Nuclear Physics B170 [FS1] (1980) 16–31

A PHYSICAL PICTURE FOR THE PHASE TRANSITIONS IN Z_N SYMMETRIC MODELS

Martin B. EINHORN[1] and Robert SAVIT[2]

Physics Department, University of Michigan, Ann Arbor, Michigan 48109, USA

Eliezer RABINOVICI

Physics Department, Hebrew University, Jerusalem, Israel

Received 14 February 1980

We show how the phase transitions in Z_N symmetric spin and gauge theories can be understood as being caused by condensations of topological excitations. For the two-dimensional Z_N periodic gaussian and vector Potts models, there are two phase transitions, the first caused by a condensation of strings (domain boundaries) and the second by an unbinding of vortices. The relationship between our picture and the double Coulomb gas representation of Kadanoff is discussed. Using our representation, we also explain the correspondence between these models and the recent theory of two-dimensional melting of Halperin and Nelson. Finally, we describe generalizations of our picture to higher dimensions and to gauge theories.

1. Introduction

Theories with a Z_N symmetry are of interest to a large class of physicists. In elementary particle physics, the importance of the center (Z_3) of the SU(3) color gauge group has been emphasized by 't Hooft [1] and Polyakov [2], who have argued that quark confinement may be characterized as triality confinement, so that studies of the phases of the Z_3 gauge theories may lead to important insights. Moreover, it may be that the phase structure of the Z_3 gauge theory in four dimensions is related to that of the Z_3 spin model in two dimensions [3, 8–10]. The Z_N gauge theory also arises naturally in a certain limit of the abelian Higgs model with charge N [4].

In condensed matter physics Z_N symmetric theories appear naturally in the problem of two-dimensional melting [5]. A crystal in two dimensions has discrete rotational invariance in the plane of the crystal. For a square lattice this symmetry is Z_4, while for a triangular lattice the symmetry is Z_6. Part of the problem of melting is associated with the restoration of full rotational symmetry which can be

[1] Work supported by the Department of Energy contract number EY-76-C-02-1112.

[2] Work supported by the National Science Foundation grant number PH78-08426.

broken to a Z_4 or Z_6 symmetry at low enough temperatures. In the problem of two-dimensional helium films [6], Z_N symmetric interactions are of interest, since the substrate on which the film is placed may have a crystalline structure which gives rise to Z_N symmetric forces acting on the helium.

For all these examples, the central issue is to understand the phase structure of the various Z_N symmetric theories. In many abelian and non-abelian theories, there are coherent excitations which have topological significance. Frequently, the phase transitions of these theories can be understood as being caused by a condensation of these topological excitations [7]. Familiar examples (reviewed briefly below) include the Ising and *x*-*y* models. In this paper, we describe the phase structure of Z_N symmetric spin and gauge theories in a similar way.

Recently, a number of papers have appeared discussing the likely phase properties of these theories [8–10]. In particular, convincing evidence has been amassed, based both on qualitative and quantitative considerations, that for $N \gtrsim 5$ there are three phases. Our considerations were motivated by this work and are consistent with their conclusions; however, our point of view is rather different. Although we shall describe the topological excitations in detail below, at this point we would like to emphasize that they are not entirely the same as the charges of the electric and magnetic Coulomb gas representation discussed by Kadanoff [11]. Rather, all the topological objects in our picture have a "magnetic" character and have a simple, immediate relationship to the original degrees of freedom of the model.

To illustrate our approach we will, in sect. 2, focus on a two-dimensional globally Z_N invariant theory which one may call the "Z_N periodic gaussian model" (PGM)*. This theory has the same topological excitations and is believed to have the same phase structure as the Z_N vector Potts model** but is somewhat more tractable. Indeed, in sect. 2 we shall sometimes regard it as a kind of simplified representation of the vector Potts model. On the other hand, the Z_N PGM is of physical interest in its own right, and it is the simplest non-trivial Z_N theory which has all of the features we wish to stress. In sect. 3 we shall describe generalizations of our picture to higher dimensions and to gauge theories.

Before discussing our picture of the Z_N theories, it will be useful to review briefly how the critical properties of the two-dimensional Ising and *x*-*y* models can be understood as condensations of topological excitations. In the case of the *x*-*y* model [13], the topological excitations are vortex points in two dimensions. An analysis of the hamiltonian for this model (using either duality transformations or other methods) indicates that the vortices at low temperatures occur in vortex–antivortex pairs bound together by a logarithmic potential. However, the entropy for such a pair is also logarithmic in their separation, and at a certain temperature, T_c, the free energy is minimized when the vortex and the antivortex are separated

* This form of representation was originally introduced for the continuous planar spin model by Berezinski [12a] and further developed by Villain [12b].

** This model also goes under the name "clock model".

by an arbitrarily large distance. Thus, above T_c the vortices behave like a plasma of free charges. This condensation into a plasma phase causes the *x-y* spin–spin correlation function, Γ, to behave like $e^{-\mu r}$ for large r, r being the separation between the spins, whereas for $T < T_c$ $\Gamma \sim r^{-\eta}$, where η is a function of T.

It is possible to understand the phase transition of the $d = 2$ Ising model in terms of topological excitations also. In this case the topological excitations are just the domain boundaries between islands of aligned spins. These domain boundaries form closed loops on the links of the dual lattice. (We always have in mind square or cubic lattices: the dual lattice is obtained by shifting the original lattice by half a lattice spacing in each direction.) This is easily seen as follows. Each link of the dual lattice crosses one and only one link of the original lattice. For each nearest neighbor pair of Ising spins which point in opposite directions color the corresponding dual lattice link red. It is then easy to see that any configuration of Ising spins on the original lattice gives rise to a configuration of closed red loops on the dual lattice. Moreover, up to a global flipping of all the spins, the correspondence is one to one.

