PHYSICAL REVIEW D VOLUME 19, NUMBER 12 15 JUNE 1979

Phase structure of discrete Abelian spin and gauge systems

S. Elitzur, R. B. Pearson, and J. Shigemitsu
Institute for Advanced Study, Princeton, New Jersey 08540
(Received 30 January 1979)

It is shown that in a two-dimensional Z_p spin system for p not too small there exists a massless phase in the middle between the ordered and disordered Ising-type phases. A similar thing happens in a four-dimensional Z_p gauge theory, where a massless QED-like phase appears between the screened and the confined phases. The existence of the middle phase is deduced logically from the existence of such a phase in the continuous O(2)-invariant models using self-duality and correlation inequalities. For the spin case the transition towards this phase is analyzed using a Kosterlitz type of renormalization group suggesting an essential singularity of the correlation length at both transition points. A Hamiltonian strong-coupling expansion up to ninth order is applied to the Z_p spin system. The results of the Padé analysis of this expansion are consistent with the phase structure described above. For $p \leq 4$ the analysis suggests two phases with a conventional singularity behavior at the transition. In the nontrivial case of $p = 3$, critical exponents are calculated and found to give good agreement with experiment. For $p \geq 5$ the analysis favors three phases with an essential singularity at the transition.

I. INTRODUCTION

It is generally accepted that the two-dimensional plane rotor model (XY model) has a rather peculiar phase structure.[1,2,3] At high enough temperature we have a massive phase with finite correlation length, qualitatively similar to the high-temperature behavior of other spin models. As the temperature goes below some critical value T_k, one finds for any $T < T_k$ a massless soft behavior with power-behaved correlation function and continuously varying exponents. This low-temperature phase is rather different from the ordered phase of the Ising model, which is a broken-symmetry massive phase. The character of the transition at T_k is also different from that of the Ising model. It was suggested by Kosterlitz[2] that the correlation length has an essential singularity at T_k of the type $\exp[C(T-T_k)^{-1/2}]$, rather than the normal algebraic singularity of the Ising model. One can go more gradually from the Ising model, which possesses a Z_2 symmetry, to the XY model by considering a series of Z_p models and sending p to infinity. The Z_p model is one in which each classical spin can form only one of the p discrete angles $\theta_m = 2\pi m/p$ with some fixed direction in the space of internal degrees of freedom. In this work we study the phase structure of these Z_p models, both for their own possible experimental importance[4] and for their role as a bridge between the two different Ising and XY behaviors.

Similar considerations apply to four-dimensional Abelian lattice gauge theories. The U(1) gauge theory (periodic QED) has a confined massive phase for high temperature in which the Wilson loop decays according to the area law.[5] For small temperatures we have the massless behavior of a free electromagnetic field with the Wilson loop decaying like the exponential of the perimeter.[6] In the massless phase external charges are expected to have a long-range Coulomb interaction. At some finite temperature a transition is expected of which only a rough qualitative understanding exists.[7,8] The Z_2 gauge theory has a similar confining high-temperature phase, but a rather different low-temperature behavior. In the lower phase[6] the Wilson loop obeys a perimeter law,[9,7] but, unlike the continuous U(1) case, there is a finite mass gap. The gauge field being massive, external charges interact only at short range, and we have a Higgs-type screened phase. Here, again, a study of Z_p gauge theory, apart from its own possible importance,[10] may supply the connection between these two different types of behavior.

At first glance one might think that the softness of the low-temperature phase of the XY model is a result of the existence of a continuous symmetry. In two dimensions a complete breaking of this symmetry with the development of a Goldstone pole at zero momentum is forbidden.[11,12] Still, in the low-temperature phase of the XY model the symmetry is locally broken over not too large regions producing long-range power-behaved correlations and a soft infrared spectrum. According to this outlook, one would expect that a Z_p model for which the symmetry is discrete will present an Ising-type phase structure with two massive phases, and that the soft phase of the XY model will appear only in the limit of a continuous symmetry when $p \to \infty$. In this work we shall argue that this picture cannot be correct. Rather, when increasing

p above some critical value p_c, a massless soft phase will appear between the two Ising-type ordered and disordered massive phases, so that the masslessness appears already at finite p for discrete symmetric models. When p is sent to infinity, the lowest-ordered phase shrinks down to zero temperature in accordance with the theorem forbidding the existence of an ordered phase for continuous symmetry,[11,12] and the remaining two phases are those of the XY model. In fact, this is not new. Jose, Kadanoff, Kirkpatrick, and Nelson found long ago such a phase structure for the XY model with an O(2)-breaking Z_p-invariant term for large enough p.[3] What we want to emphasize here is, in the first place, that this is also the behavior of the pure Z_p model in which the strength of the above-mentioned breaking term is infinite. Second, and more important, this phase structure is a logical consequence of the existence of a massless low-temperature phase in the continuously symmetric XY model. Assuming the existence of such a phase in four-dimensional U(1) gauge theory, we are able to get similar results for four-dimensional Z_p gauge models. Thus, we suggest that, for p larger than some critical value, the Z_p gauge theory in four dimensions has three phases. The low-temperature phase is a Higgs-type screened massive one with a perimeter-law behavior of the Wilson loop. The high-temperature phase is a confining massive phase presenting an area-law behavior of the Wilson loop. These are the usual types of phases expected for the Z_2 theory. For $p > p_c$ another phase has to exist in between which is QED-like, i.e., a massless phase with the Wilson loop decaying like the exponential of the perimeter and long-range Coulomb forces between external charges. For a special form of the interaction proposed by Villain,[13] both the Z_p two-dimensional spin system[3] and the four-dimensional gauge system[14] turn out to be self-dual. In that case we have an upper bound on p_c in terms of T_k, the critical temperature of the continuously symmetric model, i.e., XY model or U(1) gauge theory.

We have no further quantitative information about gauge systems. As to the spin system, by passing to the Coulomb-gas representation of Refs. 3 and 15, the Z_p symmetric renormalization-group equations proposed in Ref. 3 are rederived. On the basis of these equations we argue that, for p large enough, the two transitions leading to the middle massless phase are of the Kosterlitz type with an essential singularity of the correlation length.

In order to study the Z_p spin model more quantitatively we apply to it the high-temperature expansion method with Padé analysis. As mentioned in Ref. 16 for the XY model, better results are obtained with these methods if they are applied to a Hamiltonian system with discrete space and continuous time rather than to a Euclidean lattice system. Following this advice, we study the Hamiltonian model of a chain of Z_p spins analogous to the Euclidean two-dimensional Z_p system. It is found that the model is self-dual for any p. This is important for the strong-coupling expansion treatment, since one is then able to expand in a variable which is explicitly self-dual, and thus to increase considerably the efficiency of the method. A strong-coupling expansion of the energy gap is then constructed up to ninth order for various values of p including $p=\infty$, i.e., the XY model. For the XY model we only improve the eighth-order results of Ref. 16 by adding the ninth order. As to the finite-p cases, we find that for $p \leq 4$ the strong-coupling analysis is consistent with a normal second-order transition at the self-dual point. In fact, the cases $p=2$ and $p=4$ are trivial. For the Z_3 model we calculate the critical exponent ν at the transition and find good agreement with relevant experimental results.[4] For $p>4$ our strong-coupling results indicate a transition occurring off the self-dual point, so that by duality there have to be two transitions. We also analyze the type of singularity that the energy gap develops at the transition point and find that our results are more consistent with the Kosterlitz-type essential singularity than with a usual algebraic singularity.

In the next section we show the necessity of the existence of a massless middle phase for two-dimensional spin systems. Section III deals with four-dimensional gauge systems. In Sec. IV we deal with the Coulomb-gas representation and renormalization-group equations. The Hamiltonian is introduced and its self-duality is demonstrated in Sec. V. Section VI contains the strong-coupling expansion analysis.

II. DISCRETE vs CONTINUOUS SPIN MODELS

The XY model consists of a two-dimensional lattice of classical spins rotating in a plane with nearest-neighbor interaction. In the usual form of the model the interaction energy is proportional to the scalar product of two neighboring spins. The partition function is

$$Z(\beta)=\int_{-\pi}^{\pi}\prod_i d\theta(i)\exp\left\{\beta\sum_{i,\mu}\cos[\theta(i)-\theta(i+e_\mu)]\right\}. \tag{2.1}$$

In (2.1), i denotes a lattice site and e_μ, $\mu=1,2$ are the two lattice unit vectors. This model was studied by Kosterlitz and Thouless.[1,2] By classifying the possible excitations of the system as spin waves and vortices they showed that it has two phases. The high-temperature phase in which the vortices are frequent is a massive one, i.e., it possesses a finite correlation length. In the lower phase, where the vortices are neutralized by antivortices bound to them, the system is much more ordered in the sense that the correlation function decays like a temperature-dependent power of the distance. It is therefore a massless phase with infinite correlation length. Still, this phase is not strictly ordered in the sense of showing a finite magnetization, since the breaking of a continuous symmetry is forbidden in two dimensions.[11,12] The vortices have been shown to interact via Coulomb forces. By applying a renormalization group to this plasma of vortices, Kosterlitz[2] showed that the correlation length has an essential singularity at the critical temperature of the form $\exp[C(T-T_k)^{-1/2}]$.

A somewhat different form of an XY model was introduced by Villain[13] and studied extensively by Jose, Kadanoff, Kirkpatrick, and Nelson (JKKN).[3] In this form one replaces the Gibbs factor $\exp(\beta\cos\Delta\sigma)$ in (2.1) by a periodic Gaussian function of $\Delta\theta$ [or the Jacobi function $\theta_3(2\Delta\theta)$]. Thus,

$$Z_{XY}^V(\beta)=\int_{-\pi}^{\pi}\prod_i d\theta(i)\prod_{i,\mu}\exp[f_\beta(\theta(i)-\theta(i+e_\mu))]\,, \tag{2.2}$$

where

$$\exp[f_\beta(x)]=\sum_{m=-\infty}^{\infty}\exp[-\tfrac{1}{2}\beta(x-2\pi m)^2]\,. \tag{2.3}$$

The function $\exp(f_\beta)$ has the Fourier representation

$$\exp[f_\beta(x)]=[1/(2\pi\beta)^{1/2}]\sum_{l=-\infty}^{\infty}\exp(-l^2/2\beta)\times\exp(ilx)\,. \tag{2.4}$$

Note that for very large β (small temperature) the models (2.1) and (2.2) become very close to each other. In the Villain model (2.2), the decomposition of the configurations to spin waves and vortices becomes neater and more explicit; in fact, JKKN got for this model the same picture proposed by Kosterlitz and Thouless for the system (2.1). We have again a high-temperature massive phase and a low-temperature soft phase with a peculiar form of singularity in the transition. The model (2.2) has also a simple dual form. By substituting the representation (2.4) into (2.2) and performing the θ integrations one gets

$$Z_{XY}^V=\sum_{l_\mu(i)=-\infty}^{\infty}\exp\Big[-\frac{1}{2\beta}\sum_{i,\mu}l_\mu{}^2(i)\Big]\prod_i\delta_{\nabla_\mu l_\mu(i),0}\,. \tag{2.5}$$

