Journal of Statistical Physics, Vol. 34, Nos. 5/6, 1984

Critical Behavior of Two-Dimensional Spin Models and Charge Asymmetry in the Coulomb Gas

Bernard Nienhuis[1]

Received October 13, 1983

Many two-dimensional spin models can be transformed into Coulomb-gas systems in which charges interact via logarithmic potentials. For some models, such as the eight-vertex model and the Ashkin–Teller model, the Coulomb-gas representation has added significantly to the insight in the phase transitions. For other models, notably the XY model and the clock models, the equivalence has been instrumental for almost our entire understanding of the critical behavior. Recently it was shown that the q-state Potts model and the n-vector model are equivalent to a Coulomb gas with an asymmetry between positive and negative charges. Fieldlike operators in these spin models transform noninteger charges and magnetic monopoles. With the aid of exactly solved models the Coulomb-gas representation allows analytic calculation of some critical indices.

KEY WORDS: Two-dimensional spin models: XY, $Z(p)$, Potts, Ashkin–Teller, $O(n)$, SOS, percolation, SAW, and F model; critical behavior; critical indices; Coulomb gas; spin-wave and vortex operators; Gaussian model.

1. INTRODUCTION AND SUMMARY

This paper reviews some connections recently reported in the literature between spin models and Coulomb-gas systems on two-dimensional lattices. The phrase "Coulomb gas" (CG) here and in the remainder is used for a (lattice) gas of electric and magnetic charges, represented by integer variables e and m. The e–e and m–m interactions are Coulombic, with a strength that in two dimensions varies as the logarithm of the distance for large separations. The e–m interaction is imaginary and proportional to the

[1] Philips Natuurkundig Laboratorium, Postbus 80.000, 5600 JA Eindhoven, Netherlands.

0022-4715/84/0300-0731$03.50/0

angle of the relative position (distance vector) with an arbitrary fixed axis. A more detailed definition of the CG is given in Section 2.

In the course of the last decade a surprising variety of two-dimensional spin models has been reported to be equivalent to a CG. In an earlier article Kadanoff[1] presented a general method by which he explicitly transformed a number of discrete spin models into CG models. He thus established the CG language as a general way to describe many standard problems in two-dimensional statistical mechanics. This language has been instrumental in the calculation of the critical indices of a wide variety of models, the majority of which, however, cannot be transformed into CG systems rigorously. For many systems this method yields a framework which permits the analytic calculation of the asymptotic behavior of a large class of correlation functions. Some of these results were obtained by a direct analysis of the CG system, such as for the XY model,[2] the clock or planar Potts model,[3] and solid-on-solid (SOS) models.[4] A similar analysis of a vector CG has been very useful in providing specific predictions for the physics of melting in two dimensions.[2,5] In other cases the CG equivalence was used to show that a hierarchy of critical exponents can be parametrized by a single variable,[3,6] thus establishing a so-called extended scaling relation for any two such critical indices. Whenever one exponent of the hierarchy is known through an exact solution of a related model, all other indices follow from these extended scaling relations. This procedure has been followed with success for the Ashkin–Teller model,[7–10] the q-state Potts model,[9–13] and the $O(n)$ or n-vector model.[14,15]

This paper reviews how the use of the CG equivalence has led to the analytic calculation of critical exponents. The purpose of this review is to demonstrate a uniformity of method in a number of recent results. It is attempted by discussing a fairly wide variety of models, to give an impression of the state of the art. I hope that, between this objective and the wish to treat the subject to a satisfying depth, I have reached a happy compromise. Derivations that are central to the general discussion are given here. Of some arguments that are logically separate, only the results are given, and for details the reader is referred to some of the original papers.

Many two-dimensional spin models can be represented by CG systems in more than one way. The transformations discussed here are only those that resulted in a CG useful for further analysis of the critical behavior. The discussion is further limited by excluding vector CG models, such as are used for the description of two-dimensional melting,[2,5] and CG models with an external electric or magnetic field, used to discuss commensurate–incommensurate transitions.[16]

The status of the results is fairly summarized as follows. The calculation of critical exponents depends on qualitative assumptions, in the form

of global statements concerning the renormalization group (RG) flows. If these assumptions are valid, the results are exact; if they are violated, the results have no reason to be even approximate. Therefore the consistency between the results and approximate calculations, adds to the plausibility of the assumptions.

This paper is organized as follows: Section 2.1 defines the CG and introduces notation. In Section 2.2 the RG theory of the CG is developed, and Section 2.3 discusses some of its consequences for the phase diagram of the CG. In Section 3 representative models from a variety two-dimensional universality classes are transformed into CG systems. Where possible this is done explicitly; otherwise RG arguments are used to establish the equivalence. In Section 4 some conclusions are given concerning the accomplishments and the weaknesses of the theory.

2. THE COULOMB GAS

2.1. Description of the Model

The Coulomb gas model is defined by means of the action A ($= -H/kT$),

$$A\{e,m\} = \frac{1}{2g}\sum_{j\neq k} e_j G(r_j - r_k) e_k + \frac{g}{2}\sum_{p\neq q} m_p G(r_p - r_q) m_q + i\sum_{j,p} e_j \Phi(r_j - r_p) m_p + \sum_j \ln X(e_j) + \sum_p \ln Y(m_p) \quad (2.1)$$

The integer variables e and m represent electric and magnetic charges. The indices j and k label the sites of a lattice, while p and q represent those of the dual lattice. The square lattice is used throughout, unless stated otherwise explicitly. The vector r denotes the position of the sites. The functions $X(e)$ and $Y(m)$ control by way of fugacities the abundance of sites with a given electric or magnetic charge. The pair potentials G and Φ are long range and behave for large separation $r = (x, y)$ as

$$G(r) + i\Phi(r) \to \ln(x + iy) \quad (2.2)$$

For finite distances G is a spherically symmetric solution of

$$\nabla \cdot \nabla G(r) = 2\pi\delta_{r,0} \quad (2.3)$$

where the gradient symbol ∇ denotes the difference operator between nearest-neighbor lattice sites. It is a vector operator, and like other vector fields, it has its x (y) component on the horizontal (vertical) edges of the

lattice. For example

$$\nabla \cdot \nabla G(r) = G(r+\hat{x}) + G(r-\hat{x}) + G(r+\hat{y}) + G(r-\hat{y}) - 4G(r) \quad (2.4)$$

Here $\hat{x}$ and $\hat{y}$ are the unit lattice vectors in the x and y directions. The potential Φ now follows from the equation

$$\epsilon \cdot \nabla \Phi(r) = \nabla G(r) \quad (2.5)$$

where ϵ is the antisymmetric tensor. Equation (2.5) only has a solution with a cut or a multivalued solution, consistent with Eq. (2.2).

The partition sum Z_{cg} is now given by the summation

$$Z_{cg} = \sum_{\{e\}} \sum_{\{m\}} \exp A\{e,m\} \quad (2.6)$$

subject to the charge neutrality condition

$$\sum_j e_j = 0, \qquad \sum_p m_p = 0 \quad (2.7)$$

In order to avoid untransparant equations, I introduce a more compact notation by suppressing the explicit reference to the sites by indices. Likewise the position symbols r and the summation over the sites will be implied. With these simplifications Eq. (2.1) becomes

$$A\{e,m\} = \left[\frac{1}{2g} eGe + \frac{g}{2} mGm + ie\Phi m + \ln X(e) + \ln Y(m) \right] \quad (2.8)$$

Whenever an expression is to be summed over the lattice it will be contained in square brackets [ ] to avoid ambiguity.

2.2. Renormalization Group Analysis

The CG has a rich structure of critical behavior, which can be analyzed by means of RG theory. The parameters of the theory are the coupling constant g and the infinite hierarchy of fugacities $X(e)$ and $Y(m)$ for $e, m = \ldots, -2, -1, 0, 1, 2, \ldots$. A trivial limit of the CG is the vacuum

$$X(e) = \delta_{e,0}, \qquad Y(m) = \delta_{m,0} \quad (2.9)$$

in which all charges are suppressed, and the partition sum is 1. The RG transformation can be computed in an expansion about the vacuum in powers of the fugacities of nonzero charge. A block of L^2 sites is taken together to form a site in the renormalized lattice. Distances are measured in lattice spacings and therefore rescale by a factor $L = \exp(b)$. The asymptotic interaction expressed in the renormalized distance $r' = r/L$ is

for the electric charges

$$\frac{1}{2g}\sum_{j\neq k} e_j e_k \ln|r_j - r_k| = \frac{1}{2g}\sum_{j\neq k} e_j e_k (\ln|r'_j - r'_k| + b)$$

$$= \frac{1}{2g}\sum_{j\neq k} e_j e_k \ln|r'_j - r'_k| - \frac{b}{2g}\sum_j e_j^2 \qquad (2.10)$$

as follows from the charge neutrality condition. The $m - m$ interaction transforms in the same way, with the coupling constant g replaced by $1/g$. The fugacities, measured per new site, renormalize by a factor L^2 and by the on-site term in Eq. (2.10). The RG equations thus become

$$\frac{d}{db}X(e) = \left(2 - \frac{e^2}{2g}\right)X(e)$$

$$\frac{d}{db}Y(m) = \left(2 - \frac{m^2 g}{2}\right)Y(m) \qquad (2.11)$$

up to leading order in the fugacities. It is not necessary to calculate explicitly the higher-order terms in these recursion relations. However, it is useful to keep in mind that two charges within the confines of L^2 sites produce in the renormalized system a site with the sum of the charges, i.e., the second-order contributions to $dX(e + e')/db$ include a term $X(e)X(e')$. In the vacuum the coupling constant g does not renormalize, as follows from Eq. (2.11). This implies that Eq. (2.9) represents a line of fixed points parametrized by g; surprisingly the vacuum is a critical line. The leading contribution to the renormalization of g is second order in X and Y. Two opposite charges separated by less than the new lattice constant turn into a neutral site in the renormalized system. However, their effect remains as screening of the interaction between distant charges, which effectively changes the coupling constant. The calculation of this effect is tedious, but for the purposes of this paper only the sign is needed. From Ref. 3 I quote

$$\frac{1}{g}\frac{dg}{db} = \sum_{e=1}^{\infty} R\left(\frac{e^2}{g}\right)X(e)X(-e) - \sum_{m=1}^{\infty} R(m^2 g)Y(m)Y(-m) \qquad (2.12)$$

where $R(t) = 2\pi^2 t \exp(-\pi t/2)$.

