PHYSICAL REVIEW D VOLUME 11, NUMBER 8 15 APRIL 1975

Gauge fields on a lattice. II. Gauge-invariant Ising model

R. Balian, J. M. Drouffe, and C. Itzykson

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP No. 2, 91190 Gif-sur-Yvette, France

(Received 29 August 1974)

We study the case of a discrete local gauge group Z_2 in order to discuss the existence of a transition in dimension $d \geq 3$. We compute the critical constant for $d = 3$ and 4 and show that in three dimensions the transition is a second-order one.

I. INTRODUCTION

We have described in a previous paper[1] [hereafter referred to as (I)] a gauge theory on a lattice according to Wilson's ideas.[2] It was suggested that the system undergoes phase transitions. For a small enough coupling constant the gauge field behaves qualitatively as ordinary perturbation theory in the continuous limit would indicate, while beyond some critical coupling a new phase sets in. The long-range forces become so strongly attractive that they provide a binding mechanism for charged particles.

We want to look here more closely at the nature of the transition. To simplify as much as possible we make use of a possibility afforded by discretization, namely the introduction of a finite local gauge group instead of the usual continuous Lie local gauge group. The simplest one is Z_2, the (multiplicative) group with two elements $\{1, -1\}$. In (I) the local group was $O(n)$, $n \geqslant 2$. The present case amounts to setting $n = 1$. A drawback of this choice might be the absence of Goldstone bosons in the ordered phase. We notice, however, with reference to (I), that as the dimension d of the lattice gets large, mean-field theory hardly distinguishes $n = 1$ from $n \geqslant 2$. In particular, it predicts in all cases a first-order transition. Consequently, as an instrument to investigate the validity of mean-field theory, the present simplification is not too drastic.

With Z_2 as gauge group, we can use several devices introduced in the context of Ising models. The first of these is specific to a system with configuration variables taking values ± 1. It is a duality transformation to be described in Sec. II. Combining this duality with some mathematical results (collected in an appendix) enables one to reduce the problem in three dimensions to a standard Ising model of which much is known. It also allows one to locate exactly the critical constant in four dimensions (Sec. III). We shall also use the Griffiths-Kelly-Sherman inequalities,[3] which express the fact that strengthening "ferromagnetic" couplings can only strengthen correlations. This is explained in Sec. IV where a discussion of a global-order parameter is given.

Once some exact results are known, approximation methods, such as perturbation theory around mean-field theory in the Coulomb gauge, can be checked. If they are found reliable, one can then proceed to use them in more general cases. This is another motivation for the present work.

We recall the model. Let there be given a hypercubical lattice in d dimensions with unit spacing. We introduce a spatial cutoff by retaining N sites (eventually we let $N \to \infty$). To each link (ij) of neighboring sites we assign a variable $A_{ij} = A_{ji}$ taking the values ± 1. A set of four neighboring links is a plaquette $p = (ijkl)$. We compute a partition function

$$Z \equiv 2^{-Nd} \sum_{\{A_{ij} = \pm 1\}} \exp\left(\beta_l \sum_l A_{ij} + \beta_p \sum_p A_{ij} A_{jk} A_{kl} A_{li} \right), \tag{1.1}$$

and define a free energy as

$$F \equiv \lim_{N \to \infty} \frac{1}{N} \ln Z . \tag{1.2}$$

We wish to study the occurrence and properties of phase transitions as β_l and β_p vary. Had we insisted on presenting a gauge-invariant version of the Ising model, we would have introduced extra variables $k_i = \pm 1$ at each site and replaced the β_l term by $\beta_l \sum_l k_i A_{ij} k_j$. However, the gauge transformation $A_{ij} \to k_i A_{ij} k_j$ eliminates the k's while leaving the plaquette coupling invariant, and reduces the problem to the study of (1.1). We shall mostly be interested in the case $\beta_l = 0$, which we call the pure-gauge-field case.

II. DUALITY

Application of duality transformations to Ising-type models is well known. It has recently been further developed by Wegner.[4] Owing to this circumstance, although our presentation is slightly different, we shall be rather brief.