In (2.5) the sum is over all the integer-valued functions of the links of the lattice, and the symbol ∇_μ stands for partial finite difference in the μ direction. Solving the constraint $\nabla_\mu l_\mu=0$ by $l_\mu(i)=\epsilon_{\mu\nu}\nabla_\mu l(i)$, with $l(i)$ some integer-valued function of the lattice sites, we get

$$Z_{XY}^V=\sum_{l(i)}\exp\Big[-(1/2T')\sum_{i,\mu}(\nabla_\mu l)^2\Big]\,, \tag{2.6}$$

with $T'=\beta=1/T$. So, the dual form of (2.2) looks like a massless free-field theory with the field constrained to take only integer values. It is also known as the surface-roughening model. Since (2.6) describes the same model as (2.2), we know that it also possesses two phases. The phase of small T', which is the massive high-T phase of (2.2), is an ordered phase for the l variable of (2.6). We expect a vacuum expectation value for l, and therefore a correlation function such as $\langle\cos\alpha[l(0)-l(x)]\rangle$, with α some number, will tend to a constant nonzero value as the point x is sent to infinity. At $T'=1/T_k$ there is a transition to the high-T' phase, which is the massless lower-T phase of (2.2). Being massless and extending up to infinite T', this phase seems to have no vacuum expectation value of l. In terms of this variable this phase is disordered so that the correlation function $\langle\cos\alpha[l(0)-l(x)]\rangle$ for $0<\alpha<2\pi$ goes to zero when $|x|\to\infty$. Since there is no mass gap (i.e., finite correlation length) in this phase, that decay is expected to be like some α- and T'-dependent power of $|x|$. In fact, it can be shown by some Griffith's inequality, of the type used in the Appendix, that the model (2.6), with only integer values allowed for the field, is always more ordered than the corresponding free system at the same temperature for which the field can take any real value. Since for a free field the correlation function $\langle\cos\alpha[\phi(0)-\phi(x)]\rangle$ decays like a power of $|x|$,[3] the decay of this function for the integer-valued system (2.6) cannot be faster than a power behavior. In the high-T' phase of (2.6) the integer-value character of the field is, then, washed out by thermal fluctuation representing a qualitative free-field behavior.

So much for the properties of the continuous XY systems. The discrete Z_p-invariant model analogous to (2.1) is

$$Z_p(\beta)=\sum_{n(i)=0}^{p-1}\exp\Big\{\beta\sum_{i,\mu}\cos\frac{2\pi}{p}[n(i)-n(i+e_\mu)]\Big\}\,. \tag{2.7}$$

For $p=2$ (2.7) is the Ising model, which has two phases separated by a transition temperature T_2. For $T<T_2$, there is a completely ordered phase with finite magnetization and finite mass, unlike the lower phase of (2.1) which is massless and has no symmetry breaking. We ask ourselves, then: How can we pass from the Z_2 behavior to the Z_{XY}

behavior by changing p from 2 to infinity? One might suggest that any Z_p system shows an Ising-type behavior with a transition point T_p, such that when $p\to\infty$, $T_p\to T_k$ and the finite mass and magnetization of the low-temperature phase tend to zero. To see that this is not the case consider the discrete analog of the system (2.2),

$$Z_p^V(\beta)=\sum_{n(i)=0}^{p-1}\prod_{i,\mu}\exp\left[f_\beta\left(\frac{2\pi}{p}[n(i)-n(i+e_\mu)]\right)\right]. \qquad (2.8)$$

Substituting the representation (2.4) and summing over the $n(i)$ in (2.8), we get

$$Z_p^V(\beta)=\sum_{l_\mu(i)=-\infty}^{\infty}\prod_i\delta_{0,\nabla_\mu l_\mu(i)(\mathrm{mod}\ p)}\times\exp\left[-\tfrac{1}{2}\beta\sum_{i,\mu}l_\mu{}^2(i)\right]. \qquad (2.9)$$

The constraint $\nabla_\mu l_\mu=0$ (mod p) can be solved by

$$l_\mu(i)=\epsilon_{\mu\nu}\nabla_\nu l(i)(\mathrm{mod}\ p)\,, \qquad (2.10)$$

where $l(i)$ is some integer-valued function of the sites. Clearly, in (2.10), l_μ determines l (apart from a global constant integer) only up to multiples of p. So, for any configuration of $l_\mu(i)$ we can choose the field $l(i)$ of (2.10) such that at any site $0\le l(i)\le p-1$. Then the field $l(i)$ is fixed up to a global transformation and there exists an integer field on the links $m_\mu(i)$ such that

$$l_\mu(i)=\epsilon_{\mu\nu}[\nabla_\nu l(i)+pm_\nu(i)]\,. \qquad (2.11)$$

Substituting in (2.9), we have

$$Z_p^V(\beta)=\sum_{l(i)=0}^{p-1}\prod_{i,\mu}\sum_m\exp[-\tfrac{1}{2}\beta(\nabla_\mu l+pm)^2]\,. \qquad (2.12)$$

By (2.3), this can be expressed as

$$Z_p^V(\beta)=\sum_{l(i)=0}^{p-1}\prod_{i,\mu}\exp\left[f_{\beta'}\left(\frac{2\pi}{p}\nabla_\mu l\right)\right]=Z_p^V(\beta')\,, \qquad (2.13)$$

with

$$\beta'=p^2/4\pi^2\beta \quad\text{or}\quad T'=4\pi^2/Tp^2\,. \qquad (2.14)$$

The model (2.8) is therefore self-dual. Assuming only two Ising-type phases for any finite p implies that the transition occurs at the self-dual point $T_p=2\pi/p$. The series T_p tends, then, to 0 rather than to the Kosterlitz point T_k. If we still want to stick to the two-phase picture, we have to assume that for any finite p and $T>2\pi/p$ the system is in its high-temperature disordered phase. When $p\to\infty$, it turns out somehow that for $T<T_k$ the mass gap (i.e., inverse correlation length) tends to zero, while for $T>T_k$ it tends to a finite value. According to this view the Kosterlitz transition at T_k is created only when p is taken to infinity, separating two different limits of the same disordered phase of the discrete model.

However, this picture is also wrong. If, for $T<T_k$, the system Z_p is in its disordered phase with mass μ_p, the correlation function at that temperature falls like $\exp(-\mu_p|x|)$. If μ_p decreases to zero when p is increased it means that the system becomes more and more ordered with increasing p. It is hard to see how we can improve the ordering and the range of correlation by increasing the number of allowed configurations and, thus, the entropy. In fact, some sort of a Griffith's inequality[17] for the Villain model is proved in an appendix, implying that

$$\langle\cos[\theta(0)-\theta(x)]\rangle_p^V\ge\langle\cos[\theta(0)-\theta(x)]\rangle_{XY}^V\,, \qquad (2.15)$$

where the left-hand side of (2.15) is the correlation function of (2.8) and the right-hand side is that of (2.2) at the same β.

Thus, we are forced to discard the two-phase picture and to conclude that for p larger than some critical p_c a third phase has to appear between the two Ising-type phases. The preceding analysis also puts an upper bound on p_c. Once the self-dual point $T_p=2\pi/p$ gets below T_k the third phase must occur. We have then three phases for any p obeying

$$p>2\pi/T_k\,. \qquad (2.16)$$

For the Villain model (2.2), T_k is estimated in Ref. 3 to be

$$T_k=1.35\,, \qquad (2.17)$$

which gives three phases at least for $p>4$. If we assume only three phases, the two transition points $T_1<T_2$ separating the middle phase are, of course, mutually dual.

$$T_1=4\pi^2/p^2T_2\,. \qquad (2.18)$$

By (2.15) we have to have

$$T_2\ge T_k\,, \qquad (2.19)$$

so that

$$T_1\le 4\pi^2/p^2T_k\,. \qquad (2.20)$$

The correlation inequality (2.15) implies that the middle phase cannot be disordered and massive. It can be disordered and massless, but it can also be massive and ordered with the left-hand side of (2.15) tending to a constant at infinity. We can eliminate this latter possibility by considering the correlation of the dual variables, i.e., the integer-value field $l(i)$ of (2.6) and (2.12). The inequality (2.15) states that the angle variable θ is more ordered in the discrete model than in the continuous XY model. It is natural to expect the reversed situation for the dual variable. In fact, it is proved in the Appendix that

$$\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_{XY}^{V} \geq \langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_{p}^{V}, \qquad (2.21)$$

where the left-hand side stands for the l correlation function in (2.6) and the right-hand side for that of (2.12) at the same temperature $\beta = T'$. If, now, in the middle phase the θ variable is ordered, then, since this phase is self-dual, the l variable also has to be ordered and the right-hand side of (2.21) goes to a constant. But, for $T < T_k$, $T' > 1/T_k$ and the model (2.6) is in its disordered phase. The left-hand side of (2.21) tends, then, to zero, which contradicts the inequality. So, the middle phase must be disordered in both variables and, therefore, by (2.15), massless. We see, then, that the assumption of the existence of a low-temperature soft phase in the XY model implies the existence of such a phase in discrete models.

In fact, this is understandable if we recall two lessons that can be drawn from the existence of this massless phase in the continuous model. Looking at the form (2.2) which is periodic in the field θ, we can draw lesson (A): For low temperatures $T < T_k$ we can ignore the periodicity of the field and replace the periodic form $\exp[f_\beta(\Delta_\mu\theta)]$ by the free-field form $\exp[-(\beta/2)(\Delta_\mu\theta)^2]$ up to some finite renormalization of the temperature. Looking at the dual form (2.6), in which the field is discrete, we can deduce lesson (B): For high temperatures $T' > 1/T_k$ the discreteness of the field can be ignored and, again, it can be treated as a free continuous field. The Z_p model (2.8) has both properties. The field $n(i)$ in (2.8) is periodic and discrete. By lesson (A), for $T < T_k$ we can ignore periodicity and replace $\exp[f_\beta((2\pi/p)\Delta n)]$ by $\exp[-(4\pi^2/2p^2T)\Delta n^2]$. This is a discrete free field like that of (2.6) with effective temperature $T' = p^2T/4\pi^2$. By lesson (B), if $T' > 1/T_k$ we can also ignore the discreteness and get a free-field-like phase. To have such a phase T has to obey two restrictions. To ignore periodicity we need $T < T_k$. To ignore discreteness we need $Tp^2/4\pi^2 > 1/T_k$. These two inequalities are compatible for $p > 2\pi/T_k$, a condition identical to (2.22). In Sec. IV we shall find these two abstract properties of periodicity and discreteness materialized as two kinds of charged particles.

Up to now we considered only the Villain model (2.2) and its discrete analog (2.8). We expect the qualitative phase structure of the cosine model (2.1) and (2.7) not to differ very much. In particular, by (2.20), for p large enough, the lower transition point to the massless phase of the Villain model can be pushed arbitrarily close to zero where (2.7) and (2.8) are arbitrarily close to each other. So, at least for high enough p, the cosine model (2.7) must posses also a massless phase.

