiltonian. Place static charges $\pm g$ a distance R apart and consider their potential $V(R)$. We must first introduce some formalism to describe these charges. This will be done by inventing an angular variable $\theta(\mathbf{n})$ which resides on sites. There will be a momentum conjugate to $\theta(\mathbf{n})$, $L(\mathbf{n})$, with the property

$$[L(\mathbf{n}), \theta(\mathbf{n}')] = i\delta_{\mathbf{n},\mathbf{n}'}\,. \tag{6.80}$$

We want to make a formalism in which a charge g is a source of g units of electric flux in accord with Gauss' law. Therefore matter fields will be described by phase variables

$$\exp\{\pm i\theta(n)\}, \tag{6.81}$$

and the generator of local gauge transformations will now consist of two pieces,

$$\sum_{\pm j} L_j(\mathbf{n}) + L(\mathbf{n}), \tag{6.82}$$

so that Eq. (6.68) is generalized to

$$G(\chi) = \exp\left\{ i \sum_{\mathbf{n},j} L_j(\mathbf{n})\chi(\mathbf{n}) + i \sum_{\mathbf{n}} L(\mathbf{n})\chi(\mathbf{n}) \right\}. \tag{6.83}$$

The value in this construction lies in the result,

$$G(\chi) e^{\pm i\theta(\mathbf{n})} G^{-1}(\chi) = e^{\pm i\chi(\mathbf{n})} e^{\pm i\theta(\mathbf{n})}, \tag{6.84}$$

which should be compared with Eq. (6.18), the gauge transformation properties of an ordinary charged field. Therefore Eq. (6.84) implies that $\exp[\pm i\theta(n)]$ carries $\pm g$ units of charge.

Next we need an operator to produce a state with static charges $\pm g$. The possibility

$$\exp\{-i\theta(\mathbf{0})\} \exp\{i\theta(\mathbf{R})\}, \tag{6.85}$$

comes to mind, but it is not locally gauge invariant. To remedy this problem, we follow Schwinger and place a lattice version of $\exp(ig\int_0^{\mathbf{R}} \mathbf{A}\cdot d\mathbf{x})$ into Eq. (6.85),

$$\theta_C(\mathbf{0},\mathbf{R}) = e^{-i\theta(\mathbf{0})} \exp\left\{ i \sum_C \theta_i(\mathbf{n}) \right\} e^{i\theta(\mathbf{R})}, \tag{6.86}$$

where C is any lattice contour which runs from 0 to R. It is interesting to interpret $\theta_C(\mathbf{0},\mathbf{R})$ physically. Since

$$[L_i(\mathbf{n}), e^{\pm i\theta_j(\mathbf{n}')}] = \pm \delta_{ij}\delta_{\mathbf{n}',\mathbf{n}} e^{\pm i\theta_j(\mathbf{n}')}, \tag{6.87}$$

each operator on the contour C of Eq. (6.86) raises the eigenvalue of L_i on that link by one unit. In other words, the electric flux passing through that link has been increased by g units. Therefore the operator $\theta_C(\mathbf{0},\mathbf{R})$ produces a state with the correct amount of flux to satisfy Gauss' law. This is another way of viewing its property of local gauge invariance.

It is also interesting to determine the characteristics of that contour C which leads to a state of minimum energy. This question will certainly depend on the value of the coupling constant g. It lies at the heart of the quark confinement question. First suppose that $g^2 \gg 1$. Then the electric term in H dominates, so we can imagine treating it as the large, unperturbed piece of the Hamiltonian and account for magnetic effects in perturbation theory. The zeroth order Hamiltonian

$$aH_0 = \tfrac{1}{2}g^2 \sum_{\mathbf{n},k} L_k^2(\mathbf{n}), \tag{6.88}$$

is also trivially soluble. In the absence of static charges its ground state must have the property

$$L_k^2(\mathbf{n})|0\rangle = 0, \tag{6.89}$$

for all links. In other words, the strong coupling ground state is an eigenstate of electric flux with eigenvalue zero. Now consider the state

FIG. 36. A confining flux tube.

$$\theta_C(\mathbf{0},\mathbf{R})|0\rangle. \tag{6.90}$$

Since each link of the contour raises the value of L_k^2 to unity, that contour of shortest length gives the smallest contribution to the unperturbed Hamiltonian. This means that the flux travels directly between the charges in a tube, as pictured in Fig. 36, and the energy of the state grows linearly with R,

$$V(R) \sim |R|, \tag{6.91}$$

in agreement with our earlier analysis.

If g^2 were not large, then the magnetic effects in the Hamiltonian would become important. If we do perturbation theory around aH_0, it is clear that the lowest energy state of the $\pm g$ charges will have its flux occupying paths more complicated than that shown in Fig. 36. As g^2 is decreased to a critical value we expect the coefficient of the linear potential Eq. (6.91) to vanish abruptly. Something like this must occur because at small coupling the Hamiltonian should be well approximated by $\frac{1}{2}\int d^3x(\mathbf{E}^2+\mathbf{B}^2)$, as discussed in Eq. (6.79), and then Coulomb's law holds. In this region the flux between charges spreads out into the characteristic dipole pattern. In the lattice theory, where flux is quantized, this presumably means that very random complicated contours like that shown in Fig. 37 lead to lower energy states than the flux tube. Strong coupling expansions (Kogut *et al.*, 1976) and partition function analyses (Banks *et al.*, 1977) support these views, but really precise, convincing work is lacking.

VII. THE PLANAR HEISENBERG MODEL IN TWO DIMENSIONS

A. Introductory comments and motivation

We have remarked earlier that there are important similarities between lattice gauge theories and spin systems. Two-dimensional lattice gauge theories are equivalent to one-dimensional spin systems. In addition, four-dimensional gauge systems share many properties in common with two-dimensional spin systems. These similarities, in fact, motivated Wegner to construct the Ising gauge system (Wegner, 1971). We shall pursue this point by studying the two-dimensional planar model in detail. The same type of analysis has been partially carried out for Abelian lattice gauge theory in four dimensions (Banks *et al.*, 1977).

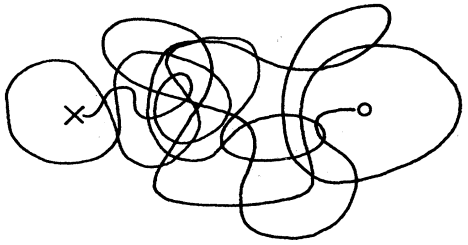


FIG. 37. Important flux configurations at weak coupling.

Let us recall some facts about the planar model. It undergoes a phase transition without the appearance of a spontaneous magnetization. Its low-temperature phase contains massless spin waves. The periodicity of its angular variable $\theta(n)$ is irrelevant in this phase. Its high-temperature phase is completely disordered, and the angular character of $\theta(n)$ is important here. All of these properties have correspondences in the Abelian gauge theory in four dimensions. The analogy runs even deeper. In an approximate renormalization group scheme the phase transition of the planar model is driven by the appearance of topological singularities—vortices—in the field $\theta(n)$. These vortices condense for $T > T_c$ and disorder the spin–spin correlation function. Similar analyses for four-dimensional Abelian gauge theories suggest that topological singularities also drive them into a disordered phase. These vortices are closely related to magnetic monopoles (Banks, *et. al.*, 1977). The condensation phenomenon leads to quark confinement in the strongly coupled phase, since electric flux cannot easily penetrate such a medium. These points will be discussed further in a later section.

The idea that condensation of field configurations having topological significance controls the phase diagrams of gauge theories appears to be more general than the specific examples we shall study here (t' Hooft, 1978). However, the planar model can be analyzed in detail, so it is a good place to learn about this phenomenon.

Much of our discussion will be rather technical. Some intuitive, physical ideas about the planar model will be discussed in the next section (Kosterlitz and Thouless, 1973). The reader is advised to refer to various review articles on two-dimensional systems (Kosterlitz and Thouless, 1977) for more extensive physical discussions.

B. The physical picture of Kosterlitz and Thouless

The Action of the two-dimensional planar model was recorded earlier,

$$S = J \sum_{\langle nm \rangle} \cos(\theta_n - \theta_m) = \tfrac{1}{2} J \sum_{\langle nm \rangle} [e^{i(\theta_n - \theta_m)} + \text{h. c.}], \tag{7.1}$$

where $\langle nm \rangle$ indicates a sum over nearest-neighbor spins on a two-dimensional square lattice. In parallel with our discussion of the Abelian gauge theory, we shall examine the spin–spin correlation function at high and low T and argue that the system has two phases. At high T the correlation function,

$$\langle e^{i\theta_0} e^{-i\theta_n} \rangle = \int \prod_m d\theta_m \, e^{i(\theta_0 - \theta_n)} \times \exp\left\{ - \frac{J}{kT} \sum_{\langle nm \rangle} \cos(\theta_n - \theta_m) \right\} \Big/ Z, \tag{7.2}$$

can be estimated using a high-temperature expansion. Since

$$\int_0^{2\pi} d\theta_m = 2\pi, \quad \int_0^{2\pi} d\theta_m \, e^{i\theta_m} = 0, \tag{7.3}$$

our usual arguments show that the first nonzero term contributing to the high-T expansion of Eq. (7.2) is of the order $(J/kT)^{|n|}$. Therefore

$$\langle \exp\{i(\theta_0 - \theta_n)\} \rangle \simeq (J/kT)^{|n|} = \exp\{ - |n| \ln(kT/J) \}. \tag{7.4}$$

Therefore the correlation function falls off exponentially in the distance between the spins for T sufficiently high.

Next consider the correlation function at low temperature. In this case the absence of significant thermal fluctuations suggests that θ_n varies slowly and smoothly throughout the system. Then the cosine in Eq. (7.1) can be expanded and only the quadratic term need be accounted for,

$$S \approx \tfrac{1}{2} J \sum_{n,i} [\Delta_i \theta(n)]^2. \tag{7.5}$$

Then the correlation function becomes,

$$\langle \exp\{i(\theta_0 - \theta_n)\} \rangle \cong \int \prod_m d\theta_m \, e^{i(\theta_0 - \theta_n)} \times \exp\left\{ - \frac{J}{2kT} \sum (\Delta\theta)^2 \right\} \Big/ Z, \tag{7.6}$$

which can be evaluated simply because it involves only Gaussian integrals. If $\Delta(n)$ is the lattice propagator for a massless field, then

$$\langle \exp\{i(\theta_0 - \theta_n)\} \rangle \sim \exp[(kT/J)\Delta(n)]. \tag{7.7}$$

For large $|n|$ the lattice propagator is well approximated by the continuum propagator,

$$\Delta(n) \approx - (1/2\pi) \ln |n|, \quad |n| \gg 1 \tag{7.8}$$

which grows slowly at large distances. Substituting into Eq. (7.7), we have

$$\langle \exp i(\theta_0 - \theta_n) \rangle \simeq \left(\frac{1}{|n|} \right)^{kT/2\pi J}. \tag{7.9}$$

This result teaches us several facts. First, it shows that the planar model never magnetizes. This follows because as $|n| \to \infty$ the expectation value of the product of two spins should approach the product of their expectation values,

$$\langle e^{i(\theta_0 - \theta_n)} \rangle \sim \langle e^{i\theta_0} \rangle \langle e^{-i\theta_n} \rangle. \tag{7.10a}$$

But Eq. (7.9) falls to zero as $|n| \to \infty$, so

$$\langle e^{i\theta_0} \rangle = 0, \tag{7.10b}$$

identically. This shows that our approximate analysis is in accord with rigorous theorems (Mermin and Wagner, 1966) which prove that continuous global symmetries cannot break down spontaneously in systems with nearest-neighbor coupling in two dimensions. Second, Eq. (7.9) shows that the theory has a line of critical points for T sufficiently small. Recall from the discussion of Sec. II that at the critical temperature of a system the spin–spin correlation function is expected to be power behaved,

$$\langle \mathbf{s}(0) \cdot \mathbf{s}(n) \rangle \sim \left(\frac{1}{|n|} \right)^{\eta}, \tag{7.11a}$$

where η is a standard critical index. For the planar

model, treated in the spin-wave approximation [Eq. (7.5)],

$$\eta = kT/2\pi J\,, \tag{7.11b}$$

and we have a fixed line. In the language of Sec. II this is a violation of the most naive universality criteria.