Geometrical duality transforms q-dimensional manifolds into $(d - q)$-dimensional ones. Let us present it for $d = 3$. We introduce a dual lattice obtained from the original one through a translation by an amount $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. There is a natural correspondence between a site, a link, a plaquette, or a cube on the original lattice, and a cube, a plaquette, a link, or a site on the dual one. For instance, to a link is associated the dual plaquette that it intersects. This construction is easily generalized to any dimension and is used as follows.

(*i*) $d=2$. We write (1.1) as

$$Z=2^{-2N}(\cosh\beta_l)^N(\cosh\beta_p)^{2N}\sum_{\{A_{ij}=\pm1\}}\prod_l(1+\tanh\beta_l A_{ij})\prod_p(1+\tanh\beta_p A_{ij}A_{jk}A_{kl}A_{li})\ , \tag{2.1}$$

expand the products, and sum over $A_{ij}=\pm1$. Nonvanishing terms are in one-to-one correspondence with configurations of P distinct plaquettes selected on the lattice. The boundary of each configuration is defined as the set of L links which belong to one and only one of the selected plaquettes. A configuration contributes a term $(\tanh\beta_l)^L(\tanh\beta_p)^P$. If a plaquette is selected, let us set $s_i=-1$ at the corresponding site i of the dual lattice and $s_i=+1$ otherwise. Obviously $P=\sum_i\frac{1}{2}(1-s_i)$ and $L=\sum_l\frac{1}{2}(1-s_i s_j)$, the summations running over sites and links of the isomorphic dual lattice. Consequently, we have

$$Z=(\tfrac{1}{4}\cosh\beta_p\cosh^2\beta_l)^N\sum_{\{s_i=\pm1\}}\exp\Big[\sum_l\tfrac{1}{2}\ln\tanh\beta_p(1-s_is_j)+\sum_i\tfrac{1}{2}\ln\tanh\beta_p(1-s_i)\Big]\ . \tag{2.2}$$

Up to a factor we recognize the partition function Z_I for an Ising model in an external field H. Letting

$$Z_I\equiv2^{-N}\sum_{\{s_i=\pm1\}}\exp\Big(\sum_l\beta_*s_is_j+\sum_i Hs_i\Big)\ ,$$
$$F_I\equiv\lim_{N\to\infty}\frac{1}{N}\ln Z_I\ , \tag{2.3}$$

we get the equality

$$F(\beta_l,\beta_p)=\tfrac{1}{2}\ln(\sinh^2 2\beta_p\ \sinh2\beta_p)+F_I(\beta_*,H)\ ,$$
$$\beta_*=-\tfrac{1}{2}\ln\tanh\beta_l,\quad H=-\tfrac{1}{2}\ln\tanh\beta_p\ . \tag{2.4}$$

As long as $H\neq0$, the system exhibits no transition, while for $H=0$ there exists a critical value β_c of β_* ($\sinh2\beta_c=1$) separating two phases: a disordered one for $\beta_*<\beta_c$ and an ordered one for $\beta_*>\beta_c$.

The condition $H\neq0$ means β_p is finite. As $\beta_p\to\infty$ we find a transition for $\tanh\beta_l=e^{-2\beta_c}$, i.e., $\beta_l=\beta_c$. This is in agreement with the discussion given in (I): No transition occurs for finite β_l, β_p. We knew already that for $\beta_l=0$ the gauge-field model is trivial for $d=2$, with $F(0,\beta_p)=\ln\cosh\beta_p$.

(*ii*) $d=3$. Repeating the previous argument we find that the coupled model is self-dual with

$$F(\beta_l,\beta_p)=\tfrac{3}{2}\ln(\sinh2\beta_l\sinh2\beta_p)$$
$$+F(-\tfrac{1}{2}\ln\tanh\beta_p,-\tfrac{1}{2}\ln\tanh\beta_l)\ . \tag{2.5}$$

Note the interchange of indices (l,p) between the two sides of this equality. We exhibit the symmetry of this self-duality by introducing the bounded variables

$$\xi_l=\ln(1+e^{-2\beta_l})\ ,$$
$$\xi_p=\ln(1+e^{-2\beta_p})\ ,$$
$$0\le\xi_l,\ \xi_p\le\ln2\ ,$$

and the function

$$f(\xi_l,\xi_p)\equiv F(\beta_l,\beta_p)-\tfrac{3}{2}\ln(1+e^{2\beta_l})(1+e^{2\beta_p})\ . \tag{2.6}$$

