Ashkin-Teller and Gross-Neveu Models: New Relations and Results

R. Shankar
J. W. Gibbs Laboratory of Physics, Yale University, New Haven, Connecticut 06520
(Received 25 February 1985)

It is shown that (i) the N-component Ashkin-Teller model in $d=2$ has continuous O(N) symmetry, (ii) its critical properties *near the decoupled Ising transition* are determined by the massive Gross-Neveu $[(\bar{\psi}\psi)^2]$ model in general and the (integrable) massless version along a line, (iii) the known results for the massless Gross-Neveu model imply that the first-order transition found as $N \to \infty$ persists down to $N=3$ along this line, (iv) the supersymmetry of the $N=3$ model implies that the leading singularity of the free energy is zero along the first-order line, and (v) the $N=4$ model along this line decouples into two identical $N=2$ models up to subleading corrections.

PACS numbers: 05.50.+q, 11.15.Ha, 12.35.Cn, 75.10.Hk

Over forty years ago Ashkin and Teller[1] introduced their model for a lattice gas in $d=2$ which exhibited Kramers-Wannier duality. It could also be viewed as a model with two Ising spins coupled through their energy densities. We are concerned with the generalization to the case with N Ising spins coupled in a symmetric way as follows:

$$Z=\sum \exp[K\sum \mathbf{s}_i\cdot\mathbf{s}_j+g\sum_{\rm nn}(\mathbf{s}_i\cdot\mathbf{s}_j)^2], \qquad (1)$$

where $\sum$ is a sum over configurations of the N Ising spins s^α ($\alpha=1\ldots N$) at each site of a two-dimensional square lattice, $\sum_{\rm nn}$ is a sum over nearest-neighbor pairs, and $\mathbf{s}\cdot\mathbf{s}'$ stands for $s_\alpha s'_\alpha$ summed over α. Recently, Fradkin[2] showed that as $N\to\infty$ the Ising transition at $g=0$, $K=K^*=[\ln(1+\sqrt{2})]/2$ extends into a first-order line for $g>0$ and a second-order line for $g<0$. For $N=2$ the model is solved along a (self-dual) line in the K-g plane by mapping it to the Baxter model.[3] In addition, Grest and Widom[4] have subjected the small-N cases to a variety of numerical and renormalization-group analyses. We are concerned here with establishing some exact results for finite N. We turn to the first of the results quoted in the abstract.

The model has O(N) symmetry.—Even though I wrote Z in terms of formal O(N) dot products this does not imply O(N) symmetry, since the group cannot act on s which takes on discrete values. To get the desired result one must rewrite Z in terms of s and the scalar field ϕ used by Fradkin, write the Ising part of Z as a Grassmann integral, and observe that the action is O(N) symmetric. There is no problem with O(N) acting on the Grassmann fields. I will present a related proof in the transfer-matrix formalism. Consider a model with couplings K_t and K_x in the two directions, with g isotropic. In the τ-continuum limit[5] defined by

$$\exp(-2K_t)=\tau/2, \quad K_x=\lambda\tau/2,$$
$$g=\tau g_0, \quad \text{with } \tau\to 0, \qquad (2)$$

the transfer matrix takes on a simple form $T=\exp(-\tau H)$, with

$$H=-\tfrac{1}{2}\sum_n\{\sigma_1^\alpha(n)+\lambda\sum_\alpha\sigma_3^\alpha(n)\sigma_3^\alpha(n+1)+g_0[\sum_\alpha\sigma_3^\alpha(n+1)]^2\}. \qquad (3)$$

We will not keep track of *analytic* c-numbers in H. For each species let us introduce two Majorana (Hermitian) Fermi operators as follows:

$$\sqrt{2}\psi_1(n)=\prod_{-\infty}^{n}\sigma_1(m)\sigma_3(n+1), \quad \sqrt{2}\psi_2(n)=\prod_{-\infty}^{n-1}\sigma_1(m)\sigma_2(n). \qquad (4)$$