Equation (2.11) suffices to calculate the critical exponents of the charge correlation functions in the vacuum. However, these correlation functions can also be computed directly from the partition sum. With the fugacities defined on every site independently the correlation functions

between two test charges are simply the derivatives of the free energy,

$$C_e^E(r_j - r_k) = \frac{d^2 \ln Z_{cg}}{dX_j(e)\, dX_k(-e)}$$
$$C_m^M(r_p - r_q) = \frac{d^2 \ln Z_{cg}}{dY_p(m)\, dY_q(-m)} \tag{2.13}$$

where the indices on X and Y refer to the sites on which they are defined. In the vacuum these correlation functions are calculated trivially

$$C_e^E(r) = \exp \frac{-e^2}{g} G(r) \simeq |r|^{-e^2/g}$$
$$C_m^M(r) = \exp - m^2 g G(r) \simeq |r|^{-m^2 g} \tag{2.14}$$

The exponents of these correlation functions agree with the eigenvalues of the RG equations (2.11).

2.3. The Phase Diagram

The recursion relations (2.11) and (2.12) will now be used to explore those sections of the phase diagram that apply to the spin models discussed in Section 3. First consider the case with only magnetic charges:

$$X(e) = \delta_{e,0} \quad \text{and} \quad Y(m) = Y(-m) > 0 \tag{2.15}$$

From Eq. (2.11) it follows that for $g > 4$ $Y(1)$ is relevant and will grow under renormalization. At the same time g diminishes [Eq. (2.12)], thus accelerating the increase of $Y(1)$. Since this flow continues beyond the limit of validity of the RG equations, one cannot from these equations alone infer the nature of this phase. It is, however, reasonable to assume that the abundance of weakly coupled magnetic monopoles will effectively screen their mutual interaction, so that the correlation functions C_m^M approach a nonzero constant at large distances.

For $g > 4$ all $Y(m)$'s are irrelevant, and therefore the system will approach the vacuum under renormalization. The physical meaning of a phase that renormalizes to the vacuum is not a phase without particles, but a phase in which all charges are so strongly bound to neutral complexes that they do not qualitatively affect the macroscopic interaction between two test particles. On a sufficiently large scale these multipoles are invisible but for their screeing effect, which is taken into account by the renormalized coupling constant. The phase boundary is therefore not precisely at $g = 4$ but at the somewhat larger value

$$g_c = 4 + Y(1)\left[8R(4)\right]^{1/2} \tag{2.16}$$

up to linear order in $Y(1)$. In the regime $g > g_c$ the interaction between monopoles is truly logarithmic, and the correlations (2.13) decay algebraically. The exponents vary with g, as in Eq. (2.14), but with the renormalized coupling constant g_R substituted for g. Though $g_R(g)$ is not known generally, at the transition clearly $g_R = 4$. At $g = g_c$, for example,

$$C_1^E(r) \simeq |r|^{-1/4} \quad \text{and} \quad C_1^M(r) \simeq |r|^{-4} \tag{2.17}$$

The singularity in the free energy when g crosses the transition is extremely weak. It can be derived from the RG equations[2,3]

$$f_s \sim \exp(-c|g - g_c|^{-1/2}) \tag{2.18}$$

with a positive constant c.

Without further analysis the behavior of a system with only electric charges can be inferred from the full symmetry of the CG for interchange between electric and magnetic particles and simultaneous inversion of the coupling constant.

Now consider a CG with both electric and magnetic particles, such that all electric charges are multiples of an elementary charge p, i.e., $X(e) = 0$ unless e is an integer multiple of p. Charge inversion symmetry is imposed, $X(e) = X(-e)$ and $Y(m) = Y(-m)$. If $p \leqslant 4$ there is no value of the coupling constant for which both the magnetic and electric fugacities are irrelevant. For such a case RG equations in low-order expansion in X and Y are clearly insufficient to discuss this model. When $p > 4$, however, there is a window in the values of g

$$4 < g < p^2/4 \tag{2.19}$$

in which both the X and Y variables are irrelevant. For small but finite values of $Y(1)$ and $X(p)$, therefore, one expects three phases: (i) $g < g_1 \simeq 4$, in which the magnetic monopoles are free; (ii) $g_1 < g < g_2 \simeq p^2/4$, in which both magnetic and electric particles are bound; and (iii) $g > g_2$, in which the electric charges are free. The precise values of g_1 and g_2 depend on the fugacities and are not universal. The exponents of the correlation functions in the intermediate phase depend on the renormalized coupling constant g_R via Eq. (2.14). At the transition values of g the RG flow is of course towards the boundaries of the window of attraction of the vacuum, Eq. (2.19), hence

$$g_R(g_1) = 4 \quad \text{and} \quad g_R(g_2) = p^2/4 \tag{2.20}$$

Consequently at the transitions the correlation-function exponents are known.

The CG models discussed above are all symmetric under inversion of the charge, i.e., $X(e) = X(-e)$ and $Y(m) = Y(-m)$. If this symmetry is

violated a richer structure of critical behavior emerges. Consider a model with only electric charges, $Y(m) = 0$ for $m \neq 0$. The fugacities $X(1)$ and $X(-1)$ are both zero, and the value of g is taken between 1 and 4, so that only $X(2)$ and $X(-2)$ are relevant. $X(4) = X(-4)$ is kept at a small positive value and the fugacities of higher charges, all irrelevant, need not be specified. Since both $X(2)$ and $X(-2)$ are relevant, only the origin of the $X(2)$-$X(-2)$ plane flows towards the vacuum under RG. This, however, is not the only critical point. When initially both $X(2)$ and $X(-2)$ are positive, they grow under renormalization. From the recursion relation for g (2.12) it follows that g also increases, thus making these fugacities all the more relevant. If, however, the initial action has $X(-2) < 0 < X(2)$, the coupling constant decreases and eventually renders the fugacities irrelevant. As a consequence $X(2)$ and $X(-2)$ begin to diminish in magnitude and the model renormalizes towards the vacuum at a rather smaller value of g_R, in particular $g_R < 1$. This implies that the first and third quadrant of the $X(2)$-$X(-2)$ plane represent quite different phases than the second and fourth quadrant, and that the axes are phase boundaries, as shown in Fig. 1. The singularity of the free energy at $X(2) = 0$ and at $X(-2) = 0$ can be

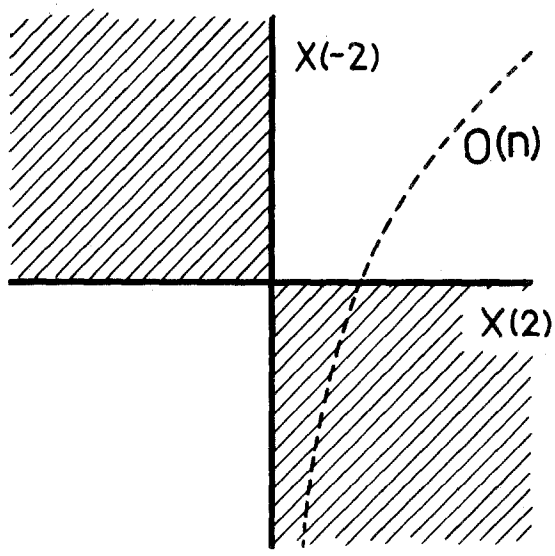

Fig. 1. The phase diagram in the $X(2)$-$X(-2)$ plane. The heavy lines are phase transitions at which the renormalized value of $X(2)$ or $X(-2)$ vanishes. The shaded region initially renormalizes away from the vacuum, but ultimately returns to it. The dashed curve is an impression of the locus of an $O(n)$ model (Section 3.6) in this parameter space, as it is transformed into a CG.

derived from the RG equations. As a consequence of the charge neutrality condition the free energy depends on the fugacities only through charge-neutral combinations, such as

$$\begin{aligned} u &= X(2)X(-2) \\ v &= X(2)^2X(-4) \\ w &= X(-2)^2X(4) \end{aligned} \tag{2.21}$$

Therefore it is useful to derive RG equations for these combinations. From Eq. (2.11):

$$\begin{aligned} \frac{du}{db} &= 4\left(1-\frac{1}{g}\right)u \\ \frac{dv}{db} &= 6\left(1-\frac{2}{g}\right)v \\ \frac{dw}{db} &= 6\left(1-\frac{2}{g}\right)w \end{aligned} \tag{2.22}$$

The free energy is a singular function of $X(2)$ as $X(2)$ vanishes, with a singular part

$$f_s \sim |X(2)|^{2/y_{X(2)}} \tag{2.23}$$

but the power depends on how $X(2)=0$ is approached. If $X(-2)=0$ is fixed while $X(2)$ vanishes, the parameters u and w are zero irrespective of $X(2)$. Therefore the RG equation for v governs the singularity in this case. Since v is quadratic in $X(2)$ its eigenvalue is twice $y_{X(2)}$:

$$y_{X(2)} = 3 - 6/g_R \tag{2.24}$$

If $X(-2)$ is proportional to $X(2)$ as it approaches zero, u, v, and w all vanish quadratically in $X(2)$. Therefore the largest eigenvalue of the RG equations (2.22), which for $g<4$ is that of u, leads to the dominant singularity. Hence in this case

$$y_{X(2)} = 2 - 2/g_R \tag{2.25}$$

Finally when $X(-2)$ is kept at a nonzero constant as $X(2)$ vanishes, w remains finite. Therefore the linearized recursion relations (2.22) yield the critical exponents of this case only when w is irrelevant, that is for $g<2$. Then the singularity is governed by the exponent

$$y_{X(2)} = 4 - 4/g_R \tag{2.26}$$

The difference between the exponents Eqs. (2.24)–(2.26) is caused by the charge neutrality condition. When positive and negative charges have

different fugacities, neutrality is an active constraint, which in this derivation is implemented by observing that the free energy depends on the fugacities only through u, v, and w.