Then this function f is symmetric with respect to the line $\xi_l+\xi_p=\ln2$:

$$f(\xi_l,\xi_p)=f(\ln2-\xi_p,\ln2-\xi_l)\ . \tag{2.7}$$

Let C be a simple closed curve on the lattice (for d arbitrary). We defined in (I) the average $\mathfrak{C}$ of the product of the link variables A_{ij} along C:

$$\mathfrak{C}\equiv\Big\langle\prod_C A_{ij}\Big\rangle=Z^{-1}2^{-Nd}\sum_{\{A_{ij}=\pm1\}}\exp\Big(\beta_l\sum_l A_{ij}+\beta_p\sum_p A_{ij}A_{jk}A_{kl}A_{li}\Big)\prod_C A_{ij}\ . \tag{2.8}$$

This is also equal to the average of the product of all plaquette variables $A_{ij}A_{jk}A_{kl}A_{li}$ for a set S of plaquettes bounded by C. For $d=3$, we find by applying the duality transformation

$$\mathfrak{C}=\Big\langle\exp\Big(-2\beta_*\sum_{S_*}s_{ij}\Big)\Big\rangle_*\ . \tag{2.9}$$

The $\langle\ \rangle_*$ average is computed with the dual coupling constants: $\beta_p\to-\frac{1}{2}\ln\tanh\beta_p$ for plaquettes and $\beta_l\to\beta_*=-\frac{1}{2}\ln\tanh\beta_p$ for links of the dual lattice (to which are associated variables $s_{ij}=\pm1$). Formula (2.9) allows an interpretation of $\mathfrak{C}$ as

$$\mathfrak{C}=\exp[-(\mathfrak{F}-\mathfrak{F}')]\ , \tag{2.10}$$

where $\mathfrak{F}$ is the free energy of the dual model, and $\mathfrak{F}'$ the similar quantity obtained by reversing the sign of the coupling constants on the links of the dual model belonging to S_*, that is, all the links intersecting the surface S.

(*iii*) $d=4$. In this case a cube is dual to a link, a plaquette to a plaquette. We can define a free energy $F_*(\beta_{*l}, \beta_{*p})$ for a dual model by

$$\exp(NF_*) = 2^{-6N} \sum_{\{s_{ijkl}=\pm 1\}} \exp\left(\beta_{*l} \sum_p s_{ijkl} + \beta_{*p} \sum_c \prod_{\alpha=1}^{6} s_{i_\alpha j_\alpha k_\alpha l_\alpha}\right). \tag{2.11}$$

Each plaquette $(ijkl)$ of the dual lattice carries a variable $s_{ijkl} = \pm 1$. These are combined six by six along the faces of three-dimensional cubes to give the interaction $\prod_{\alpha=1}^{6} s_{i_\alpha j_\alpha k_\alpha l_\alpha}$. We defined $\beta_{*l} = -\frac{1}{2}\ln\tanh\beta_p$, $\beta_{*p} = -\frac{1}{2}\ln\tanh\beta_l$. Between the original model and the dual one, duality yields the relation

$$F(\beta_l, \beta_p) = 2\ln\tfrac{1}{2}\sinh 2\beta_l + 3\ln 2\sinh 2\beta_p + F_*(\beta_{*l}, \beta_{*p}) . \tag{2.12}$$

III. CRITICAL COUPLINGS IN THREE AND FOUR DIMENSIONS

Duality has given a complete solution in two dimensions. In particular, we have recovered the fact that the pure gauge system undergoes no transition. The results of the previous section will enable us to show that a transition occurs in three and more dimensions. In the pure gauge system, the critical values of β_p are given by

$$d=3, \quad \beta_c = 0.7613 ,$$
$$d=4, \quad \beta_c = 0.4407 . \tag{3.1}$$

Let us see how these values are obtained.