Although this analysis of the low-T region of the model is rather informal and is at the same level as our discussion of the Abelian gauge theory at low T, it suggests that the planar model has two phases. What then is the nature of the system's phase transition? An intriguing answer was provided by the seminal work of J. M. Kosterlitz and D. J. Thouless (Kosterlitz and Thouless, 1973). They suggested that the periodicity of the variable θ_n allowed for singular spin configurations—vortices—to appear in the system at sufficiently high temperatures, and that they disorder the spin–spin correlation function. They considered the Gaussian approximation Eq. (7.5) to the model, but they supplemented it with a lattice cutoff and retained the periodicity of its Action. That Action leads to the equation of motion

$$\nabla^2\theta = 0 \quad (0 < \theta < 2\pi)\,. \tag{7.12}$$

These authors also pointed out that solutions to Eq. (7.12) could be labeled by their winding number,

$$\oint \nabla\theta \cdot d\mathbf{l} = 2\pi q, \quad q = 0, \pm 1, \pm 2, \ldots\,. \tag{7.13}$$

In the spin-wave analysis leading to Eq. (7.9) we ignored the periodicity problem, so we effectively only dealt with the $q = 0$ sector. A spin configuration having $q = 1$ is shown in Fig. 38. It is clear that such a configuration disorders the system significantly, since it stirs the spins over the entire two-dimensional plane. There is a famous plausibility argument (Kosterlitz and Thouless, 1973) suggesting that vortices are irrelevant at low T but drive a phase transition at a moderate T_c. For $T \geq T_c$ it also suggests that the ground state of the system is a vortex condensate. The argument begins by evaluating the Action of a vortex (Fig. 38) using Eq. (7.4),

$$S \cong \pi J \ln(R/a)\,, \tag{7.14}$$

where R is the linear dimension of the two-dimensional world and a is the lattice spacing. Thus the vortex Action diverges as $R \to \infty$. In addition, the fact that S diverges as $a \to 0$ reminds us that a vortex is a singular solution to the continuum version of Laplace's equation. One might think that since S diverges as $R \to \infty$ vortex configurations would be irrelevant to the thermodynamics of the planar model. However, to estimate their importance, consider the free energy of a vortex,

$$F = \text{Action} - \text{Temperature} \times \text{Entropy}\,. \tag{7.15}$$

We must estimate the entropy of a free vortex. The entropy is just the logarithm of the multiplicity of the configuration. Since a vortex is completely specified by the location of its origin, its multiplicity is just the number of sites of the system,

$$\text{Entropy} = k \ln(R/a)^2\,. \tag{7.16}$$

Therefore

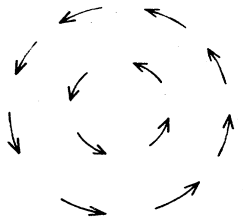


FIG. 38. A vortex.

$$F = (\pi J - 2kT)\ln(R/a)\,. \tag{7.17}$$

The competition between the logarithms of Eqs. (7.14) and (7.16) leads us to the interesting conclusion that as T is increased one reaches a point

$$T_c = \pi J/2k\,, \tag{7.18}$$

where F vanishes. Therefore, for all temperatures equal to and greater than T_c, we expect vortices to blend into the ground state—vortex condensation.

We can now summarize the Kosterlitz–Thouless picture of the phases of the model. At low temperature spin waves exhaust the relevant spin configurations of the theory. Spin–spin correlation functions fall off slowly with distance. Free vortices do not exist, but bound states of vortex–antivortex pairs can occur. They do not disorder the system significantly since they affect spins only over small regions. As the temperature is raised, the size of the vortex–antivortex bound states grows until T_c is reached where it diverges. Then free vortices exist and the ground state is a vortex condensate of indefinite global vorticity. The system is disordered and the spin-spin correlation function falls exponentially.

C. The planar model in the periodic Gaussian approximation

We wish to make the physical picture of the phases of the planar model quantitative. To do this we shall introduce a slightly simpler model, the periodic Gaussin planar model, which shares the same symmetries with the Action equation (7.1). To motivate the new model note that the periodic character of the Action equation (7.1) permits us to write a Fourier series for each bond,

$$\exp\{-\beta[1 - \cos(\theta_n - \theta_m)]\} = \sum_{l=-\infty}^{\infty} \exp[il(\theta_n - \theta_m)] I_l(\beta)\,, \tag{7.19}$$

where $I_l(\beta)$ is the Bessel function of imaginary argument. If β were large (low temperature), $I_l(\beta)$ would be well approximated by a Gaussian and Eq. (7.19) would become

$$\exp\{-\beta[1 - \cos(\theta_n - \theta_m)]\} \approx (1/\sqrt{2\pi\beta}) \sum_{l=-\infty}^{\infty} \exp[il(\theta_n - \theta_m)] \times \exp(-l^2/2\beta)\,. \tag{7.20a}$$

Note that the right-hand side of this expression has the same essential ingredients as the planar model based on the cosine interaction. It has the global Abelian

symmetry and it preserves the periodicity in the θ_n variables. Therefore, although it is numerically close to the original planar model only for low temperatures, we can consider the right-hand side of Eq. (7.20a) at all temperatures and study its phases. Since it shares the same general characteristics with the planar model, the two models should have essentially identical phase diagrams. It is called the periodic Gaussian model (Villain, 1975; R. Savit, 1977). Of course, the real reason we study it is that it can be analyzed elegantly, while the original cosine model has only been studied numerically.

How well does the periodic Gaussian model represent the physics of the planar model? If we accepted the strongest universality argument, then we would believe that their critical behaviors should be identical, since the models share the same symmetries. However, critical lines and nonuniversal behavior are common in two-dimensional physics, so this argument is not convincing. We shall see that in the periodic Gaussian approximation spin waves and vortices do not interact. In the planar model there are more severe nonlinearities which could effect the critical indices. It is not known, however, if this really happens. This and related points are subjects of active research. Experiments may, in fact, resolve these uncertainties. As will be discussed in the next section, the approximate renormalization group calculations of the periodic Gaussian model (Kosterlitz, 1974) predict that the critical index η governing the spin–spin correlation function is 1/4 at the critical temperature. A low-energy theorem (Nelson and Kosterlitz, 1977) relates it to the critical behavior of superfluid helium films, and the data (Webster *et al.*, 1979) support this prediction. It would be interesting if other predictions of the theory could be confronted with experiment.

Let us develop the formalism of the periodic Gaussian model. We shall see that it is possible to separate the spin waves of the model from its vortices without any additional approximations. We shall also see that the vortex sector of the model is equivalent to the two-dimensional Coulomb gas. This point will shed light on the original spin system, since the phases of the Coulomb gas can be understood physically. To begin, write Eq. (7.20) with more elegant notation,

$$\exp\{-\beta[1-\cos(\Delta_\mu\theta(r))]\} \to \frac{1}{\sqrt{2\pi\beta}}\sum_{n_\mu(r)=-\infty}^{\infty}\exp(in_\mu\Delta_\mu\theta)\times\exp(-n_\mu^2/2\beta) \tag{7.20b}$$

where r labels the two-dimensional lattice and $n_\mu(r)$ is a vector field. Substituting into the partition function,

$$Z=\int\prod_{r'}d\theta(r')\times\prod_{r,\mu}\sum_{n_\mu(r)=-\infty}^{\infty}\exp(in_\mu\Delta_\mu\theta)\exp(-n_\mu^2(r)/2\beta), \tag{7.21}$$

where we have dropped an overall multiplicative constant. The integrals over $\theta(r)$ are now trivial. Some thought shows that each integral generates a constraint,

$$\Delta_\mu n_\mu(r)=0. \tag{7.22}$$

Now,

$$Z=\prod_{r,\mu}\sum_{n_\mu(r)=-\infty}^{\infty}\delta_{\Delta_\mu n_\mu(r),0}\exp(-n_\mu^2(r)/2\beta). \tag{7.23}$$

It is best to solve the constraint explicitly,

$$n_\mu(r)=\varepsilon_{\mu\nu}\Delta_\nu n(r), \tag{7.24}$$

where $n(r)$ is a scalar, integer-valued field on the lattice. Equation (7.24) expresses the usual fact that a divergence-free vector field is a pure curl. The partition function becomes

$$Z=\sum_{n(r)=-\infty}^{\infty}\exp\left\{-(1/2\beta)\sum_{r,\mu}[\Delta_\mu n(r)]^2\right\}, \tag{7.25}$$

because

$$n_\mu(r)n_\mu(r)=[\Delta_\mu n(r)]^2. \tag{7.26}$$

So, we have transformed our original model into that of a nearest-neighbor coupled integer-valued field subject to the temperature $T^*=T^{-1}$. Equation (7.25) is the "interface roughening" model (Chiu and Weeks, 1976) of crystal growth, but it is not directly useful to us because at low temperatures many terms in the sum over $n(r)$ must be kept in evaluating Eq. (7.25). Clearly we would like to replace the integer-valued field $n(r)$ by an ordinary scalar field $\phi(r)$ and still preserve the periodic character of the original model. The necessary manipulation is provided by the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty}g(n)=\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}d\phi g(\phi)\,e^{2\pi im\phi}, \tag{7.27}$$

where g is an arbitrary function. The sum over the integer variable m insures that the periodicity of the original Action is preserved. Applying this identity to Eq. (7.25) gives

$$Z=\int_{-\infty}^{\infty}\prod_r d\phi(r)\sum_{m(r)=-\infty}^{\infty}\exp\left\{-(1/2\beta)\sum_{r,\mu}(\Delta_\mu\phi)^2+2\pi i\times\sum_r m(r)\phi(r)\right\}. \tag{7.28}$$

As we analyze this expression further we shall be able to identify $\phi(r)$ with the spin waves of the original model and $m(r)$ with its vortices! Equation (7.28) is pleasantly simple: the vortices act as sources for the spin waves, which are ordinary massless fields. The identification of the $m(r)$ with vortex variables is made clearer if we integrate out the spin waves in Eq. (7.28). This integration is a standard Gaussian, so we obtain without difficulty

$$Z=Z_{\rm sw}\sum_{m(r)=-\infty}^{\infty}\exp\left\{-2\pi^2\beta\sum_{r,r'}m(r)G(r-r')m(r')\right\}, \tag{7.29}$$

where $Z_{\rm sw}$ is the spin-wave contribution produced by the Gaussian integrations and $G(r-r')$ is the lattice propagator for a massless field. Since spin waves alone do not drive a phase transition, we shall not discuss $Z_{\rm sw}$. The lattice propagator $G(r)$ satisfies

$$\Delta^2G(r)=\delta_{r,0}, \tag{7.30}$$

where Δ^2 is a discrete form of the second derivative. For example, we take

$$\Delta^2 = \Delta_x^2 + \Delta_y^2, \tag{7.31a}$$

where

$$\Delta_x^2 f(r) = f(r+x) + f(r-x) - 2f(r). \tag{7.31b}$$

Then one can easily verify that

$$G(r) = \int_{-\pi}^{\pi} \frac{dk_x}{2\pi} \int_{\pi}^{\pi} \frac{dk_y}{2\pi} \frac{e^{ikr}}{(4 - 2\cos k_x - 2\cos k_y)}. \tag{7.32}$$

Let us examine some of the properties of $G(r)$. For large r it is well approximated by the continuum massless scalar propagator of two dimensions (Spitzer, 1964)

$$G(r) \approx -(1/2\pi)\ln(r/a) - 1/4, \quad (|r| \gg 1). \tag{7.33}$$

Therefore the variables $m(r)$ interact among themselves through a logarithmic potential. One can easily check that vortex spin configurations as sketched in Fig. 38 experience the same force law: vortices of opposite vorticity experience an attractive logarithmic potential, and vortices of the same vorticity repel one another through a logarithmic potential. Next consider the self-mass of the field $m(r)$. This comes from the $r = r'$ piece of the sum in Eq. (7.29). It is easy to see from Eq. (7.32) that $G(0)$ is infrared divergent,

$$G(0) \approx (1/2\pi)\ln(R/a), \tag{7.34}$$

where R is the linear dimension of the two-dimensional plane. We recognize this as the Action equation (7.14) of an isolated vortex. Of course, this infrared problem means that we should treat Eqs. (7.14) and (7.32) quite carefully. It is best to decompose $G(r)$ into two pieces, one of which is infrared finite and the other is not,

$$G(r) = G'(r) + G(0), \tag{7.35a}$$

where

$$G'(r) = \int_{-\pi}^{\pi} \frac{dk_x}{2\pi} \int_{-\pi}^{\pi} \frac{dk_y}{2\pi} \frac{(e^{ikr} - 1)}{[4 - 2\cos k_x - 2\cos k_y]}$$

$$G(0) = \int_{-\pi}^{\pi} \frac{dk_x}{2\pi} \int_{-\pi}^{\pi} \frac{dk_y}{2\pi} \frac{1}{[4 - 2\cos k_x - 2\cos k_y]}. \tag{7.35b}$$

Substituting into the partition function, we have

$$Z = Z_{sw} \sum_{m(r)} \exp\left\{ -2\pi^2 \beta G(0) \left[\sum_r m(r) \right]^2 \right\}$$

$$\times \exp\left\{ -2\pi^2 \beta \sum_{r,r'} m(r) G'(r - r') m(r') \right\}. \tag{7.36}$$

The exponential singles out those configurations which are "neutral,"

$$\sum_r m(r) = 0. \tag{7.37}$$

Other configurations do not contribute to Z. Therefore we can write

$$Z = Z_{sw} \sum_{m(r)}{}' \exp\left\{ -2\pi^2 \beta \sum_{r,r'} m(r) G'(r - r') m(r') \right\}, \tag{7.38}$$

where the prime on the sum means "neutral configurations of vortices only." It is enlightening to replace G' by the explicit formula Eq. (7.33). That asymptotic form is quite good even for small $|r| \approx 1$. Then Eq. (7.38) becomes

$$Z = Z_{sw} \sum_{m(r)}{}' \exp\left\{ (\pi^2\beta/2) \sum_{r,r'}{}' m(r) m(r') + \pi\beta \sum_{r,r'}{}' m(r) \ln(|r - r'|/a) m(r') \right\}, \tag{7.39}$$

where the primes on the sums over sites mean that the $r = r'$ terms is omitted. The first term in the exponential can be simplified, since the neutrality condition Eq. (7.37) implies

$$0 = \sum_{r,r'} m(r) m(r') = \sum_{r,r'}{}' m(r) m(r') + \sum_r m^2(r). \tag{7.40}$$

Now,

$$Z = Z_{sw} \sum_{m(r)}{}' \exp\left\{ -(\pi^2\beta/2) \sum_r m^2(r) + \pi\beta \sum_{r,r'}{}' m(r) \ln(|r - r'|/a) m(r') \right\}. \tag{7.41}$$

The first term in the braces gives the chemical potential of each vortex and the second gives the logarithmic interactions between different ones.

