Journal of Statistical Physics, Vol. 29, No. 4, 1982

On the Solution of Some Vertex Models Using Factorizable S Matrices

R. Shankar[1]

Received July 7, 1982

The intimate connection between factorizable S matrices and some vertex models in two dimensions (to be reviewed here) is exploited to show that the knowledge of the S matrix not only allows us to define a solvable vertex model *à la* Zamolodchikov, but often to write down the free energy by inspection. The prototype for discussion is Baxter's eight-vertex model generated by Zamolodchikov's Z_4 S matrix. The method is then applied to a hitherto unsolved 19-vertex model, based on the isospin-1 S matrix of Zamolidchikov and Fateev, and agreement is checked to fourth order in a perturbation series. The possibility of molding other problems like the q-state Potts model into this framework is considered.

KEY WORDS: Factorizable S matrices; Baxter model; Yang–Baxter relations; Lee–Yang and Suzuki–Fisher theorems; Potts model; Temperley–Lieb equivalence.

1. INTRODUCTION

In this article I discuss in some detail a problem that was handled rather succinctly in a letter.[1] It concerns the intimate relation between two seemingly unrelated problems: the determination of factorizable S matrices in 1 + 1 dimensions and the solution of certain vertex models of statistical mechanics on a two-dimensional lattice. (I hasten to assure readers familiar with either or neither problem that both will be reviewed here.)

Research supported in part by NSF grant No. INT 8117361.

[1] Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, Laboratoire Propre du Centre National de la Recherche Scientifique, associé à l'Ecole Normale Supérieure et à l'Université de Paris Sud, Postal address: 24, rue Lhomond, 75231 Paris Cedex 05; and J. W. Gibbs Laboratory, Yale University, New Haven, Connecticut 06520 U.S.A.

0022-4715/82/1200-0649$03.00/0

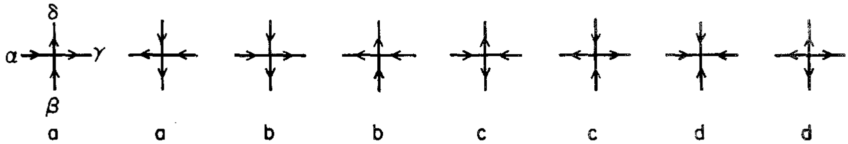

Fig. 1. The allowed vertices. Vertices related by reversal of all arrows are given the same weight. In the S-matrix context these correspond to the processes $\alpha(\Theta/2)+\beta(-\Theta/2) \to \gamma(\Theta/2)+\delta(-\Theta/2)$.

Before plunging into any details let us understand schematically the nature of these problems and their interrelationship, taking as a concrete example the Z_4 S matrix[2] of A. B. Zamolodchikov (A. B. Z.) and Baxter's eight-vertex model.[3] Baxter considers a lattice, on the bonds of which are placed arrows pointing up or down or to the right or left. Let us mean by a vertex at any site, the state of the arrows on the four bonds attached to it. Of the $2^4 = 16$ possible vertices, let us allow just the 8 shown in Fig. 1 and assign Boltzmann weights a, b, c, and d to vertices 1, 3, 5, and 7 and also to 2, 4, 6, 8. The sum over all allowed configurations of the lattice defines a partition function Z. The partition function per site $z = Z^{1/N^2}$ (on an $N \times N$ lattice) was found by Baxter, in the limit $N \to \infty$. For a nice review of vertex models see Lieb and Wu.[4]

Zamolodchikov considers the determination of an S matrix for particles of charge $Q = \pm 1$, the latter being conserved modulo 4 (hence the name Z_4). This he does, not starting with some Lagrangian, but from "general" principles which provide functional equations for the S-matrix elements. He *postulates* elasticity and *factorizability* (i.e., the N-body S matrix must be a product of $N(N-1)/2$ two-body S matrices). There are eight two-body amplitudes, in one-to-one correspondence with the vertices in Fig. 1 if we let the south and west (or north and east) arrows denote the values of Q for the incoming (or outgoing) particles. Imposing invariance under charge reversal, we are left with four amplitudes which we call S_a, S_b, S_c, and S_d in obvious notation, i.e.,

$$S = \begin{array}{c} \\ ++ \\ -- \\ +- \\ -+ \end{array} \begin{array}{c} \begin{array}{cccc} ++ & -- & +- & -+ \end{array} \\ \left[\begin{array}{cccc} S_a & S_d & 0 & 0 \\ S_d & S_a & 0 & 0 \\ 0 & 0 & S_b & S_c \\ 0 & 0 & S_c & S_b \end{array}\right] \end{array} \tag{1.1}$$