In three dimensions the coupled model is self-dual. If one sets $\beta_l = 0$ in (2.5), infinities occur on the right side while the left side is obviously finite. The required cancellations are exhibited on the form (2.7) which reduces to

$$f(\ln 2, \xi_p) = f(\ln 2 - \xi_p, 0) . \tag{3.2}$$

Thus, the study of the pure gauge model is equivalent to the study of the coupled system in the limit $\beta_{*p} \to \infty$ ($\xi_{*p} = 0$). We then expect the gauge field to reduce to a pure gauge, as discussed in (I). This is indeed true, since

$$\tilde{Z} \equiv Z(\beta_{*l}, \beta_{*p})(\cosh\beta_{*p})^{-3N}\big|_{\tanh\beta_{*p}=1} = (\cosh\beta_{*l})^{3N} \sum_{\substack{\{s_{ij}=\pm 1\} \\ \{s_{ij}s_{jk}s_{kl}s_{li}=1\}}} \prod_l (1 + \tanh\beta_{*l} s_{ij}) . \tag{3.3}$$

The last sum is taken under the constraints that the product $s_{ij}s_{jk}s_{kl}s_{li}$ for every plaquette is equal to 1. A theorem (see Appendix) states that the general solution of these constraints is $s_{ij} = s_i s_j$, with s_i defined up to an overall sign. Consequently, we obtain

$$\tilde{Z} = \tfrac{1}{2} \sum_{\{s_i=\pm 1\}} \exp\left(\beta_{*l} \sum_l s_i s_j\right) , \tag{3.4}$$

which means that for $d=3$ the pure-gauge-field model is related to the Ising model:

$$F(\beta_l = 0, \beta_p) = -\tfrac{1}{2}\ln 2 + \tfrac{3}{2}\ln\sinh 2\beta_p + F_I(-\tfrac{1}{2}\ln\tanh\beta_p) , \tag{3.5}$$

where F_I is the free energy of the three-dimensional Ising model. The Ising model is known to have a unique second-order transition in three dimensions. The techniques for the proof are based on the arguments of Griffiths and Peierls[3] to be discussed in the next section. The value β_I at which F_I is singular is obtained numerically from the high-temperature expansion[5] and is given by

$$\frac{1}{2\beta_I} = 2.255\,16 . \tag{3.6}$$

From the relation $-\frac{1}{2}\ln\tanh\beta_c = \beta_I$, we find the value given in (3.1).

We turn to the case $d=4$. We use Eqs. (2.12) and (2.11) and set $\beta_l = 0$ or $\tanh\beta_{*p} = 1$. This amounts to sum over plaquette variables s_{ijkl} constrained by $\prod_{\alpha=1}^{6} s_{i_\alpha j_\alpha k_\alpha l_\alpha} = 1$, the product running over the faces of every cube. The same theorem quoted in the Appendix states that, up to a gauge transformation, one has then $s_{ijkl} = s_{ij}\, s_{jk}\, s_{kl}\, s_{li}$. Taking into account this gauge arbitrariness, we sum freely over the variables s_{ij} and divide by the "volume" of the gauge group. By inspection of (2.11) it is seen that, as we let β_l approach zero or β_{*p} approach infinity, we can extract an infinite term from F_* (proportional to β_{*p}). Using then the above expression for s_{ijkl}, we recover the original pure-gauge-field model. The final result is expressed as a self-duality formula:

$$F(0, \beta_p) = 3\ln\sinh 2\beta_p + F(0, -\tfrac{1}{2}\ln\tanh\beta_p). \tag{3.7}$$

This is a remarkable result,[4] analogous to the

Kramers-Wannier duality for the Ising model in two dimensions. If we assume that the gauge field undergoes a unique transition for $d=4$ as it did for $d=3$, then the critical constant follows from (3.7):

$$\beta_c = -\tfrac{1}{2}\ln\tanh\beta_c \ \text{ or } \ \sinh 2\beta_c = 1\,. \tag{3.8}$$

This yields the value quoted in (3.1).

The trend exhibited by (3.1) (to which we can add $\beta_c=\infty$ for $d=2$) is a clear decrease of the critical value as d increases. In fact, we expect a behavior in $1/d$ for d large.