Some caution should be taken with the application of these results, Eqs. (2.24)–(2.26), to actual models. The renormalized coupling constant g_R is generally not known explicitly. But also the phase boundaries in Fig. 1 are located where the renormalized $X(2)$ or $X(-2)$ vanish, rather than their initial values. That these criteria do not coincide is due to the nonlinear terms in the RG equations, such as a term $X(2)X(-4)$ in the renormalization of $X(-2)$.

3. TRANSFORMATIONS OF SPIN MODELS

3.1. The XY Model

In spite of a proof of the absence of long-range order[17] in the two-dimensional XY model, numerical results have been known to indicate a phase transition at finite temperature.[18] Kosterlitz and Thouless[2] for the first time gave a quantitative description of this new kind of phase transition which has since borne their name (KT). Their papers pioneered the use of the CG as a convenient way to describe two-dimensional models. Later Villain[19] modified the XY model so that it could be transformed into a CG rigorously.

The model is generally defined to have on each site i of a lattice a unit vector $(\cos\theta_i, \sin\theta_i)$, interacting via the nearest-neighbor Hamiltonian

$$\frac{-H}{kT} = \sum_{\langle j,k \rangle} V(\theta_j - \theta_k) \tag{3.1}$$

The function $V(\theta)$ is analytic, periodic with period 2π, and monotonic between its maximum and minimum at $\theta = 0$ and π, respectively. The most obvious choice for $V(\theta)$ is $g\cos\theta$, but following Villain I will take the form

$$\exp V(\theta) = \sum_{n=-\infty}^{\infty} \exp \frac{-g}{4\pi}(\theta - 2\pi n)^2 \tag{3.2}$$

With this definition of V the partition sum in the compact notation of Section 2 becomes

$$Z_{xy} = \sum_{\{k\}} \int_0^{2\pi} d\{\theta\} \exp\left[-\frac{g}{4\pi}(\nabla\theta - 2\pi k)^2 \right] \tag{3.3}$$

Here k is an integer vector field residing on the nearest-neighbor links of the lattice. By means of partial summation of the implied sum in the

exponent this partition sum can be written

$$Z_{xy} = \sum_{\{k\}} \int_0^{2\pi} d\{\theta\} \exp\left[\frac{g}{4\pi} \theta \nabla \cdot \nabla \theta - g\theta \nabla \cdot k - \pi g k \cdot k \right] \tag{3.4}$$

The integral over θ extends from 0 to 2π, but since all dependence on θ is periodic, the limits can be changed into $-\infty$ and ∞, with no change but a simple overall, albeit infinite, constant. The remaining Gaussian integral over θ can be performed, making use of Eq. (2.3), with the result

$$Z_{xy} \simeq \sum_{\{k\}} \exp\left[-\frac{g}{2} (\nabla \cdot k) G (\nabla \cdot k) - \pi g k \cdot k \right] \tag{3.5}$$

apart from a multiplicative constant, analytic in g. Rewriting the last term in the exponent gives the expression

$$Z_{xy} \simeq \sum_{\{k\}} \exp - \frac{g}{2} \left[(\nabla \cdot k) G (\nabla \cdot k) + k \cdot (\nabla \cdot \nabla G k) \right] \tag{3.6}$$

which can be simplified to

$$Z_{xy} \simeq \sum_{\{k\}} \exp\left[\frac{g}{2} (\nabla \times k) G (\nabla \times k) \right]$$

and again apart from a multiplicative constant to

$$Z_{xy} \simeq \sum_{\{m\}} \exp\left[\frac{g}{2} m G m \right] \tag{3.7}$$

This completes the transformation of the XY model defined by Eqs. (3.1) and (3.2) to a simple CG with only one kind of charges. The particular form (3.7) is identical to the definition Eq. (2.1) with $Y(m) = 1$, and $X(e) = \delta_{e,0}$.

It would be interesting to see the meaning of the charges m in terms of the original θ variables. Therefore consider a situation in which all m variables are zero except in the origin. A corresponding configuration of k is one where all the k_x, that is, the k variables on the horizontal bonds, vanish and the k_y are 1 on the positive x axis.

$$k_x = 0, \qquad k_y = \delta_{y,0}\{1 + \operatorname{sgn}(x)\}/2 \tag{3.8}$$

The average configuration of θ with this k can be found easily, since for Gaussian integrals the average and most probable configurations are the same. Maximizing the integrand of Eq. (3.3) leads to the equation

$$\nabla \cdot \nabla \theta = 2\pi \nabla \cdot k \tag{3.9}$$

which is solved by

$$\theta_j = \Phi(r_j) + \text{const} \tag{3.10}$$

This configuration, called a vortex, is characterized by the fact that a traveler round the origin finds the spin $s = (\cos\theta, \sin\theta)$ on his way going

round once as well, as he returns to his starting point. A typical example of a vortex is the configuration where all the spins point inward or outward from the origin, $s = -\hat{r}$ or $s = \hat{r}$. An antivortex is the configuration associated with a charge $m = -1$ and has the form $\theta_j = -\Phi(r_j) + \text{const.}$

In Section 2.3 it is shown that the CG equation (3.7) has a phase at large g, where the charges (the vortices) are bound into dipole pairs, and another phase at small g where the vortices are free to move individually. A single phase transition characterized by the weak singularity in the free energy equation (2.8), the KT transition, separates the two phases. It is clear that a phase in which vortices are abundant must be disordered in the θ-variables. The interpretation of the low-temperature (large g) phase is less obvious. As it renormalizes towards the vacuum, one expects algebraic decay of some correlation functions, in particular of the vortex–antivortex correlation function. A more complete discussion of this phase is deferred to Section 3.2, where an ordering field in the XY model is included in the transformation to the CG.

3.2. Clock and Solid-on-Solid Models

The XY Hamiltonian [Eqs. (3.1) and (3.2)] is generalized with an on-site term which may represent the effect of a magnetic field:

$$\frac{-H}{kT} = \sum_{\langle j,k \rangle} V(\theta_j - \theta_k) + \sum_j W(p\theta_j) \tag{3.11}$$

The function $W(\theta)$ is periodic with period 2π, and is typically chosen $\cos(\theta)$. The effect of this term physically is best understood in terms of the two-component spin $(\cos\theta, \sin\theta)$, as representing the magnetic moment of an atom. For $p = 1$ W acts like a magnetic field, aligning the spins in a fixed direction. Higher values of p may simulate the effect of the surrounding crystal, which makes it energetically favorable for the spins to point in the direction of one of the crystallographic axes. The main result of W in any case is to break the full circular symmetry, leaving only the p-fold symmetry of a regular polygon. Following Villain[19] I take

$$\exp W(\theta) = \sum_{n=-\infty}^{\infty} \exp(-in\theta + n^2 \ln x) \tag{3.12}$$

This definition is an alternative way to write Eq. (3.2) with g replaced by $4\pi/\ln(x)$. The parameter x which varies between 0 and 1 is a measure of the strength of W. If x is small it acts like a weak field proportional to x. At $x = 1$ the potential W is infinitely strong so that θ can assume only integer multiples of $2\pi/p$. In this limit with only discrete values of θ the model is known as the Z_p model or p-state clock model. The complete partition sum

is now

$$Z = \sum_{\{k\}} \sum_{\{n\}} \int_0^{2\pi} d\{\theta\} \exp\Big[\frac{-g}{4\pi} (\nabla\theta - 2\pi k)^2 - ipn\theta + n^2 \ln(x) + (\nabla \times k)^2 \ln(y) \Big] \quad (3.13)$$

In addition to the symmetry breaking field another term is included, proportional to the square of the curl of k, which above has been recognized as the vorticity of θ. Thus by varying the parameter y one is able to control the concentration of vortices. By setting $y = 1$ the additional term vanishes so that the model reduces to a clock model ($x = 1$) or an XY model with a field ($0 < x < 1$) or without a field ($x = 0$). In the other extreme $y = 0$ the vortices are totally suppressed so that the periodic nature of θ is lost. This limit deserves some closer attention.