III. GAUGE SYSTEMS

The behavior of the discrete spin model was derived from the assumption of the existence of a massless low-temperature phase in the continuous model using correlation inequalities to compare the two cases. Since the existence of such a phase is expected also for four-dimensional continuous Abelian gauge theories,[5,6,8] we can get similar results for discrete gauge systems. The U(1) gauge model has the partition function

$$Z(\beta) = \int_{-\pi}^{\pi}\prod_{i,\mu} d\theta_\mu(i)\exp\Big[\beta\sum_{i,\mu>\nu}\cos\theta_{\mu\nu}(i)\Big], \qquad (3.1)$$

with $\theta_{\mu\nu}(i) = \theta_\mu(i) + \theta_\nu(i+e_\mu) - \theta_\mu(i+e_\nu) - \theta_\nu(i)$. In (3.1) $\theta_\mu(i)$ is defined on the link connecting the site i to the site $i+e_\mu$ in a four-dimensional cubic lattice, and $\theta_{\mu\nu}(i)$ is the usual plaquette variable.[5] For large $T = 1/\beta$ the system is confined and massive; for small temperatures $T < T_g$ one expects a massless QED-like behavior.

Again, the cosine form of the interaction in (3.1) can be replaced by the Villain form

$$Z^V(\beta) = \int_{-\pi}^{\pi}\prod_{i,\mu} d\theta_\mu(i)\prod_{i,\mu>\nu}\exp[f_\beta(\theta_{\mu\nu}(i))], \qquad (3.2a)$$

with the f_β defined in (2.3). This model was studied by Banks, Meyerson, and Kogut (BMK).[7] The qualitative picture of the two phases described above is expected to hold also for (3.2a). In particular, for large β it becomes identical to (3.1). As in the previous section, (3.2a) has a simple dual form which is a free noncompact gauge theory with the gauge field constrained to take on integer value

$$Z^V(\beta) = \sum_{l(i)}\exp\Big[\frac{-1}{2\beta}\sum_{i,\mu>\nu}(\nabla_\mu l_\nu - \nabla_\nu l_\mu)^2\Big]. \qquad (3.2b)$$

[Because of the noncompactness of the gauge group which is Z, (3.2b) has actually no meaning unless some gauge-fixing method is adopted.] The low-β phase of (3.2b), which is the high-temperature phase of (3.2a), is massive and screened due to the discreteness of the field. As β, the temperature of (3.2b) exceeds $1/T_g$, the discreteness is washed out by thermal fluctuations and we find the massless QED-like behavior.

The Z_p analog of (3.1) is

$$Z_p(\beta) = \sum_{s_\mu(i)=0}^{p-1}\exp\Big[\beta\sum_{i,\mu>\nu}\cos\frac{2\pi}{p}(\nabla_\mu s_\nu - \nabla_\nu s_\mu)\Big]. \qquad (3.3)$$

The Z_2 model has been studied in several works.[18,9,6] The high-temperature phase is confined and massive like that of the U(1) model. The low-temperature phase is also massive, being dual to the high-temperature phase. Thus, this is some kind of a Higgs phase in which the long-range interaction of charges is screened. We can ask, as we did in the previous section: How do we get the QED phase of the U(1) model as a limit of the Z_p model when $p\to\infty$? Here, again, by passing to the Villain model we shall be able to exclude the possibility of the Higgs phase of the Z_p model turning gradually to the QED phase of U(1) with the mass $\mu_p\to 0$ and the transition temperature $T_p\to T_g$. The Villain Z_p model is

$$Z_p^V(\beta)=\sum_{s_\mu(i)=0}^{p-1}\prod_{i,\mu>\nu}\exp\left[f_\beta\left(\frac{2\pi}{p}(\nabla_\mu s_\nu-\nabla_\nu s_\mu)\right)\right]. \quad (3.4)$$

Substituting the representation (2.4) and summing over $n_\mu(i)$, we have

$$Z_p^V(\beta)=\sum_{l_{\mu\nu}(i)=-\infty}^{\infty}\prod(\delta_{0,\nabla_\mu l_{\mu\nu}(\mathrm{mod}\ p)})\times\exp\left[-\frac{1}{2\beta}\sum_{i,\mu>\nu}l_{\mu\nu}{}^2(i)\right], \quad (3.5)$$

where, in (3.5), $l_{\nu\mu}=-l_{\mu\nu}$. The constraint $\nabla_\mu l_{\mu\nu}=0\ (\mathrm{mod}\ p)$ is solved by

$$l_{\mu\nu}=\epsilon_{\mu\nu\alpha\beta}\nabla_\alpha l_\beta(\mathrm{mod}\ p)\,. \quad (3.6)$$

Again, adding any multiple of p to l_β does not affect $l_{\mu\nu}$ in (3.6), so we can choose $0\le l_\beta\le p-1$ at each dual link. Even that fixes l_β only up to a gauge transformation

$$l'_\alpha=l_\alpha+\nabla_\alpha l(\mathrm{mod}\ p)\,, \quad (3.7)$$

with $l(i)$ some arbitrary integer-site function. The number of the gauge transformations (3.7) is p^N, with N the total number of lattice sites. So, by summing in (3.5) over l_α instead of $l_{\mu\nu}$, one multiplies the partition function by the irrelevant constant factor p^N. Up to such a factor we have

$$Z_p^V(\beta)=\sum_{l_\alpha(i)=0}^{p-1}\prod_{i,\mu>\nu}\sum_m\exp\left[-\frac{1}{2\beta}(\nabla_\mu l_\nu-\nabla_\mu l_\nu+pm)^2\right], \quad (3.8)$$

which is by (2.3)

$$Z_p^V(\beta)=\sum_{l_\alpha(i)=0}^{p-1}\prod_{i,\mu>\nu}\exp\left[f_{\beta'}\left(\frac{2\pi}{p}(\nabla_\mu l_\nu-\nabla_\nu l_\mu)\right)\right]=Z_p^V(\beta')\,, \quad (3.9)$$

where $\beta'=p^2/4\pi^2\beta$ or $T'=4\pi^2/Tp^2$.

We see that the discrete Villain gauge model is self-dual, obeying the same duality relation as the two-dimensional spin model. This fact was recognized in Ref. 14. So, if there are only two phases in the Z_p model, the transition point T_p is equal to $2\pi/p$ and thus tends to zero rather than to T_g. Again, by the same reasoning as in the preceding section, we can exclude the possibility of a single confined phase extending from $T=2\pi/p$ up to $T\to\infty$. If this is the case we have to assume that the average of the Wilson loop for $2\pi/p<T<T_g$, which has to decay like the area of the loop, tends towards a perimeter behavior when $p\to\infty$. This strange possibility that increasing p will improve the ordering of the system can be excluded by a correlation inequality analogous to (2.15). Choose some oriented closed loop in the lattice and define $J_\mu(i)$ by

$$J^\mu(i)=\begin{cases}1 \text{ if the loop contains the link } i\to i+e_\mu\\ -1 \text{ if the loop contains the link } i+e_\mu\to i\\ 0 \text{ otherwise.}\end{cases}$$

It is shown in the Appendix that

$$\langle\cos\textstyle\sum J_\mu\theta_\mu\rangle_p^V\ge\langle\cos\sum J_\mu\theta_\mu\rangle_{U(1)}^V. \quad (3.10)$$

The left-hand side of (3.10) is the average Wilson loop of the model (3.4), and the right-hand side is the same average for the model (3.2a) at the same temperature. By (3.10), if for any $T<T_g$ the U(1) system (3.2a) obeys a perimeter law, so must the Z_p system (3.4). The transition from the high-temperature area-law behavior to the perimeter behavior in the Z_p model must, then, occur at some $T_2\ge T_g$. If $p>2\pi/T_g$, then T_2 cannot be identical to the self-dual point $2\pi/p$. By duality there has to exist another transition at a point T_1 dual to T_2, namely,

$$T_1=4\pi^2/p^2T_2\,. \quad (3.11)$$

For any $p>2\pi/T_g$ we must, then, have a middle phase. Unfortunately, there is no good estimate for T_g. The rough estimate of Ref. 7, $T_g\approx 1.9$, suggests three phases at least for any $p>3$. We have seen that in the middle phase the Wilson loop, i.e., the average of $\cos[(2\pi/p)\sum J_\mu(i)s_\mu(i)]$ with the J_μ of (3.10) and s_μ of (3.4), has to decay like the exponential of the perimeter. Since this phase is self-dual, the variable dual to s_μ, i.e., the field l_μ of (3.8), has also to obey a perimeter law. 't Hooft has argued[10] that such a situation, in which two mutually dual variables obey the perimeter decay law, is possible only if there is no mass gap. According to this argument, the middle phase must be a massless QED-like phase. The same discussion at the end of the preceding section about the role of periodicity and discreteness of the fields applies also here. The description of these properties as various charged currents will be briefly discussed in the next section.

The cosine model (3.3) is expected not to differ very much qualitatively from (3.4). In particular, as in the spin case, by choosing p large enough we can push the lower transition point T_1 and, hence, the massless phase arbitrarily close to zero. At such small temperatures (3.3) and (3.4) are identical, so the model (3.3) must have three phases at least for large p.

IV. THE COULOMB-GAS REPRESENTATION OF THE Z_p VILLAIN MODELS

As Kadanoff has shown,[14] many two-dimensional problems in statistical physics can be rewritten as Coulomb-gas problems. One then typically finds a system with two kinds of interacting "charges" represented by integer-valued variables $N(\vec{r})$ and $M(\vec{R})$. In this section we consider the Coulomb-gas picture of the Z_p spin theories. This representation turns out to be useful for discussing the phase diagrams of these models. In particular, the disordered massless phase, which we have argued must exist for $p > p_c$, will reveal itself in a straightforward way. Rewriting the spin-spin correlation function

$$F_p(\vec{\rho}) \equiv \langle \exp\{i(2\pi/p)[n(0) - n(\vec{\rho})]\}\rangle$$

in the Coulomb-gas language, one finds three qualitatively distinct types of behavior as the temperature is varied. For $p^2T \ll 1$, one has

$$F_p(\vec{\rho})_{\overrightarrow{|\vec{\rho}|\to\infty}} \text{const};$$

for $T \gg 1$, one has $F_p(\vec{\rho}) \to 0$ exponentially; and, if p is sufficiently large, one finds a third middle phase in which $F_p(\vec{\rho})$ falls off as a power. A Kosterlitz renormalization-group analysis indicates that this third phase is stable for $p > 4$.