IV. GLOBAL ORDER PARAMETER

From now on we set $\beta_l=0$. In order to analyze further the nature of the transition, it is interesting to find a quantity with a qualitatively discontinuous behavior. Due to gauge invariance the choice of such an order parameter is not straightforward. For instance, if we look at the Green's function

$$\langle \{A_{ij}A_{jk}A_{kl}A_{li}\}\{A_{mn}A_{np}A_{pq}A_{qm}\}\rangle_{\text{connected}}$$

pertaining to two plaquettes far apart, we expect it to decrease exponentially with the distance both for β_p small and β_p large. However, Wilson[2] has suggested that the average $\mathfrak{C}$ along the closed curve C defined in (2.8) may be used to define ordering. We recall from (I) that for $d=2$, $-\ln\mathfrak{C}$ is proportional to $|S|$, the area of the set S of plaquettes enclosed by C. For β_p large enough, it was also made plausible that if $d\geq 3$, $-\ln\mathfrak{C}$ increases like the length $|C|$ of C. We prove the following result for the present model.

Theorem: Let $d\geq 3$ and $|S|$ denote the minimal area enclosed by C; if β_p is small enough, there exist two positive constants a_1 and a_2 such that

$$a_2 \leq -\frac{\ln\mathfrak{C}}{|S|} \leq a_1\,. \tag{4.1}$$

Set $\beta_l=0$ and expand both numerator and denominator of (2.8) in powers of $t=\tanh\beta_p$. Let a diagram D be a set of plaquettes chosen on the lattice (which we take at first as finite). The boundary of D, noted ∂D, is the set of links which belong to an odd number of plaquettes of D. If ∂D is empty, we say that D is closed. Denoting by $|D|$ the number of plaquettes of D, we have

$$\mathfrak{C}=\frac{\mathcal{P}}{\mathfrak{Q}}\,,\quad \mathcal{P}=\sum_{D:\,\partial D=C} t^{|D|}\,,\qquad \mathfrak{Q}=\sum_{D:\,\partial D=\emptyset} t^{|D|}\,. \tag{4.2}$$

The sum in the denominator $\mathfrak{Q}$ includes the empty diagram $D=\emptyset$ which gives a contribution equal to 1. Among the diagrams involved in the numerator $\mathcal{P}$ of (4.2), let us distinguish the family of *irreducible* diagrams $\tilde{D}$, defined as follows. A diagram with boundary C is called irreducible if it does not contain any closed subset of plaquettes. When a diagram with boundary C is reducible, it may be decomposed at least in one fashion into an irreducible part and a closed part, having no plaquettes in common. This is easily seen by repeatedly stripping the diagram from closed parts.

The first part of the proof consists in showing that $\mathfrak{C}$ is smaller than the contribution $\tilde{\mathcal{P}}$ of irreducible diagrams:

$$\mathfrak{C}<\tilde{\mathcal{P}}=\sum_{\substack{\tilde{D}:\,\partial\tilde{D}=C\\ \tilde{D}\ \text{irreducible}}} t^{|\tilde{D}|}\,. \tag{4.3}$$

Consider the product $\tilde{\mathcal{P}}\mathfrak{Q}$. It contains the contribution to $\mathcal{P}$ of irreducible diagrams, coming from the term 1 in $\mathfrak{Q}$. It also contains the contribution to $\mathcal{P}$ of each reducible diagram counted as many times as this diagram may be decomposed into an irreducible part (contribution to $\tilde{\mathcal{P}}$) and a closed part (contributing to $\mathfrak{Q}$), i.e., at least once. Finally, the product $\tilde{\mathcal{P}}\mathfrak{Q}$ generates additional terms which do not appear in $\mathcal{P}$, and for which some plaquettes are repeated twice. All terms are positive, since we have "attractive" interactions. Therefore $\tilde{\mathcal{P}}\mathfrak{Q}$ is larger than $\mathcal{P}$ and (4.3) holds.