This Coulomb gas representation suggests the character of the phase diagram discussed in the previous section. If β is very large, then the chemical potential term in Eq. (7.41) suppresses the vortices very effectively, leaving only the spin waves behind. In fact, the strong logarithmic potential between vortices suggests that any vortex–antivortex pairs which might populate the ground state are tightly bound together. However, as the temperature is raised vortices are not suppressed significantly. Once $\beta = J/kT$ becomes of order unity one would naively expect vortices to become important configurations of the system. Since the interaction between them are long range, they cannot be thought of simply as free excitations. In fact, they form a plasma, and a screening length is generated dynamically (Kosterlitz, 1974). It will require a renormalization group analysis to see this.

D. Renormalization group analysis and the theory's critical region

The long-range logarithmic forces between vortices imply that a detailed understanding of the model requires dealing with many-body effects. One tool which was developed for this type of problem is the renormalization group (Wilson and Kogut, 1974). Before delving into the details of the periodic Gaussian model, we shall review the strategy of this approach.

When we write down the local Action of a spin system or field theory, we see clearly how degrees of freedom interact over distances of the order of the lattice spacing a. However, frequently we are interested in the long-distance characteristics of the theory or, equivalently, its low-energy content. The original lattice Action does not tell us anything transparent about these character-

istics of the theory. The trouble is that cooperative effects occur. Some details of the lattice Action may be important to the interactions between widely separated degrees of freedom and others may not be. Cooperative, many-body effects determine which elements of the lattice Action are important. To understand the critical region of a theory where the correlation length is huge, we need a description of the theory formulated directly in terms of its long-distance characteristics. We might generate such a description from the original lattice Action by averaging out some of its short-distance degrees of freedom. This averaging procedure may be difficult to carry out in practice, and the effective description of the system may be quite complicated if cooperative effects are really important. The renormalization group approach does the averaging in small steps so that each computation is tractable. A sequence of effective Actions is then generated by *iteration*. Large amplification or deamplification effects, which are expected when we begin with a lattice Action and end with an effective Action which describes the bulk properties of the theory, are obtained from the iteration procedure. The aim of such computations is to uncover the long-distance physics in the original model. The final effective Action might not have an appearance similar to the original Action, but it has the same physical content.

We shall carry out this program for the periodic Gaussian planar model. We shall work in momentum space, using continuum physics methods, and shall systematically integrate out the high-frequency parts of the theory. In effect, we shall obtain an effective Action for the low-frequency (or long-distance) content of the theory. The phase diagram of the effective Action will be clear almost by inspection.

To begin, return to Eq. (7.28),

$$Z=\int_{-\infty}^{\infty}\prod_r d\phi(r) \times\sum_{m(r)}\exp\left\{-1/2\beta\sum_{r,\mu}(\Delta_\mu\phi)^2+2\pi i\sum_r m(r)\phi(r)\right\}. \tag{7.28}$$

Let us first anticipate some of the effects of integrating out the high-frequency components of this expression. Equation (7.41) shows the effects of a complete integration over $\phi(r)$ while leaving the vortex field unintegrated—a chemical potential term and long-range interactions of the vortices occurred. It is convenient to anticipate the generation of a vortex chemical potential when the high-frequency components of the spin waves are integrated out, by incorporating such a term into Eq. (7.28) from the start (José *et al.*, 1977),

$$Z(y)=\int_{-\pi}^{\pi}\prod_r d\phi(r)\sum_{m(r)}\exp\left\{-1/2\beta\sum_{r,\mu}(\Delta_\mu\phi)^2 +\ln y\sum_r m^2(r)+2\pi i\sum_r m(r)\phi(r)\right\}, \tag{7.42}$$

where y includes the chemical potential observed in Eq. (7.41),

$$y=y_0\exp(-\pi^2\beta/2), \tag{7.43}$$

and y_0 is an adjustable parameter we shall discuss further below. The original model is defined by $y_0=1$. To be consistent one must equate the lattice propagator with the pure logarithm,

$$G'(r)\to-(1/2\pi)\ln(|r|/a), \tag{7.44}$$

since the $-1/4$ piece of Eq. (7.33) is already accounted for in Eq. (7.43). We make these definitions so that we can compare with published accounts of the resulting renormalization group (Kosterlitz, 1974). Note that the Action is now parametrized by two quantities: the temperature and the chemical potential of a vortex. The renormalization group analysis will produce effective Actions having the same form as Eq. (7.42) but with different parameters: an effective temperature and an effective chemical potential. Let us anticipate how the effective chemical potential will behave as a function of temperature. At low T we expect vortices to be suppressed. This means that although they can show up in tightly bound pairs of zero vorticity, they will not affect the long-distance features of the theory. It is reasonable, therefore, to expect the effective chemical potential of a vortex to grow large as more and more high frequencies are integrated out, if the temperature is small. Then only the $m(r)=0$ term of the partition function would be significant. However, as T is raised we expect to find a temperature T_c above which vortices populate the ground state. This should be reflected in the tendency of the effective chemical potential to decrease as more iterations of the renormalization group are done for $T>T_c$. We should be able to test for these qualitatively different types of behavior by studying $Z(y)$ for small y: at low T an initial small value of y should iterate to zero, while above a critical temperature an initially small y should grow. All of the analysis which follows in this section is aimed at establishing these points.

Note that if y is very small then Eq. (7.42) can be simplified. The point is that if $y\ll 0$ only the terms $m(r)=0,\pm1$ need be kept in the sum over vortices. Then

$$\begin{aligned}\sum_{m(r)=0,\pm1}\exp\{\ln y m^2(r)+2\pi i m(r)\phi(r)\} &\cong 1+\exp(\ln y)(\exp[2\pi i\phi(r)]+\text{h. c.})\\ &\approx 1+2y\cos[2\pi\phi(r)]\\ &\approx\exp\{2y\cos[2\pi\phi(r)]\}.\end{aligned} \tag{7.45}$$

The partition function becomes

$$Z(y)\approx\int\prod_r d\phi(r) \times\exp\left\{-\frac{1}{2\beta}\sum_{r,\mu}(\Delta_\mu\phi)^2+2y\sum_r\cos(2\pi\phi(r))\right\}. \tag{7.46}$$

To restore conventional field-theoretic notation, rescale $\phi(r)\to\phi(r)/\sqrt{\beta}$ so

$$Z(y)\approx\int\prod_r d\phi(r) \times\exp\left\{-\frac{1}{2}\sum_r(\Delta_\mu\phi)^2+2y\sum_r\cos(2\pi\sqrt{\beta}\,\phi(r))\right\}, \tag{7.47}$$

which we recognize as the quantum sine–Gordon theory. This is a convenient form for developing a renormalization group of the theory. We shall work directly in momentum space and regulate the theory with a momentum cutoff Λ. We shall then integrate out high frequencies of the theory, i.e., frequencies near but lower than Λ, and find an effective Action for the lower frequency modes of the theory. This approach (Raby and Ukawa, 1977) complements the more familiar real space renormalization group analysis done using the Coulomb gas representation of the theory (Kosterlitz, 1974).

To begin, define a momentum space cutoff version of Eq. (7.47). All Euclidean momenta p must have magnitude less than Λ,

$$Z_\Lambda = \int_{0<p<\Lambda} \mathfrak{D}\phi(p)\{\exp(-S[\phi_\Lambda])\}, \tag{7.48}$$

where $\mathfrak{D}\phi(p)$ denotes a functional integral over all the momentum components of the field,

$$S[\phi_\Lambda] = \int d^2x\left\{-\frac{1}{2}\,\phi_\Lambda(x)\Delta^2\phi_\Lambda(x) - \mu\cos[2\pi\sqrt{\beta}\,\phi_\Lambda(x)]\right\} \tag{7.49}$$

and

$$\phi_\Lambda(x) = \int_{0<p<\Lambda} \frac{d^2p}{(2\pi)^2}\, e^{ipx}\phi(x), \quad \phi^*(p) = \phi(-p). \tag{7.50}$$

The parameter μ in Eq. (7.49) is just $2y/a^2$.

To integrate out the high-frequency components of $\phi_\Lambda(x)$, split the momentum range $0<p<\Lambda$ into a low-frequency part $0<p<\Lambda'$ and a high-frequency part $\Lambda'<p<\Lambda$. Define a field which only has momentum components in the lower slice,

$$\phi_{\Lambda'}(x) = \int_{0<p<\Lambda'} \frac{d^2p}{(2\pi)^2}\, e^{ipx}\phi(p). \tag{7.51}$$

Then the high-frequency part of ϕ_Λ will be

$$h(x) = \phi_\Lambda(x) - \phi_{\Lambda'}(x) = \int_{\Lambda'<p<\Lambda} \frac{d^2p}{(2\pi)^2}\, e^{ipx}\phi(p). \tag{7.52}$$

Next, organize Z_Λ in the same way. The kinetic energy piece of the Action is easily split,

$$\begin{aligned}\frac{1}{2}\int d^2x\,\phi_\Lambda\Delta^2\phi_\Lambda &= -\frac{1}{2}\int \frac{d^2p}{(2\pi)^2}\,p^2\phi^2(p)\\ &= -\frac{1}{2}\int_{0<p<\Lambda'} \frac{d^2p}{(2\pi)^2}\,p^2\phi^2(p)\\ &\quad -\frac{1}{2}\int_{\Lambda'<p<\Lambda} \frac{d^2p}{(2\pi)^2}\,p^2\phi^2(p)\\ &= \frac{1}{2}\int d^2x\,\phi_{\Lambda'}\Delta^2\phi_{\Lambda'} + \frac{1}{2}\int d^2x\,h\Delta^2h.\end{aligned} \tag{7.53}$$

Therefore, the partition function becomes

$$Z_\Lambda = \int_{0<p<\Lambda'} \mathfrak{D}\phi(p)\exp\left\{\frac{1}{2}\int d^2x\,\phi_{\Lambda'}\Delta^2\phi_{\Lambda'}\right\}Z', \tag{7.54a}$$

where

$$Z' = \int_{\Lambda'<p<\Lambda} \mathfrak{D}\phi(p)\exp\left\{\frac{1}{2}\int d^2x\,h\Delta^2h + \mu\int d^2x\cos[2\pi\sqrt{\beta}\,(\phi_{\Lambda'}+h)]\right\}. \tag{7.54b}$$

Since $\mu \propto y$ which is assumed to be very small, we shall evaluate Z' in perturbation theory. Let us introduce the notation for averages of quantities θ in the high-momentum slice,

$$\langle\theta\rangle_h = \int_{\Lambda'<p<\Lambda} \mathfrak{D}\phi(p)\exp\left\{\frac{1}{2}\int d^2x\,h\Delta^2h\right\}\theta \Big/ \int_{\Lambda'<p<\Lambda} \mathfrak{D}\phi(p)\exp\left\{\frac{1}{2}\int d^2x\,h\Delta^2h\right\}. \tag{7.55}$$

Then Z' becomes

$$Z' = 1 + \mu\int d^2x\langle\cos[2\pi\sqrt{\beta}\,(\phi_{\Lambda'}+h)]\rangle_h + \frac{1}{2}\,\mu^2\int d^2x\,d^2y\,\langle\cos\{2\pi\sqrt{\beta}\,[\phi_{\Lambda'}(x)+h(x)]\}\cos\{2\pi\sqrt{\beta}\,[\phi_{\Lambda'}(y)+h(y)]\}\rangle_h + \cdots, \tag{7.56}$$

where an unimportant overall multiplicative constant has been dropped. The expectation values in Eq. (7.56) are easy to evaluate, since they are taken with respect to a free, cutoff field. For example,

$$\begin{aligned}\langle\cos[2\pi\sqrt{\beta}\,(\phi_{\Lambda'}+h)]\rangle_h &= \tfrac{1}{2}\langle\exp[2\pi i\sqrt{\beta}\,(\phi_{\Lambda'}+h)] + \text{h. c.}\,\rangle_h\\ &= \tfrac{1}{2}\exp(2\pi i\sqrt{\beta}\,\phi_{\Lambda'})\langle\exp(2\pi i\sqrt{\beta}h)\rangle_h + \text{h. c.} = \exp\{-2\pi^2\beta G_h(0)\}\cos[2\pi\sqrt{\beta}\,\phi_{\Lambda'}(x)],\end{aligned} \tag{7.57}$$

where $G_h(x)$ is the free propagator for the high-frequency components of ϕ_Λ,

$$G_h(x) = \int_{\Lambda'<p<\Lambda} \frac{d^2p}{(2\pi)^2}\, e^{ipx}\,\frac{1}{p^2}. \tag{7.58}$$