We start from the partition function of Eq. (2.8),

$$Z = \left(\prod_{\vec{r}} \sum_{n(\vec{r})=0}^{p-1}\right)\left(\prod_{r,\mu} \sum_{m_\mu(\vec{r})=-\infty}^{\infty}\right) \exp\left\{-\tfrac{1}{2}\beta \sum_{\vec{r},\mu}\left[\frac{2\pi}{p}\nabla_\mu n(\vec{r}) - 2\pi m_\mu(\vec{r})\right]^2\right\}. \tag{4.1}$$

The integer-valued variables $n(\vec{r})$ can be replaced by continuous variables $p/2\pi\theta_\mu(\vec{r})$ using

$$\frac{2\pi}{p}\sum_{n=0}^{p-1}\delta(\theta - 2\pi n/p) = \sum_{N=-\infty}^{\infty}\exp[ipN\theta] \quad (0 \leq \theta < 2\pi), \tag{4.2}$$

$$Z = \left(\prod_{\vec{r}} \sum_{N(\vec{r})=-\infty}^{\infty}\right)\left(\prod_{\vec{r},\mu} \sum_{m_\mu=-\infty}^{\infty}\right)\left[\prod_{\vec{r}} \int_0^{2\pi}\frac{d\theta(\vec{r})}{2\pi/p}\right]\exp\left[ip\sum_{\vec{r}}N(\vec{r})\theta(\vec{r})\right]\exp\left\{-\frac{\beta}{2}\sum_{\vec{r},\mu}[\nabla_\mu\theta(\vec{r}) - 2\pi m_\mu(\vec{r})]^2\right\}. \tag{4.3}$$

Half the summations $\sum_{m_\mu=-\infty}^{\infty}$ over the link variables m_μ [e.g., all the variables $m_x(r)$ in the x direction] can be used to extend the range of the $\theta(\vec{r})$ integrations. The remaining $m_\mu(\vec{r})$'s can then be expressed in terms of integers $M(\vec{R})$ defined on dual lattice sites:

$$M(\vec{R}) = \epsilon_{\mu\nu}\nabla_\nu m_\mu(\vec{r}) = \text{curl around the square containing the dual site } \vec{R}$$

or (4.4)

$$m_\mu(\vec{r}) = \epsilon_{\mu\nu}n_\nu(\vec{n}\cdot\nabla)_{\vec{r}}^{-1}M(\vec{R}).$$

In (4.4) we have introduced

$\vec{n}$ = unit vector,

$(\vec{n}\cdot\nabla)_{\vec{r}}^{-1}$ = discrete line integral in the $\vec{n}$ direction

[e.g., for $\vec{n} = \hat{x}$, $(\vec{n}\cdot\nabla)_{\vec{r}}^{-1}f(\vec{R}) = \sum_{R_x < r_x + 1/2} f(R_x, R_y = r_y + \frac{1}{2})$]. We now have

$$Z = \left(\prod_{\vec{R}} \sum_{M(\vec{R})=-\infty}^{\infty}\right)\left(\prod_{\vec{r}} \sum_{N(\vec{r})=-\infty}^{\infty} \int_{-\infty}^{\infty}\frac{d\theta(\vec{r})}{2\pi/p}\right)$$
$$\times \exp\left[ip\sum_{\vec{r}}N(\vec{r})\theta(\vec{r})\right]\exp\left\{-\frac{\beta}{2}\sum_{\vec{r},\mu}[\nabla_\mu\theta(\vec{r}) - 2\pi\epsilon_{\mu\nu}n_\nu(\vec{n}\cdot\nabla)_{\vec{r}}^{-1}M(\vec{R})]^2\right\}. \tag{4.5}$$

The $\theta(r)$ integrations can be performed yielding (up to constant factors)

$$Z = \left(\prod_{\vec{R}} \sum_{M(\vec{R})=-\infty}^{\infty}\right)\left(\prod_{\vec{r}} \sum_{N(\vec{r})=-\infty}^{\infty}\right)\delta\left(\sum_{\vec{r}}N(\vec{r})\right)\delta\left(\sum_{\vec{R}}M(\vec{R})\right)\exp\left[-\sum_{\vec{r}\neq\vec{r}'}\frac{p^2}{4\pi\beta}N(\vec{r})G'(\vec{r}-\vec{r}')N(\vec{r}')\right]$$
$$\times\exp\left[-\sum_{\vec{R}\neq\vec{R}'}\frac{(2\pi)\beta}{2}M(\vec{R})G'(\vec{R}-\vec{R}')M(\vec{R}')\right]\exp\left[ip\sum_{\vec{r},\vec{r}'}N(\vec{r})G(\vec{r}-\vec{r}')\nabla'_\mu\epsilon_{\mu\nu}n_\nu(n\cdot\nabla)_{\vec{r}'}^{-1}M(\vec{R})\right], \tag{4.6}$$

where

$$G'(\vec{r}) = 2\pi[G(\vec{r}) - G(0)] \tag{4.7}$$

and $G(\vec{r})$ is the lattice Green's function satisfying

$$-\Big(\sum_{\pm\hat{e}} \delta_{\vec{r},\hat{e}} - 4\delta_{\vec{r},\vec{0}}\Big)G(\vec{r}) \equiv -\nabla^2 G(\vec{r}) = \delta_{\vec{r},\vec{0}}\,. \tag{4.8}$$

A good approximation for $G'(\vec{r})$, valid for all $|\vec{r}|$, is given by

$$G'(\vec{r}) \sim -\ln|\vec{r}| - \pi/2\,. \tag{4.9}$$

So,

$$Z = \Big(\prod_{\vec{R}} \sum_{M(\vec{R})=-\infty}^{\infty}\Big)\Big(\prod_{\vec{r}} \sum_{N(\vec{r})=-\infty}^{\infty}\Big)\delta\Big(\sum_{\vec{r}} N(\vec{r})\Big)\delta\Big(\sum_{\vec{R}} M(\vec{R})\Big)\exp\Big[-\frac{p^2}{8\beta}\sum_{\vec{r}} N^2(\vec{r})\Big]\exp\Big[-\frac{\pi^2\beta}{2}\sum_{\vec{R}} M^2(\vec{R})\Big]$$
$$\times\exp\Big[\frac{p^2}{4\pi\beta}\sum_{\vec{r}\neq\vec{r}'} N(\vec{r})\ln|\vec{r}-\vec{r}'|N(\vec{r}')\Big]\exp\Big[\pi\beta\sum_{\vec{R}\neq\vec{R}'} M(\vec{R})\ln|\vec{R}-\vec{R}'|M(\vec{R}')\Big]\exp\Big[ip\sum_{\vec{r},\vec{R}} N(\vec{r})\theta(\vec{r}-\vec{R})M(\vec{R})\Big]\,, \tag{4.10}$$

where

$$\theta(\vec{r}-\vec{R}) = \sum_{R'_x=R_x}^{\infty} \nabla_{r_y} G'(r_x - R'_x - \tfrac{1}{2}, r_y - R_y - \tfrac{1}{2})\,.$$

The partition function (4.10) characterizes a neutral system containing two types of "charges." Considering their origin, one might call the N's "discreteness" variables and the M's the "periodicity" or "vortex" variables. The N-N and M-M interactions are Coulombic. The complicated-looking N-M potential $\theta(\vec{r}-\vec{R})$ has been studied in Refs. 3 and 15. It measures the angular position of $\vec{R}$ relative to $\vec{r}$. In the Coulomb-gas picture self-duality follows from the invariance of Eq. (4.10) with respect to the replacements

$$\beta \to p^2/4\pi^2\beta \quad\text{and}\quad N \leftrightarrow M\,. \tag{4.11}$$

Consider next the correlation function

$$F_p(\rho) = \langle \exp\{i(2\pi/p)[n(\vec{0}) - n(\vec{\rho})]\}\rangle$$
$$= Z^{-1}\Big(\prod_{\vec{r}} \sum_{n(r)=0}^{p-1}\Big)\Big(\prod_{r,\mu} \sum_{m_\mu=-\infty}^{\infty}\Big)\exp\Big\{-\frac{\beta}{2}\sum_{\vec{r},\mu}\Big[\frac{2\pi}{p}\nabla_\mu n(\vec{r}) - 2\pi m_\mu(\vec{r})\Big]^2\Big\}\exp\Big\{i\frac{2\pi}{p}[n(\vec{0}) - n(\vec{\rho})]\Big\}\,. \tag{4.12}$$

Going through the same steps as before, one ends up with essentially the same expression as (4.10) for the numerator in (4.12), with the $N(\vec{r})$'s replaced by

$$N(\vec{r}) \to N'(\vec{r}) = \begin{cases} N(\vec{r}), & \vec{r}\neq\vec{0} \text{ or } \vec{\rho} \\ N(\vec{r}) + 1/p, & \vec{r} = \vec{0} \\ N(\vec{r}) - 1/p, & \vec{r} = \vec{\rho}\,. \end{cases} \tag{4.13}$$

So,

$$F_p(\vec{\rho}) = Z^{-1}\Big(\prod_{\vec{r}} \sum_{N(\vec{r})=-\infty}^{\infty}\Big)\Big(\prod_{\vec{R}} \sum_{M(\vec{R})=-\infty}^{\infty}\Big)\delta\Big(\sum_{\vec{r}} N(\vec{r})\Big)\delta\Big(\sum_{\vec{R}} M(\vec{R})\Big)\exp\Big[-\frac{p^2}{8\beta}\sum_{\vec{r}} N'^2(\vec{r})\Big]$$
$$\times\exp\Big[-\frac{\pi^2\beta}{2}\sum_{\vec{R}} M^2(\vec{R})\Big]\exp\Big[\frac{p^2}{4\pi\beta}\sum_{\vec{r}\neq\vec{r}'} N'(\vec{r})\ln|\vec{r}-\vec{r}'|N'(\vec{r}')\Big]$$
$$\times\exp\Big[\pi\beta\sum_{\vec{R}\neq\vec{R}'} M(\vec{R})\ln|\vec{R}-\vec{R}'|M(R')\Big]\exp\Big[ip\sum_{\vec{r},\vec{R}} N'(\vec{r})\theta(\vec{r}-\vec{R})M(\vec{R})\Big]\,, \tag{4.14}$$

with $N'(\vec{r})$ given by (4.13).

For $T \gg 1$ ($\beta \ll 1$) only the terms with all $N(\vec{r}) = 0$ will contribute to Eq. (4.14), and $F_p(\vec{\rho})$ becomes identical to the correlation function for the Villain XY model. It falls off exponentially in this $T \gg 1$ regime. Having set all the discreteness variables $N(r)$ equal to zero, the theory has lost all information that could distinguish it from a continuous XY model.

For $T \ll 1$ one can set all the $M(\vec{R})$'s equal to zero and one obtains

$$F_p(\vec{\rho}) = Z^{-1} \sum_{N(\vec{r})=-\infty}^{\infty} \exp\left[-\sum_{\vec{r}} \frac{p^2}{8\beta} N'(\vec{r})^2\right] \exp\left[\frac{p^2}{4\pi\beta} \sum_{\vec{r}\neq\vec{r}'} N'(\vec{r}) \ln|\vec{r}-\vec{r}'| N'(\vec{r}')\right]. \tag{4.15}$$

By undoing some of the previous transformations, Eq. (4.15) becomes

$$F_p(\vec{\rho}) = Z^{-1}\left[\prod_{\vec{r}} \int_{-\infty}^{\infty} d\theta \sum_{N(\vec{r})=-\infty}^{\infty}\right] \exp\left[ip\sum_{\vec{r}} N(\vec{r})\theta(\vec{r})\right] \exp\{i[\theta(\vec{0})-\theta(\vec{\rho})]\} \exp\left\{-\frac{\beta}{2}\sum_{\vec{r},\mu}[\nabla_\mu \theta(\vec{r})]^2\right\}$$

$$\to Z^{-1}\left(\prod_{\vec{r}} \sum_{s(\vec{r})=-\infty}^{\infty}\right) \exp\left\{-\frac{\beta}{2}\sum_{\vec{r},\mu}\left[\frac{2\pi}{p}\nabla_\mu s(\vec{r})\right]^2\right\} \exp\left\{i\frac{2\pi}{p}[s(\vec{0})-s(\vec{\rho})]\right\}. \tag{4.16}$$

One recognizes again the surface-roughening model [Eq. (2.6)], this time with β the right side up. For $p^2 T \ll 1$, Eq. (4.16) leads to

$$F_p(\rho) \to \text{finite constant}. \tag{4.17}$$

The two phases discussed above exist for all values of p. If there is only one transition point separating them it will be at the self-dual point $\beta = p/2\pi$. As we stressed in the previous section we would then run into difficulties with the correlation inequalities once $p > 4$. We had to argue that there are at least three phases for these $Z(p>4)$ models. Can we see this from Eq. (4.14)?