The vector field k, having no curl, can be written as the gradient of some integer scalar field, say s. It is then convenient to introduce a new variable $\theta' = \theta - 2\pi s$, and to replace the integral over θ and the sum over s jointly by an integral over θ':

$$\sum_{\{s\}} \int_0^{2\pi} d\{\theta\} \to \int_{-\infty}^{+\infty} d\{\theta'\} \quad (3.14)$$

The partition sum is then

$$Z = \sum_{\{n\}} \int_{-\infty}^{\infty} d\{\theta'\} \exp\Big[\frac{-g}{4\pi} (\nabla\theta')^2 - ipn\theta' + n^2 \ln(x) \Big] \quad (3.15)$$

noting that $\exp(2\pi ipns) = 1$. If $x = 0$ the variables n are forced to be zero, and the remaining partition sum is simply a Gaussian integral over θ'. Thus in the limit $x = y = 0$, the model defined by Eq. (3.13) is the Gaussian model. At $x = 1$ the sum over n results in periodic δ functions for the θ' variables. The integration then reduces to a sum over the now discrete θ' variables, interacting via a Gaussian. The resulting model, appropriately called the discrete Gaussian model, is an example of a class of models, known as solid-on-solid (SOS) models. In general SOS models have discrete, unbounded variables and a Hamiltonian which depends on the gradient of these variables, such that large gradients are suppressed. They are used to describe the physics of crystal surfaces.

Thus the partition sum (3.13) includes the XY model in the limit $x = 0$, $y = 1$, a Z_p model for $x = y = 1$, the Gaussian model at $x = y = 0$, and the discrete Gaussian model for $x = 1$, $y = 0$.

Now I proceed to transform the model to a CG. First observe that the integrand-summand of Eq. (3.13) is invariant under the gauge transforma-

tion

$$\begin{aligned} \theta \to \theta' &= \theta + 2\pi s \\ k \to k' &= k + \nabla s \end{aligned} \tag{3.16}$$

where the s are integer variables on each site of the lattice. Now choose s such that the horizontal component of k' vanishes, and accordingly replace the sum over k by a sum over s and over the vertical component of k':

$$\sum_{\{k\}} \to \sum_{\{s\}} \sum_{\{k'\}} \tag{3.17}$$

The integral is turned into a Gaussian integral by subsequently performing the replacement Eq. (3.14). The partition sum now has the form

$$Z = \sum_{\{k'\}} \sum_{\{n\}} \int_{-\infty}^{+\infty} d\{\theta'\} \exp\Big[\frac{g}{4\pi} \theta' \nabla \cdot \nabla \theta' - \theta'(ipn + g\nabla \cdot k) - \pi g k \cdot k + n^2 \ln(x) + (\nabla \times k')^2 \ln(y) \Big] \tag{3.18}$$

Performing the Gaussian integral, making use of the inverse of the Laplacian Eq. (2.3) leads to the result

$$Z = Z_G \sum_{\{n\}} \sum_{\{k'\}} \exp\Big[\frac{-1}{2g} (ipn + g\nabla \cdot k') G (ipn + g\nabla \cdot k') - \pi g k \cdot k + n^2 \ln(x) + (\nabla \times k')^2 \ln(y) \Big] \tag{3.19}$$

where Z_G is the partition integral of the Gaussian model proper. The terms bilinear in k' can be expressed in $m = \nabla \times k'$ as in Eqs. (3.5)–(3.7), and the remaining k' dependence likewise with the use of Eq. (2.5):

$$\frac{Z}{Z_G} = \sum_{\{m,n\}} \exp\Big[\frac{p^2}{2g} nGn + \frac{g}{2} mGm + ipm\Phi n + n^2 \ln(x) + m^2 \ln(y) \Big] \tag{3.20}$$

This completes the transformation of the spin model Eq. (3.13) into a CG of which the electric charges e are integer multiples of p. Equation (3.20) can be identified with Eq. (2.8) by defining

$$Y(m) = y^{(m^2)}$$

and

$$\begin{aligned} X(e) &= x^{(n^2)} \qquad \text{for} \quad e = np \\ &= 0 \qquad\quad \text{otherwise} \end{aligned} \tag{3.21}$$

The critical behavior of the CG has been discussed in Section 2.3. It is relatively straightforward to translate these findings into SOS, XY and Z_p

spin language. An electric charge e corresponds to a factor $\exp(ie\theta)$, called a spin-wave excitation, and a magnetic monopole m corresponds to a vortex with vorticity m. This is the main key to reach some conclusions for the critical behavior of these planar spin models from the discussion in Section 2.

An XY model corresponds to the CG (3.20) with $x = 0$ and $y = 1$, that is a CG without electric charges, but with magnetic monopoles. In its high-temperature phase the vortices are free, and the model is disordered. In the low-temperature phase the CG renormalizes towards the vacuum, where the correlation functions are known. In particular the spin–spin correlation

$$\langle \cos(\theta_j - \theta_k) \rangle = \langle \exp(i\theta_j - i\theta_k) \rangle \simeq |r_j - r_k|^{-1/g_R} \tag{3.22}$$

The fact that this correlation function approaches zero at large distances, implies that there is no genuine long-range order,[17] or that $\langle \cos\theta \rangle = 0$. At very low temperatures where y is strongly irrelevant, g_R does not differ appreciably from g. At the KT transition, because the vortices are marginal, $g_R = 4$ [Eq. (2.17)]. Correlation functions with higher periodicity q

$$\langle \cos(q\theta_j - q\theta_k) \rangle \sim |r_j - r_k|^{-q^2/g_R} \tag{3.23}$$

Therefore even though g_R is not known exactly, relations between exponents of different correlation functions and their value at the transition follow from this approach exactly.

The discrete Gaussian model transforms into a similar CG model, now with $x = 1$, $y = 0$, with electric rather than magnetic charges. The phase diagram is therefore similar to that of the XY model, but here it is the high-temperature phase that renormalizes to the CG vacuum. In the low-temperature phase the spin waves are relevant and consequently $\langle \cos(q\theta) \rangle$ (with $q < p$) acquires a nonzero expectation value. This indicates that one of the discrete values available to θ predominates throughout the lattice, i.e., the model is ordered. The corresponding correlation function does not vanish at infinite distances, in contrast with its behavior in the high-temperature phase, given by Eq. (3.23). The transition is at the point where the eigenvalue of the smallest electric charge p changes sign,

$$g_R = p^2/4 \tag{3.24}$$

There are no vortices in the discrete Gaussian model per se, but they may be introduced by giving the parameter y a small nonzero value. In the high-temperature phase ($g_R < p^2/4$) the free energy is a nonanalytic function of y with a singular part

$$f_s \sim y^{4/(4-g_R)} \tag{3.25}$$

as follows from the RG equations (2.11).

The Z_p or p-state clock model, $x = y = 1$, is transformed into a CG with both electric and magnetic particles, the former only with charges that are multiples of p. In the high-temperature extreme the vortices effectively disorder the system, while in the low-temperature limit the presence of the relevant spin-wave excitations bring the model in an ordered state. Some phase transition must separate these two phases. If $p \leqslant 4$ there is no value of g_R for which both the vortices and the spin waves are irrelevant. Therefore the phase transitions of the Z_2, Z_3, and Z_4 models, otherwise known as the Ising, three-state Potts, and Ashkin–Teller models, respectively, are not governed by the vacuum of this CG. Fortunately these models are equivalent to other CG models (Sections 3.3 and 3.4) in which the transition is more conveniently located. When $p > 4$ there is a regime

$$4 < g_R < p^2/4 \tag{3.26}$$

in which both x and y are irrelevant. This leads to the existence of an intermediate phase which is neither genuinely ordered, nor disordered. Correlation functions in this phase typically decay algebraically. The spin–spin correlation function, Eq. (3.22), has an exponent which varies between $4/p^2$ in the low- and $1/4$ in the high-temperature extreme of the intermediate phase.

When $x = y = 0$ the partition sum Eq. (3.13) represents a Gaussian model, and transforms into a CG in the vacuum. This agrees with the well-known fact that the Gaussian model has no phase transition and algebraically decaying correlation function at all temperatures.

The conclusions presented above for the XY, Z_p, and discrete Gaussian models, clearly illustrate that the transformation to a CG is a useful tool to analyze critical behavior. However, the transformation of the model (3.13) into the CG (3.20) could be done explicitly, only by virtue of the quadratic form of the Hamiltonian. Many models that do not have this property describe the same physics nonetheless. An SOS model in which the energy is linear in the difference between neighboring site variables, or in which differences of more than a single unit are not allowed, cannot be transformed simply into CG models. However the Hamiltonian may be expanded in powers of the gradient of θ, the leading term of which is Gaussian, and a RG transformation can be computed in leading order of the remaining terms. Since the non-Gaussian terms all contain more than two gradient operators they are irrelevant at all temperatures. By thus using RG arguments SOS, clock, and XY models with other than Villain potentials can be formally brought into a form in which a transformation to a CG is feasible. The results of these techniques may be accepted with a certain amount of reservation. RG transformations that can be calculated only in low-order expansion away from the Gaussian model, alias the CG

vacuum, are used at finite distance away from these models. It is implicitly assumed that the RG flows connect the CG vacuum and the model under investigation uninhibited by other fixed points. The beauty of the argument is that provided this qualitative assumption, quantitatively exact conclusions can be made.