Since these quantities will occur frequenctly below, define

$$A(x) = \exp[-2\pi^2\beta G_h(x)], \tag{7.59}$$

so Eq. (7.57) becomes

$$\langle\cos[2\pi\sqrt{\beta}\,(\phi_{\Lambda'}+h)]\rangle_h = A(0)\cos[2\pi\sqrt{\beta}\,\phi_{\Lambda'}(x)]. \tag{7.60}$$

Consider next the connected part of the third term in Eq. (7.56),

$$\langle\cos\{2\pi\sqrt{\beta}\,[\phi_{\Lambda'}(x)+h(x)]\}\cos\{2\pi\sqrt{\beta}\,[\phi_{\Lambda'}(y)+h(y)]\}\rangle_h - \langle\cos\{2\pi\sqrt{\beta}[\phi_{\Lambda'}(x)+h(x)]\}\rangle_h\langle\cos\{2\pi\sqrt{\beta}[\phi_{\Lambda'}(y)+h(y)]\}\rangle_h. \tag{7.61a}$$

Calculating as in Eq. (7.57), we find that Eq. (7.61a) equals

$$\tfrac{1}{2}A^2(0)[A^2(x-y)-1]\cos 2\pi\sqrt{\beta}[\phi_{\Lambda'}(x)+\phi_{\Lambda'}(y)]+\tfrac{1}{2}A^2(0)[A^{-2}(x-y)-1]\cos 2\pi\sqrt{\beta}[\phi_{\Lambda'}(x)-\phi_{\Lambda'}(y)]\,. \tag{7.61b}$$

We can simplify this expression if we consider only that region of real space–time where it is appreciable. The propagator $G_h(x-y)$ should be non-negligible only for $|x-y|<1/\Lambda'$ if the momentum slicing is done smoothly enough. Therefore the presence of $[A^{-2}(x-y)-1]$ in Eq. (7.61b) allows us to expand the difference,

$$\phi_\Lambda(x)-\phi_\Lambda(y)\cong\xi\partial\phi_\Lambda(z)\,, \tag{7.62a}$$

where

$$\xi=x-y,\quad z=\tfrac{1}{2}(x+y)\,. \tag{7.62b}$$

Now Eq. (7.61b) becomes approximately

$$\tfrac{1}{2}A^2(0)[A^2(\xi)-1]\cos[4\pi\sqrt{\beta}\,\phi_\Lambda(z)] + \tfrac{1}{2}A^2(0)[A^{-2}(\xi)-1]\{1-\tfrac{1}{2}4\pi^2\beta[\xi\partial\phi_{\Lambda'}(z)]^2\}\,. \tag{7.63}$$

Finally, Eqs. (7.57) and (7.63) can be substituted into (7.56). If one defines various constants,

$$a_1=\int d^2\xi[A(\xi)-1]\,,$$
$$a_2=\int \xi^2 d^2\xi[A^{-2}(\xi)-1]\,, \tag{7.64}$$
$$a_3=\int d^2\xi[A^2(\xi)-1]\,,$$

carries out the integrations in Eq. (7.56), and finally substitutes back into Eq. (7.54), then

$$Z_\Lambda\cong\exp\left\{\frac{1}{4}\mu^2A^2(0)a_3\int d^2z\right\}\int_{0<p<\Lambda'}\mathfrak{D}\phi(p)\,. \exp\left\{-\left(1+\frac{1}{2}\pi^2\beta\mu^2A^2(0)a_2\right)\int\frac{1}{2}(\partial\phi_{\Lambda'})^2d^2z + \mu A(0)\int\cos[2\pi\sqrt{\beta}\,\phi_{\Lambda'}(z)]d^2z\right\}, \tag{7.65}$$

where we have ignored induced interaction terms which are higher order in μ than the term $\mu A(0)\cos(2\pi\sqrt{\beta}\,\phi_{\Lambda'})$ retained. If we rescale the field $\phi_{\Lambda'}$

$$\phi_{\Lambda'}\to\sqrt{1+\tfrac{1}{2}\pi^2\beta\mu^2A^2(0)a_2}\;\phi_{\Lambda'}\,, \tag{7.66}$$

and rescale momenta $p\to p/2$, then Eq. (7.65) becomes

$$Z_\Lambda(\mu,\beta)=\exp\left\{\tfrac{1}{4}\mu^2A^2(0)a_3\int d^2z\right\}Z_\Lambda(\mu',\beta')\,, \tag{7.67}$$

where the new parameters are given by

$$\mu'=A(0)\mu\,,$$
$$\beta'=\beta/[1+\tfrac{1}{2}\pi^2\beta\mu^2A^2(0)a_2]\,. \tag{7.68}$$

In addition, the change in the overall scale of the partition function means that the free energy density $F'=-\ln Z'/\int d^2x$ has changed,

$$F=F'-\tfrac{1}{4}\mu^2A^2(0)a_3\,. \tag{7.69}$$

In conclusion, Eq. (7.67) provides us with a new parametrization of the original model. The physics at the point (μ,β) is the same as the physics at (μ',β'). The calculation has shown that, when high-frequency fluctuations are integrated out, an effective Action is obtained which differs from the original Action by (1). wave-function renormalization, Eq. (7.66), (2). coupling constant renormalization, Eq. (7.68), (3). ground-state energy shift, Eq. (7.69). The form of the effective Action is the same as the original Action only because of the assumption $\mu\ll 1$. If μ were not infinitesimal, interaction effects of the form $\cos[4\pi\sqrt{\beta}\,\phi_{\Lambda'}]$ would occur in Z' and could not be ignored.

To understand the implications of Eq. (7.68) it is convenient to consider an infinitesimal high-momentum slice $\Lambda'\approx\Lambda$. Then Eq. (7.68) will reduce to differential equations. The high-energy slice will be defined by

$$\Lambda-d\Lambda<p<\Lambda\,. \tag{7.70}$$

Now the propagator $G_h(x)$ and the constants a_1, a_2, and a_3 will simplify. For example,

$$G_h(x)\quad\int_{\Lambda-d\Lambda<p<\Lambda}\frac{d^2p}{(2\pi)^2}\,e^{ipx}\,\frac{1}{p^2}=\frac{d\Lambda}{\Lambda}\,\frac{1}{(2\pi)}\,J_0(\Lambda|x|)\,, \tag{7.71}$$

where we have done an angular integration to expose the Bessel function. Unfortunately, Eq. (7.71) is not good. We have employed a sharp momentum space cutoff [Eq. (7.70)] and found that the propagator in real space behaves like a Bessel function. But Bessel functions do not fall off rapidly as their argument increases. Therefore, the naive estimates leading to Eq. (7.63) do not go through. This flaw can be solved by considering smooth momentum space slicing (Wilson and Kogut, 1974). We shall not need to execute such a procedure here, but we need to know that such a calculation can be done in principle. Then $G_h(x)$ is really a short-ranged function. Let us write Eq. (7.71) in the generic form,

$$G_h(x)\to(d\Lambda/\Lambda)\,(1/2\pi)\,\text{“}J_0(|\Lambda|x)\text{”}, \tag{7.72}$$

where the quotes indicate a function obtained from a smooth momentum space slicing calculation. With the same procedure the quantity a_2 becomes

$$a_2\to 4\pi^2\beta\Lambda^{-5}d\Lambda\;\alpha_2\,, \tag{7.73a}$$

where

$$\alpha_2=\int d\rho\,\rho^3\text{“}J_0(\rho)\text{”} \tag{7.73b}$$

is a dimensionless number dependent on the slicing procedure.

Differential equations for μ and β follow. Consider Eq. (7.68) for the parameter μ,

$$\delta\mu=\mu'-\mu=\mu[A(0)-1]\,. \tag{7.74}$$

Since $A(0)$ becomes

$$A(0)\approx 1-2\pi^2\beta G_h(0) \approx 1-2\pi^2\beta(d\Lambda/\Lambda)(1/2\pi) \tag{7.75}$$

for an infinitesimal slice, the differential equation for μ is

$$d\mu=-\pi\beta\mu(d\Lambda/\Lambda)\,. \tag{7.76}$$

The differential equation for β follows in a similar fashion,

$$d\beta = -2\pi^4\alpha_2\mu^2\beta^3(d\Lambda/\Lambda^5)\,. \tag{7.77}$$

To compare with other derivations of this renormalization group, Eqs. (7.76) and (7.77), we pass to a real-space cutoff $a=\Lambda^{-1}$, so

$$\begin{aligned} d\mu &= -\pi\beta\mu(da/a)\,,\\ d\beta &= -2\pi^4\alpha_2\mu^2\beta^3a^3da\,. \end{aligned} \tag{7.78}$$

Recall from Eq. (7.42) that the natural variables for the problem are β and $y=\mu a^2$. The differential equation for y reads

$$\begin{aligned} dy &= a^2d\mu + \mu da^2\,,\\ &= -\pi\beta\mu(da/a)a^2 + 2\mu ada\,,\\ &= -\mu a(\pi\beta - 2)da\,. \end{aligned} \tag{7.79}$$

Inspecting the equations for $d\beta$ and dy, we note that

$$y=0,\ \beta = 2/\pi\,, \tag{7.80}$$

is a fixed point of the coupled differential equations. Therefore, define

$$x = \pi\beta - 2\,. \tag{7.81}$$

A little algebra shows that the quantities x and y satisfy the differential equations,

$$\begin{aligned} dy^2 &= -2xy^2\frac{da}{a}\,,\\ dx &= -16\pi^2\alpha_2y^2\frac{da}{a}\,, \end{aligned} \tag{7.82}$$

in the vicinity of the fixed point Eq. (7.80). Finally, by changing the scale of "a," the second of these equations can be simplified. The first equation is unaffected by a change of scale, since it involves only the scale-invariant differentials dy^2/y^2 and da/a. So, finally Eq. (7.82) becomes

$$\begin{aligned} dy^2 &= -2xy^2\frac{da}{a}\,,\\ dx &= -y^2\frac{da}{a}\,. \end{aligned} \tag{7.83}$$

We have worked hard to put the renormalization group equations into this form because they have been thoroughly studied in a related problem (Anderson *et al.*, 1970).

We can now understand the character of these equations and their implications for the phase diagram of the planar model. Note that there is a trivial first integral of Eq. (7.83),

$$x^2 - y^2 = \text{const.} \tag{7.84}$$

The trajectories of the differential equation in the (x,y) plane are, therefore, hyperbolas, as shown in Fig. 39. First, consider trajectories in the region of the plane labeled $T<T_c$. They flow into the x axis and stop in accord with Eq. (7.83). Since any two points on the same trajectory give the same physics, all trajectories in this region of the plane define theories which are equivalent to ones having $y=0$. But a model having

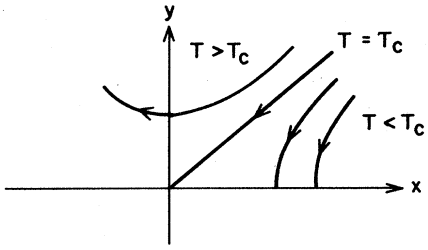


FIG. 39. The renormalization group trajectories of the planar model.

$y=0$ consists just of spin waves, as Eq. (7.47) shows! Therefore, we have rediscovered the *fixed line* of the planar model—all of these trajectories flow to the positive x axis. The trajectory labeled $T=T_c$ is the boundary between those which flow to $y=0$ and those which do not. To the left of this boundary the trajectories flow toward larger and larger y values. Therefore, since y measures the importance of vortices in the system, we have indeed discovered that vortices drive the phase transition of the planar model and give rise to a qualitatively distinct high-temperature phase. The point $(x,y)=(0,0)$ labels the end of the fixed line and is therefore associated with the critical temperature.

To analyze Eq. (7.84) more precisely, recall that in the original model

$$x = \pi\beta - 2 = (\pi J/kT) - 2\,,\quad y \sim \exp(-\pi^2J/2kT)\,. \tag{7.85}$$

The critical temperature is then determined by the condition

$$x^2 - y^2 = 0\,, \tag{7.86a}$$

which becomes

$$\pi J/kT_c - 2 \sim \exp(-\pi^2J/2kT_c)\,. \tag{7.86b}$$

Therefore, the nonvanishing of the right-hand side gives a correction to our earlier, naive estimate of T_c, $T_c \approx \pi J/2k$. Since the correction behaves as $\exp(-1/T)$, it is due to vortex interactions which were neglected in the earlier, intuitive discussion. If we expand x^2-y^2 about zero using Eq. (7.85), we can find the constant of Eq. (7.84),

$$x^2 - y^2 \cong -c\left(\frac{T-T_c}{T_c}\right)\,, \tag{7.87}$$

where c is a positive quantity ≈ 2.1 (Kosterlitz, 1974).

A more detailed analysis of these differential equations allows one to calculate the correlation length in the critical region. If one accepts the plausible assumption that trajectories labeled by $T>T_c$ flow to infinite temperature, then one can show (Kosterlitz, 1974)

$$\xi(T) \underset{T>T_c}{\sim} \exp\left\{b\left(\frac{T_c}{T-T_c}\right)^{1/2}\right\}\,. \tag{7.88}$$

Therefore $\xi(T)$ diverges with an essential singularity and the usual power-law hypotheses for critical singularities, as reviewed in Sec. II.A, do not hold in this model. However, correlation scaling does apply. For example, dimensional analysis suggests that the free energy density, measured in units of kT, should behave as

$$f/kT \sim \xi^{-2}(T), \tag{7.89}$$

in the critical region, and this result does follow from the differential equation (7.69). This result shows that the free energy and all its derivatives are continuous in the critical region. And finally, the spin–spin correlation function is predicted to behave as

$$\langle \exp i(\theta_0 - \theta_n)\rangle \sim \left(\frac{1}{|n|}\right)^{\eta}, \tag{7.90}$$

at the critical temperature T_c, and the index η is 1/4 as in the Ising model. This last result agrees with experiment (Webster *et al.*, 1979).