Consider the situation where

$$\exp(-\pi^2\beta/2) \ll 1 \quad \text{and} \quad \exp(-p^2/8\beta) \ll 1. \tag{4.18}$$

We may then set both the $N(\vec{r})$'s and $M(\vec{R})$'s equal to zero and

$$F_p(\vec{\rho}) \to \exp(-1/4\beta)\exp(-\ln|\vec{\rho}|/2\pi\beta) \propto |\vec{\rho}|^{-1/2\pi\beta}. \tag{4.19}$$

So, if (4.18) holds, $F_p(\vec{\rho})$ shows qualitatively different behavior from the $T \gg 1$ and $p^2 T \ll 1$ cases and we take this as evidence for a third phase. The power-law behavior (4.19) is characteristic of a disordered massless phase. To establish the stability of the middle phase we study the effects of having nonzero $N(\vec{r})$'s and $M(\vec{R})$'s in the dilute-gas approximation [$N(\vec{r}), M(\vec{R}) = 0, \pm 1$]. Let the parameters x, y, and y_p be defined initially by

$$x = \pi\beta, \quad y = 2\pi\exp(-\pi^2\beta/2), \quad y_p = 2\pi\exp(-p^2/8\beta). \tag{4.20}$$

Using the same methods and approximations employed in Ref. 2 to analyze the XY model, one can obtain recursion formulas for these parameters. Upon rescaling the lattice spacing $\tau \to \tau + d\tau$, one finds that the partition function reproduces itself (up to constant factors that contribute to the free energy) provided one makes the replacements

$$\begin{aligned} x \to \tilde{x} &= x - \left(x^2 y^2 - \frac{p^2}{4} y_p^2\right)\frac{d\tau}{\tau}, \\ \frac{p^2}{4x} \to \left(\frac{\tilde{p}^2}{4x}\right) &= \frac{p^2}{4x} - \left[\left(\frac{p^2}{4}\right)^2 \frac{y_p^2}{x^2} - \frac{p^2}{4} y^2\right]\frac{d\tau}{\tau}, \\ y \to \tilde{y} &= y + (2-x) y \frac{d\tau}{\tau}, \\ y_p \to \tilde{y}_p &= y_p + \left(2 - \frac{p^2}{4x}\right) y_p \frac{d\tau}{\tau}, \end{aligned} \tag{4.21}$$

or, equivalently,

$$\begin{aligned} \frac{dx}{dl} &= \frac{p^2}{4} y_p^2 - x^2 y^2, \\ \frac{dy}{dl} &= (2-x) y, \quad (l \equiv \ln\tau) \\ \frac{dy_p}{dl} &= \left(2 - \frac{p^2}{4x}\right) y_p. \end{aligned} \tag{4.22}$$

Equations (4.22) are identical to those obtained by JKKN.[3] This should not come as a surprise, since the question of whether a "spin-wave phase" is stable against perturbations due to symmetry-breaking fields and vortices is just the question we are asking ourselves here. If one finds that for a range of initial x values both y and y_p iterate towards zero, we will have shown that the massless phase of our Z_p models is stable. To avoid any possible confusion, we note that the spin-wave phase in the present treatment enters the theory in a somewhat different manner from in JKKN. There, one was concerned about the effect of symmetry-breaking fields on the spin-wave phase of the planar model and an external term $\sum_{\vec{r}} h_p \cos p\theta(\vec{r})$ was added to the planar model action. In their model a Z_p theory would emerge, strictly speaking, only in the limit $h_p \to \infty$. In the present case we always stay within the Z_p model. The "spin-wave phase" appears as a legitimate phase for these discrete models through Eqs. (4.14), (4.19), and the arguments which led to a power-law falloff of the correlation function.

After making this reinterpretation we can take over many of the results obtained by JKKN. Equation (4.22) shows us that there is a line of fixed points for $y^{\rm eff}=0$, $y_p^{\rm eff}=0$. It is stable for

$$2 \leq x^{\rm eff} \leq p^2/8 . \tag{4.23}$$

Equation (4.23) can only be satisfied if $p \geq 4$. We refer the reader to JKKN for further analysis of Eq. (4.22), in particular, to their intriguing Fig. 1 [recall that we may reinterpret $h_p({\rm JKKN}) \sim y_p$ = small].

One can also argue that the two major results emerging from the original work of Kosterlitz on the XY model,[1,2] i.e., the behavior of the correlation length and the value of the critical exponent η at the critical point, are valid for the $Z(p={\rm large})$ models. Consider linearizing Eq. (4.22) about the point $x=2$. Then, introducing $\xi \equiv x-2$, one obtains

$$\begin{aligned}\frac{d\beta}{dl}&=\frac{p^2}{4}{y_p}^2-4y^2 ,\\ \frac{dy}{dl}&=-\xi\cdot y ,\\ \frac{dy_p}{dl}&=\left(2-\frac{p^2}{8}\right)y_p+\frac{p^2}{16}\xi y_p .\end{aligned} \tag{4.24}$$

Except for the case $p=4$, which has been treated in detail by JKKN[3] and by Kadanoff,[15] we have not been able to solve (4.24) in closed form to obtain the upper critical temperature for the Z_p models. Because of the correlation inequalities we expect $T^c(Z_p)$ to approach T^c_{XY} monotonically from above as p increases. Furthermore, for large p and $x \simeq 2$, y_p will start out much smaller than y and the third equation in (4.24) tells us that y_p decays approximately as $\sim \exp[-(p^2/8-2)l]$. The dependence of ξ and y on $l(=\ln\tau)$ should then be similar to in the XY model. So, although the critical temperature will be shifted, we expect the correlation length in the Z_p models to exhibit a Kosterlitz-type essential singularity as T_c is approached from above (for large p). In the next section we turn to the Hamiltonian version of Z_p models, and one important objective there will be to investigate how the high-temperature mass gap vanishes at the critical point. We will argue that the strong-coupling series favors a Kosterlitz-type behavior rather than an Ising type for $p>4$. Following again closely the analysis of Kosterlitz,[2] one can calculate corrections to the correlation function Eq. (4.19) in the dilute-gas approximation. One finds

$$F(\rho)={\rm F}_{\rm SW}(\rho)F_{N,M}(\rho) , \tag{4.25}$$

where $F_{\rm SW}=c\,|\vec\rho\,|^{1/2\pi\beta}=c/|\vec\rho\,|^{1/2x}$ and $F_{N,M}(\rho)$ obeys the scaling equation

$$\begin{aligned}F_{N,M}(\rho,\tau)&=F_{N,M}(\rho,\tau+d\tau)\exp\left[-\tfrac{1}{2}\left(y^2-\frac{p^2}{4x^2}{y_p}^2\right)\frac{d\tau}{\tau}\ln\left|\frac{\rho}{\tau}\right|\right]\\ &=F_{N,M}(\rho,\tau+d\tau)\exp\left[-\tfrac{1}{2}dx^{-1}\ln\left|\frac{\rho}{\tau}\right|\right].\end{aligned} \tag{4.26}$$

Scaling τ up to order ρ and working to first order in $\xi_i=x_{\rm initial}-2$,

$$\begin{aligned}F(\rho)&\sim 1/|\rho|^{1/4(1+\xi/2)}\exp\left\{\tfrac{1}{8}\int_{\xi_i}^{\xi(\rho)}d\xi'[\ln\rho-\ln\tau(\xi')]\right\}\\ &\sim\exp[-\ln\rho/4(1+\xi_i/2)]\exp[\tfrac{1}{8}\ln\rho(\xi(\rho)-\xi_i)]\exp\left[-\tfrac{1}{8}\int_{\xi_i}^{\xi(\rho)}d\rho'\ln\tau(\rho')\right],\end{aligned}$$

$$F(\rho)\sim\frac{1}{\rho^{1/4}}\exp\left\{\frac{1}{8}\left[\xi(\rho)\ln\rho-\int_{\xi_i}^{\xi(\rho)}d\rho'\ln\tau(\rho')\right]\right\}. \tag{4.27}$$

$\xi(\rho)$ and $\tau(\rho')$ in the exponent of Eq. (4.27) are determined by the solution to the linearized equations (4.24) at the critical point. Just as in the XY model we expect the whole exponential factor to give at most a logarithmic dependence on ρ. So, the exponent η defined as $F(\vec\rho)\sim 1/|\rho|^\eta$ $(\rho\to\infty)$ approaches $\frac{1}{4}$ at the higher critical temperature. A similar analysis about $x=p^2/8$ gives $\eta=4/p^2$ at the lower T_c.

One can also construct the analog of a "Coulomb gas" representation for the four-dimensional Z_p gauge theories. Instead of a neutral system with two types of interacting scalar charges, one finds a partition function describing a gas of "electric" current loops interacting with another gas of "monopole" current loops.

Consider the Wilson loop around the contour c,

$$\exp[-W(c)]=Z^{-1}\left[\prod_{\vec{r},\nu}\sum_{S_\nu(r)=0}^{p-1}\right]\left[\prod_{\vec{r},\mu<\nu}\sum_{m_{\mu\nu}=-\infty}^{\infty}\right]\exp\left\{-\frac{\beta}{2}\sum_{r,\mu<\nu}\left[\frac{2\pi}{p}S_{\mu\nu}-2\pi m_{\mu\nu}\right]^2\right\}\exp\left[i\sum_{r,\nu}\frac{2\pi}{p}S_\nu(r)J_\nu(r)\right], \tag{4.28}$$

where $S_{\mu\nu}=\nabla_\mu S_\nu-\nabla_\nu S_\mu$;

$$J_\nu(r)=\begin{cases}+1 & \text{if the link } (r,r+\hat{\nu}) \text{ is on } c,\\ -1 & \text{if the link } (r+\hat{\nu},r) \text{ is on } c,\\ 0 & \text{otherwise.}\end{cases}$$

Using the notation and techniques developed in Banks *et al.*,[7] one obtains an expression analogous to Eq. (4.6) for the spin theories:

$$\exp[-W(c)]=Z^{-1}\left[\prod_{r,\nu}\sum_{N_\nu(r)=-\infty}^{\infty}\right]\left[\prod_{R,\nu}\sum_{M_\nu(R)=-\infty}^{\infty}\right]\left[\prod_{r}\delta(\nabla_\nu N_\nu)\right]\left[\prod_{R}\delta(\nabla_\nu M_\nu)\right]$$
$$\times\exp\left[-\frac{p^2}{2\beta}\sum_{r,r',\nu}\left(N^\nu(r)+\frac{J^\nu}{p}\right)G(r-r')\left(N^\nu(r')+\frac{J^\nu}{p}\right)\right]$$
$$\times\exp\left[-\tfrac{1}{2}\beta(2\pi)^2\sum_{R,R,\nu}M^\nu(R)G(r-R')M^\nu(R')\right]\exp\left[ip2\pi\sum_{r,r'\nu}\left(N^\nu(r)+\frac{J^\nu}{p}\right)G(r-r')\nabla'_\alpha\epsilon_{\nu\mu\alpha\beta}n^\mu(n\cdot\nabla)^{-1}_{r'}M_\beta(R)\right]. \tag{4.29}$$

$G(r)=$ the four-dimensional Coulomb potential. Since both the N_ν and M_ν currents are conserved, the summations will be over two types of closed current loops.