I have not shown the Klein factors (which involve just products of σ_1's over all sites) that make different species anticommute since they drop out of H. In terms of the fermion operators obeying

$$\{\psi^\alpha(n),\psi^\beta(m)\}=\delta^{mn}\delta^{\alpha\beta}, \qquad (5)$$

$$H=-i\sum_n\boldsymbol{\psi}_1(n)\cdot\boldsymbol{\psi}_2(n+1)+i\lambda\sum_n\boldsymbol{\psi}_1(n)\cdot\boldsymbol{\psi}_2(n)+g_0\sum_n[2i\boldsymbol{\psi}_1(n)\cdot\boldsymbol{\psi}_2(n)]^2. \qquad (6)$$

The free energy per site of the Ashkin-Teller model (ATM) is the ground-state energy per site of H.

The O(N) symmetry of the model follows (since group action can be unitarily implemented on the isovector fields ψ) and must be unbroken as $d=2$. Hereafter, the arrows on ψ will be dropped. The range of α will be clear from the context.

The critical region.—We know that when $g_0=0$, the above H has a vanishing gap and the ground-state energy $E_0(\lambda)$ becomes singular at $\lambda=1$. In the vicinity of this point, i.e., in the scaling region, one may construct a continuum theory which will capture the infrared behavior when the mass $\to 0$. One must simply choose λ to vary with the lattice spacing a $(=1/\Lambda)$ as $\lambda(a)=1+ma$, or $d\lambda/d\ln a=\beta(\lambda)=\lambda$ to obtain an a-independent limit as $a\rightarrow 0$. With g_0 small but nonzero, one finds[4]

$$\beta(\lambda)=\lambda+O(\lambda,g_0),\quad \beta(g_0)=4(N-2)g_0^2/\pi+O(g_0^3), \tag{7}$$

which means that g_0 must vanish with a as $1/\ln(1/a)$. In terms of

$$\Psi=\frac{\psi}{\sqrt{a}},\quad m_0=\frac{(\lambda-1)}{a},\quad p=-i\frac{d}{dx},\quad \gamma^0=\beta=\begin{bmatrix}0 & i\\ -i & 0\end{bmatrix},\quad \alpha\equiv\gamma^5\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},\quad \bar{\Psi}\Psi=\Psi^T\beta\Psi, \tag{8}$$

the continuum Hamiltonian is

$$H_c=H/a$$
$$=\int dx\{\tfrac{1}{2}[\Psi^T(\alpha p+\beta m_0)\Psi]-g_0(\bar{\Psi}\Psi)^2\}. \tag{9}$$

[In obtaining H_c from H all higher derivatives have been dropped. While these are irrelevant and can be eliminated by momentum space renormalization, they will modify m_0 and g_0 before leaving; i.e., $m_0a=\lambda-1+O(g)$, etc. However, this point can be ignored unless one wants to transcribe the phase diagram from the m_0-g_0 plane to the λ-g_0 plane.]

In the path-integral formalism H_c would be associated with a Lagrangean density

$$L=\tfrac{1}{2}[\bar{\Psi}(i\partial\!\!\!/-m_0)\Psi]+g_0(\bar{\Psi}\Psi)^2, \tag{10}$$

which, for $g_0>0$, we recognize to be the L for the O(N) Gross-Neveu model[6] (GNM) with a (bare) mass term. Under a chiral transformation, $\Psi\rightarrow\gamma^5\Psi$, $\bar{\Psi}\Psi$ changes sign. The spectrum, and in particular the ground-state energy, must be invariant under $m_0\rightarrow -m_0$:

$$E_0(m_0,g_0)=E_0(-m_0,g_0). \tag{11}$$

Thus if the Ising transition evolves into a single transition line, it must be the line $m_0=0$. Of course, the order of the transition and its very existence are still open. Now notice that for large $|m_0|$, $\langle\bar{\Psi}\Psi\rangle$ will be nonzero and oppose m in sign. Thus, $\bar{\Psi}\Psi$ is the order parameter and m_0 is the external field. On the line $m_0=0$, the derivative of $E_0(m_0,g_0)$ with respect to $|m_0|$ is $\langle\bar{\Psi}\Psi\rangle$ in the field theory and is the internal energy in the ATM. Therefore a nonzero $\langle\bar{\Psi}\Psi\rangle$ in the GNM, i.e., the spontaneous breakdown of chiral symmetry, corresponds to a first-order transition. Gross and Neveu established that this happens for $N\rightarrow\infty$. I will now show that the first-order transition persists at $N=4$. Later I will use known S-matrix results for this model to argue that the first-order transition persists down to $N=3$. Of course, at $N=2$ we know that the transition is second order with continuously varying exponents, a result that follows from the mapping of the ATM to the Baxter line or the equivalence of the $N=2$ GNM to the massless Thirring model.[7]

The case N=4.—We will now follow Witten[8] and bosonize the model. First we form two Dirac fields from the four Majorana fields,

$$\Psi_{\rm I}=(\Psi_1+i\Psi_2)/\sqrt{2},\quad \Psi_{\rm II}=(\Psi_3+i\Psi_4)/\sqrt{2}, \tag{12}$$

and bosonize the two Dirac fields by the usual rules,[9]

$$i\bar{\Psi}\partial\!\!\!/\Psi=\tfrac{1}{2}(\partial_\mu\phi)^2,\quad \bar{\Psi}\Psi=c\Lambda\cos[(4\pi)^{1/2}\pi\phi],$$
$$\bar{\Psi}\gamma_\mu\Psi=\epsilon_{\mu\nu}\partial_\nu\phi/\sqrt{\pi}, \tag{13}$$

where c is a constant, to get

$$L=\tfrac{1}{2}(\partial_\mu\phi_{\rm I})^2+\tfrac{1}{2}(\partial_\mu\psi_{\rm II})^2+c^2\Lambda^2\{\cos[(4\pi)^{1/2}\phi_{\rm I}]+\cos[(4\pi)^{1/2}\phi_{\rm II}]\}^2. \tag{14}$$

Using now the following identity for a Dirac field,

$$(\bar{\Psi}\Psi)^2=-2(\bar{\Psi}\gamma_\mu\Psi)^2=(2/\pi)(\partial_\mu\phi)^2, \tag{15}$$

rescaling ϕ to ϕ' so as to make the coefficient of the free-field term $\frac{1}{2}$, and redefining $\phi_\pm=(\phi_{\rm I}'\pm\phi_{\rm II}')/\sqrt{2}$, we get two sine-Gordon (SG) systems,

$$L=(\tfrac{1}{2})(\partial_\mu\phi'_+)^2+\Lambda^2c^2g_0\cos\beta\phi'_+ +(+\rightarrow -),$$
$$\beta=[8\pi/(1+16g_0/\pi)]^{1/2}. \tag{16}$$

We know from S-matrix and Bethe-*Ansatz* calculations[10,11] that as $\beta^2\rightarrow 8\pi^-$ (recall g_0 tends to 0 with a) the spectrum has only kinks, solitons s and antisolitons $\bar{s}$, and that they form an isospinor doublet. We will now see how all this implies chiral-symmetry breakdown.