Now these domain boundaries have a finite energy per unit length. At low temperatures the dominant configurations are small closed loops occurring with a fairly small density (i.e., at low temperatures most spins point in the same direction). However, a closed loop of a given total length has an entropy proportional to its length (modulo logarithms) [14]. Thus there is a competition in the free energy of a domain boundary between the energy and the entropy, both of which are proportional to L. At a certain temperature the minimum of the free energy shifts from $L = 0$ to $L = \infty$ (in this crude approximation), and it becomes favorable to create relatively many closed loops of arbitrary size. But this is just a description of the disordered phase of the Ising model—i.e., many misaligned nearest neighbor pairs. Thus the phase transition can be thought of as being produced by a condensation of the domain boundaries which are the topological excitations of this model*.

In the rest of this paper we shall proceed to understand the phase transitions of Z_N symmetric theories in an analogous way. Because Z_N is intermediate between U(1) and Z_2, we will find that the topological excitations have characteristics of those found in U(1) theories as well as in Ising-like systems. This gives rise to rather interesting conclusions about the kinds of configurations that are likely to dominate Z_N symmetric theories in various phases.

The outline of the rest of the paper is as follows. In sect. 2 we discuss the Z_N PGM in two dimensions. Subsect. 2.1 is devoted to a brief review of the phase properties of this model. In subsect. 2.2 we describe the topological excitations of

* Using this picture, a crude approximation to the string entropy consists of supposing that the number of configurations is given by a non-backtracking, but otherwise unrestricted, random walk. Balancing the entropy and energy for a string of length L, one calculates a critical temperature $\beta = \frac{1}{2} \ln 3 \approx 0.55$, which is rather close to the exact value of $\beta = \frac{1}{2} \ln(1 + \sqrt{2}) \approx 0.44$.

the model and write the partition function directly in terms of these topological variables. In subsect. 2.3 we show how the phase transitions of the model can be understood as condensations of the topological excitations of the model. In sect. 3 we present several comments which include the application of our picture to Z_N spin theories in higher dimensions and to gauge theories, the relationship of our picture to the double Coulomb gas picture of Kadanoff [11] and the close relationship of our work to theory of two-dimensional melting presented by Halperin and Nelson [5].

2. The two-dimensional Z_N spin model

2.1. REVIEW OF PHASE PROPERTIES

The two-dimensional Z_N PGM and the two-dimensional vector Potts model are believed to have the same phase structure. In this section we will describe these models and will review what this phase structure is expected to be. For simplicity we will generally work in the context of the Z_N PGM, but one should bear in mind also the vector Potts model.

In the Z_N vector Potts model one represents the two-dimensional spin vector by a complex phase, $e^{i(2\pi/N)q}$ with $0 \leqslant q \leqslant N-1$, at each site of a square lattice. The partition function is

$$Z = \sum_{\{q=0\}}^{N-1} \exp\left(\beta \sum_{\langle\,\rangle} \cos\left[\frac{2\pi}{N}\Delta_\mu q_i\right]\right), \tag{2.1}$$

where i is a two-dimensional vector labelling the lattice sites, and $\Delta_\mu q_i \equiv q_i - q_{i-\hat{\mu}}$ (for simplicity, we drop the vector notation on i). The sum in the exponent is over all nearest neighbor pairs on the lattice, and β is the inverse temperature. This theory is invariant under the global rotation $q_i \to q_i + a$, where a is an integer, and also under the "local" transformation $q_i \to q_i - N\Lambda_i$ where Λ_i is an integer. This local symmetry just expresses the fact that the hamiltonian in (2.1) is periodic.

A more mathematically tractable model having essentially the same structure is the Z_N PGM defined by the partition function

$$Z = \sum_{\{q=0\}}^{N-1} \sum_{\{l=-\infty\}}^{\infty} \exp\left[-\tfrac{1}{2}\beta \sum \left(\frac{2\pi}{N}\Delta_\mu q_i + 2\pi l_{\mu;\,i}\right)^2\right]. \tag{2.2}$$

Once again there is an integer valued variable, q_i, on each site of the lattice as well as an integer valued variable $l_{\mu;\,i}$ on each link of the lattice. The sum in the exponent is over all lattice links. After summing over all $l_{\mu;\,i}$ from $-\infty$ to $+\infty$, it is clear that the logarithm of the argument of the sum over $\{q\}$ will be invariant under $q_i \to q_i + a$, as well as $q_i \to q_i - N\Lambda_i$. The second, local invariance is precisely the result of summing over the $l_{\mu;\,i}$. Thus (2.2) considered as a function of the q_i's has the same symmetry as (2.1). (As a function of q_i and $l_{\mu;\,i}$, (2.2) has something like a local gauge symmetry with $l_{\mu;\,i}$ playing the role of a vector potential.) In the limit that $N \to \infty$, (the *x-y* model limit) (2.2) approximates (2.1) at sufficiently low temperatures (up to an overall field-independent factor).