Equation (4.29) permits us to argue again that once p is sufficiently large, there will be a range of T values for which both the N_ν and M_ν current-loop gases are dilute. This gives rise to the middle phase, discussed in Sec. III, with the characteristics of the lower phase of compact U(1) lattice gauge theory. According to the picture of BMK,[7] the phase transition to the confining phase (or the Higgs phase) occurs when the M_ν loops (the N_ν loops) become arbitrarily large and dense.

V. HAMILTONIAN Z_p MODELS

Instead of an Euclidean approach we may study the Hamiltonian

$$H(\lambda)=-\sum_{i=1}^{N}\left[\cos\left(\frac{2\pi}{p}L_i\right)+\lambda\cos\left(\frac{2\pi}{p}(M_i-M_{i-1})\right)\right], \tag{5.1}$$

where $M_{N+i}\equiv M_i$ and the operators L_i, M_i have as their spectrum Z_p, the integers modulo p, and $[L_i,M_j]=-i(p/2\pi)\delta_{ij}$. In the limit $p\to\infty$ we may write $(2\pi/p)M_i=\theta_i$, which takes on values in the range $(0,2\pi)$, and expand the cosine in the first term to obtain

$$H(\lambda)=-N+(2\pi^2/p^2)\sum_i L_i^2-\lambda\sum_i\cos(\theta_i-\theta_{i-1}), \tag{5.2}$$

which is the Hamiltonian for an O(2) model. We choose as the standard form for the O(2) model Hamiltonian

$$H(x)=\sum_{i=1}^{N}[L_i^2-x\cos(\theta_i-\theta_{i-1})], \tag{5.3}$$

which suggests for all p the definition

$$H(\lambda)=-N+[1-\cos(2\pi/p)]H_p(x);$$
$$x=\lambda[1-\cos(2\pi/p)]^{-1}, \tag{5.4}$$

$$H_p(x)=\sum_{i=1}^{N}\left[\frac{1-\cos\left(\frac{2\pi}{p}L_i\right)}{1-\cos\left(\frac{2\pi}{p}\right)}-x\cos\left(\frac{2\pi}{p}(M_i-M_{i-1})\right)\right]. \tag{5.5}$$

Equation (5.5) is the form of the Hamiltonian for which we will later give explicit results, since in several respects it is the most natural form for comparing different p's. However, $H(\lambda)$ is important since it obeys the duality

$$H(\lambda)=\lambda H(1/\lambda) \tag{5.6}$$

(in the sense that their spectra are the same). The simple relation (5.6) requires some additional boundary conditions, as we shall see from a derivation of it. In a basis in which the L_i are diagonal $H(\lambda)$ may be written

$$H(\lambda)=-\sum_{i=1}^{N}\left[\cos\left(\frac{2\pi}{p}L_i\right)+\tfrac{1}{2}\lambda(R_iR_{i-1}^\dagger+\text{H.c.})\right], \tag{5.7}$$

where

$$R_i^\dagger |l_i\rangle = |l_i+1\rangle \pmod p\,,$$
$$R_i |l_i\rangle = |l_i-1\rangle \pmod p\,. \qquad (5.8)$$

States in Fock space are characterized by a set of integers $(l_1,\ldots,l_N)$, and the space is decomposed into sectors where $l=\sum_{i=1}^N l_i \pmod p$ is a conserved quantum number. An entirely equivalent labeling of states can be made in terms of $(\tilde l_1,\tilde l_2,\ldots,\tilde l_N)$, where $\tilde l_1=l_1$, $\tilde l_2=l_1+l_2,\ldots,\ \tilde l_N=l_1+\cdots+l_N=l$. In terms of the $\tilde l_i$ and $\tilde R_i$, $\tilde R_i^\dagger$, which act on the $\tilde l$'s in the same fashion as (5.8), the Hamiltonian can be written

$$H(\lambda)=-\left[\cos\left(\frac{2\pi}{p}\tilde L_1\right)+\tfrac{1}{2}\lambda(\tilde R_1\tilde R_2\cdots\tilde R_{N-1}+\text{H.c.})\right] -\sum_{i=2}^{N}\left[\cos\left(\frac{2\pi}{p}(\tilde L_i-\tilde L_{i-1})\right)+\lambda\cos\left(\frac{2\pi}{p}\tilde M_{i-1}\right)\right], \qquad (5.9)$$

where

$$\cos\left(\frac{2\pi}{p}\tilde M_i\right)=\tfrac{1}{2}(\tilde R_i+\tilde R_i^\dagger)\,. \qquad (5.10)$$

Aside from the first term at the boundary, the duality (5.6) is immediate. In order to treat the boundary condition correctly we must treat each l separately. For a given l we trivially write

$$\cos\left(\frac{2\pi}{p}\tilde L_1\right)=\cos\left[\frac{2\pi}{p}(\tilde L_1-\tilde L_N+l)\right], \qquad (5.11)$$

after which we observe that the operator $\tilde R_1\tilde R_2\cdots\tilde R_{N-1}$ has essentially the same effect as $\tilde R_N^\dagger$ and that this substitution has no effect on the spectrum of H. The replacement (5.11) can be written $\cos[(2\pi/p)(\tilde L_1-\tilde L_0)]$ if we require $\tilde L_{N+i}=\tilde L_i+l$ as our boundary condition for the $\tilde L$'s, which is the additional boundary condition we mentioned.

The system described by the Hamiltonian (5.1) will have critical behavior if for some value of λ the mass or energy gap between the ground state and first excited state vanishes. The duality (5.6) applies equally well to linear combinations of the eigenvalues H, so that the mass gap satisfies

$$m(\lambda)=\lambda m(1/\lambda)\,, \qquad (5.12)$$

and, if $m(\lambda)=0$, then so must $m(1/\lambda)$. In the case of an ordinary second-order phase transition with a single critical point this must occur at $\lambda=1$. In the case of a Kosterlitz-Thouless transition there will be a region of λ for which $m=0$ with its end points reciprocals of each other. A further result of (5.12) is to suggest a functional form, which is explicitly self-dual, for $m(\lambda)$,

$$m(\lambda)=(1+\lambda)f(u),\quad u=4\lambda/(1+\lambda)^2\,. \qquad (5.13)$$

The variable u conformally maps the positive real line in λ to the line segment [0,1] in u, and has very beneficial effects on the behavior of series expansions, as we shall later see.

The analog of the correlation inequalities for the Euclidean Z_p models would be in the present case

$$m_{p'}\le m_p,\quad p\le p'\,. \qquad (5.14)$$

There is an ambiguity associated with choosing coupling constants at which to make the comparison which, in the absence of a proof of an inequality for the Hamiltonian case, we do not resolve here. Rather, we suggest that $H_p(x)$ defined in (5.5) is a plausible parametrization of the p dependence, certainly valid for large p, and probably a reasonable approximation for small p. (An alternative, to scale by $2\pi^2/p^2$, is nearly the same for $p\gtrsim 4$.) It is most likely that the O(2) model has a Kosterlitz-Thouless phase transition at $x\cong 1.7$, at which point its mass gap vanishes.[16] If we express in terms of x the critical coupling ($\lambda=1$) suggested by a second-order transition for various p we obtain

$$\begin{aligned} p&=2, & x&=0.5,\\ &=3, & &=0.666\,67,\\ &=4, & &=1.0,\\ &=5, & &=1.447\,21,\\ &=6, & &=2.0. \end{aligned} \qquad (5.15)$$

The existence of an inequality of the form (5.14) would rule out a second-order phase transition for $p\ge 6$. In fact, we know that the cases $p=2$ and $p=4$ are trivial with $m_2=(1-2x)$, $m_4=(1-x)$ corresponding to second-order transitions with $\nu=1$. The case $p=5$ appears undecided and could conceivably have a second-order transition in contrast to the Euclidean theories, where $p>4$ are indicated as having Kosterlitz-Thouless transitions. In the next section by means of a Padé analysis of perturbation series for $m_p(x)$ we will attempt to determine the nature of the phase transitions directly.

VI. SERIES ANALYSIS

The mass gap $m_p(x)$ for the Hamiltonian (5.5) may be computed perturbatively as a power series in x. In Table I are presented the coefficients, through order x^9, of the series for several values of p which we have computed using a method based on explicitly constructing the fourth-order perturbed wave function. By studying the coefficients of the series we may try to learn the location and character of the singularity where the mass gap vanishes. We treat separately the cases $p=2,4$, which are simple and correspond to

TABLE I. Strong-coupling expansion coefficients for the "mass gap" $m_p(x)=\sum_{n=1}^{\infty} m_n x^n$.

n	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=12$	$p=\infty$
0	1	1	1	1	1	1	1
1	−2	−1	−1	−1	−1	−1	−1
2	0	$-\frac{3}{4}$	0	0.059 017 0	$\frac{1}{12}$	0.116 025 4	$\frac{1}{8}$
3	0	$\frac{9}{16}$	0	0.039 957 5	5/144	$3.141\,108\times10^{-2}$	$\frac{1}{32}$
4	0	−179/192	0	0.012 643 2	17/864	$1.640\,440\times10^{-2}$	$1.438\,802\times10^{-2}$
5	0	1099/1152	0	$9.795\,97\times10^{-3}$	357/57 600	$5.839\,254\times10^{-3}$	$6.002\,061\times10^{-3}$
6	0	−15 865/13 824	0	$-4.186\,77\times10^{-4}$	$2.270\,785\times10^{-3}$	$6.518\,928\times10^{-4}$	$2.261\,496\times10^{-4}$
7	0	163 717/221 184	0	$5.474\,67\times10^{-3}$	$1.762\,122\times10^{-3}$	$9.528\,411\times10^{-4}$	$6.957\,991\times10^{-4}$
8	0	−4 564 375/ 15 925 498	0	$-2.409\,10\times10^{-3}$	$2.019\,777\times10^{-4}$	$-2.382\,668\times10^{-4}$	$-1.750\,279\times10^{-4}$
9	0	−117 156 563/ 117 964 800	0	$3.222\,22\times10^{-3}$	$5.393\,014\times10^{-4}$	$1.492\,168\times10^{-4}$	$7.031\,156\times10^{-5}$

Ising models, $p=3$ which has a second-order transition, and $p\geq 5$, which have Kosterlitz-Thouless transitions.