In summary, this model illustrates how topological singularities in a lattice spin system can drive a phase transition. The final conclusion is that the low-temperature phase is given by our naive, smooth-field expectations—it is exhausted by spin waves (aside from some finite renormalization effects). The high-temperature phase consists of a vortex condensate and has short-range spin–spin correlations.

It is interesting that a periodic Gaussian version of Abelian lattice gauge theory in four dimensions can also be defined and analyzed along these lines (Banks *et al.*, 1977). In that theory the topological singularities consist of closed lines of vortex singularities. It appears that at low T these vortex singularities are bound together into neutral nets and do not affect the long-distance characteristics of the system substantially. Then our naive expectation that the theory reduces to the ordinary free electromagnetic field should hold. However, at larger T the nets of vortex singularities should grow into infinitely long loops which disorder the system and lead to quark confinement. Unfortunately, reliable renormalization group analyses have not yet been done on the theory, so although these points are plausible they are not proved.

VIII. NON-ABELIAN LATTICE THEORIES

A. General formulation of the SU(2) theory

It is not difficult to extend our general considerations to non-Abelian local symmetries. The physics here is much more exciting and, hopefully, is relevant to the world of strong interactions.

Let us introduce the idea of a local non-Abelian symmetry in a geometric fashion, following the original ideas of Yang and Mills (Yang and Mills, 1954). Consider a cubic lattice in d-dimensional Euclidean space–time. Let there be a "frame of reference" at each site. If the local symmetry group is SU(2), the frame of reference is three dimensional and refers to the internal symmetry (color) indices of the gauge group. Suppose that these frames can be oriented arbitrarily from site to site, so that color cannot be compared at different points in space–time. For example, the theory might have colorful quanta which can hop from site to site and whose color is measured relative to the local frames of reference. We want to construct the Action of the theory so that it is invariant to changes in the orientation of the local color frames of reference. The existence of a colorful gauge field arises from this symmetry principle in a rather natural way (Yang and Mills, 1954).

Consider two nearest-neighbor frames, one at site n and the other at site $n+\mu$. Their relative orientation is specified by a rotation matrix which lives on the link between them,

$$U_\mu(n) = \exp[iB_\mu(n)], \quad U_{-\mu}(n+\mu) = U_\mu^{-1}(n), \tag{8.1a}$$

where

$$B_\mu(n) = \tfrac{1}{2} a g \tau_i A_\mu^i(n) \tag{8.1b}$$

and the color index i is summed from 1 to 3. In Eq. (8.1b) we are anticipating the connection of this formalism to ordinary continuum non-Abelian gauge fields. A gauge transformation will be a local rotation of a frame of reference. For example if the frames at sites n and $n+\mu$ are rotated, the $U_\mu(n)$ transforms as

$$[U_\mu(n)]_{ij} \to \sum_{kl} \{\exp[-i\tfrac{1}{2}\tau_m \chi^m(n)]\}_{ik} \times \{\exp[i\tfrac{1}{2}\tau_{m'}\chi^{m'}(n+\mu)]\}_{jl} [U_\mu(n)]_{kl} \tag{8.2a}$$

where the SU(2) matrix

$$\exp[-i\tfrac{1}{2}\tau_m \chi^m(n)], \tag{8.2b}$$

describes the local rotation at site n. One can also visualize the transformation law [Eq. (8.2a)] in terms of the quantum-mechanical spherical top. Then $U_\mu(r)$ is the rotation matrix describing the relation between the space-fixed coordinate system and the body-fixed coordinate system (Kogut and Susskind, 1975).

Consider the theory with only gauge field degrees of freedom, $U_\mu(n)$. A locally gauge-invariant Action can be written down following the motivation discussed for Ising and Abelian theories,

$$S = -(1/2g^2) \sum_{n,\mu\nu} \mathrm{tr} U_\mu(n) U_\nu(n+\mu) \times U_{-\mu}(n+\mu+\nu) U_{-\nu}(n+\nu) + \text{h.c.} \tag{8.3}$$

It is easy to check that the Action incorporates the local symmetry equation (8.2a). In addition, the link variables $B_\mu(n)$ enter the Action only through rotation matrices, so their range of variation is naturally compact $(0 < B_\mu(n) < 4\pi)$. This point should be contrasted with our discussion of Abelian lattice gauge theory. There we constructed a theory based on phase variables, and periodicity in the variable $agA_\mu(n)$ followed. This was not forced upon us. However, for non-Abelian theories it is natural to form a lattice version with group elements $U_\mu(n)$. Since $SU(n)$ groups are compact, the bounded character of $B_\mu(n)$ follows.

Let's understand Eq. (8.3) in more detail. The first feature we should establish is that for classical, smooth fields it reduces to the ordinary Action for non-Abelian gauge fields. To begin, write Taylor expansions for the gauge fields,

$$\begin{aligned} B_\nu(n+\mu) &\cong B_\nu(n) + a\partial_\mu B_\nu(n) \\ B_{-\mu}(n+\mu+\nu) &= -B_\mu(n+\nu) \cong -[B_\mu(n) + a\partial_\nu B_\mu(n)] \\ B_{-\nu}(n+\nu) &= -B_\nu(n) . \end{aligned} \tag{8.4}$$

Then the matrix product of four U's around a directed plaquette becomes

$$U_\mu(n)U_\nu(n+\mu)U_{-\mu}(n+\mu+\nu)U_{-\nu}(n+\nu)$$
$$\approx \exp(iB_\mu)\exp[i(B_\nu + a\partial_\mu B_\nu)]$$
$$\times \exp[-i(B_\mu + a\partial_\nu B_\mu)]\exp(-iB_\nu)\,. \tag{8.5}$$

Now apply the Baker–Hausdorff formula,

$$e^x e^y = e^{x+y+(1/2)[x,y]+\cdots}\,. \tag{8.6}$$

We shall not need the terms denoted by dots in Eq. (8.6) because in our application they will be multiplied by many powers of the lattice spacing and will not contribute in the continuum limit. So,

$$U_\mu(n)U_\nu(n+\mu)U_{-\mu}(n+\mu+\nu)U_{-\nu}(n+\nu) \simeq \exp\{i(B_\mu + B_\nu + a\partial_\mu B_\nu) - \tfrac{1}{2}[B_\mu,B_\nu]\}\exp\{-i(B_\mu + B_\nu + a\partial_\nu B_\mu) - \tfrac{1}{2}[B_\mu,B_\nu]\}$$
$$\approx \exp\{ia(\partial_\mu B_\nu - \partial_\nu B_\mu) - [B_\mu,B_\nu]\} = \exp[ia^2 g\mathcal{F}_{\mu\nu}] \tag{8.7}$$

where we have defined in the last step,

$$\mathcal{F}_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu,A_\nu]\,, \tag{8.8}$$

where $A_\mu = \frac{1}{2}\tau_i A_\mu^i$. $\mathcal{F}_{\mu\nu}$ is the standard Yang–Mills field strength. Note that the commutator $[A_\mu,A_\nu]$ appears naturally in $\mathcal{F}_{\mu\nu}$ as a consequence of local gauge invariance. For smooth, classical fields we can assume that $a^2 g\mathcal{F}_{\mu\nu} \ll 1$ and simplify Eq. (8.7) further,

$$\operatorname{tr}\exp(ia^2 g\mathcal{F}_{\mu\nu}) = \operatorname{tr}\{1 + ia^2 g\mathcal{F}_{\mu\nu} - \tfrac{1}{2}a^4 g^2 \mathcal{F}_{\mu\nu}^2 + \cdots\}$$
$$= \operatorname{tr} 1 - \tfrac{1}{2}a^4 g^2 \operatorname{tr}\mathcal{F}_{\mu\nu}^2 + \cdots \tag{8.9}$$

The tr 1 term has no dynamics in it and can be dropped. The term linear in $\mathcal{F}_{\mu\nu}$ vanished when the trace was taken. Using the commutator

$$[\tau_i,\tau_j] = 2i\varepsilon_{ijk}\tau_k\,, \tag{8.10}$$

we find

$$\operatorname{tr}\mathcal{F}_{\mu\nu}^2 = \tfrac{1}{2}(\partial_\mu A_\nu^k - \partial_\nu A_\mu^k - g\varepsilon_{kij}A_\mu^i A_\nu^j)^2\,. \tag{8.11}$$

Now the Action becomes

$$S \approx \frac{1}{2g^2}\int \frac{d^4x}{a^4} a^4 g^2 \frac{1}{2}(\partial_\mu A_\nu^k - \partial_\nu A_\mu^k - g\varepsilon_{kij}A_\mu^i A_\nu^j)^2 + O(a^2)\,. \tag{8.12}$$

So

$$S = \frac{1}{4}\int d^4x (\mathcal{F}_{\mu\nu}^i)^2\,, \tag{8.13a}$$

where

$$\mathcal{F}_{\mu\nu}^i = \partial_\mu A_\nu^i - \partial_\nu A_\mu^i - g\varepsilon^{ijk}A_\mu^j A_\nu^k\,. \tag{8.13b}$$

The reader should recognize Eq. (8.13a) as the usual classical Euclidean Action of pure Yang–Mills fields.

B. Special features of the non-Abelian theory

To discuss the continuum limit of the field theory equation (8.3) we must treat $B_\mu(n)$ as a legitimate fluctuating variable. One might guess, however, that for weak coupling the classical analysis should be a good guide to the physics of the theory. This was the case for the planar model in two dimensions. There we found that for $T < T_c$ the vortices were irrelevant, and the theory was well described by massless spin waves. Presumably, an analogous result holds for Abelian lattice gauge theory in four dimensions. However, there are good reasons to believe that the classical Action is *never* a good guide to the non-Abelian models! This is known to be true for non-Abelian spin models in two dimensions, as we shall discuss later. The reason for the dramatic difference between Abelian and non-Abelian theories lies in coupling-constant renormalization. We saw that for $T < T_c$, the effective temperature of the planar model did not renormalize significantly. However, above the critical point, the effective temperature rose as coarser and coarser lattices were considered. It was tempting to assume that for $T > T_c$ the renormalization group trajectories flowed to an infinite temperature fixed point (Kosterlitz, 1974). In non-Abelian theories perturbation theory calculations show that even for weak coupling the renormalization group trajectories flow toward stronger and stronger coupling (t' Hooft, 1972; Politzer, 1973; Gross and Wilczek, 1973). This suggests that the non-Abelian theories do not have a distinct weak coupling phase of massless particles and Coulomb-like forces. Rather, for all coupling the theory resembles the high-temperature phase of the Abelian models, in which a mass gap develops dynamically and the system is disordered.

The property by which the renormalization group trajectories of non-Abelian gauge theories flow toward strong coupling is called "asymptotic freedom" (t' Hooft, 1972; Politzer, 1973; Gross and Wilczek, 1973). Consider two versions of the theory, one with a space–time cutoff "a_0" and another with space–time cutoff "a." We can think of a_0 and a as lattice spacings, although the calculations are done using the continuum formulation of the model. In order that the physics of the two formulations be identical, their coupling constants must be related,

$$g^2(a) = \frac{g_0^2}{1 + (Cg_0^2/2\pi)\ln(a_0/a)}\,, \tag{8.14}$$

where C is a positive group-theoretic constant. Note that as $a \to 0$,

$$g^2(a) \to 2\pi/C\ln(a_0/a) \to 0\,, \tag{8.15}$$

so the continuum limit ($a \to 0$) of the theory is at zero coupling! Since the interacting theory is weakly coupled at short distances, detailed comparison between theory and experiment are possible for processes which are only sensitive to the short-distance features of strong interactions. These comparisons have been very successful, and such detailed successes have strengthened many theorists' belief that non-Abelian gauge fields provide the basic hadronic force.

The long-distance properties of the theory are not well understood. Equation (8.14) suggests that the effective coupling must be large when the theory is formulated with a large space–time cutoff. If this is the

case, one would expect the theory to have a rich spectrum of bound states. In addition, the force law between static quarks which belong to the fundamental representation of the gauge group could be confining. Certainly the strong coupling lattice theory, Eq. (8.3), has these properties.