3.3. The F Model and the BCSOS Model

The F model[20] is one of a class of two-dimensional models, known as six-vertex (6V) and eight-vertex (8V) models. These models are defined as follows. Place an arrow on each of the edges of a square lattice, such that the number of arrows pointing into any vertex is even. This allows eight configurations of arrows round a given vertex, as shown in Fig. 2a. Each of these configurations is given a weight, $W_1, W_2, \ldots, W_8$, and the partition sum of the model is the sum over all possible arrow configurations, of the product of all the vertex weights. For general values of W_j this model is called the 8V model. A subclass of these, the 6V models, satisfy the constraint that the number of arrows pointing into any vertex be equal to the number of outward pointing arrows. This condition is called the ice rule from the interpretation of the arrows as hydrogen bonds with an asymmetric charge distribution. Local neutrality at each vertex requires the exclusion of vertex configurations 7 and 8. The F-model satisfies the ice rule, $W_7 = W_8 = 0$, and has $W_1 = W_2 = W_3 = W_4$ and $W_5 = W_6$. Its solution by Lieb[21] was followed by a long list of generalizations,[22,23] which also were solved exactly. The F-model has a transition at $W_5 = 2W_1$, where the free energy is singular in the same way as described by Eq. (2.18).

The symmetric 8V model, with weights invariant under simultaneous inversion of all the arrows, has been solved by Baxter.[24] The F-model forms a critical subspace of this more general model. The free energy is a singular function of W_7 at $W_7 = W_8 = 0$, with an exponent that varies with

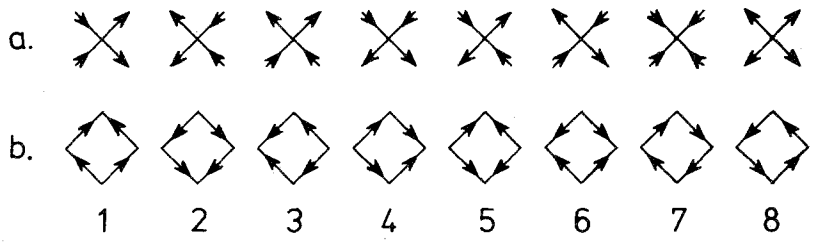

Fig. 2. Eight vertex (a) and BCSOS (b) configurations. The configurations of row b are formed by turning the edges of row a each about its center over $\pi/2$ to the left.

the other weights:

$$f_s \sim |W_7|^{2/y_{8V}} \quad \text{with} \quad y_{8V} = \frac{4}{\pi} \cos^{-1}\left(\frac{W_5}{2W_1}\right) \tag{3.27}$$

at the F model

$$W_1 = W_2 = W_3 = W_4, \quad W_5 = W_6, \quad \text{and} \quad W_7 = W_8 \sim 0 \tag{3.28}$$

Van Beijeren[25] observed that the F model is equivalent to an SOS model for the surface of a bcc crystal, which he named the BCSOS model. Turning all the edges and arrows of the model over a straight angle to the left produces a new configuration of arrows on the edges of the dual lattice. In Fig. 2b the result of this transformation is shown. The ice rule, which implies that the original arrow field has no divergence, guarantees that the turned arrows are free of curl. Therefore the latter arrow configuration is the gradient of some scalar variable, say θ, on the sites of the dual lattice. Between nearest-neighbor sites the θ variables always differ by the same amount, say $\pi/2$, and second neighbors may be equal or differ by π. The lattice thus naturally decomposes into two sublattices, on which the θ variables assume even and odd multiples of $\pi/2$, respectively. This is a natural representation of the surface of a bcc crystal projected in the $(0, 0, 1)$ direction, the sublattices corresponding with the cubic and body-centered positions, respectively. Note further that the configurations 7 and 8, excluded in the F model, correspond to vortices in the BCSOS model (Fig. 2b).

As the F model is identical to an SOS model, the RG arguments at the end of Section 3.2 suggest that it be equivalent to a CG. This statement can be supported by the following circumstantial evidence. Just like the electric CG, the F model has two phases separated by a phase transition of infinite order [Eq. (2.18)]. In the high-temperature phase of the F model the vortex weights W_7 and W_8 have a critical exponent which varies with temperature [Eq. (3.27)], just like the exponent of the monopole fugacity in the electric CG [Eq. (3.25)]. The final test is the value of this exponent at the transition, which according to the analysis in Section 2.2 is universal. The θ variables of the BCSOS model on the scale chosen above, assume integer multiples of $\pi/2$. This corresponds to $p = 4$ in the model (3.13), and with a critical g_R of 4 [Eq. (3.24)]. The arrow configurations 7 and 8 have vorticity 1 and -1, corresponding to four steps of $\pi/2$ each (see Fig. 2b). The critical exponent of these vortices is precisely zero at the transition, from Eqs. (3.24)–(3.25), in accord with the result of Baxter's exact solution, Eq. (3.27).

The agreement between the exact solution of the F model and the results from the CG analysis justifies the assumption that the RG flows indeed connect the models, unobstructed by intermediate fixed points.

Once the connection is accepted the 8V exponent [Eq. (3.27)] determines the renormalized coupling constant g_R as a function of the vertex weights:

$$g_R = \frac{8}{\pi} \sin^{-1}\left(\frac{W_5}{2W_1}\right) \tag{3.29}$$

This result is useful even for the F and 8V models as the correlation function of these models are not known otherwise, in spite of the exact calculations of their free energy. In addition the CG equivalence allows the student of these models to explore the parameter space in regions where no exact solution is available.

3.4. The Ashkin–Teller Model

The Z_4 model, originally introduced by Ashkin and Teller,[26] is generally named after these authors. In an Ising representation its Hamiltonian is

$$\frac{-H}{kT} = \sum_{\langle j,k \rangle} K s_j s_k (t_j t_k + 1) + L t_j t_k \tag{3.30}$$

where the spins s and t are 1 or -1 and the sum is over nearest-neighbor pairs of sites of a square lattice L. This Hamiltonian is related to Eq. (3.13) with $p = 4$, by the identification $s_j = \cos(\theta_j) + \sin(\theta_j)$ and $t_j = \cos(2\theta_j)$. By a dual transformation[27,10] on the s variables while keeping the t's fixed, one obtains an Ising representation of a 6V model.[28] This dual model consists of t variables on the original lattice L and dual variables of s, say $\tilde{s}$, on the dual lattice D. On the combined lattice LD, as shown in Fig. 3, the $\tilde{s}$ and t variables are nearest neighbors of each other. The equivalence of a 6V model is realized by placing arrows on the edges of LD pointing from $\tilde{s}$ to t or from t to $\tilde{s}$ if $\tilde{s} = t$ or $\tilde{s} \neq t$, respectively. This produces a curl-free arrow configuration of the kind exemplified in Fig. 2b. The edges of L and D form the diagonals of the faces of LD, as shown in Fig. 3. On the squares of which the horizontal diagonal is an edge of L (D) the weights are labeled with superscript a (b). Thus the weights of the 6V model dual to the Ashin–Teller model (3.30) are

$$\begin{aligned} W_1^a &= W_2^a = W_3^b = W_4^b = \sinh 2K \\ W_3^a &= W_4^a = W_1^b = W_2^b = \exp -2L \\ W_5^a &= W_6^a = W_5^b = W_6^b = \cosh 2K \\ W_7^a &= W_8^a = W_7^b = W_8^b = 0 \end{aligned} \tag{3.31}$$

Let the θ variables on the sublattice L of this BCSOS model, assume integer multiples of π, and on the sublattice D odd multiples of $\pi/2$. Then the

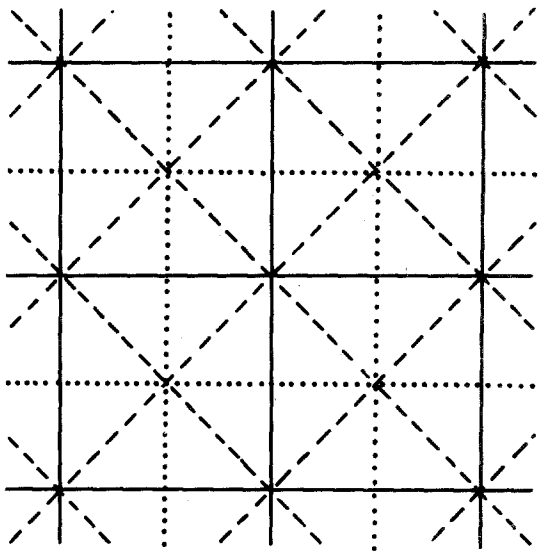

Fig. 3. The square lattice L (solid) and its dual D (dotted) together form a square lattice LD (dashed). The faces of LD center on the edges of L and D.

staggering of the weights (3.30) can be very naturally expressed in θ by a term in the Hamiltonian given by

$$t_{\mathrm{AT}} \cos\left(\frac{\theta_j + \theta_k + \theta_p + \theta_q}{2} \right) \tag{3.32}$$

with

$$t_{\mathrm{AT}} = 2L + \ln(\sinh 2K)$$

where the subscripts j and k are sites on L while p and q label sites on D. A more complete discussion of terms of this kind is found in Refs. 8 and 12. On the line $\sinh(2K) = \exp(-2L)$ the coefficient of the term (3.32) vanishes, and the model is identical to an F model. Since the F model is critical for $W_5 < 2W_1$, this may well be a critical line in the Ashkin–Teller model, depending on the relevance of t_{AT}. The term (3.32) represents the presence of four charges of $1/2$ or $-1/2$, on the corners of the square. Under RG transformations the most relevant contribution of the term (3.32) is of the form $\cos(2\theta)$, corresponding to the total charge 2 or -2 in the CG. The fugacity of such charge has an exponent $2 - 2/g_R$, which is relevant for $g_R > 1$. Therefore the line $t_{\mathrm{AT}} = 0$ is indeed a critical line in the Ashkin–Teller model with a singular part of the free energy

$$f_s \sim |t_{\mathrm{AT}}|^{2/y_{\mathrm{AT}}} \qquad \text{with} \quad y_{\mathrm{AT}} = 2 - 2/g_R \tag{3.33}$$

and

$$g_R = \frac{8}{\pi} \sin^{-1}\left(\frac{\coth 2K}{2} \right)$$

Other operators in the Ashkin–Teller model can also be expressed in CG variables,[8,9] but for the present review Eq. (3.33) sufficiently demonstrates the predictive power of the method.