Consider Q_{12} and Q_{34}, which are the generators of commuting O(4) rotations in the 1-2 and 3-4 planes. They are given by the space integrals of $\bar{\Psi}_{\rm I}\gamma^0\Psi_{\rm I}$ and $\bar{\Psi}_{\rm II}\gamma^0\Psi_{\rm II}$. Under bosonization they become $[\phi_i'(\infty)-\phi_i'(-\infty)]/\sqrt{\pi}$, $i=$ I or II. Given that ϕ'_+ and ϕ'_- change by $\pm 2\pi/(8\pi^-)^{1/2}$ between $x=-\infty$ and $+\infty$ in the kink states, and the relation of these fields to $\phi_{\rm I}$ and $\phi_{\rm II}$, we see that the latter change by $\pm\sqrt{\pi}/2$. Thus the solitons in ϕ'_+ and ϕ'_- have charges $(\pm\frac{1}{2},\pm\frac{1}{2})$. That is, they are isospinors. [In terms of the generators Q_R and Q_L of $O(4)=SU(2)_R\otimes SU(2)_L$ given by $Q_{R/L}=Q_{12}\pm Q_{34}$, the kinks of ϕ_+ have charges $(\pm\frac{1}{2},0)$ while the kinks of ϕ_- have charges $(0,\pm\frac{1}{2})$. It is in this remarkable way that the two decoupled SG equations at $\beta^2=8\pi^-$ maintain the

O(4) symmetry.] Let us now note that for the changes in $\phi_{I/II}$ mentioned above $\overline{\Psi}\Psi = \{\cos[(4\pi)^{1/2}\phi_I] + \cos[(4\pi)^{1/2}\phi_{II}]\}$ changes sign between $x = \pm\infty$. *Thus isospinors are solitons connecting the two broken symmetric vacua.* Their existence which was shown above, therefore, implies chiral-symmetry breaking. Witten[8] has shown that for any N isospinors correspond to kinks linking vacua with opposite values for $\langle \overline{\Psi}\Psi \rangle$.

The decoupling of the $N=4$ GNM into two SG equations translates into the decoupling of the $N=4$ ATM into two $N=2$ ATM's. To show this one fermionizes the SG equations back to massive $N=2$ GNM's. One finds the following parameters λ_2 and g_2 in the $N=2$ models as $g_0 \to 0$:

$$\lambda_2 - 1 \simeq cg_0, \quad g_2 \simeq -\pi/32 + g_0/2. \tag{17}$$

Whereas the β function near $g_2 = 0$ vanished identically, it is nonzero near $-\pi/32$ since g_2 is tied to g_0. All of this reflects the fact that the anomalous dimension of $\cos\beta\phi$ in the SG is 2 as $\beta^2 \to 8\pi$. Grest and Widom raised this possibility, but for $g_0 < 0$.

The case N=3.—At the Lagrangean level Witten has shown[8] that the model has an unexpected supersymmetry, by bosonizing two of the three Ψ's. I will argue that chiral symmetry is broken while supersymmetry is not. This implies that along the first-order line the leading singularity in the free energy is 0 as g_0 tends to 0. (This point is quite subtle and will be discussed in a longer version of this paper.) The correlation length, of course, diverges like $\exp(\text{const}/g_0)$. There are other such examples, like the ones considered by Domany[12] or Crombrugge and Rittenberg,[13] though only in the latter is supersymmetry the obvious cause.

I will rely on the S-matrix calculations performed on the GNM to establish the above results as well as to argue that chiral symmetry is broken for all higher N. Now the most satisfactory way to get the S matrix would be to solve H by something like the Bethe *Ansatz* and get the S matrix from the in and out states. While all of this is possible for the SG or chiral Gross-Neveu (GN) models[14] it has not been possible here. There is an alternative originated by Zamolodchikov and Zamolodchikov[10] and formalized by Karowski *et al.*[15] for getting just the S matrix. (The S matrix obtained this way for the SG and chiral GN models agrees with the Bethe-*Ansatz* method. We also saw how well it did at $\beta^2 8\pi^-$ in connection with the $N=4$ case). In the GNM it goes as follows. One assumes along with Zamolodchikov and Zamolodchikov (ZZ)[16] that the model has an infinite number of conservation laws. (After their work, Witten[8] explicitly constructed one of these nontrivial, local, conserved charges. I have argued[17] that given one such charge, an infinite number follow.) One then assumes an isovector multiplet of *massive* fermions. (That is, one assumes chiral-symmetry breaking.) One looks for an elastic two-body S matrix with O(N) symmetry, crossing symmetries, analyticity, unitarity, and factorizability (the condition for the N-body S matrix to be the product of two-body S matrices). One finds only two choices: with and without bound states. One chooses the former and assigns the other to the O(N) Heisenberg model. From the study of the isovector with its bound states and so on, ZZ extract the following formula for the complete mass spectrum at any N:

$$M_n = 2M\sin(n\pi/N-2),$$
$$n = 1, 2, \ldots, (N-1)/2. \tag{18}$$

At each n there are O(N) antisymmetric tensors of rank n, $n-2, \ldots, 1$, or 0, the states of even (odd) n being bosons (fermions). The last two are the isovector and scalar, respectively. In the SG case where such a formula occurs [with $\pi/N-2$ replaced by $(\beta^2/16)/(1-\beta^2/8\pi)$] M is the kink mass and M_n are bound-state masses. Here too, M is the mass of the isospinor kinks connecting vacua with opposite values of $\langle \overline{\Psi}\Psi \rangle$. This was established by Shankar and Witten[18] for even N by considering the isospinor S matrix, looking for the pole structure, and regaining Eq. (18). (For $N=4$ the full S matrix was found.)

Having found an S matrix this way one must ask if it really is the S matrix for the GNM. Here is the evidence: At large N the spectrum coincides with the WKB result of Dashen, Hasslacher, and Neveu.[19] If one follows their guess (based on a SG analogy) and replaces N by $N-2$ in the large-N formulas one gets Eq. (18). Next, ZZ have verified that the S matrix expanded out to lowest nontrivial order in $1/N$ agrees with the field theory calculation in the same limit. Coming down in N, I have shown[20] by manipulating the Lagrangean that at $N=8$ the $n=1$ isovector and isospinor are degenerate as a result of a triality symmetry. This agrees with Eq. (18) at $N=8$, $n=1$. At $N=4$, the S matrix found with Witten exhibits the decoupling between the left and right isospinors that we saw at the Lagrangean level. Also M_1, the mass of the $n=1$ isovector which would be a right-left bound state must now be $2M$ since they have decoupled. One finds that this is so in Eq. (18). Thus we see the S matrix is doing well down to $N=4$ and capturing any peculiarities that arise at special values of N. This confirms among other things chiral-symmetry breaking which was assumed in the derivation.

In view of this overwhelming evidence we are driven to believe that the *mass formula,* Eq. (18), is correct for all $N > 2$. [At $N=2$ it obviously shows a pathology in the factor $(N-2)^{-1}$.] This implies that supersymmetry is unbroken at $N=3$ since Eq. (18) precludes a massless isoscalar Goldstone fermion that would accompany symmetry breakdown. The sole particles are the isospinors and their presence is a testimony to chiral-symmetry breakdown. (If we turn to the SG model and recall the correspondence between N and β that was mentioned earlier we see that $N > 4$ corresponds to the region $\beta^2 < 4\pi$ where the s and $\bar{s}$ attract and form bound states, while the range $4\pi < \beta^2 < 8\pi$ containing only s and $\bar{s}$ corresponds to $2 < N < 4$.)

The mass formula tells us something about the very S-matrix calculations that led to it. Notice that as we lower N, the number of bound states decreases and at $N = 4$ even the GN particle corresponding to Ψ is absent. If one blindly continues the isovector-isovector S matrix to $N = 3$ one sees pathologies. If one knows that M in Eq. (18) is the isospinor mass one sees that the isovector has become unstable to decay into isospinors.[21] This is what led Witten and myself to work on the isospinor S matrices for small N. The mass formula is, however, stable under the exit of the isovector (or any higher-rank tensor when it becomes unstable) because it can be derived from the isospinors. This is why we believe it at $N = 3$ though no one has explicitly determined it at $N = 3$. It spells trouble only at $N = 2$ where the very notion of an isospinor breaks down. [I have pointed out in Ref. 17 the intimate connection between this model at any N and the group O(N). In particular, the decoupling at $N = 4$ corresponds to the factorization of O(4) into SU(2) ⊗ SU(2) and the triality at $N = 8$ corresponds to the triality symmetry of the O(8) Dynkin diagram.]