For general N the model (2.2) (and (2.1)) is thought to have two phase transitions as a function of β [8–10]. At low temperatures (large β) the system is in a ferromagnetic state. Define $s_i = e^{i(2\pi/N)q_i}$ and $G(r) = \langle s_i s_j \rangle$, $r = |\underset{\sim}{i} - \underset{\sim}{j}|$. Then, in this phase $G(r) \to \text{const.}\ (\neq 0)$ as $r \to \infty$, and $\langle s_i \rangle \neq 0$. At a certain value of β, β_1, there is a phase transition and the system passes into a phase characterized by $\langle s_i \rangle = 0$ and $G(r) \sim r^{-\eta}$ as $r \to \infty$ with η a function of β. Thus, in this phase the ferromagnetic order disappears, but the system is not totally disordered since $G(r)$ falls to zero only like a power and not like an exponential. The power-law behavior of $G(r)$ signals the appearance of a massless excitation in this phase, in close analogy with the low-temperature phase of the *x-y* model. As the temperature is increased still further one encounters another phase transition at $\beta = \beta_2$. For $\beta < \beta_2$ the system is in a true disordered state with $\langle s_i \rangle = 0$, and $G(r) \sim e^{-\mu r}$ for large r. For $N < N_c$ ($N_c \approx 5$) [8–10] this scenario breaks down and there is only one phase transition connecting the low- and high-temperature phases described above. The middle (massless) phase disappears.

Because the Z_N PGM is self-dual, the two phase transitions are related to each other by $\beta_1 \beta_2 = N^2/4\pi^2$. As $N \to \infty$ we should recover the U(1) PGM, which is expected to have the same phase structure as the *x-y* model. From these and other considerations, we anticipate that $\beta_2 \to \beta_v$, the critical point of the U(1) periodic gaussian model, and $\beta_1 \to \infty$ as $N \to \infty$. That is, the low-temperature, ferromagnetic phase of the Z_N model disappears, and we are left with a massless phase extending from $T = 0$ to $T_v \equiv \beta_v^{-1}$ (in which U(1) vortices of equal and opposite vorticity are bound in pairs) followed by a massive disordered phase (in which U(1) vortices are unbound) for $T > T_v$. The simplest dependence of β_1 and β_2 on N which realizes this behavior is $\beta_1 \approx N^2/(4\pi^2\beta_v)$ and $\beta_2 \approx \beta_v$ for large enough N.

2.2. TOPOLOGICAL EXCITATIONS

We turn now to a discussion of the excitations of the Z_N symmetric theories with topological significance. For our first intuitive arguments it is most convenient to couch the discussion in terms of the complex spins of the vector Potts model, (2.1). Precisely analogous excitations are also present in the model (2.2) as we shall see explicitly below.

First we note that since Z_N is a discrete symmetry we can have closed domain boundaries separating islands of aligned spins in analogy with the Ising model. Imagine, for example, a nearest neighbor pair connected by a lattice link in direction μ of Z_N spins which are rotated by a relative angle $\Delta_\mu q$ (e.g., by $2\pi/N$). We may associate a piece of domain boundary $j_\nu = \varepsilon_{\nu\mu} \Delta_\mu q$ with the dual lattice link which crosses the link joining these two spins. As in the two-dimensional Ising model it is clear that we will be able to generate *closed* domain boundaries on the dual lattice using this prescription, since $\Delta_\nu j_\nu = 0$. But these domain boundaries differ in two important respects from those of the Ising model. First, the boundaries can have different strengths (or flux). If the orientations of two nearest neighbors differ by an angle whose magnitude is $2\pi h/N$, where h is an integer and $0 < 2h/N \leqslant 1$, then the domain boundary separating them has strength h. One can also consider a boundary of strength h to be a superposition of h boundaries of strength one. Second, for $N \geqslant 3$ the domain boundaries are orientable; that is, they have an arrow on them. Consider for instance a region, R, of spins oriented at an angle $2\pi/N$ surrounded by spins pointing at an angle zero. R is thus circumscribed by a domain boundary of unit flux. If we now rotate all spins in R so that they point at an angle $-2\pi/N$ we will still have R circumscribed by a boundary of unit flux, but it will have the opposite sense. Thus we must associate an arrow with a domain boundary. Such a phenomenon does not occur in the Ising (Z_2) case.

Since for $N \geqslant 3$ the spins are complex phases, we might suppose that the Z_N theory also contains vortex-like excitations, in addition to the strings (domain boundaries) described above. Suppose, in analogy to the *x-y* model, that we wish to place a vortex with vorticity one at some site on the dual lattice. We then require that $\sum \Delta_\mu (2\pi/N) q_i = 2\pi$ for any closed loop surrounding the vortex, in the notation of (2.1). What is a minimum energy spin configuration which satisfies this requirement? Recall that for the *x-y* model a minimum energy vortex configuration divides the required 2π rotation more or less equally among all nearest neighbor pairs as we transverse a circle of any radius with the vortex at the center. Thus spins a distance r away from the vortex center are rotated from their neighbors by an angle $\sim 1/r$. The fact that the spins can be rotated from their neighbors by an infinitesimal amount gives rise to an energy for a single *x-y* model vortex which diverges logarithmically with the size of the system.

Return now to the Z_N case. Consider an area with diameter of order N lattice spacings surrounding the location of our Z_N vortex. Within this region the Z_N spins can divide the required 2π rotation equally among themselves as we traverse a circle surrounding the vortex. This minimizes the energy density here, and thus in this region it is difficult to tell the difference between a Z_N vortex and an *x-y* model vortex, in that the typical minimum energy spin configurations look the same this close to the vortex center. Outside of this region, however, the Z_N spins can no longer equally divide the rotation by 2π, and the best we can do is illustrated in fig. 1. Our vortex looks like a bicycle wheel with a hub of diameter $O(N)$ lattice spacings. The spokes are domain boundaries separating approximately wedge-shaped regions of aligned spins. As we cross a domain boundary the spins change their orientation by $2\pi/N$, so that traversing the N spokes the spins will have rotated by 2π. (Notice that the arrows on the domain boundaries indicate that the spin rotation always has the same sense.)