3.5. The q-State Potts Model

One of the many generalizations of the Ising model is the Potts model[29] defined by the Hamiltonian

$$\frac{-H}{kT} = \sum_{\langle j,k \rangle} J\delta_{s_j,s_k} \tag{3.34}$$

where the spins s can assume q different values. The model has been mapped onto a 6V model by Temperley and Lieb.[30] For the details of this transformation the reader is referred to a pedagogical rederivation by Baxter *et al.*[31] Here I just use the equivalence to establish a connection of the Potts model via the BCSOS model to the CG. The resulting 6V model is defined by the weights

$$\begin{aligned} W_1^a &= W_2^a = W_3^b = W_4^b = w \\ W_3^a &= W_4^a = W_1^b = W_2^b = 1 \\ W_6^a &= W_5^b = \exp(-iv) + w\exp(iv) \\ W_5^a &= W_6^b = \exp(iv) + w\exp(-iv) \end{aligned} \tag{3.35}$$

with $\exp(J) = 1 + 2w\cos 2v$ and $q = 2 + \cos 4v$. The superscripts a and b are defined as for Eq. (3.31). Once the model has been transformed to the 6V model the number of states q need no longer be a natural number but can assume real values.

The critical point of the Potts model has long been known[29] by duality and is given by $w = 1$. At this value of w the 6V model (3.35) reduces to the F model. As in the Ashkin–Teller model, when $w \neq 0$, the term (3.32) gets a nonzero coefficient of the BCSOS Hamiltonian. In addition the weights W_5 and W_6 are no longer equal and differ between the sublattices. It is not difficult to convince oneself that this difference is precisely taken care of by an additional term in the Hamiltonian proportional to

$$i\sin\left(\frac{\theta_j + \theta_k + \theta_p + \theta_q}{2} \right) \tag{3.36}$$

where the sites j, k, p, and q are again the corners of an elementary square.

This term leaves the weights W_1 through W_4 unchanged, and modifies W_5^a and W_6^b by an equal, imaginary amount and W_6^a and W_5^b by the same amount in the opposite direction. The terms (3.32) and (3.36) are quite similar. Both represent the presence of a complex of charged particles of total electric charge 2 or -2. However, (3.36) introduces a difference between $X(2)$ and $X(-2)$ whereas (3.32) only contributes to their sum. This implies that in the equivalent CG the fugacities of the positive and negative charges are unequal. The consequences of such asymmetry are discussed at the end of Section 2.3, and can be readily applied here. Since at the critical point of the Potts model, that is at the F model, both $X(2)$ and $X(-2)$ vanish, the candidates for the thermal exponent of the Potts model are, from Eqs. (2.24) and (2.25), $3-6/g_R$ and $2-2/g_R$. There is to my knowledge no *a priori* way to tell which of these two is correct, since there seems to be no simple way of deciding whether the renormalized value of $X(-2)$ is zero. At $q=0$ (resistor network[32]) and $q=2$ (Ising model), however, the Potts exponent is known exactly and at $q=3$ by universality (hard hexagon model[33]). These values agree with a thermal Potts exponent of the kind (2.24)

$$y_{P,t} = 3 - 6/g_R$$

with

$$q = 2 + 2\cos\left(\frac{\pi}{2} g_R\right) \quad \text{and} \quad 2 \leqslant g_R \leqslant 4 \tag{3.37}$$

This is the value that corresponds to the case described by Eq. (2.24), and implies that the renormalized value of $X(-2)$ is zero even away from the critical point of the Potts model. Numerical estimates of this exponent[34] clearly support this value also at intermediate values of q. However, based on the leading exponent alone, one cannot exclude the possibility that, while $X(2)$ is linear in $w-1$, $X(-2)$ varies as $(w-1)^3$ [by dual symmetry $(w-1)^2$ is forbidden]. The subspace of BCSOS parameters, which is equivalent to the Potts model, is of a special symmetry, invariance under permutation of the q states. The universality hypothesis implies that subspaces of higher symmetry are preserved under renormalization. In the phase diagram Fig. 1 the only RG flow line that meets the origin tangential with the $X(2)$ axis is the $X(2)$ axis itself. Therefore unless universality is violated in this case, I conclude that the renormalized value of $X(-2)$ vanishes in the Potts model at all temperatures. The curious significance of this statement is that the Potts subspace is a critical manifold in the space of more general BCSOS models, even where the Potts model itself is not critical. The long-range correlations associated with the singularity of the Potts subspace must be impossible to express in Potts variables. In retrospect this is not that surprising, since the operators that have critical

fluctuations are those associated with breaking the special symmetry. These operators naturally cannot be expressed in Potts variables, in which the symmetry is inherent.

A suitably generalized Potts model[12,34] exhibits besides critical also tricritical behavior. Analyzed by means of the same CG equivalence[12] this tricritical behavior is due to an asymmetry between the $e = +4$ and -4 charges. At the tricritical point the $+4$ charges are completely expelled from the renormalized CG, while the -4 charges still remain. The leading tricritical exponent is given by the same equation (3.37) as the critical exponent, now on the branch with $4 \leqslant g_R \leqslant 6$. The second tricritical exponent governing the crossover singularities is, analogous to Eq. (2.26)

$$y_{P,t2} = 4 - 16/g_R \tag{3.38}$$

Den Nijs[13] recently showed that the spin–spin correlation function of the q-state Potts model translates into a charge correlation function between fractional charges. Curiously the value of these charges is a continuous function of g_R and therefore of q. The magnetic eigenvalue follows from the exponent of these correlation functions, and is

$$y_{P,m} = 1 + \frac{3}{2g_R} + \frac{g_R}{8} \tag{3.39}$$

The values of the critical and tricritical exponents for integer q are listed in Table I. This table is a clear illustration of the predictive power of this theory, since more than half of the numbers in the table were not previously known from other sources. Most of these exponents apply at least in

Table I. Critical and Tricritical Exponents of the q-State Potts Model[a]

q	0	1	2	3	4
$y_{P,t}$	0	$\frac{3}{4}$	1	$\frac{6}{5}$	$\frac{3}{2}$
$y_{P,m}$	2	$\frac{91}{48}$	$\frac{15}{8}$	$\frac{28}{15}$	$\frac{15}{8}$
$y_{P,t1}$	2	$\frac{15}{8}$	$\frac{9}{5}$	$\frac{12}{7}$	$\frac{3}{2}$
$y_{P,t2}$	$\frac{4}{3}$	1	$\frac{4}{5}$	$\frac{4}{7}$	0
$y_{P,m}$	2	$\frac{187}{96}$	$\frac{77}{40}$	$\frac{40}{21}$	$\frac{15}{8}$

[a] For the critical point the leading thermal, $y_{P,t} = 1/\nu$, and magnetic, $y_{P,m} = 2 - \eta/2$, eigenvalues are shown. For the tricritical transition the leading thermal, $y_{P,t1} = 1/\nu$, the next-to-leading thermal, $y_{P,t2} = \phi/\nu$, and the leading magnetic $y_{P,m} = 2 - \eta/2$ eigenvalues are listed.

principle to actual phase transitions in adsorbed monolayers.[35] The values at $q = 1$ describe two-dimensional percolation processes.[32]

3.6. The $O(n)$ Model

Another two-dimensional spin model of interest is the isotropic n-vector model or $O(n)$ model.[14] This class of models has n-component spins for variables, interacting via a ferromagnetic Hamiltonian invariant under simultaneous rotation of the spins. In many representations the partition is expressed in a form in which n is simply a parameter and may then be given non-natural values. In the limit $n = 0$ this partition sum then describes the scaling behavior of a long polymer in solution.[36] The sequence of transformations from $O(n)$ models to the CG has been performed only for a limited class of models, of which the Hamiltonian is peculiarly designed so as to make the mapping possible.[15] The partition integral for an $O(n)$ model on the honeycomb lattice is defined by

$$Z_{O(n)} = \int \prod_i Q(s_i)\, d^n s_i \prod_{\langle j,k \rangle} (1 + K\, s_j \cdot s_k) \tag{3.40}$$

The first product is over the sites and the second over nearest-neighbor pairs of sites of the honeycomb lattice. This model is an $O(n)$ model if the weight function $Q(s)$ is isotropic. However, for the derivation that follows it is sufficient if $Q(s)$ is invariant under cubic transformations, i.e., permutation or inversion of the spin components. The weight function and the length of the n-component spins s are normalized so that $\int Q(s)\, d^n s = 1$ and $\int s^2 Q(s)\, d^n s = n$. The partition sum (3.40) can be expanded in a sum over all diagrams consisting of closed rings on the honeycomb lattice[35]:

$$Z_{O(n)} = \sum_G K^L n^c \tag{3.41}$$

where c is the number of rings in the diagram G and L their total length.