A. $p=2,4$

For $p=2$ the mass gap for all x is given by

$$m_2(x) = |1-2|x|| , \tag{6.1}$$

and, for $p=4$,

$$m_4(x) - |1-1|x|| . \tag{6.2}$$

The systems undergo a second-order (anti) ferromagnetic transition at $x=(-)\frac{1}{2}$ or $(-)1$, respectively. In both cases, if we express the mass gap (for $x>0$) in terms of the self-dual variable u introduced in (5.13) we find

$$m(x)/(1+\lambda) = (1-u)^{1/2} , \tag{6.3}$$

since $(1-u)\propto(1-x)^2$ near $x,u=1$. In later examples we will find the mass gap index in terms of singularities in the u plane and if this singularity is at 1 we need to multiply it by 2. In this case, as is well known, $\nu=1$, the Ising result.

B. $p=3$

For $p=3$, the erratic behavior of the series for $m_3(x)$ shows the existence of singularities competing with the expected one at $\lambda=1$. Since the map to the self-dual variable u takes the unit λ disk, in which $m(x)$ should converge, to the cut u plane with a cut from $u=1$ to $u=\infty$, the series for $f(u)$ should be more regular. In Table II we see that this is so. We are now in a position to try to determine the dominant singularity in $f(u)$. Padé approximates to the logarithmic derivative of $f(u)$ give consistent evidence for a pole at $u=1.0001$ with residue 0.422 ± 0.001, from which we conclude that the $p=3$ system has a second-order phase transition which must occur at exactly $u=1$. This allows an improved estimate of the residue from either the Padé approximate to $(u-1)(d/du)\ln f(u)$ evaluated at $u=1$ or from extrapolation of the coefficients to the series for $(d/du)\ln f(u)$ fit to a polynominal in $(1/n)$. Both methods give the residue as 0.420 ± 0.001. From the approximate residue we may try to determine the critical amplitude A, where

$$f(u) \cong A(1-u)^{0.420\pm0.001}(1+\cdots) \tag{6.4}$$

from the sequence of approximations

$$A_n = f_n \Big/ \binom{n-1.420\pm0.001}{n} . \tag{6.5}$$

Again fitting A_n to a polynomial in $1/n$, we obtain $A\cong 0.991\pm0.003$. Thus, we find the approximate analytic form for $m(x)$

TABLE II. Series for $f_3(u)=\sum_{n=0}^{\infty} f_n u^n$.

n	f_n		
0	1	=	1.0
1	$-\frac{5}{12}$	=	−0.416 666
2	$-\frac{1}{8}$	=	−0.125 000
3	$-25/2^7 3$	=	−0.065 104
4	$-10\,385/2^{10}3^5$	=	−0.041 735
5	$-133\,393/2^{11}3^7$	=	−0.029 782
6	$-14\,630\,857/2^{15}3^9$	=	−0.022 684
7	$-1\,117\,897\,763/2^{20}3^{10}$	=	−0.018 054
8	$-396\,739\,114\,315/2^{24}3^{13}$	=	−0.014 832
9	−0.012 480 257 071 6		

$$m(x) \cong 0.991(1-3x/2)^{0.840}(1+3x/2)^{0.160}(1+\cdots), \tag{6.6}$$

valid for x near $\frac{2}{3}$.

There exists an experimental value for the specific heat index $\alpha \cong 0.35 \pm 0.02$ for He_4 films on graphite, which are thought to have Z_3 symmetry.[4] From the scaling relation $\nu = 1 - \alpha/d$, this gives an experimental value of $\nu = 0.825 \pm 0.010$ or, conversely, we predict $\alpha = 0.320 \pm 0.004$. These numbers are in quite reasonable agreement with each other.

C. $p \geq 5$

For $p \geq 5$ the behavior found is quite different from that for $p \leq 4$. As in the case $p=3$, the coefficients for $p=5,6,12$ shown in Table I are fairly erratic, which suggests the same transformation as before, and this does greatly improve the series. If as before we construct Padé approximates to the logarithmic derivatives, we find that there is strong evidence for an important cut on the real $u > 1$ axis and that the residue of the pole closest to 1 is not stable in the Padé table. The Kosterlitz renormalization-group equations give

$$m(x) \sim \exp[-c/(x_c - x)^{1/2}], \tag{6.7}$$

which has a logarithmic derivative

$$d\ln m(x)/dx = -(c/2)(x_c - x)^{-3/2}. \tag{6.8}$$

This still has a cut, but a second logarithmic derivative gives

$$d\ln[d\ln m(x)/dx]/dx = -\tfrac{3}{2}(x - x_c)^{-1}, \tag{6.9}$$

which is a simple pole with a residue determined from the singularity in the exponential. This suggests that one should study the Padé approximates to the second logarithmic derivatives to the series $f(u)$. The location of the pole close to the origin will give the critical point and its residue the exponent for a generalized form of the singularity

$$m(x) \simeq \exp[c/(x_c - x)^{\sigma}],$$
$$(LD)^2 m(x) \simeq -(1+\sigma)/(x - x_c) + \cdots. \tag{6.10}$$

We summarize in Table III the results of this analysis for the cases $p=5,6,12$. For the cases $p=6,12$ the behavior found is roughly consistent with $\sigma = \frac{1}{2}$, while for $p=5$ we find evidence that σ is converging to $\sigma \cong 0.22$. This is not conclusive, but we feel that this is interesting since it opens up the possibility for a wider class of behavior than is allowed by the Kosterlitz renormalization group, which always gives $\sigma = \frac{1}{2}$. While the numerical results are not as precise as those for $p=3$, they are an improvement over the direct analysis of the series in terms of the variable x. In summary, for $p \geq 5$ we find strong evidence for an essential zero in the mass gap, but find a more general behavior than is indicated by the Kosterlitz renormalization group, namely, σ is a function of p.

TABLE III. Analysis of the Padé approximates to the second logarithmic derivatives of $f(u)$. σ is the exponent defined in Eq. (6.10).

p	Padé $[n,m]$	Pole location $[u]$	$[x]$	σ
5	2,3	0.9996	1.3905	0.2358
	3,2	0.9996	1.3905	0.2360
	3,3[a]	0.9783	1.0756	−0.5225
	3,4	0.9985	1.3393	0.2218
	4,3	0.9985	1.3393	0.2222
6	2,3[a]	$1.0178 \pm 0.0182i$	...	...
	3,2	0.9762	1.4654	0.3176
	3,3	0.9794	1.4979	0.3681
	3,4	0.9802	1.5066	0.3841
	4,3	0.9806	1.5110	0.3912
12	2,3	0.6103	1.7267	0.5174
	3,2	0.6103	1.7267	0.5173
	3,3	0.6105	1.7276	0.5200
	3,4	0.6155	1.7504	0.7221
	4,3	0.6126	1.7371	0.5673

[a] Spurious.

Added note. After completing this work we heard about various other groups which had reached similar conclusions. These are J. L. Cardy,[19] A. Guth, A. Ukawa, and P. Windey, D. Horn, M. Weinstein, and S. Yankielowicz,[20] and Al. B. Zamolodchikov.[21]

ACKNOWLEDGMENTS

We are grateful to T. Banks for many helpful conversations and illuminating suggestions. A lot of help was gained by us from the advice and encouragement of J. B. Kogut. Discussing the subject with H. Bergknoff has contributed much to our understanding. We acknowledge B. Simon for teaching us how to use correlation inequalities. Finally, we thank the Institute for Advanced Study for its hospitality. This research was supported by the Department of Energy under Grant No. EY-76-S-02-2220.

APPENDIX: CORRELATION INEQUALITIES

The correlation function in the Z_p Villain model is

$$\langle\cos[\theta(0)-\theta(x)]\rangle_p^V=\frac{1}{Z_p^V}\sum_{n(i)=0}^{p-1}\prod_{i,\mu}\exp\left[f_\beta\left(\frac{2\pi}{p}[n(i)-n(i+e_\mu)]\right)\right]\exp\left[\frac{2\pi i}{p}\sum Q(i)n(i)\right], \tag{A1}$$

where Z_p^V is defined in (2.8) and

$$Q(i)=\delta_{i,0}-\delta_{i,x}. \tag{A2}$$

Substituting the representation (2.4) of e^{f_β} and summing over $n(i)$, one gets a form similar to (2.9),

$$\langle\cos[\theta(0)-\theta(x)]\rangle_p^V=\frac{1}{Z_p^V}\sum_{l_\mu(i)=-\infty}^{\infty}\sum_{m(i)=-\infty}^{\infty}\prod_i\delta_{0,\nabla_\mu l_\mu(i)+pm(i)+Q(i)}\exp\left(-\frac{1}{2\beta}\sum l_\mu^{\,2}\right). \tag{A3}$$

The correlation function of the Villain XY model is

$$\langle\cos[\theta(0)-\theta(x)]\rangle_{XY}^V=\frac{1}{Z_{XY}^V}\int_{-\pi}^{\pi}\prod_i d\theta(i)\prod_{i,\mu}\exp[f_\beta(\theta(i)-\theta(i+e_\mu))]\exp\left[\sum Q(i)\theta(i)\right], \tag{A4}$$

with Z_{XY}^V that of (2.2). Instead of (A3) we have for this case

$$\langle\cos[\theta(0)-\theta(x)]\rangle_{XY}^V=\frac{1}{Z_{XY}^V}\sum_{l_\mu(i)=-\infty}^{\infty}\prod_i\delta_{0,\nabla_\mu l_\mu(i)+Q(i)}\exp\left(-\frac{1}{2\beta}\sum l_\mu^{\,2}\right). \tag{A5}$$

Define an interpolating correlation function $\langle\cos(\theta(0)-\theta(x))\rangle_h$ by

$$\langle\cos[\theta(0)-\theta(x)]\rangle_h=\frac{1}{Z_h}\sum_{l_\mu(i)=-\infty}^{\infty}\sum_{m(i)=-\infty}^{\infty}\left(\prod\delta_{0,\nabla_\mu l_\mu(i)+pm(i)+Q(i)}\right)\exp\left[-\frac{1}{2\beta}\sum_{i,\mu}l_\mu^{\,2}-\frac{1}{h}\sum_i m^2(i)\right], \tag{A6}$$

Z_h in (A6) being the same sum with $Q\equiv 0$. For $h=0$, (A6) is identical to (A5), i.e., the XY correlation. For $h\to\infty$, (A6) tends to (A4) giving the Z_p correlation. To prove (2.15) it is sufficient, then, to show that for any $h>0$ $(\partial/\partial h)\langle\cos[\theta(0)-\theta(x)]\rangle_h\geq 0$. Taking the derivative, we get

$$\begin{aligned}Z_h^{\,2}h^2(\partial/\partial h)\langle\cos[\theta(0)-\theta(x)]\rangle_h=&\sum_{l_\mu,m}\sum_{l'_\mu m'}\left(\prod_i\delta_{0,\nabla_\mu l_\mu+pm+Q}\delta_{0,\nabla_\mu l'_\mu+pm'}\right)\\&\times\exp\left[-\frac{1}{2\beta}\sum(l_\mu^{\,2}(i)+l_\mu'^{\,2}(i))\right]\exp\left[-\frac{1}{h}\sum(m^2+m'^2)\right]\sum_k[m^2(k)-m'^2(k)].\end{aligned} \tag{A7}$$

Define now the new variables

$$\begin{aligned}&\lambda_\mu(i)=l_\mu(i)+l'_\mu(i),\quad \lambda'_\mu(i)=l_\mu(i)-l_\mu(i),\\&\nu(i)=m(i)+m'(i),\quad \nu'(i)=m(i)-m'(i).\end{aligned} \tag{A8}$$