An analogous result had been established[6] in the context of the Ising model, for which the two-point correlation function was shown to be bounded by the sum of all self-avoiding walks. The present diagrams appear as two-dimensional extensions of the Ising ones: Plaquettes replace links, and the boundary contour C replaces the two end points. Irreducible diagrams, which are the self-avoiding walks in the Ising case, have here a more complicated topology, since the plaquettes of the two dimensional irreducible manifold are not naturally ordered as are the links of a walk.

In the second part of the proof, we provide an upper bound for the number $\tilde{n}_k$ of irreducible diagrams made of exactly k plaquettes and bordered by C. Since k is at least equal to the minimal area $|S|$ enclosed by C, we have

$$\tilde{\mathcal{P}}=\sum_{k\geq|S|}\tilde{n}_k t^k\,. \tag{4.4}$$

Let us define an iterative process designed to generate at least all irreducible diagrams. For this purpose, we number once for all the links of the lattice. The construction starts from the contour C. We pick along C the link of lowest rank, and select a plaquette p_1 adjacent to this link: There are $2d-2$ such possible choices. We now define a new contour C_1 along which we shall add the second plaquette:

$$C_1=C\Delta\partial p_1\,.$$

This is the symmetric difference between the contour C and the boundary of p_1, i.e., the set of links which belong either to C or to ∂p_1 but not to both. The plaquette p_2 is chosen among those which border the link of lowest rank of C_1, and so on. At each step of this iterative process, we perform the following operations:

(i) Identify along C_{q-1} the link of lowest rank,
(ii) select a plaquette p_q adjacent to this link, and
(iii) introduce a new contour $C_q = C_{q-1}\Delta\partial p_q$.

This construction stops at some finite stage if the resulting contour is empty. We thus obtain a finite ordered set of plaquettes having C as boundary. We denote it by $\hat{D}$. Some of these sets $\hat{D}$ are genuine diagrams contributing to $\mathcal{P}$, but it may happen that such a set contains some plaquette more than once and thus cannot occur as a diagram D. If the ordering is ignored, a given set of plaquettes may, of course, be obtained several times. Let $\hat{n}_k$ be the number of sets $\hat{D}$ with k plaquettes (distinct or not). At each step there are $2d-2$ possibilities for adding a plaquette. Thus, we find at most $(2d-2)^k$ ordered sets at stage k (the construction might indeed have stopped before). Among them, those for which $C_k = \emptyset$ are obviously only a subclass, and therefore

$$\hat{n}_k < (2d-2)^k .$$

We can improve our bound by modifying the definition of the sets $\hat{D}$ (and correspondingly of their number $\hat{n}_k$); we exclude the sets with overlapping plaquettes, by changing the rule (ii). If the link of lowest rank on C_{q-1} belongs to C, there are still at most $2d-2$ possible choices for p_q. If, however, it does not, we have already selected an odd number of plaquettes adjacent to it, and hence the number of choices is then at most $2d-3$. Since at least $k-|C|$ steps involve such a link, we obtain now the bound

$$\hat{n}_k < (2d-2)^{|C|}(2d-3)^{k-|C|}. \qquad (4.5)$$

It remains to show that any irreducible diagram $\tilde{D}$ is obtained (at least once) as an ordered set $\hat{D}$. This will imply $\tilde{n}_k < \hat{n}_k$. For this purpose, given an irreducible diagram $\tilde{D}$, let us order its k plaquettes by using the above procedure. Instead of performing our choice of plaquettes over all those of the lattice, we restrict this choice to those of $\tilde{D}$, keeping otherwise the same rules. At stage q, we note that C_{q-1} is the boundary of the remaining plaquettes of $\tilde{D}$. Hence, each link belonging to C_{q-1} borders an odd number of remaining plaquettes, and thus step (ii) is always possible, unless C_{q-1} is empty. This, however, cannot happen for $q \leq k$, k being the number of plaquettes of $\tilde{D}$; otherwise the remaining plaquettes would form a closed subset, and $\tilde{D}$ would be reducible. Thus, indeed, we have $\tilde{n}_k < \hat{n}_k$.