Given our experience with simpler theories, we can discuss the non-Abelian lattice Action easily. Since it incorporates a local symmetry, Elitzur's theorem applies to it, so the theory cannot have a local order parameter. Wegner's construction of a gauge-invariant correlation function generalizes to continuous non-Abelian groups. Consider a closed, directed contour C and the operator

$$\prod_C \exp[iB_\mu(n)]. \tag{8.16}$$

We can study its expectation value,

$$\left\langle \prod_C \exp[iB_\mu(n)] \right\rangle = Z(C)/Z\,, \tag{8.17a}$$

where

$$Z(C) = \int \prod_{n,\mu} dB_\mu(n) \exp(-S) \prod_C \exp[iB_\mu(n)] \tag{8.17b}$$

and S is the lattice Action equation (8.3). Our usual high-temperature expansion methods show that for g large enough the area law applies,

$$\left\langle \prod_C \exp[iB_\mu(n)] \right\rangle \approx \exp[-F(g^{-2})A]\,, \tag{8.18}$$

where A is the area of a minimal surface enclosed by C and $F(g^{-2})$ is a well-defined, finite function. The physical interpretation of this correlation function is the same as the Abelian theory—it gives the force law between heavy, static quarks. Of course now the quarks carry a non-Abelian charge, color. Assigning the quarks to the fundamental representation of SU(2) color means that under local color gauge transformations they transform as

$$\psi_i(n) \to [\exp\{i\tfrac{1}{2}\tau_l \chi^l(n)\}]_{ij}\psi_j(n)\,. \tag{8.19}$$

Then to make a bilocal operator such as $\sum_i \bar{\psi}_i(n+\mu)\psi_i(n)$ locally gauge invariant, we must follow Schwinger's construction and form the operator

$$\sum_{ij} \bar{\psi}_i(n+\mu)[e^{iB_\mu(n)}]_{ij}\psi_j(n)\,. \tag{8.20}$$

So, as a quark hops from site n to $n+\mu$ it must be accompanied by a gauge field rotation matrix on the intervening link. Therefore, if one specifies that the quark hop around the closed contour C, the gauge field piece of the amplitude is clearly Eq. (8.17b). Finally, if a rectangular closed contour (Fig. 27) is considered, we obtain the quark–antiquark potential,

$$V(R) = -\lim_{T\to\infty} (1/T) \ln\left\langle \prod_C e^{iB_\mu(n)} \right\rangle, \tag{8.21}$$

by essentially repeating the arguments which led to Eq. (5.42) for the Abelian theory. Then Eq. (8.18) gives the linear confining potential

$$V(R) \sim |R|\,. \tag{8.22}$$

In summary, the lattice Action has simple properties at strong coupling which may be properties of the strong interactions. The continuum quantum theory has simple short-distance characteristics which compare well with experiment. Are these two theories related? It is hoped that they lie on the two ends of one renormalization group trajectory and that the theory of non-Abelian gauge fields enjoys asymptotic freedom on fine lattices and confinement on coarse ones. The complexity of the theory has prevented a constructive proof of this point, the Holy Grail of the subject. However, strong coupling expansions have been used to search for possible phase transitions in the intermediate coupling region of the theory, and none were found (Kogut, Sinclair, and Susskind, 1976). It appeared that the only critical point of the theory resides at $g=0$, but the analysis was not strong enough to be convincing. An approximate recursion relation due to A. B. Migdal (Migdal, 1975) also suggests that the theory resides in just one phase. Some remarks about his results will be made in a later section. Much more work (and inspiration) is needed on this crucial subject.

Throughout these lectures parallels have been drawn between lattice gauge theories in d dimensions and spin systems in $d/2$ dimensions. Such connections also exist for non-Abelian theories. First, it is easy to see that two-dimensional SU(2) lattice gauge theory is equivalent to the SU(2)×SU(2) Heisenberg spin system in one dimension. Choose the "temporal gauge" $B_0(n)=0$. Then the lattice Action reduces to

$$S = -(1/2g^2)\sum_n \mathrm{tr} U_x(n) U_x^{-1}(n+\hat{\tau}) + \mathrm{h.c.}\,, \tag{8.23}$$

or

$$S = \sum_x \left\{ -\frac{1}{2g} \sum_\tau \mathrm{tr} U_x(n) U_x^{-1}(n+\hat{\tau}) + \mathrm{h.c.} \right\} \tag{8.24}$$

where sites of the lattice are labeled $n=(x,\tau)$. Therefore the original Action breaks down into many copies of a nearest-neighbor coupled one-dimensional spin model. Since the U matrices transform according to SU(2)×SU(2), the spin model is the SU(2)×SU(2) Heisenberg chain. And, since SU(2)×SU(2) is identical to $O(4)$, we can visualize Eq. (8.24) as copies of a chain of nearest-neighbor coupled four-dimensional unit vectors. To make this point explicit, note that since $U_x(n)$ can be represented as a 2×2 unitary matrix it can be parametrized in terms of Pauli matrices τ_i,

$$U_x(n) = \sigma(n) + i\boldsymbol{\tau}\cdot\boldsymbol{\pi}(n) \tag{8.25}$$

where $(\sigma,\boldsymbol{\pi})$ are four real fields. The condition that $U_x(n)$ be unitary is

$$\sigma^2(n) + \boldsymbol{\pi}^2(n) = 1\,. \tag{8.26}$$

So, consider a four-dimensional unit vector on each link,

$$\mathbf{s}(n) = \begin{pmatrix} \sigma \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}, \quad |\mathbf{s}(n)|^2 = 1\,. \tag{8.27}$$

Using the identity

$$\mathrm{tr}U_x(n)U_x^{-1}(n+\hat{\tau}) = 2[\sigma(n)\sigma(n+\hat{\tau}) + \boldsymbol{\pi}(n)\cdot\boldsymbol{\pi}(n+\hat{\tau})]$$
$$= 2\mathbf{s}(n)\cdot\mathbf{s}(n+\hat{\tau}) \tag{8.28}$$

the Action becomes

$$S = \sum_x \left\{ -(2/g^2) \sum_\tau \mathbf{s}(n)\cdot\mathbf{s}(n+\hat{\tau}) \right\}, \tag{8.29}$$

as claimed. The one-dimensional spin system is disordered at all temperatures. Its exponentially decaying correlation function implies the area law and confinement for the two-dimensional gauge system. The details of this construction are essentially the same as for the Ising gauge theory, Eq. (5.40).

It is more interesting to consider the similarities between SU(n) gauge theories in four dimensions and SU(n)×SU(n) spin systems in two dimensions. Striking similarities between these two systems have been emphasized by A. M. Polyakov (1975) and A. B. Migdal (1975). Recall some of the characteristics of the two-dimensional spin systems. First, they are asymptotically free. This will be shown in the next section, where the weak coupling renormalization group will be obtained. Second, they cannot have a local order parameter because a continuous global symmetry cannot break down spontaneously in two dimensions. Third, they develop a mass gap dynamically, and their scattering amplitudes contain only massive excitations (Zamolodchikov and Zamolodchikov, 1977). It is not known if non-Abelian gauge theories in four dimensions have this last property, although it is strongly hoped for. Because of these similarities and interesting properties, it is worth our while to consider the spin systems in more detail.

C. Renormalization group analysis of $O(n)$ spin systems in two dimensions

It is interesting and not difficult to understand asymptotic freedom and dynamical mass generation in non-Abelian spin systems in two dimensions. One aim of this section is to understand asymptotic freedom in terms of the geometry of the non-Abelian group. For easy presentation consider the $O(3)$ model, where one can visualize the local spin and easily see that the local curvature of the sphere implies asymptotic freedom.

To develop the theory's weak coupling renormalization group, we shall integrate out the high-frequency components of the local spin variable $\mathbf{s}(n)$ and find an effective Action for its low-frequency components. For weak coupling, the effective Action will have the same form as the original, but the effective temperature will be increased. The partition function of the model is

$$Z = \int \prod_n d\mathbf{s}(n) \exp\left\{ -\frac{1}{2g}\int \partial_\mu \mathbf{s}\cdot\partial_\mu \mathbf{s}\, d^2r \right\}, \tag{8.30}$$

where the spin is constrained,

$$|\mathbf{s}(r)|^2 = 1. \tag{8.31}$$

We shall be using continuum space–time notation throughout this discussion, and will form a renormalization group using momentum space cutoffs as in Sec. VII.D. Since the integrations of the renormalization group preserve the $O(3)$ symmetry of the model, the effective Action can only have terms of the form

$$S' = (1/2g')\partial_\mu \mathbf{s}\cdot\partial_\mu \mathbf{s} + (1/2f')(\partial_\mu \mathbf{s}\cdot\partial_\mu \mathbf{s})^2 + \cdots. \tag{8.32}$$

Our aim is to compute g'. In weak coupling the higher order terms in Eq. (8.32) are not significant.

Since we want to expose the geometry underlying the asymptotic freedom of the model, we shall parametrize the calculation judiciously. In general each component of $\mathbf{s}$ has high- and low-frequency components. However, if one considers only weak coupling, two components, s_1 and s_2, can be chosen to be slowly varying and large, $O(1)$, while the third component, s_3, is rapidly varying and small, $O(\sqrt{g})$. In the course of the calculation we shall see that this arrangement is possible. Since the spin satisfies the constraint of Eq. (8.31) we can parametrize it,

$$s_1 = \sqrt{1-s_3^2}\sin\theta$$
$$s_2 = \sqrt{1-s_3^2}\cos\theta. \tag{8.33}$$

Then the Lagrangian becomes

$$\mathcal{L} = \frac{1}{2g}(\partial_\mu \mathbf{s})^2 = \frac{1}{2g}\left[(1-s_3^2)(\partial_\nu\theta)^2 + \frac{(\partial_\mu s_3)^2}{1-s_3^2} \right]. \tag{8.34}$$

If we anticipate that, for small g, $s_3 \sim O(\sqrt{g})$ we can approximate $\mathcal{L}$ by

$$\mathcal{L} = (1/2g)[(\partial_\mu s_3)^2 + (1-s_3^2)(\partial_\mu\theta)^2 + s_3^2(\partial_\mu s_3)^2 + \cdots]. \tag{8.35}$$

Finally, it is convenient to rescale s_3 to eliminate the overall factor of g,

$$h(r) = s_3(r)/\sqrt{g}, \tag{8.36}$$

so

$$\mathcal{L} = \tfrac{1}{2}[(\partial_\mu h)^2 + [(1/g) - h^2](\partial_\mu\theta)^2 + gh^2(\partial_\mu h)^2 + \cdots]. \tag{8.37}$$

This expression makes the virtues of this parametrization clear. As h fluctuates, it affects the contribution of the θ field to the Action. However, fluctuations in θ do not react back onto h. These facts justify the earlier comments stating that for the purposes of weak coupling renormalization s_3 could be treated as small and rapidly fluctuating, while s_1 and s_2 are large and slowly varying.

Now consider the theory with a momentum space cutoff $|p| < \Lambda$. Introduce a high-momentum slice, $\Lambda' < |p| < \Lambda$, and a low-momentum slice, $0 < |p| < \Lambda'$, and organize the computation of the partition function accordingly,

$$Z = \int_{0<p<\Lambda} \mathfrak{D}\theta(p) \exp\left\{ -\frac{1}{2g}\int (\partial_\mu\theta)^2 d^2x \right\}$$
$$\times \int_{\Lambda'<p<\Lambda} \mathfrak{D}h(p) \exp\left\{ -\frac{1}{2}\int (\partial_\mu h)^2 d^2x \right\} \tag{8.38}$$
$$\times \exp\left\{ \frac{1}{2}\int h^2(\partial_\mu\theta)^2 d^2x - \frac{1}{2}\int gh^2(\partial_\mu h)^2 d^2x + \cdots \right\}.$$

For small g the second term in the last exponential can be dropped. Then the integral over $h(p)$ is a pure Gaussian and it can be computed exactly,

$$\int_{\Lambda'<p<\Lambda} \mathfrak{D}h(p) \exp\left\{-\frac{1}{2}\int (\partial_\mu h)^2 d^2x\right\} \exp\left\{\frac{1}{2}\int h^2(\partial_\mu\theta)^2 d^2x\right\}$$
$$= N \exp\left\{\frac{1}{2}\int \langle h^2\rangle_h (\partial_\mu\theta)^2 d^2x\right\}$$
$$= N \exp\left\{\frac{1}{2}\int (\partial_\mu\theta)^2 d^2x \int_{\Lambda'}^{\Lambda} \frac{d^2p}{(2\pi)^2}\frac{1}{p^2}\right\}$$
$$= N \exp\left\{\frac{1}{4\pi}\ln(\Lambda'/\Lambda)\int (\partial_\mu\theta)^2 d^2x\right\}, \tag{8.39}$$

where the overall constant N is not important. Putting this result back into Eq. (8.38) we see that the θ term of the effective Action is

$$\mathcal{L}' = \tfrac{1}{2}[(1/g) - (1/2\pi)\ln(\Lambda/\Lambda')](\partial_\mu\theta)^2 + \cdots . \tag{8.40}$$

Comparing this with the θ term in the original Action, $(1/2g)(\partial_\mu\theta)^2 + \cdots$, we find coupling-constant renormalization,

$$(1/g') = (1/g) - (1/2\pi)\ln(\Lambda/\Lambda') . \tag{8.41}$$

In making this identification we have used the fact discussed in Eq. (8.33) that the $O(3)$ symmetry restricts the form of the effective Action. So, although the parametrization used here obscures the $O(3)$ symmetry of the effective Action, the reciprocal of the coefficient of the $(\partial_\mu\theta)^2$ term is necessarily the effective coupling constant of the effective Action equation (8.32).