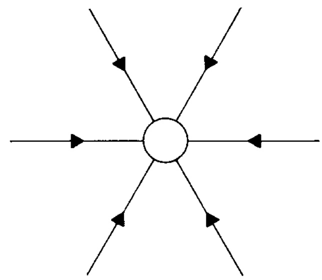

Fig. 1. A single Z_N vortex for the case $N=6$.

Since the domain boundaries have finite energy per unit length, the energy of a single Z_N vortex diverges linearly with the size of the system. Thus single vortex configurations should be suppressed, particularly at low temperatures. However, we can also create vortex–antivortex pairs with a finite energy as in the *x-y* model. Such a pair is shown in fig. 2. The strings joining the vortices are, again, the domain boundaries which have a finite energy per unit length. This gives rise to a linear potential between the vortices. This should be compared with the logarithmic potential between *x-y* model vortices.

Figs. 1 and 2 suggest that the domain boundaries can somehow be considered as currents, the divergences of which are associated with the vortices. We will now show that the strings (domain boundaries) and vortices form a complete set of variables in terms of which the partition function can be written.

For the Z_N PGM, this is very simple because the vortices and domain boundaries are easily expressed in terms of the original degrees of freedom. First we define a vector $j_{\mu;\,i}$

$$j_{\mu;\,i} \equiv \varepsilon_{\mu\nu}(\Delta_\nu q_i + N l_{\nu;\,i}) \,. \tag{2.3a}$$

The $j_{\mu;\,i}$ are naturally associated with links of the dual lattice. Note that

$$\left(\frac{2\pi}{N}\Delta_\mu q_i + 2\pi l_{\mu;\,i}\right)^2 = \frac{4\pi^2}{N^2}(\varepsilon_{\mu\nu} j_{\nu;\,i})^2 = \frac{4\pi^2}{N^2}(j_{\mu;\,i})^2 \,. \tag{2.3b}$$

Consequently, the partition function (2.2) becomes simply

$$Z = \sum_{\{q\}} \sum_{\{l\}} \exp\left[-\frac{2\pi^2\beta}{N^2} \sum_{\langle\,\rangle} (j_{\nu;\,i})^2\right], \tag{2.4}$$

where the sum in the exponent is now over dual lattice links.

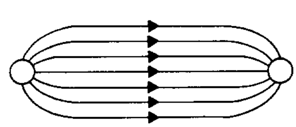

Fig. 2. A Z_N vortex–antivortex pair for the case $N=7$.

The vector $j_{\mu;\,i}$ is the topological current. The current itself represents the domain boundaries between Z_N spins of different orientation and its divergences

$$\Delta_\mu j_{\mu;\,i} = N\varepsilon_{\mu\nu}\Delta_\mu l_{\nu;\,i} \tag{2.5}$$

are the vortices (fig. 1). (The extra factor of N should cause no worry: it is clear that when $\Delta_\mu j_{\mu;\,i} = N$ the Z_N spins rotate through 2π around the dual lattice thus producing a vortex of vorticity one.) Indeed, if we define $\theta_i = (2\pi/N)q_i$, then in the limit $N \to \infty$, θ_i becomes a continuous spin angle, and eq. (2.2) becomes the partition function for the periodic gaussian approximation of the x-y model whose vortex field is given by $\varepsilon_{\mu\nu}\Delta_\mu l_{\nu;\,i}$.

We were able to make an immediate association between the original degrees of freedom of the Z_N PGM and its topological currents because the Z_N periodicity of the PGM is explicitly represented by a field, $l_{\mu;\,i}$. In other Z_N theories of interest, for example the vector Potts model, this is not the case, and it is not possible to express the theory in terms of its topological excitations by direct substitution. However, for such theories one can use a more general procedure involving duality transformations to write the theory in terms of its topological excitations. We now illustrate this method for the Z_N PGM. applying the standard duality transformation [7] to the model (2.2) we arrive at the dual form

$$Z = \sum_{r_i=0}^{N-1} \sum_{m_{\nu;i}=-\infty}^{\infty} \exp\left[-\frac{N^2}{8\pi^2\beta} \sum_{\langle\,\rangle} \left(\frac{2\pi\Delta_\nu r_i}{N} + 2\pi m_{\nu;i} \right)^2 \right], \tag{2.6}$$

where $r_i, m_{\nu;i}$ are integer fields defined on the vertices and links respectively of the dual lattice. Note that the Z_N PGM is self-dual [15]. We may write this as a Fourier series

$$Z = \sum_{j_{\nu;i}=-\infty}^{\infty} \sum_{r_i=0}^{N-1} \exp\left[-\frac{2\pi^2\beta}{N^2} \sum_{\langle\,\rangle} j_{\nu;i}^2 \right] \exp\left[i2\pi \sum_{\langle\,\rangle} \frac{\Delta_\nu r_i}{N} j_{\nu;i} \right]. \tag{2.7}$$

Thus, $j_{\nu;\,i}$ is a vector field associated with links of the dual lattice, and $\Delta_\nu j_\nu$ is conjugate to the dual spins r. Performing the sum on r_i yields the constraint

$$\Delta_\nu j_{\nu;\,i} = 0 \qquad \text{mod } N\,. \tag{2.8}$$

The general solution of the constraint equation is given precisely by eq. (2.3a). [Indeed, substituting (2.3a) into (2.7) gives back (2.2).]