In order to make the connection with the CG I will show that the $O(n)$ model (3.41) is equivalent to a suitably defined SOS model. The necessary sequence of transformations is given starting from a 6V model on the Kagomé lattice, shown in Fig. 4. This lattice has three types of vertices, labeled A, B, and C. The arrow configurations as numbered in Fig. 2a, at a vertex of type A have weights

$$W_1, W_2, \ldots, W_6 = 1, 1, b, b, \exp(2iu), \exp(-2iu) \tag{3.42}$$

On the vertices of type B and C the weights (3.42) are assigned to the

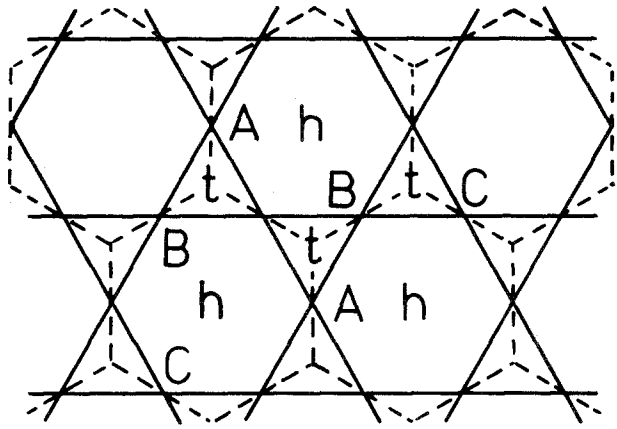

Fig. 4. The Kagomé lattice shown by solid lines. Vertices of types A, B, and C differ by rotation over $\pm 2\pi/3$. The dashed lines indicate the honeycomb lattice. The symbols t and h mark the positions of the variables θ^t and θ^h of the KSOS model.

configurations of Fig. 2a rotated over $\pm 2\pi/3$. The same process by which the square 6V model is transformed to the BCSOS model (see Fig. 2b), maps the 6V model defined here onto an SOS model (the KSOS model for quick reference), of which the variables reside on the faces of the Kagomé lattice. The variables on the hexagons θ^h and on the triangles θ^t assume even and odd multiples of $\pi/2$, respectively, such that two adjacent θ^t and θ^h always differs by precisely one such unit. The weights (3.42) of the KSOS model correspond to an interaction defined by the following rule: A weight b is assigned to each nearest-neighbor pair of θ^h with unequal members, and a weight $\exp(iu)$ or $\exp(-iu)$ to each pair of adjacent θ^t and θ^h if $\theta^t > \theta^h$ or $\theta^t < \theta^h$, respectively. Since these rules do not include an interaction of the θ^t among themselves, these variables can be traced out. This process leaves an effective interaction among three adjacent θ^h. Such triplet can be in one of three energetically different configurations: (i) They are equal. In this case the θ^t in the center can be in two states, and the effective weight is $2\cos(3u)$. (ii) One of the θ^h is larger than the other two. The center θ^t must now assume the intermediate value, and the weight is $b\exp(iu)$. (iii) One θ^h is smaller than the other two, with weight $b\exp(-iu)$. The result is an SOS model on a triangular lattice (the TSOS model), in which adjacent variables may be equal or different by π.

The partition sum of this model can be expressed as a low-temperature expansion, in powers of b. The ground state, with all θ^h equal, has a weight $2\cos(3u)$ per triplet, which it is convenient to divide out. The expansion

parameter is then

$$K = \frac{b}{2\cos(3u)} \tag{3.43}$$

A natural diagram representation of the states of this model is given by the domain walls between regions of different θ^h. These walls form closed rings on the honeycomb lattice (see Fig. 4). To calculate the weight of a given ring, follow it all the way round, keeping the higher value of θ^h on the left-hand side. At the vertices of the honeycomb lattice, the domain wall turns either to the right or to the left. At a vertex where the wall turns left, it has a weight $K\exp(iu)$, and where it turns right the weight is $K\exp(-iu)$. Since each domain wall is closed and does not intersect itself, the number of right turns N_r and the number of left turns N_l satisfy $|N_l - N_r| = 6$. The domain walls can thus be of two types with $N_l - N_r = 6$, or with $N_l - N_r = -6$. Therefore the partition sum is performed in two steps, first a sum over all possible configurations of closed domain walls on the lattice, and then for a given such configuration a sum over their type. Since the type of each wall can be decided independently of the other walls, the second sum is trivial and leaves the partition sum

$$Z_{\mathrm{TSOS}} = (2\cos 3u)^{2N} \sum_G K^L (2\cos 6u)^c \tag{3.44}$$

N is the total number of variables, and, for each wall configuration c is the number of rings and L their combined length. Obviously the partition sums $Z_{O(n)}$ and Z_{TSOS} are equivalent if

$$n = 2\cos(6u) \tag{3.45}$$

This completes the transformation of the KSOS model (3.42) into the $O(n)$ model (3.40), and suggests a connection with the CG.

The weights (3.42) are similar to the weights (3.35), corresponding to the Potts model. The same arguments as in Section 3.5 show that in the CG model equivalent with this KSOS model, the fugacities $X(2)$ and $X(-2)$ are generally unequal and nonzero. A distinct difference between the KSOS model and the BCSOS model, is the absence of symmetry between the θ^t and the θ^h variables. The weights W_1 and W_2 thus are not equivalent with W_3 and W_4. Therefore it is not evident from symmetry at what value of b the fugacities $X(2)$ and $X(-2)$ vanish, unlike the case of the analogous BCSOS model. In order to determine the locus of the $O(n)$ model in the parameter space of the CG, I now generalize some of the results derived in Sections 3.3 and 3.5.

The Potts model [Eq. (3.34)] on the triangular lattice is equivalent[31]

with a KSOS model, with weights

$$W_1, W_2, \ldots, W_6 = 1, 1, b, b, \exp(2iv) + b\exp(-iv), \exp(-2iv) + b\exp(iv) \tag{3.46}$$

on the A vertices, and for the rotated configurations on the B and C vertices. The parameters b and v are related to the Potts parameters by

$$\exp(J) = 1 + \frac{\cos(3v)}{b} \tag{3.47}$$

and

$$q = 2 + 2\cos(6v)$$

At its critical point, $b = 2\cos(3v)$ the Potts model reduces to the F model on the Kagomé lattice, of which Baxter has demonstrated[23,38] that it can be transformed into the square F model. Hence also for the triangular Potts model the fugacities $X(2)$ and $X(-2)$ both vanish at the critical point. Thus of the analysis of Section 3.5 all conditions that could be lattice dependent are nonetheless satisfied for the triangular Potts model. Therefore, in keeping with the analysis at the end of Section 2.3, the 6V model (3.46) forms a critical manifold in a larger parameter space. The $O(n)$ model is therefore critical at its intersection with the Potts model, and its transition is computed readily from the coincidence of (3.46) and (3.42),

$$b = -2\cos(3v) \tag{3.48}$$

and

$$u = \pi/2 - 2v$$

which expressed in n and K [Eq. (3.43) and (3.45)] is

$$K_c = \left[2 + (2-n)^{1/2}\right]^{-1/2} \tag{3.49}$$

The renormalized coupling constant g_R at the F model follows as in Section 3.3 from the 8V exponent, which is given for the Kagomé F model and the CG, respectively, by

$$y_{8V} = \frac{6v}{\pi} = 2 - \frac{g_R}{2} \tag{3.50}$$

Since the number of states q of the Potts model is a label of symmetry and universality classes, RG flow lines in the Potts model are lines of constant q. Consequently, the relation (3.37) between q and g_R holds in the entire Potts model (3.46) and not only in the F model.

The thermal $O(n)$ exponent is then that of $X(-2)$ at finite values of $X(2)$ [Eq. (2.24)]

$$y_{O(n),t} = 4 - 4/g_R \tag{3.51}$$

g_R follows from Eqs. (3.45), (3.48) and (3.50)

$$n = -\cos(\pi g_R) \tag{3.52}$$

with $1 < g_R < 2$. Note that the product $X(2)^2X(-4)$, which is finite at the transition, is irrelevant for these values of g_R.

The magnetic exponent of the $O(n)$ model follows from the spin–spin correlation function $\langle s_j \cdot s_k \rangle$. It is readily verified that the diagrams of this quantity are walks on the honeycomb lattice from site j to k, possibly combined with rings. An example of such diagram is shown in Fig. 5. In the SOS language these diagrams correspond with the vortex–antivortex correlation function, of magnetic charge $1/2$, as the walk represents a domain wall of step π. However, the peculiar weights of the TSOS model associate energy with the curvature of the walk, with the result that the diagrams Figs. 5a and 5b have different weight. This problem can be repaired by adding the spin-wave operator

$$\exp\left(\frac{-6iu\theta}{\pi}\right) \tag{3.53}$$

to both vortex operators. For each full turn of the walk about one of its ends, this point is raised or lowered by π, as shown in Fig. 5. The spin-wave (3.53) then gets multiplied by a factor $\exp(-6iu)$ or $\exp(6iu)$, precisely opposite to the factor from the additional turns. Some diagrams contributing to the spin–spin correlations function contain rings surrounding both end points. When such ring contributes the factor $\exp(6iu)$, its interior is

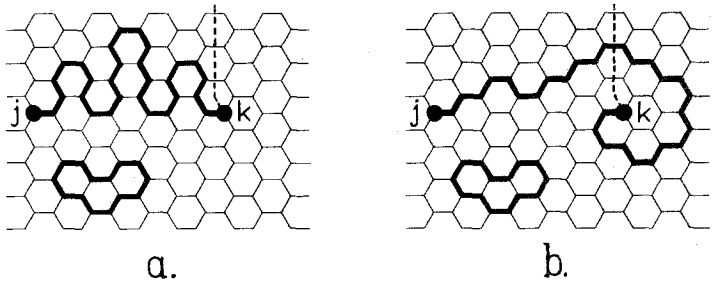

Fig. 5. Two diagrams (a and b) that contribute to $\langle s_j \cdot s_k \rangle$ with equal weight. A path (dashed curve) connects site k to the boundary of the lattice. Since the path in diagram b crosses one domain wall, the height of site k differs from that in diagram a by one step of π. This difference does not depend on the particular dashed path chosen, provided the same path is taken in diagram a and b.

raised by π, and the two spin-wave operators both contribute an extra $\exp(-6iu)$, so that the total becomes $\exp(-6iu)$. Thus the spin waves replace the factor $\exp(6iu)$ by $\exp(-6iu)$ and vice versa, so that the total weight of the surrounding ring, being the sum of the two is conveniently left invariant.