It is best to convert Eq. (8.41) to a differential equation by letting the high-momentum slice, $\Lambda'<p<\Lambda$, be infinitesimal. Setting $\Lambda' = \Lambda + \delta\Lambda$, we have

$$\ln(\Lambda/\Lambda') = -\ln(1+\delta\Lambda/\Lambda) \approx -\delta\Lambda/\Lambda \tag{8.42a}$$

and

$$(1/g') - (1/g) = d(1/g) = -(1/g^2)dg . \tag{8.42b}$$

Therefore

$$\Lambda\frac{dg}{d\Lambda} = -\frac{1}{2\pi}g^2 . \tag{8.43}$$

Writing this in terms of a real space cutoff such as a variable lattice spacing,

$$a\frac{dg}{da} = \frac{1}{2\pi}g^2 . \tag{8.44}$$

Equations (8.43) and (8.44) show that the theory is asymptotically free. Equation (8.43) states that the effective coupling is a decreasing function of the momentum cutoff. Alternatively, Eq. (8.44) means that a smaller coupling must be used with a finer space–time cutoff than with a coarse one. Therefore, assuming that the effective coupling increases indefinitely with a, Eq. (8.44) implies that the theory's continuum limit is at $g=0$. It would also mean that the theory's long-distance properties are described by a strongly coupled Action. That Action would clearly have a finite correlation length, so the theory would be free of massless spin waves. Other analyses suggest this point (Zamolodchikov and Zamolodchikov, 1977).

Before leaving this topic, let us discuss the physical reason for the asymptotic freedom of this model. In computing the effective Action we integrated out the high-frequency, small-amplitude fluctuations in the spin variable $\mathbf{s}$. This led to a description of the model in terms of a new spin variable $\mathbf{s}'$. Since the sphere has local curvature, the averaging effectively reduces the magnitude of the spin variable so that $|\mathbf{s}'|<1$. To restore the condition that the effective spin has unit length requires a rescaling $\mathbf{s}' \to \mathbf{s}'/|s'|$ which can be absorbed into a change in the coupling constant $g \to g/|s'|^2$. Therefore the effective coupling has *increased* as a result of integrating out the high frequencies.

It is particularly interesting that the $O(3)$ model has significant coupling-constant renormalization even at infinitesimal g. This occurs because even infinitesimal fluctuations of the spin variables experience the local curvature of the sphere. In other words, the compact character of the group is apparent even in weak coupling. This should be contrasted with the planar model. There small fluctuations are not sensitive to the fact that $\theta(n)$ is an angular variable. Therefore the low-temperature perturbative calculations are the same as those of an ordinary free field and show no coupling-constant renormalization. It is only at finite temperatures, where nonperturbative vortex variables are considered, that coupling-constant renormalization and a phase transition are discovered. The relevance of the vortex variables at sufficiently high T is an indication that spins are now winding through their entire orbits and the periodicity of the $\theta(n)$ variables is affecting the dynamics. And once the boundedness of the θ variable becomes important, it is clear from our intuitions gained from the $O(3)$ model that the planar model will experience coupling-constant renormalization and that its effective temperature will grow.

The generalization of Eq. (8.44) to the $O(N)$ model is clear,

$$a\frac{dg}{da} = \frac{1}{2\pi}(N-2)g^2 . \tag{8.45}$$

It is $N-2$ because this is the number of directions in which the spin can experience rapid, small fluctuations which are *perpendicular* to its average, slow motion around the group space. Only the perpendicular fluctuations contribute to coupling-constant renormalization. Fluctuations in the direction of the spin's average, slow motion have the same character as fluctuations of a planar spin model, which, as we have seen, experiences no coupling-constant renormalization at low temperature.

The standard notation for weak coupling renormalization equations such as Eq. (8.45) is

$$a\frac{dg}{da} = \beta(g) , \tag{8.46}$$

where β is the Callan–Symanzik function (Symanzik, 1970; Callan, 1970). In the language used in this discussion, positive β is asymptotic freedom.

As mentioned above, the asymptotic freedom of the $O(n)$ models suggests that they develop a mass gap dynamically. Recall that the planar model develops a gap for temperatures above the Kosterlitz–Thouless phase transition, and that for $T \approx T_c$ it behaves as

$$m(T) \underset{T\to T_c^+}{\sim} \exp\left\{-b\left(\frac{T_c}{T-T_c}\right)^{1/2}\right\}. \tag{8.47}$$

It is easy to obtain the analogous formula for asymptotically free theories (Lane, 1974; Gross and Neveu, 1974). Simply observe that the mass gap is a physical quantity which we can hold fixed as the cutoffs of the underlying field theories are changed. Therefore

$$\frac{d}{da}m=0\,. \tag{8.48}$$

Trivial indeed. One can also write m in the form

$$m=(1/a)F(g)\,, \tag{8.49}$$

by dimensional analysis. By combining these last two equations we can determine m as a function of g. Substituting Eq. (8.49) into Eq. (8.48) gives

$$-\frac{1}{a^2}F(g)+F'(g)\frac{dg}{da}=0\,. \tag{8.50}$$

Using Eq. (8.46) this becomes

$$\frac{d}{dg}F(g)=F(g)/\beta(g)\,, \tag{8.51}$$

which can be integrated immediately,

$$F(g)=\exp\left\{\int\frac{dg'}{\beta(g')}\right\}. \tag{8.52}$$

If we choose $g^2\ll 1$, we can evaluate Eq. (8.52) using the lowest order expression [Eq. (8.45)] for $\beta(g)$,

$$F(g)=\exp\left\{-\frac{2\pi}{N-2}\frac{1}{g}\right\}. \tag{8.53}$$

Therefore the mass gap measured in units of the lattice spacing depends nonanalytically on the coupling constant. Such an effect could never be discovered using ordinary perturbation theory. It implies that the critical singularities of such models are not the traditional power laws. In this case the singular functions of the critical region should be expressed as powers of the correlation length itself.

It is interesting to consider an asymptotically free theory which has a rich mass spectrum m_i, $i=1,2,\ldots$. The $O(n)$ spin systems are not in this class because they have repulsive forces. Anyway, if a richer theory's critical point lies at $g=0$ and if it has no intrinsic mass scales, then the ratios of its dynamically generated masses m_i must be pure numbers,

$$m_i/m_j=c_{ij}\,. \tag{8.54}$$

The mass ratios can only depend on the group of the theory. The beauty of this result is that once a mass scale is set, all the masses of the theory are determined with no free parameters. It is believed that the four-dimensional non-Abelian gauge theory of strong interactions involves massless gauge fields and a doublet of (essentially) massless quarks. Then the low-energy end of the hadronic mass spectrum should satisfy Eq. (8.54). Therefore many of those curious numbers of the Particle Data Table should be predicted with no freedom!

D. Results from the Migdal recursion relation

How deep and reliable are the analogies between SU(n) ×SU(n) spin systems in two dimensions and SU(n) gauge theories in four dimensions? Exact mappings or even approximate mappings with calculable corrections, have not been derived. One of the obstacles to such a program has been the complexity of the link variables $U_\mu(n)$ and the constraints among plaquette variables following from local gauge invariance. It has already been pointed out in the discussion of Abelian lattice gauge theories that plaquette variables satisfy constraints such as Eq. (5.58). The non-Abelian analogs of such constraints are much more difficult to deal with. In the context of a renormalization group calculation, one would begin with an Action written in terms of plaquette variables, and integrate out various link variables to generate an effective Action written in terms of new plaquette variables, which satisfy the essential constraints. A practical scheme of this sort has not appeared. However, A. B. Migdal proposed some time ago (Migdal, 1975) a recursive scheme which does handle gauge degrees of freedom as well as spin degrees of freedom. It is not exact and it is not known how to calculate corrections to it. However, it is extremely clever and applies at all values of the coupling. It produces results which in many cases agree rather well with results obtained by other techniques. For example, it predicts the $g^2\ll 1$ coupling-constant renormalization equations of $O(n)$ spin systems to within 30%. Its estimates for some of the critical exponents of the two-dimensional Ising model are within 10% of the exact results. However, it has many limitations and pitfalls which have been discussed by Kadanoff (Kadanoff, 1976b).

The Migdal scheme is particularly interesting here because it relates gauge theories in four dimensions to spin systems in two dimensions in just the fashion anticipated,

4-D Gauge System		2-D Spin System
Z_N	⟷	Z_N
U(1)	⟷	U(1)
SU(N)	⟷	SU(N)×SU(N)

The correspondence means, among other things, that the coupling-constant renormalization problems in the gauge and spin systems are essentially identical. Also, the phase diagrams are related with the correspondence

$$\text{confinement} \longleftrightarrow \text{disorder}\,.$$

The SU(N)×SU(N) spin systems are predicted to have only one phase, so the scheme predicts confinement for non-Abelian gauge theories for all couplings! Detailed relations among critical exponents also follow. For example,

$$2\nu_{\text{gauge}}=\nu_{\text{spin}} \tag{8.55}$$

for the mass gap exponent.

These are extremely encouraging results for a lattice gauge enthusiast. Perhaps more exacting analyses can be made and will sharpen these correspondences.

IX. PARTING COMMENTS

The topics discussed in these lectures are for the most part "old hat." We shall close this review with a few remarks concerning current research developments.

A. M. Polyakov is initiating an ambitious program to formulate and solve lattice gauge theories in three and four dimensions by developing deep correspondences to two-dimensional spin systems (A. M. Polyakov, 1979). As mentioned earlier in the text, two-dimensional $O(n)$ Heisenberg spin systems are soluble in the continuum limit for $n \geq 3$. The reason for this lies in the fact that these theories have "hidden symmetry," which leads to an infinite number of conservation laws. In two dimensions these conservation laws prohibit particle production in scattering processes, and this simplicity leads to solubility (Zamolodchikov and Zamolodchikov, 1977). Such conservation laws cannot exist in nontrivial four-dimensional theories (Coleman and Mandula, 1967). However, Polyakov has suggested that a nontrivial extension of the conservation laws to non-Abelian pure gauge fields in four dimensions exists and may lead to an elegant, closed solution. Polyakov views non-Abelian gauge theories as chiral fields defined on closed loops in real space–time. We saw in the text that Abelian lattice gauge theory could be viewed in an analogous fashion—the physical space of states of the theory consists of closed loops of electric flux. Recall how this came about. The τ-continuum Hamiltonian of the theory was

$$H = \tfrac{1}{2}a^3 \sum_{\mathbf{n},k} E_k^2(\mathbf{n}) - (1/g^2 a) \sum_{\mathbf{n},ik} \cos\theta_{ik}(n) , \tag{9.1a}$$

where θ_{ij} is related to the magnetic field,

$$\theta_{jk} = a^2 g B_i \quad (ijk \text{ cyclic}) . \tag{9.1b}$$

The physical space of states is locally gauge invariant or, as discussed in Sec. VI, satisfies Gauss's law. Therefore only closed loops of electric flux are permissible. The loops must be closed because the theory has no sources or sinks of flux. In strong coupling only the electric term in Eq. (9.1a) is significant, and the energy of a closed loop of flux is proportional to its length. The magnetic term in Eq. (9.1a) allows the loops to fluctuate but always leaves them closed. Strong coupling expansions for the energies of such states, "glueballs" or "boxitons," then resemble mass gap calculations discussed earlier in the text (Kogut *et al.*, 1976).

The idea which Polyakov abstracts from this formulation of the theory is that phase factors defined on closed paths are the basic quantities of the theory,

$$\exp\left(ig \oint_C A(\mathbf{x}) \cdot d\mathbf{x}\right) . \tag{9.2}$$

The Hamiltonian form of the non-Abelian theory leads one to a similar perspective. In this case we have chiral fields defined on loops,

$$\prod_C \exp\{(i/2)ag\boldsymbol{\tau} \cdot \mathbf{A}_i\} , \tag{9.3}$$

as the operator which generates the physical space of locally gauge-invariant states. Next Polyakov explores the properties of chiral fields defined on the space of closed contours and finds intriguing similarities to chiral fields defined on points in two-dimensional space–time. He has proposed an infinite set of conservation laws for the gauge theories and connections to fermionic string models. Some of this work is incomplete, so it is not appropriate to discuss it further here. It is an extremely imaginative approach with some very high goals.

In the text we limited our discussion to gauge theories without matter fields. We discussed methods of labeling the theories' possible phases. Of course, such theories are unrealistic. It is believed, however, that if the pure non-Abelian gauge theory confines quarks then the theory with dynamical quark fields added will have its physical spectrum exhausted by color singlets. It is important to study this problem directly and map out the phase diagrams of models with matter fields. It is also most interesting to place the matter fields in the fundamental representation of the gauge group as in quantum chromodynamics (Fritzsch *et al.*, 1973). If this is done, the traditional method of labeling the phase diagram fails. In fact, the interquark potential can never rise linearly in such theories. This is simply a consequence of screening. Consider the theory at strong coupling, place a static quark–antiquark pair into the system, and separate them a distance R. Then in the lowest-energy state dynamical quarks will materialize near the static quarks and screen their color locally, as depicted in Fig. 40. The interquark potential then has a short range.