In view of our intuitive discussion initiating this section, we see that the current $j_{\nu;\,i}$ itself represents the topological excitations, i.e., domain boundaries between regions having different orientations of Z_N spins, and its non-trivial divergences are precisely the vortices.

2.3. PHASE TRANSITIONS AND TOPOLOGICAL EXCITATIONS

We will now describe the phase properties of the model, (2.2) or (2.4) in terms of the topological current and its divergence. Since the Z_N symmetry is discrete, we expect that at very low temperatures ($\beta \gg 1$) the system will be in a ferromagnetic state with $\langle e^{i(2\pi/N)q_i} \rangle \neq 0$. In terms of the current, $j_{\mu;\,i}$ this phase will have a relatively small density of relatively small closed domain boundaries, as well as a small density of tightly bound vortex–antivortex pairs. Recall that the vortices are bound by a linear potential. (The current, $j_{\mu;\,i}$, has a finite energy per unit length so that the energy between two divergences connected by strings must grow linearly with the separation.) Note also that the strings binding the vortices will want to proceed from the vortex to the antivortex in almost the shortest way possible. On the other hand, the flux will prefer to be spread across a transverse region of order N lattice spacings, since the exponent in (2.4) depends on the (flux)^2. It is thus energetically favorable to have no more than one unit of flux on any lattice link. From these considerations it is clear that vortex pairs separated by a distance L take at least N times as much energy to be produced as a closed domain boundary of perimeter L. So for low temperatures closed strings should dominate in size and number over vortex–antivortex pairs.

As the temperature is increased the size and density of closed domain boundaries and vortex–antivortex pairs will increase. We now ask whether either one or both of these kinds of excitations are able to condense into a plasma-like phase. Insofar as the intervortex potential is linear the vortices will be prevented from unbinding since the entropy of a vortex pair is only logarithmic (we will come back to this below). However, we can anticipate a condensation of the closed domain boundaries. Let us restrict our attention to domain boundaries of unit flux. Higher flux boundaries will be present but are energetically disfavored, at least at low temperatures. If we ignore the orientability of the domain boundaries, then we have a situation very much like that of the two-dimensional Ising model discussed earlier. The entropy of the domain boundaries is given by a kind of modified random walk, and so we anticipate a phase transition at some temperature $T_1 = \beta_1^{-1}$. If we approximate the number of allowed configurations of a domain boundary as a non-backtracking random walk then we can approximate the free energy of such a string by

$$F(L) \approx \frac{2\pi^2\beta}{N^2} L - L \ln 3 \,. \tag{2.9}$$

For $\beta > \beta_1 \simeq N^2 \ln 3/2\pi^2$, $F(L)$ has its minimum at $L = 0$, while for $\beta < \beta_1$ the minimum of $F(L)$ is at $L = \infty$. This crude argument indicates a condensation of the domain boundaries at a temperature $\beta_1 \approx N^2 \ln 3/2\pi^2$.

We do not expect that this simple estimate of β_1 will be very accurate. First, in (2.9) we have neglected modifications to the random walk which more closely describes the actual situation. For instance, we should not allow walks which tread on a previously used link. (More precisely, such repeated walks have a higher energy.) Restrictions like these will reduce the entropy for long strings which means that the real value of T_1 will be higher than our estimate. A second point, related to the one we have just mentioned, is that we have completely neglected the orientability of the strings. Including the orientability adds a length independent factor to the entropy of a closed string (i.e., the number of possible configurations is multiplied by 2 independent of the length of the string) and so at the level of the argument leading to (2.9) does not change the estimate of β_1. However, in the real system the strings do overlap and interact and the energy associated with their overlaps does depend on their relative orientation and so could affect the value of β_1.

In a moment we will return to the discussion of the numerical value of β_1, but now let us ask how the system behaves for $\beta<\beta_1$. For $\beta<\beta_1$ the free energy (2.9) favors configurations with a relatively high density of strings of arbitrary length. This means that it is very easy for the q_i to change by one unit as we move from lattice site to lattice site. Now, we are interested in the large-distance structure of the theory, and so since in this phase the q_i can easily vary by one unit as we move to neighboring lattice sites we expect that important configurations in, say, (2.2) will include those in which the q_i's differ by arbitrarily large amounts over sufficiently long distances. Thus, for the purposes of determining the large distance structure, we should be able to replace the discrete sum over the q_i's by a continuous integral over q_i from $-\infty$ to ∞*.

Of course for any non-zero β there will be corrections to this approximation. These can be computed by using the identity

$$\sum_{q=-\infty}^{\infty} = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{d}q\, \mathrm{e}^{i2\pi qk} \tag{2.10}$$

and keeping successively higher values of $|k|$ in the sum on the right hand side. But for $\beta<\beta_1$ these corrections will affect only the short-distance structure of the theory; the long-distance behavior will be determined by the $k=0$ term**.

With this in mind, we approximate (2.2) for $\beta<\beta_1$ as

$$Z \simeq \sum_{\{l\}}{}' \int_{-\infty}^{\infty} \mathrm{D}q \exp\left[-\frac{2\pi^2\beta}{N^2}\sum_{\langle\,\rangle}(\Delta_\mu q_i + N l_{\mu;i})^2\right], \qquad \beta<\beta_1, \tag{2.11}$$

* Another way to say this is to imagine performing a real-space renormalization group procedure in which we group together spins in successively larger blocks. Since the nearest neighbor spins can freely vary by one unit, we expect that large enough blocks will include arbitrarily large values of the block spin variable with reasonably high probability. Thus, in this phase it should be a good large distance approximation to replace the discrete q_i's by continuous variables.