Note that the constraints $\nabla_\mu l_\mu+pm+Q=\nabla_\mu l'_\mu+pm'=0$ imply $\nabla_\mu\lambda_\mu+p\nu+Q=\nabla_\mu\lambda'_\mu+p\nu'+Q=0$. We can now express (A7) as a sum over λ,ν variables instead of l and m. However, the sum over λ_μ and λ'_μ is not independent since, by (A8), $\lambda_\mu(i)$ and $\lambda'_\mu(i)$ ought to have the same parity. The same is true for ν and ν'. We can still make the sum independent by including products of factors of the form $\frac{1}{2}[1+(-1)^{\lambda_\mu+\lambda'_\mu}]$ and $\frac{1}{2}[1+(-1)^{\nu+\nu'}]$ for every link and site, eliminating all the terms for which the parities are not the same. We get then (with N being the number of lattice site)

$$\begin{aligned}2^{3N}Z_h^{\,2}h^2\frac{\partial}{\partial h}\langle\cos[\theta(0)-\theta(x)]\rangle_h=&\sum_{\lambda_\mu\nu}\sum_{\lambda'_\mu\nu'}\left(\prod_i\delta_{0,\nabla_\mu\lambda_\mu+p\nu+Q}\delta_{0,\nabla_\mu\lambda'_\mu+p\nu'+Q}\right)\sum_k\nu(k)\nu'(k)\\&\times\exp\left(-\frac{1}{4\beta}\sum_{i,\mu}\lambda_\mu^{\,2}-\frac{1}{2h}\sum_i\nu^2\right)\exp\left(-\frac{1}{4\beta}\sum\lambda_\mu'^{\,2}-\frac{1}{2h}\sum\nu'^2\right)\\&\times\prod_{i,\mu}[1+(-1)^{\lambda_\mu(i)+\lambda'_\mu(i)}]\prod_i[1+(-1)^{\nu(i)+\nu'(i)}].\end{aligned} \tag{A9}$$

Expanding the products over links and sites in the last factors of (A9) gives

$$2^{3N}Z_h^{\,2}h^2\frac{\partial}{\partial h}\langle\cos[\theta(0)-\theta(x)]\rangle_h=\sum_k\sum_S\sum_L\Bigg\{\sum_{\lambda_\mu\lambda'_\mu}\sum_{\nu\nu'}\Big(\prod\delta_{0,\nabla_\mu\lambda_\mu+p\nu+Q}\Big)\nu(k)$$
$$\times\exp\Big(-\frac{1}{4\beta}\sum\lambda_\mu^{\,2}-\frac{1}{2h}\sum\nu^2\Big)(-1)^{[\Sigma_{(j,\mu)\in L}\lambda_\mu(j)+\Sigma_{j\in S}\nu(j)]}$$
$$\times\Big(\prod_i\delta_{0,\nabla_\mu\lambda'_\mu+\nu+Q}\Big)\nu'(k)\exp\Big(-\frac{1}{4\beta}\sum\lambda_\mu'^{\,2}-\frac{1}{2h}\sum\nu'^2\Big)(-1)^{[\Sigma_{(j,\mu)\in L}\lambda'_\mu(j)+\Sigma_{j\in S}\nu'(j)]}\Bigg\}\,, \tag{A10}$$

where in (A10) S is any subset of the lattice sites and L any subset of the links. Clearly, the expression inside the curly brackets in (A10) is a complete square and thus non-negative. As a sum of non-negative terms we conclude that

$$\frac{\partial}{\partial h}\langle\cos[\theta(0)-\theta(x)]\rangle_h\geq 0\,, \tag{A11}$$

establishing (2.15). If instead of the h derivative of (A6) we take the derivative with respect to β, we get (A7) with $\sum(m^2-m'^2)$ replaced by $\sum(l_\mu^{\,2}-l_\mu'^{\,2})$, which can be shown to be non-negative by just the same argument. This proves the obvious fact that reducing the temperature improves the correlation. Changing in (A6) the meaning of the m variables from a two-dimensional site field to four-dimensional link field, changing also the l field in (A6) to be a four-dimensional plaquette variable, and replacing Q of (A6) by J_μ defined in (3.10), we get the same results for the gauge system (3.2) and (3.4). So, the above analysis proves also (3.10).

All these inequalities have been known for a long time for the cosine model (2.1) and (2.13).[18] However, for the analysis of Secs. II and III we needed their generalization for the self-dual Villain models which were presented above.

To show the opposite inequality for the dual variable, i.e., (2.21), we proceed in a similar way. The correlation function of the dual variable in the Villain XY model is, by (2.6),

$$\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_{XY}^{V}=\frac{1}{Z_{XY}^{V}}\sum_{l(i)}\exp\Big[-\frac{1}{2\beta}\sum_{i,\mu}(\nabla_\mu l)^2\Big]\cos\Big(\frac{2\pi}{p}\sum Q(i)l(i)\Big)\,. \tag{A12}$$

This is the limit $h\to 0$ of the quantity

$$\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_h=\frac{1}{Z_h}\sum_{l(i)}\sum_{m_i(i)}\Big(\prod_{i,\mu}\exp\Big[-\frac{1}{2\beta}(\nabla_\mu l(i)+pm_\mu(i))^2\Big]\Big)$$
$$\times\exp\Big[-\frac{1}{h}\sum m_\mu^{\,2}(i)\Big]\cos\Big(\frac{2\pi}{p}\sum Q(i)l(i)\Big)\,, \tag{A13}$$

where Z_h is the same sum with $Q\equiv 0$. In the limit $h\to\infty$, (A13) becomes identical to the right-hand side of (2.21), i.e., the average of $\cos\{(2\pi/p)[l(0)-l(x)]\}$ in the dual of the Villain Z_p model, Eq. (2.12). Since for $h=\infty$ the summand in (A13) is periodic in l with period p, the fact that the l sum in (A13) is unlimited while in (2.12) it is confined to $0\leq l\leq p-1$ introduces only an infinite multiplicative constant, which cancels between the numerator and the denominator of (A13). Thus, to prove (2.21) it is sufficient to show that $(\partial/\partial h)\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_h\leq 0$ for any $h>0$. Taking the derivative gives

$$Z_h h^2\frac{\partial}{\partial h}\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_h=\sum_{l(i)m_\mu(i)}\sum_{l'm'_\mu}\exp\Big[-\frac{1}{2\beta}\sum_{i,\mu}[(\nabla_\mu l+pm_\mu)^2+(\nabla_\mu l'+pm'_\mu)^2]-\frac{1}{h}\sum_{i,\mu}(m_\mu^{\,2}+m_\mu'^{\,2})\Big]$$
$$\times\Big(\sum_{k,\mu}[m_\mu^{\,2}(k)-m_\mu'(k)]\Big)\cos\Big(\frac{2\pi}{p}\sum Q(i)l(i)\Big)\,. \tag{A14}$$

Equation (A14) can be symmetrized with respect to primed and unprimed variables by replacing the last factor $\cos[(2\pi/p)\sum Ql]$ by $\frac{1}{2}[\cos((2\pi/p)\sum Ql)-\cos((2\pi/p)\sum Ql')]$. Define again the sum and difference variables

$$\lambda(i)=l(i)+l'(i),\quad \lambda'=l-l',$$
$$\nu_\mu=m_\mu+m'_\mu,\qquad \nu'_\mu=m_\mu-m'_\mu. \tag{A15}$$

Using the identity

$$\tfrac{1}{2}(\cos x-\cos y)=-\sin\tfrac{1}{2}(x+y)\sin\tfrac{1}{2}(x-y)\,, \tag{A16}$$

we can express the summand of (A14) in terms of the new variables as

$$-\sum_k\left\{\exp\left[-\sum_{i,\mu}\left(\frac{1}{4\beta}(\nabla_\mu\lambda+p\nu_\mu)^2+\frac{1}{2h}\nu_\mu{}^2\right)\right]\nu_\mu(k)\sin\left(\frac{\pi}{p}\sum Q\lambda\right)\right.$$
$$\left.\times\exp\left[-\sum\left(\frac{1}{4\beta}(\nabla_\mu\lambda'+p\nu'_\mu)^2+\frac{1}{2h}\nu'^2_\mu\right)\right]\nu'_\mu(k)\sin\left(\frac{\pi}{p}\sum Q\lambda'\right)\right\}. \tag{A17}$$

If we could sum (A17) freely on $\lambda\nu$ and $\lambda'\nu'$ we would see immediately that the expression in the curly brackets is a complete square. Actually, we should not sum over them independently because of the constraint that λ and λ' must have the same parity. This constraint, however, can be dealt with exactly in the same way as was done in (A9), expressing the $lm_\mu l'm'_\mu$ sum of (A17) as a sum of non-negative quantities. As a result of the minus sign in the identity (A16), we reach now the opposite result that

$$(\partial/\partial h)\langle\cos\{(2\pi/p)[l(0)-l(x)]\}\rangle_h\leq 0\,, \tag{A18}$$

which proves (2.21).

[1]J. M. Kosterlitz and D. J. Thouless, J. Phys. C 6, 1181 (1973); V. L. Berezinskii, Zh. Eksp. Teor. Fiz. 59, 907 (1970) [Sov. Phys. JETP 32, 493 (1971)]; 61, 1144 (1971) [34, 610 (1972)].
[2]J. M. Kosterlitz, J. Phys. C 7, 1046 (1974).
[3]J. V. José, L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, Phys. Rev. B 16, 1217 (1977).
[4]M. Bretz, Phys. Rev. Lett. 38, 501 (1977).
[5]K. G. Wilson, Phys. Rev. D 10, 2445 (1974).
[6]E. Fradkin and L. Susskind, Phys. Rev. D 17, 2637 (1978).
[7]T. Banks, J. B. Kogut, and R. Meyerson, Nucl. Phys. B129, 493 (1977); R. Savit, Phys. Rev. Lett. 39, 55 (1977); Phys. Rev. B 17, 1340 (1978).
[8]M. Peskin, Ann. Phys. (N.Y.) 113, 122 (1978).
[9]R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D 11, 2098 (1975).
[10]G. 't Hooft, Nucl. Phys. B138, 1 (1978).
[11]N. D. Mermin, J. Math. Phys. 8, 1061 (1967).
[12]S. Coleman, Commun. Math. Phys. 31, 259 (1973).
[13]J. Villain, J. Phys. (Paris) 36, 581 (1975).
[14]A. Casher, DESY report 1978 (unpublished).
[15]L. P. Kadanoff, J. Phys. A 11, 1399 (1978); Brown University report, 1978 (unpublished).
[16]C. J. Hamer, J. B. Kogut, and L. Susskind, Phys. Rev. D 19, 3091 (1979).
[17]J. Ginibre, Commun. Math. Phys. 16, 310 (1970).
[18]F. G. Wegner, J. Math. Phys. 12, 2259 (1971).
[19]J. L. Cardy, UCSB report, 1978 (unpublished).
[20]D. Horn, M. Weinstein, and S. Yankielowicz, following paper, Phys. Rev. D 19, 3715 (1979).
[21]Al. B. Zamolodchikov, Dubna Report No. E 17-11305, 1978 (unpublished).