To study the phase diagrams of gauge theories with matter fields, it is instructive to return to Ising lattice gauge theory and incorporate into it Ising matter fields (Wegner, 1971). The degrees of freedom of this model have the group structure of interest, and the discreteness of the variables permits easy, reliable analysis. The theory is defined as follows: on sites there are Ising matter fields

$$\sigma(r) = \pm 1 , \tag{9.4}$$

and on links there are Ising gauge fields

$$U_\mu(r) = \pm 1 . \tag{9.5}$$

The gauge-invariant Action is

$$S = \beta \sum_{r,\mu} \sigma(r) U_\mu(r) \sigma(r+\mu) + K \sum_{r,\mu\nu} U_\mu(r) U_\nu(r+\mu) U_\mu(r+\nu) U_\nu(r) . \tag{9.6}$$

The local gauge symmetry simultaneously flips the matter field at r and all the link gauge fields emanating from r. The reader familiar with lattice versions of quantum chromodynamics will recognize the first term in Eq. (9.6) as the Ising form of a gauge-invariant quark kinetic energy.

Consider this theory's phase diagram in the (β,K) plane (Fradkin and Shenker, 1979). Two limiting cases are easy:

(1) $K = \infty$. The gauge field is frozen, so the Action reduces to the Ising model. For $d \geq 2$ this is a two-phase system having $\langle\sigma(r)\rangle \neq 0$ for $\beta > \beta_c$ and $\langle\sigma(r)\rangle = 0$ otherwise. The β_c critical point is shown on the right-

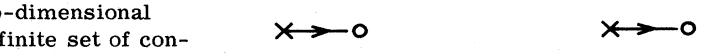


FIG. 40. Screening the long-range confining potential.

hand vertical axis of Fig. 41. In field theoretic jargon, the magnetized phase of the Ising model displays the "Higgs mechanism" because the vacuum expectation value of the charge-bearing field is nonzero.

(2) $\beta=0$. Only the Ising gauge theory remains in this case. It confines for $K<K_c$ and does not for $K>K_c$. It is plotted on the lower horizontal axis of Fig. 41.

The other boundaries of the phase diagram are trivial and are free of critical points.

It is more difficult to describe the interior of the phase diagram. First we should ask whether the critical points at $(\beta,K)=(0,K_c)$ and $(\beta,K)=(\beta_c,\infty)$ are the endpoints of critical lines which extend into the figure. Consider the critical point K_c of the pure gauge system. It is not hard to see that it extends into the phase diagram, because if β is small the matter fields' primary effect is just to renormalize the temperature K_c slightly (Wegner, 1971). This is a lattice version of vacuum polarization familiar from ordinary quantum electrodynamics. To see this, organize the calculation of the partition function by first doing the sum over matter fields. For small β this is done using high-temperature expansion methods. Define an effective Action for the gauge fields (Fradkin and Shenker, 1979)

$$\exp\{S_{\rm eff}[U_\mu(r)]\}=\sum_{\sigma(r)=\pm1}\exp\left\{\beta\sum\sigma U\sigma+K\sum UUUU\right\}. \tag{9.7}$$

Use the usual identity,

$$\exp(\beta\sigma U\sigma)=\cosh\beta(1+\sigma U\sigma\tanh\beta) \tag{9.8}$$

and note that the sum over $\sigma(r)$ picks out only closed loops of link variables U. The smallest closed loop is just a plaquette, so

$$\exp\{S_{\rm eff}[U_\mu(r)]\}\sim e^{K\Sigma UUUU}\left\{1+(\tanh\beta)^4\prod UUUU+\cdots\right\}$$

$$\sim e^{K\Sigma UUUU}e^{(\tanh\beta)^4\Sigma UUUU}$$

$$\sim\exp\left\{(K+\tanh^4\beta)\sum UUUU\right\}. \tag{9.9}$$

So, the effective Action is another pure gauge model with an effective coupling,

$$K_{\rm eff}=K+\tanh^4\beta,\quad(\beta\ll1). \tag{9.10}$$

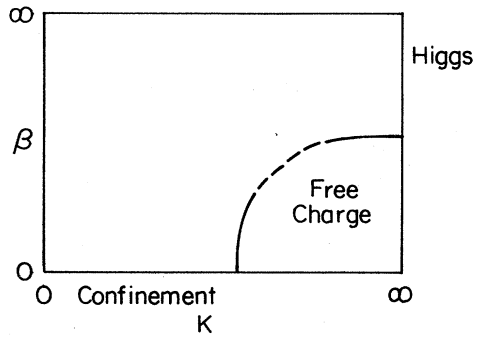


FIG. 41. Phase diagram of Ising gauge theory with matter fields.

This theory has a critical point at $K_{\rm eff}=K_c$, or

$$K=K_c-\tanh^4\beta. \tag{9.11}$$

Therefore coupling to matter fields has shifted the transition point to a larger value of the electric charge ($K=1/2e^2$) as one's physical intuition of vacuum polarization would suggest (Stack, 1978).

Similar arguments show that the spin system's transition at $(\beta,K)=(\beta_c,\infty)$ extends into the phase diagram. In three-dimensional systems this follows from the fact that the theory is self-dual (Wegner, 1971), and in four dimensions it follows from the fact that the theory is dual to a higher gauge theory (Wegner, 1971) which can be analyzed along the lines discussed above.

Although these arguments are reliable only near the edges of the phase diagram, it is reasonable to guess that the two critical lines meet inside the diagram as shown in Fig. 41. Then the lower right portion of the phase diagram would be isolated. Presumably free charges could exist in this phase.

The other portion of the diagram is very interesting. It is bounded on the right vertical side by the Higgs mechanism and on the lower left horizontal side by confinement. Although these regions of the phase diagram differ in many quantitative aspects, one can show that they are not separated by a critical surface. In other words, they are continuously connected, and no symmetry criterion or phase transition separates them. The proof of this fact begins by observing that the theory is trivial on the two boundaries $K=0$ and $\beta=\infty$. Then convergent expansion methods imply that no critical point can exist within a strip of finite width bordering this wedge-shaped boundary (Fradkin and Shenker, 1970). The free energy is analytic in this strip, so the Higgs mechanism and confinement phases are connected. This result depends crucially on the fact that the matter field belongs to the fundamental representation of the group, so that no impurities of smaller charge can be placed into the system to measure the force law. If the matter fields were not in the fundamental representation, then the $\beta\to\infty$ limit of the theory would not be trivial, and the Higgs and confinement boundaries could be separated by a critical line (Einhorn and Savit, 1978).

Fig. 41 is the simplest phase diagram which can be drawn that is consistent with the analysis done to date. The results discussed here are not particular to Ising systems. They apply equally well to lattice theories with non-Abelian gauge groups and scalar matter fields. Unfortunately, the extension to theories with real fermions remains unclear.

This article has devoted almost no space to one of the major goals of lattice gauge theory—the calculation of the mass spectrum of strongly interacting particles. The reason is that, although considerable efforts have gone into this direction, only rather crude results have been achieved. At present several research groups are attempting to develop reliable calculational tools for such projects. Strong coupling expansions, Monte Carlo integration methods, and renormalization groups are all being studied. Considerable dirty numerical work will probably be necessary before a good quantitative understanding of these theories is obtained.

Hopefully there will be considerable progress in lattice gauge theory in the next few years.

ACKNOWLEDGMENT

The author thanks the Physics Department of the University of Illinois for providing the opportunity to present a preliminary version of this article in the form of a special topic graduate course.

REFERENCES

Abers, E. S., and B. W. Lee, 1973, Phys. Rev. C **9**, 1.
Anderson, P. W., 1976 (unpublished).
Anderson, P. W., G. Yuval, and D. R. Hamann, 1970, Phys. Rev. B **1**, 4464.
Balian, R., J. M. Drouffe, and C. Itzykson, 1975, Phys. Rev. D **11**, 2104.
Banks, T., R. Myerson, and J. Kogut, 1977, Nucl. Phys. B **129**, 493.
Baym, G., 1969, *Lectures in Quantum Mechanics* (Benjamin, New York).
Callan, C., 1970, Phys. Rev. D **2**, 1541.
Chui, S. T., and J. D. Weeks, 1976, Phys. Rev. B **14**, 4978.
Coleman, S., and J. Mandula, 1967, Phys. Rev. **159**, 1251.
Creutz, M., 1977, Phys. Rev. D **15**, 1128.
Creutz, M., L. Jacobs, and C. Rebbi, Phys. Rev. Lett. **42**, 1390 (1979).
Domb, C., 1974, in *Phase Transitions and Critical Phenomena*, Vol. 3, edited by C. Domb and M. S. Green (Academic, London).
Einhorn, M., and R. Savit, 1978, Phys. Rev. D **10**, 2583.
Elitzur, S., 1975, Phys. Rev. D **12**, 3978.
Elitzur, S., R. B. Pearson, and J. Shigemitsu, 1979, IAS preprint.
Feynman, R. P., 1948, Rev. Mod. Phys. **20**, 367.
Fisher, M. E., 1967, Rep. Prog. Phys. **30**, 615.
Fradkin, E., and L. Susskind, 1978, Phys. Rev. D **17**, 2637.
Fradkin, E., and S. Shenker, 1979, SLAC preprint.
Fritzsch, H., M. Gell-Mann, and H. Leutwyler, 1973, Phys. Lett. B **47**, 365.
Gross, D. J., and A. Neveu, 1974, Phys. Rev. D **10**, 3235.
Gross, D. J., and F. Wilczek, 1973, Phys. Rev. Lett. **30**, 1343.
Hamer, C. J., and J. B. Kogut, 1979, IAS preprint.
Jordan, P., and E. Wigner, 1928, Z. Phys. **47**, 631.
Jose, J., L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, 1977, Phys. Rev. B **16**, 1217.
Kadanoff, L. P., 1976*a*, in *Phase Transitions and Critical Phenomena*, Vol. 5*A*, edited by C. Domb and M. S. Green (Academic, London).
Kadanoff, L. P., 1976*b*, Ann. Phys. (N.Y.) **100**, 359.
Kadanoff, L. P., 1977, Rev. Mod. Phys. **49**, 267.
Kadanoff, L. P., and H. Ceva, 1971, Phys. Rev. B **3**, 3918.
Kogut, J. B., 1976, in *Many Degrees of Freedom in Particle Theory*, edited by H. Satz (Plenum, New York).
Kogut, J. B., D. K. Sinclair, and L. Susskind, 1976*a*, Nucl. Phys. B **114**, 199.
Kogut, J. B., and L. Susskind, 1975, Phys. Rev. D **11**, 395.
Kosterlitz, J. M., 1974, J. Phys. C **7**, 1046.
Kosterlitz, J. M., and D. J. Thouless, 1973, J. Phys. C **6**, 118.
Kosterlitz, J. M., and D. J. Thouless, 1977 (to be published).
Kramers, H. A., and G. H. Wannier, 1941, Phys. Rev. **60**, 252.
Lane, K., 1974, Phys. Rev. D **10**, 1353.
Mermin, N. D., and H. Wagner, 1966, Phys. Rev. Lett. **17**, 1133.
Messiah, A., 1958, *Quantum Mechanics*, Vol. I (Wiley, New York).
Migdal, A. A., 1975a, Zh. Eksp. Teor. Fiz. **69**, 810.
Migdal, A. A., 1975b, Zh. Eksp. Teor. Fiz. **69**, 1457.
Nelson, D. R., and J. M. Kosterlitz, 1977, Phys. Rev. Lett. **39**, 1201.
Pfeuty, P., 1970, Ann. Phys. (NY) **57**, 79.
Politzer, H. D., 1973, Phys. Rev. Lett. **30**, 1346.
Polyakov, A. M., 1975a, Phys. Lett. B **59**, 79.
Polyakov, A. M., 1975*b*, Phys. Lett. B **59**, 82.
Polyakov, A. M., 1979 (submitted to Phys. Lett. B).
Raby, S., and A. Ukawa, 1978 (unpublished).
Savit, R., 1977, Phys. Rev. Lett. **39**, 55.
Schultz, T. D., D. C. Mattis, and E. H. Lieb, 1964, Rev. Mod. Phys. **36**, 856.
Schwinger, J., 1958, Proc. Natl. Acad. Sci. U.S.A. **44**, 956.
Schwinger, J., 1959, Phys. Rev. Lett. **3**, 296.
Spitzer, F., 1964, *Principles of the Random Walk* (Van Nostrand, Princeton).
Stack, J., 1978 (unpublished).
Stanley, H. E., 1971, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University, New York).
Stanley, H. E., and T. A. Kaplan, 1966, Phys. Rev. Lett. **17**, 913.
Streater, R. F., and A. S. Wightman, 1964, *PCT, Spin and Statistics and All That* (Benjamin, New York).
Symanzik, K., 1970, Commun. Math. Phys. **49**, 424.
't Hooft, G., 1972, unpublished remarks at the Marseilles Conference on Gauge Theories.
't Hooft, G., 1978, Nucl. Phys. B **138**, 1.
Villain, J., 1975, J. Phys. C **36**, 581.
Webster, E., G. Webster, and M. Chester, 1979, Phys. Rev. Lett. **42**, 243.
Wegner, F., 1971, J. Math. Phys. **12**, 2259.
Wilson, K. G., 1974, Phys. Rev. D **14**, 2455.
Wilson, K. G., and J. Kogut, 1974, Phys. Rep. C **12**, 75.
Wortis, M., 1974, in *Phase Transitions and Critical Phenomena*, Vol. 3, edited by C. Domb and M. S. Green (Academic, London).
Yang, C. N., and R. L. Mills, 1954, Phys. Rev. **96**, 1605.
Zamolodchikov, A., and Al. Zamolodchikov, 1978, ITEP preprint 35.