** The terms with $k\neq 0$ correspond, in the language of Kadanoff (ref. [11]) to electric vortices.

where the prime on the sum over $\{l\}$ means a sum over $\{l\}$ that produces distinct configurations of $\varepsilon_{\mu\nu}\Delta_\mu l_{\nu;\,i}$ as discussed in subsect. 2.2. We can now do the gaussian integral over the q_i to obtain

$$Z = Z_0 \sum_{\{m\}} \exp\left[2\pi^2\beta \sum_{i,j} m_i V_{ij} m_j - c m_i^2\right]. \tag{2.12}$$

The sum over m_i is a sum over integer valued variables $m_i = \varepsilon_{\mu\nu}\Delta_\nu l_{\mu;\,i}$ on the sites of the dual lattice. From the discussion following (2.5) these are also the divergences of the topological current, i.e., $m_i = (1/N)\Delta_\mu j_{\mu;\,i}$. The potential $V_{ij} \propto \ln|i-j|$ for $|i-j| \geqslant 1$, and c is a positive constant. The factor Z_0 is just the gaussian integral in (2.11) with all $l_\mu = 0$.

Eqs. (2.11) and (2.12) are different forms of the periodic gaussian approximation to the two-dimensional x-y model [12]. We can now easily understand the long-distance structure of the system for $\beta < \beta_1$. First, we note that as a result of the domain boundaries becoming floppy for $\beta < \beta_1$, the intervortex potential has been softened from linear to logarithmic. This effect is easy to understand in terms of our previous discussion of the x-y model. Remember that in the x-y model the vortex potential is logarithmic because the x-y model spins are continuous and can differ in orientation infinitesimally from their neighbors. The Z_N spins on the other hand can take on only a finite number of orientations. However, in the phase in which the Z_N spin domain boundaries are floppy, one can imagine performing a thermal average over all possible configurations of domain boundaries, and so generating infinitesimal relative rotations of nearest neighbor spins in an average sense. In the language of the renormalization group, we imagine making block spins as we scale to larger and larger distances. So long as all relative Z_N spin orientations occur with reasonable probability (as we expect for $\beta < \beta_1$), then the block spin variable should take on an increasing number of possible orientations as the size of the block is increased, generating finally an effectively U(1) symmetric spin. [Note that this argument is not valid for the Ising case where the block spin variable is expected to become a one component (real) continuous field rather than a two component (complex) field.] Finally, we note that, using (2.11), we can compute $\langle s_i s_j \rangle$. For $\beta > \beta_1$ we argued that this would approach a non-zero constant as $r = |i-j| \to \infty$. But for $\beta < \beta_1$ we expect that this will approach zero as $r \to \infty$. That is does so is well-known from previous studies of the x-y model. The rate at which the correlation function decays for large r (whether algebraically or exponentially) will depend on β and on the x-y model-like phase in which we find ourselves.

The phase in which we find ourselves depends on the behavior of the vortices. Since for $\beta < \beta_1$ the vortex–vortex potential is logarithmic rather than linear, the vortices can now dissociate to form a plasma in a manner familiar from the x-y model. Following Kosterlitz and Thouless [13] one can easily approximate this dissociation temperature by writing down the free energy for a vortex–antivortex pair and noting the temperature above which the entropy dominates over the energy. More sophisticated calculations of this dissociation temperature have also been done [16]. In terms of our variable β, the best value seems to be $\beta=\beta_2\approx 0.741$. Now, if $\beta_2<\beta_1$, then our theory will have three phases; a low-temperature ferromagnetic phase with small closed domain boundaries and vortex–antivortex pairs tightly bound by a linear potential, an intermediate phase in which the domain boundaries have condensed, but the vortices are still bound (albeit only logarithmically) and finally a completely disordered high-temperature phase in which the vortices unbind and condense to form a plasma of free charges. In the middle phase, $\beta_2<\beta<\beta_1$, $G(r)=\langle s_i s_j\rangle \sim r^{-\eta}$, and this phase thus has a kind of massless excitation, while for $\beta<\beta_2$, $G(r)\sim e^{-\mu r}$, indicating that in this phase a mass has been generated. If $\beta_1<\beta_2$, then the massless phase will be absent: by the time the vortex potential has changed from linear to logarithmic due to the strings becoming floppy, the vortices already have sufficient entropy to unbind and so they condense at the same temperature as the domain boundaries. From (2.7) we see that β_1 depends on N, while β_2 does not. Writing

$$\beta_1=\frac{N^2 f}{2\pi^2}\approx\frac{N^2\ln 3}{2\pi^2}\ ,$$

we see that we should have three phases for $N>N_c=2\pi(\beta_2/2f)^{1/2}\approx 3.65$, while for $N<N_c$ we expect only two phases.

These results are in qualitative agreement with the work of refs. [8–10], although our crude estimate for β_1 incorrectly suggests two phase transitions for $N=4$ which is clearly wrong. But more important for our picture is that the dependence of β_1 on N is the same as that deduced in refs. [8–10], (see also subsect. 2.1) but from different arguments. Finally, we remark that since the model (2.2) is self-dual, the two phase transitions (assuming there are only two) must satisfy $\beta_1\beta_2=N^2/4\pi^2$. Using our rough approximation for β_1 gives instead

$$\beta_1\beta_2\simeq\frac{N^2}{4\pi^2}\left(\frac{2\ln 3}{1.35}\right)\simeq\frac{N^2}{4\pi^2}1.63\ ,$$

which is another measure of the crudeness of our approximation for β_1. (A more accurate estimate of N_c may be obtained by enforcing duality. Assuming $\beta_2\approx\beta_{xy}\approx 0.741$, then $\beta_1=N^2/4\pi^2\beta_2$. But we have $\beta_1>\beta_2$ only for $N>2\pi\beta_{xy}\approx 4.66$.)