Taking into account this result with (4.3), (4.4), and (4.5), we obtain

$$\mathfrak{C} < \sum_{k \geq |S|} (2d-2)^{|C|}(2d-3)^{k-|C|}t^k = \left(\frac{2d-2}{2d-3}\right)^{|C|} \frac{[(2d-3)t]^{|S|}}{1-(2d-3)t}, \qquad (4.6)$$

provided that

$$t < \frac{1}{2d-3}. \qquad (4.7)$$

If the curve C gets very large in such a way that $|C|/|S| \to 0$, we have established the left inequality (4.1).

In order to get the other inequality (4.1), we need the Griffiths-Kelly-Sherman (GKS) result[3] which we recall for completeness. Consider N sites with variables $\sigma_i = \pm 1$ attached to each one. Let $\Lambda = \{R, S, \ldots\}$ be the family of all subsets of sites and write $\sigma_R \equiv \prod_{i\in R}\sigma_i$. Define

$$\langle\sigma_R\rangle = \left[\sum_{\{\sigma_i=\pm1\}} \sigma_R \exp\left(\sum_{S\in\Lambda} J_S\sigma_S\right)\right] \times \left[\sum_{\{\sigma_i=\pm1\}} \exp\left(\sum_{S\in\Lambda} J_S\sigma_S\right)\right]^{-1}. \qquad (4.8)$$

Then if $J_R \geq 0$ for all R, GKS state that

$$\langle\sigma_R\rangle \geq 0, \qquad \langle\sigma_R\sigma_S\rangle \geq \langle\sigma_R\rangle\langle\sigma_S\rangle. \qquad (4.9)$$

To apply this result to $\mathfrak{C}$ we write it as

$$\mathfrak{C} = \left\langle \prod_S A_{ij}A_{jk}A_{kl}A_{li} \right\rangle, \qquad (4.10)$$

where the product runs over a minimal set S of plaquettes with $\partial S = C$ and area of $S = |S|$. Applying inequality (4.9) we find

$$\mathfrak{C} \geq \langle A_{ij}A_{jk}A_{kl}A_{li}\rangle^{|S|}, \qquad (4.11)$$

in fact, no matter what the dimension or the value of β is. This establishes the right inequality in (4.1) and the theorem is proved.

It is likely that (4.1) can be strengthened to yield

$$-\frac{\ln\mathfrak{C}}{|S|} \to \text{constant}$$

for β_p small enough and C going regularly to infinity. From (4.7) we derive that, if β_c is the largest value for which (4.1) holds, then

$$\tanh\beta_c \geq \frac{1}{2d-3}. \qquad (4.12)$$

Thus for $d=2$ there is no transition as expected, while (4.12) is easily verified for $d=3$ and 4 using the critical constants given in (3.1). For d large we expected $\beta_c \sim 1/d$, which is again in agreement with (4.12).

It would be nice to complete this theorem by proving that for β_p large enough and $d \geq 3$, $-\ln \mathfrak{C}$ behaves like $|C|$. In fact, it is sufficient to establish it for $d=3$ (where a transition is known to occur). An argument based on inequality (4.9) will then show that it also applies to $d>3$. Using duality, this amounts to studying a small β_* property for a corresponding three-dimensional Ising problem. Perturbation expansion of $\mathfrak{F}-\mathfrak{F}'$ in Eq. (2.10) in powers of β_* shows that, to a finite order, $-\ln \mathfrak{C}$ is indeed proportional to $|C|$. Although strong indications exist that the result holds beyond perturbation theory, we have not been able to find a completely satisfactory proof. This is unfortunate, for it would have demonstrated the existence of a transition in any dimension $d \geq 3$, with $-(\ln \mathfrak{C})/|S|$ as an order parameter. The order of the transition remains questionable: We have shown above that a second-order transition exists for $d=3$, but mean-field predictions seem to indicate first-order transitions for $d \geq 4$.

APPENDIX

We sketch some results of cohomology on a lattice analogous to similar properties of differential forms in the continuous case.