To conclude, I have used the fact that in the vicinity of the pure Ising transition the ATM may be related to the massive Gross-Neveu model and that the temperature (measured from the pure Ising transition) and energy-energy coupling of the former become the bare mass and g_0 of the latter. In the field theory the order parameter is $\bar{\Psi}\Psi$; the mass plays the role of the external ordering field. Thus the transition is first order if there is a nonzero $\bar{\Psi}\Psi$ in the massless GNM, i.e., chiral-symmetry breaking. Since the massless GNM is integrable, a lot is known about it. Using the available information I have argued that the first-order transition seen at large N persists down to $N = 3$ and that at $N = 3$ the leading term in the free energy vanishes along the first-order line right up to the pure Ising point. Conversely, it is satisfying to know where the GNM comes from.[22]

I am grateful to numerous colleagues, especially M. Kardar, for discussions. This research was funded by the Alfred P. Sloan Foundation and U.S. Department of Energy Grant No. DE-AC 02 76 ERO 3075.

[1]J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).

[2]E. Fradkin, Phys. Rev. Lett. **53**, 1967 (1984).

[3]F. Wegner, J. Phys. C **5**, L131 (1972); F. Y. Wu and Y. K. Wang, J. Math. Phys. (N.Y.) **17**, 439 (1976); R. J. Baxter, Ann. Phys. (N.Y.) **70**, 193 (1972).

[4]G. Crest and M. Widom, Phys. Rev. B **24**, 6508 (1981).

[5]E. Fradkin and L. Susskind, Phys. Rev. D **17**, 2637 (1978).

[6]D. Gross and A. Neveu, Phys. Rev. D **10**, 3235 (1974).

[7]L. P. Kadanoff and A. C. Brown, Ann. Phys. (N.Y.) **121**, 318 (1979); J. R. D. de Felicio and R. Koberle, Phys. Rev. B **25**, 511 (1982).

[8]E. Witten, Nucl. Phys. B **142**, 285 (1978).

[9]S. Coleman, Phys. Rev. D **11**, 2088 (1975).

[10]A. B. Zamolodchilov and Al. B. Zamolodchikov, Ann. Phys. (N.Y.) **120**, 253 (1979).

[11]H. Bergnoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979); L. A. Takhtajan and L. D. Fadeev, Theor. Math. Phys. **21**, 1046 (1975).

[12]E. Domany, Phys. Rev. Lett. **52**, 871 (1984).

[13]M. de Crombrugge and V. Rittenberg, Ann. Phys. (N.Y.) **151**, 99 (1983).

[14]B. Berg and P. Weisz, Nucl. Phys. B **146**, 205 (1979); V. Kurak and J. A. Swieca, Phys. Lett. **82B**, 289 (1979); N. Andrei and J. Lowenstein, Phys. Lett. **91B**, 401 (1980).

[15]M. Karowski, H. J. Thun, T. T. Troung, and P. Weisz, Phys. Lett. **67B**, 321 (1977).

[16]A. B. Zamolodchilov and Al. B. Zamolodchikov, Nucl. Phys. **133**, 525 (1978).

[17]R. Shankar, Phys. Lett. **102B**, 257 (1981).

[18]R. Shankar and W. Witten, Nucl. Phys. **B141**, 349 (1978).

[19]R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **12**, 2443 (1975).

[20]R. Shankar, Phys. Rev. Lett. **46**, 379 (1981).

[21]A. B. Zamolodchikov, private communication.

[22]See also M. Ogilvie, J. Phys. A **14**, L519 (1981), who derives the bosonized form of the GNM at even N from a generalized Coulomb gas.