We close this section by noting that since (2.2) is self-dual our entire picture can be reversed and we could discuss the phase transitions as condensations of electric vortices (replacing the magnetic strings) and electric domain boundaries (replacing the magnetic vortices). We have emphasized the totally magnetic picture, because given the original form of the theory, these variables are the most natural ones*.

* It is important to note that this same change from magnetic strings and vortices to electric vortices and strings is also possible in the vector Potts model even though that model is not self-dual. The [illegible]ce of self-duality means that the interactions among the electric variables will not be the same interactions among the magnetic variables. See also ref. [7].

3. Comments

3.1. Z_N LATTICE GAUGE THEORIES AND Z_N SPIN MODELS IN HIGHER DIMENSIONS

Following our analysis of the two-dimensional Z_N spin model, we should be able to easily describe the topological excitations of Z_N spin models in higher dimensions, and of Z_N lattice gauge models, and perhaps say something about their phase properties. It should be clear that in d dimensions the topological excitation of the vector Potts model will be closed $d-1$ dimensional domain boundaries (the analogue of our closed strings) and N open $d-1$ dimensional manifolds which co-terminate on a $d-2$ dimensional (closed) manifold (the analogue of our vortex-antivortex pair connected by N strings). The Z_N lattice gauge theory in $d+1$ dimensions has the same topological excitations as the Z_N vector Potts model in d dimensions. This follows the general pattern expected from the study of duality for abelian theories [7], and is simply related to the fact that the gauge theory interactions are defined on a 2-dimensional surface (elementary lattice plaquettes), while the spin theory has interactions defined on lattice links which are one-dimensional. Thus, in a fixed number of space dimensions, the dimension of the domain boundaries and their divergences (the "vortex"-like excitations) are reduced by one for the gauge theory.

Using the fact that the periodic gaussian form of the four-dimensional Z_N lattice gauge theory is self-dual, the authors of refs. [8–10] were able to generalize the arguments used for the two-dimensional Z_N periodic gaussian spin model to show that for $N > N_c$ there must also be at least two phase transitions for the four-dimensional Z_N lattice gauge theory*. From our point of view, the first (lower temperature) phase transition should correspond to a condensation of the closed two dimensional domain boundaries** taking us from a ferromagnetic to a massless phase, while the second phase transition should correspond to the condensation of the open bounded manifolds. Unfortunately, it is not easy to make quantitative arguments in this case, because of the difficulty of computing the entropy of a two-dimensional sheet. Although not much is known about random two-dimensional flopping, qualitative arguments have been presented [18] which make this picture quite plausible.

Another case of interest is the 3-dimensional Z_N spin model, which by duality can be transformed into the three-dimensional Z_N lattice gauge theory. Again we have two types of excitations with topological significance. From the spin model point of view they are closed two-dimensional surfaces and open, bounded two-dimensional surfaces, while from the gauge model point of view they are closed

* The beautiful Monte Carlo studies of the four-dimensional Z_N lattice gauge theories of ref. [17] are a strong confirmation of the picture of two phase transitions for $N > 5$.

** Note that these are not really domain boundaries in the sense that they do not enclose a region of four-space. They only enclose a region of three-space orthogonal to the direction defined by an excited vector potential. They are also gauge invariant.

strings and strings with monopoles on the ends. The latter picture is easier to analyze, and using it one would conclude that for general N there is only one phase transition. In the low-temperature phase the string free energy has its minimum at L = string length = 0, and so there is a linear binding potential between monopole-antimonopole pairs. At the phase transition, when the strings become floppy, the monopole-antimonopole potential is softened to $\propto 1/r$, and so the monopole-antimonopole pairs unbind at the same temperature.

3.2. RELATION TO THE ELECTRIC AND MAGNETIC COULOMB GAS REPRESENTATION

Kandanoff and others [8–11] have shown that the two-dimensional Z_N spin system has a representation in terms of a gas of interacting electric and magnetic charges. Similarly, the four-dimensional lattice gauge theory has a representation in terms of a soup of interacting electric and magnetic current loops [8–10]. The relationship between this representation and our picture is quite straightforward.

To help us understand this relationship, consider the periodic gaussian approximation to the two-dimensional *x-y* model. This model has topological excitations which are vortex points interacting through a logarithmic potential. On the other hand, this model is dual to the two-dimensional discrete gaussian model. The "topological excitations" of the discrete gaussian model are the closed domain boundaries (strings) between islands of different values of the discrete gaussian field. Thus we see that in two dimensions we can replace a set of vortices by a set of strings. This is precisely analogous to the way in which the representation of Kadanoff differs from our picture. By undoing one of the duality transformations necessary to get the Coulomb gas representation, we are able to replace a set of electric charges by a set of magnetic strings. Of course both representations retain the magnetic charges which become the *x-y* model vortices in the limit $N \rightarrow \infty$. In a similar way, we can replace the electric current loops of the four-dimensional Z_N gauge theory by magnetic sheets to get our representation in terms of magnetic sheets and magnetic sheets terminating on magnetic loops rather than electric loops and magnetic loops.