On a hypercubical lattice in d dimensions we define $d+1$ sets $\mathfrak{L}_p$ of functions with values in the group $Z_2 \equiv \{1, -1\}$. The set $\mathfrak{L}_0$ contains functions defined at each site, the set $\mathfrak{L}_1$ functions on links, $\mathfrak{L}_2$ on plaquettes, $\mathfrak{L}_3$ on cubes and so on. Sites, links, plaquettes, cubes, ... are simplexes of dimension 0, 1, 2, 3, ... In each $\mathfrak{L}_p$ a privileged element e assigns the value $+1$ to all p-dimensional simplexes. A product of two elements in $\mathfrak{L}_p$ is the function that assigns to each simplex the product of the values of the given elements: e is a unity for this product. Each element is idempotent and $\mathfrak{L}_p$ is a group. Let ∂ be a map from $\mathfrak{L}_p$ to $\mathfrak{L}_{p+1}$ $(0 \leq p \leq d-1)$ defined as follows. For each $\varphi \in \mathfrak{L}_p$, $\partial\varphi$ assumes on a $(p+1)$-dimensional simplex a value equal to the products of the values of φ on the p-dimensional simplexes of its boundary. For instance, if $\varphi \in \mathfrak{L}_1$, $\partial\varphi_{ijkl} = \varphi_{ij}\varphi_{jk}\varphi_{kl}\varphi_{li}$. It is clear that $\partial(\varphi\psi) = \partial\varphi\partial\psi$, $\partial e = e$, and $\partial(\partial\varphi) = e$. It is convenient to define $\mathfrak{L}_{-1} = Z_2$ and extend the definition of ∂ for an element $\pm$ of $\mathfrak{L}_{-1}$ as the constant function which assigns this value at each site.

Theorem: if $\varphi \in \mathfrak{L}_p$ and $\partial\varphi = e$, then $\varphi = \partial\psi$, for some $\psi \in \mathfrak{L}_{p-1}$. Of course, ψ is arbitrary to the extent that it can be multiplied by an element of the form $\partial\psi'$, $\psi' \in \mathfrak{L}_{p-2}$ (a generalized gauge transformation).

The most elementary case is with $\varphi \in \mathfrak{L}_0$: $\partial\varphi = e$ means $\varphi = \text{const}$, which by its very definition means $\varphi = \partial\psi$, $\psi \in \mathfrak{L}_{-1}$.

A general proof of the theorem is not very instructive. Let us rather discuss as an example the case $d=3$, $p=2$. On an ordinary cubic lattice we have a function φ with value ± 1 for each plaquette, such that the product of its values on the faces of each cube is $+1$. By duality, this provides us with a function $\tilde{\varphi}$ defined on the links of a dual lattice, such that the product of the six values corresponding to the six links incident on a site is $+1$. Let us mark all the links where $\tilde{\varphi}$ assumes the value -1. Because of the aforementioned condition this set can be decomposed (perhaps not uniquely) into elementary closed circuits. Each of those can be considered as the boundary of a surface made of plaquettes. This choice involves, of course, a large arbitrariness. We define a function $\tilde{\psi}$ equal to -1 on these plaquettes and $+1$ otherwise. Consider the product of the values of $\tilde{\psi}$ on the four plaquettes having a fixed link in common. An even number of these plaquettes carries a value $\tilde{\psi} = -1$ if this link does not belong to the closed circuits, and an odd one if it does. Thus the product in question is precisely equal to the value of $\tilde{\varphi}$ on this link. If we now return to the original lattice, there corresponds to $\tilde{\psi}$ a function $\psi \in \mathfrak{L}_1$ and the property just proved means $\partial\psi = \varphi$.

---

[1]R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D 10, 3376 (1974). This paper, to which we refer as (I), contains the notations and motivations for the present one.

[2]K. G. Wilson, Phys. Rev. D 10, 2445 (1974).

[3]R. B. Griffiths, J. Math. Phys. 8, 478 (1967); D. G. Kelly and S. Sherman, J. Math. Phys. 9, 466 (1968); R. E. Peierls, Phys. Rev. 54, 918 (1938). For an illuminating review of the mathematical aspects of the Ising models see Ginibre, in Cargèse Lectures in Physics, 1973 (Gordon and Breach, New York, to be published).

[4]F. J. Wegner, J. Math. Phys. 12, 2259 (1971).

[5]M. Fisher and D. S. Gaunt, Phys. Rev. 133, 225 (1964).

[6]M. Fisher, Phys. Rev. 162, 480 (1967).