The magnetic exponent is therefore that of a combination of a vortex and a spin wave, the latter of which has a charge continuously dependent on n. The resulting magnetic exponent is

$$y_{O(n),m} = 1 + \frac{3g_R}{8} + \frac{1}{2g_R} \tag{3.54}$$

The exponents (3.53) and (3.56) reproduce the well-known results for the Ising ($n = 1$) and XY ($n = 2$) models, and for the polymer limit they give the numbers $\nu = 3/4$ and $\gamma = 43/32$. The exponent ν is the same as that calculated by Flory,[39] but his mean-field method has the character of an approximation. The agreement is probably a coincidence.

Since at the $O(n)$ transition $X(-2)$ changes sign at a finite value of $X(2)$, in the low-temperature phase these two fugacities have unequal sign. Recall that Fig. 1, where the $O(n)$ model is indicated by a dashed line, shows that the region with $X(2)X(-2) < 0$ is critical and renormalizes ultimately to the vacuum, with a $g_R < 1$. In the $O(n)$ model this value of the renormalized coupling constant is simply another branch of Eq. (3.52), and thus known as a function of n. The magnetic exponent (3.56), with this value of g_R substituted, describes the power law decay of the correlation function and the singularity of the free energy as function of the magnetic field, in the low-temperature phase. For $1 < n < 2$ the exponent $y_{O(n),m} < 2$, which means that the model is not genuinely ordered, but has a second-order transition in the field. This is similar to the low-temperature phase of the XY model, but the magnetic exponent is universal and does not depend on temperature. For $n = 1$ the low-temperature exponent is 2, in agreement with the finite value of the spontaneous magnetization. For $n < 1$ $y_{O(n),m} > 2$, so that the spin–spin correlation function grows instead of decays with distance, and the magnetization is infinite at zero field. This nonphysical situation is likely the result of the nonphysical values of n, and of no consequence. However, the physics of a two-dimensional polymer solution in the semidilute regime,[40] which is described by the low-temperature phase of the $O(n = 0)$ model, probably reflects this unusual behavior.

4. CONCLUSIONS

A connection between the CG and a number of spin models in two dimensions has been established and utilized for the calculation of critical

behavior. The CG model is like a language in which many of the standard problems in two-dimensional statistical mechanics may be phrased. Similar uses have been made of SOS models and the one-dimensional electron gas.[41]

The CG as a computational tool has by no means been exhausted by the calculations sampled in this reivew. For the models discussed here many more critical indices can be calculated of a variety of symmetry breaking fields and the like, in addition to the asymptotic behavior of multiple correlation functions.[42] Many models are not discussed here.

The question which models that as yet have not had the pleasure of an intimate connection with the CG can be treated with these techniques is not easy to answer. It is a basic weakness of the theory that the transformations used to make the connection between the spin model and the CG are rarely found on purpose. It is far from clear what can and cannot be done by these mappings. The mechanism of the phase transition in the CG models equivalent with the Z_p, Ashkin–Teller, Potts, or $O(n)$ models, are driven by strikingly dissimilar mechanisms. Nevertheless the Ising model is a special case of all four of these classes. This suggests that an insight in the physical nature of the phase transition of a model is a poor guide to find, whether, and where in the parameter space, the model can be mapped onto a CG.

An obvious disadvantage of the method is its utter dependence on RG and universality assumptions. On the other hand some of the results reviewed in this paper are an independent confirmation of universality. For instance the three-state Potts critical exponents coincide with those of the hard hexagon problem,[33] and the Ising tricritical indices agree with those of the hard square model.[43] If the outcome of a calculation like those in Section 3 of this paper strongly disagree with approximates or estimates, a suspicion of the underlying assumptions is justified. If, however, the results are known to be approximately correct, this can be taken as evidence that the assumptions are valid and the results in fact exact. Calculations of this kind cannot give approximate answers except by mere coincidence.

REFERENCES

1. L. P. Kadanoff, *J. Phys. A* **11**:1399 (1978).
2. J. M. Kosterlitz and D. J. Thouless, *J. Phys. C* **6**:1181 (1973); 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**:12 (1977).
4. H. J. F. Knops, *Phys. Rev. Lett.* **39**:766 (1977).
5. B. I. Halperin and D. R. Nelson, *Phys. Rev. Lett.* **41**:121 (1978), and *Phys. Rev. B* **19**:2457 (1979); A. P. Young, *Phys. Rev. B* **19**:1855 (1979).
6. L. P. Kadanoff and A. C. Brown, *Ann. Phys. (N.Y.)* **121**:318 (1979).
7. J. Ashkin and E. Teller, *Phys. Rev.* **64**:178 (1943).
8. H. J. F. Knops, *Ann. Phys. (N.Y.)* **128**:448 (1980).

9. H. J. F. Knops and L. W. J. den Ouden, *Ann. Phys. (N.Y.)* **138**:155 (1982).
10. B. Nienhuis, E. K. Riedel, and M. Schick, *Phys. Rev. B* **27**:5625 (1983).
11. R. B. Potts, *Proc. Cambridge Phil. Soc.* **48**:106 (1952).
12. B. Nienhuis, *J. Phys. A* **15**:199 (1982).
13. M. P. M. den Nijs, *Phys. Rev. B* **27**:1674 (1983).
14. H. E. Stanley, *Phys. Rev. Lett.* **20**:589 (1968).
15. B. Nienhuis, *Phys. Rev. Lett.* **49**:1062 (1982).
16. V. L. Pokrovsky and A. L. Talapov, *Phys. Rev. Lett.* **42**:65 (1979).
17. N. D. Mermin and H. Wagner, *Phys. Rev. Lett.* **17**:1133 (1966); N. D. Mermin, *J. Math. Phys.* **8**:1061 (1967).
18. H. E. Stanley, *Phys. Rev. Lett.* **20**:150 (1968).
19. J. Villain, *J. Phys. (Paris)* **36**:581 (1975).
20. F. Rys, *Helv. Phys. Acta* **36**:537 (1963).
21. E. H. Lieb, *Phys. Rev. Lett.* **18**:1046 (1967).
22. E. H. Lieb and F. Y. Wu, in *Phase Transitions and Critical Phenomena*, C. Domb and M. S. Green, eds. (Academic Press, London, 1972), Vol. 1.
23. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
24. R. J. Baxter, *Phys. Rev. Lett.* **26**:832 (1971).
25. H. van Beijeren, *Phys. Rev. Lett.* **38**:993 (1977).
26. J. Ashkin and E. Teller, *Phys. Rev.* **64**:178 (1943).
27. K. A. Kramers and G. H. Wannier, *Phys. Rev.* **60**:252 (1941); F. Y. Wu and Y. K. Wang, *J. Math. Phys.* **17**:439 (1976).
28. L. P. Kadanoff and F. Wegner, *Phys. Rev. B* **4**:3989 (1971).
29. R. B. Potts, *Proc. Cambridge Phil. Soc.* **48**:106 (1952).
30. H. N. V. Temperley and E. H. Lieb, *Proc. R. Soc. London Ser. A* **322**:251 (1971).
31. R. J. Baxter, S. B. Kelland, and F. Y. Yu, *J. Phys. A* **9**:397 (1976).
32. E. M. Fortuin and P. Kasteleyn, *Physica* **57**:536 (1972).
33. R. J. Baxter, *J. Phys. A* **13**:L61 (1980).
34. M. P. M. den Nijs, *Physica* **95A**:449 (1979); B. Nienhuis, E. K. Riedel, and M. Schick, *J. Phys. A* **13**:L31 (1980).
35. E. Domany, M. Schick, J. S. Walker, and R. B. Griffiths, *Phys. Rev. B* **18**:2209 (1978); E. Domany and M. Schick, *Phys. Rev. B* **20**:3828 (1979).
36. P. G. de Gennes, *Phys. Lett.* **38A**:339 (1972).
37. E. Domany, D. Mukamel, B. Nienhuis, and A. Schwimmer, *Nucl. Phys.* **B190** [FS3]:279 (1981).
38. R. J. Baxter, *Phil. Trans. R. Soc. London* **289**:315 (1978).
39. P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, New York, 1953).
40. J. des Cloizeaux, *Phys. Rev. A* **10**:1665 (1974).
41. See for instance M. P. M. den Nijs, *Phys. Rev. B* **23**:6111 (1981).
42. B. Nienhuis and L. P. Kadanoff, to be published.
43. D. A. Huse, *Phys. Rev. Lett.* **49**:1121 (1982); R. J. Baxter and P. A. Pearce, *J. Phys. A* **16**:2239 (1983).