3.3. RELATION TO TWO-DIMENSIONAL MELTING

Recently Halperin and Nelson [5] have suggested that the melting of a two-dimensional crystal occurs via two separate phase transitions. Each phase transition can be understood as being caused by the condensation of a different kind of topological defect of the crystal. Roughly, their picture can be described in the following way. First, they identify two kinds of defects, dislocations and disclinations. At low temperatures the dislocations are bound in pairs by a logarithmic potential, while the disclinations are bound by a potential which grows like the square of the separation. Next, they define separate correlation functions, Γ_{T} and

Γ_R which measure the degree of translational and rotational symmetry. For the triangular (square) lattice which they consider Γ_R has the same form as the spin–spin correlation function for our $d = 2$ spin theory with a Z_6 (Z_4) symmetry. In the low temperature phase of the crystal, $\Gamma_T \sim r^{-p}$ and $\Gamma_R \to \text{const}$ ($\neq 0$) as $r \to \infty$. Thus we have discretely broken rotational symmetry and a translational symmetry which would be broken in the usual way except that we are in two dimensions. As we raise the temperature, the dislocations suddenly unbind and condense into a plasma phase. In this middle phase $\Gamma_T \sim e^{-\mu r}$ and $\Gamma_R \sim r^{-\eta}$ as $r \to \infty$. Thus we have restored translational symmetry, but we still have long-range rotational order, although not of the ferromagnetic type (again, because we are in two dimensions). In addition, the plasma of dislocations screens the potential between the disclinations making it logarithmic and setting the stage for the next phase transition which occurs when the disclinations unbind. In the third, high-temperature phase, Γ_T continues to fall exponentially at large distances and also $\Gamma_R \sim e^{-mr}$ so we have a complete restoration of both translational and rotational symmetry.

In our model we have no simple analogue of the translational degrees of freedom of the crystal, but our Z_N spins are clearly analogous to the crystal's rotational degrees of freedom. Thus both systems have a Z_N symmetry ($N = 4$ for a square crystal and $N = 6$ for a triangular crystal). In both systems there are two phase transitions (which are thought to be essential singularities), and the large-distance behaviors of Γ_R in the crystal and of the spin-spin correlation function in our Z_N models are the same in each of the three phases. Furthermore, in both systems each phase transition is caused by the unbinding and condensation of one of two types of topological excitations. Finally, in both systems the unbinding of the topological excitations that gives rise to the second (high-temperature) phase transition is only possible because their binding energy is screened by the plasma of topological excitations whose condensation caused the first (low-temperature) phase transition.

Of course, the two systems are not identical. In particular, the energetics of analogous configurations in the two systems are not the same. This gives rise to several differences, one of which is that even in the case of a square crystal with a Z_4 symmetry one expects two separate phase transitions for melting while for the Z_4 vector Potts or Villain model there is only one. However, because the underlying rotational symmetry of the two theories is the same their phase behavior is quite analogous.

One of us (R.S.) would like to thank T. Banks, A. Schwimmer, T. Witten, and especially R. Pearson for very stimulating discussion. R.S. is also grateful to the Nuclear Physics Department of the Weizmann Institute of Science and the Theory Division of CERN for their hospitality while part of this work was being done. E.R. appreciates the support of Lawrence Berkeley Laboratory, where a portion of this work was performed.

References

[1] G. 't Hooft, Nucl. Phys. B138 (1978) 1; B153 (1979) 141
[2] A.M. Polyakov, Trieste preprint (1977)
[3] A.A. Migdal, JETP (Sov. Phys.) 69 (1975) 810, 1957;
L.P. Kadanoff, Rev. Mod. Phys. 49 (1977) 267
[4] T. Banks and E. Rabinovici, LBL preprint (1979);
D. Horn, M. Weinstein and S. Yankielowicz, Phys. Rev. D19 (1979) 3715
[5] B. Halperin and D. Nelson, Phys. Rev. Lett. 42 (1978) 121; Phys. Rev. B19 (1979) 2457
[6] M. Bretz, J. de Phys. C6 (1978) 1348, and references therein
[7] R. Savit, Univ. of Michigan preprint UM HE 79-8 (1979), Rev. Mod. Phys. (April, 1980), to appear
[8] S. Elitzur, R. Pearson and J. Shigemitsu, Phys. Rev. D19 (1979) 3698
[9] J. Cardy, UCSB preprint no. TH-30 (1978)
[10] A. Ukawa, P. Windey and A. Guth, Phys. Rev. D21 (1980) 1013
[11] L. Kadanoff, J. Phys. A11 (1978) 1399
[12] (a) V.L. Berezinski, ZhETF (USSR) 59 (1970) 907; 61 (1972) 1144 [JETP (Sov. Phys.) 32 (1971) 493; 34 (1972) 610];
(b) J. Villain, J. de Phys. C36 (1975) 581
[13] J. Kosterlitz and D. Thouless, J. de Phys. C6 (1973) 1181
[14] J.M. Hammersley, Proc. Camb. Phil. Soc. 53 (1957) 642; 57 (1961) 516
[15] A. Casher, Nucl. Phys. B151 (1979) 353;
C.P. Korthals Altes, Nucl. Phys. B142 (1978) 315
[16] J. Jose, L. Kadanoff, S. Kirkpatrick and D. Nelson, Phys. Rev. B16 (1977) 1217;
J. Kosterlitz, J. Phys. C7 (1974) 1046
[17] M. Creutz, L. Jacobs and C. Rebbi, Phys. Rev. D20 (1979) 1915
[18] M. Einhorn and R. Savit, Phys. Rev. D19 (1979) 1198