The problem is to find the S_i. One can show that these are meromorphic in

Θ, the relative rapidity (more on this later). A. B. Z. shows that factorizability is self-consistent only if at each Θ

$$S_a : S_b : S_c : S_d = \frac{\operatorname{sn}[2\eta(1 + (i\Theta/\pi)), k]}{\operatorname{sn}(2\eta, k)} : \frac{-\operatorname{sn}(2\eta i\Theta/\pi, k)}{\operatorname{sn}(2\eta, k)}$$
$$: 1 : k \operatorname{sn}(2\eta i\Theta/\pi, k)\operatorname{sn}[2\eta(1 + i\Theta/\pi), k] \qquad (1.2)$$

where η and k are free parameters (corresponding to coupling constants of the underlying field theory, whatever it may be) and $\operatorname{sn}(z, k)$ are Jacobian elliptic functions of argument z and modulus k.[5] Thus only one S_i, say S_c, needs to be found. The unitarity equation

$$S(\Theta)S^T(-\Theta) = I \qquad (1.3)$$

and the crossing equation (more on this later)

$$S_c(\Theta) = S_c(i\pi - \Theta) \qquad (1.4)$$

provide two functional equations for $S_c(\Theta)$. A. B. Z. found a unique "minimal" solution (with restrictions on its poles and zeros). Many S-matrices have been found this way and eventually linked to some underlying field theory like the sine-Gordon model.[6]

Now for the connection between the two problems. The first, due to Zamolodchikov, is that if one defines an eight-vertex model with weights obeying

$$a : b : c : d = S_a : S_b : S_c : S_d \qquad (1.5)$$

then it is solvable in the sense that

$$[T(\Theta, \eta, k), T(\Theta', \eta, k)] = 0 \qquad (1.6)$$

T being the transfer matrix. This would not surprise readers familiar with Baxter's work. But A. B. Z. shows that the result is general[2]: *any* vertex model with vertex weights in the same ratio as the elements of any factorizable two-body S matrix [as in Eq. (1.5)] is solvable in the sense of Eq. (1.6). Besides this result, firmly established in Ref. 2, A. B. Z. also noticed another remarkable coincidence in the Z_4 eight-vertex case: if one considers the eight-vertex model with weights equal to (and not just proportional to) the S_i themselves, then $z(S) = 1$ for a range of parameters called the principal region or PR ($c > a + b + d$, all positive). In other words, if one takes Baxter's formula for $z(a, b, c, d)$ which has the form (in the PR)

$$z(a, b, c, d) = c/f(\Theta, \eta, k)$$

one finds that $f(\Theta,\eta,k)$ *is just* $S_c(\Theta,\eta,k)$. Consequently

$$\begin{aligned} z(S_a,S_b,S_c,S_d) &= S_c z(S_a/S_c, S_b/S_c, 1, S_d/S_c) \\ &= S_c z(a/c, b/c, 1, d/c) \\ &= \frac{S_c}{c} z(a,b,c,d) \\ &= \frac{S_c}{c}\cdot\frac{c}{f} = \frac{S_c}{f} = 1 \end{aligned} \tag{1.7}$$

where I have repeatedly used the fact that (i) if *all* the weights are multiplied by a factor ρ, $Z \to \rho^{N^2} Z$ and $z \to \rho z$; (ii) Eq. (1.5) implies that $a/c = S_a/S_c$, $b/c = S_b/S_c$, etc.

The present investigation was performed to answer the following questions:

(i) Is it possible to *explain* the coincidence pointed out by Zamolodchikov, i.e., to *derive* the result $z(S) = 1$ (in the PR)?

(ii) Is this a general phenomenon, i.e., is $z(S) = 1$ (in the corresponding PR) for other vertex models defined by other S matrices in the same way?

The answer is affirmative, granted some assumptions. Before going into these or the derivation, let us recognize the importance of the result $z(S) = 1$: it implies that not only does the knowledge of S allow us to define a *solvable* vertex model as A. B. Z. observed, *it also allows us to write down the answer at once*, in the PR. This is done by referring to the equalities leading up to Eq. (1.7): if $a, b, c, d, e, \ldots$ are the vertex weights in the same ratio as S-matrix amplitudes $S_a, S_b, S_c, S_d, S_e, \ldots$ etc.; then

$$\begin{aligned} z(a,b,c,\ldots) &= \frac{c}{S_c} z(S_a, S_b, S_c, \ldots) \\ &= \frac{c}{S_c} = \frac{a}{S_a} = \frac{b}{S_b} \cdots \end{aligned} \tag{1.8}$$

There is another way to state this result. Consider a vertex model with weights $\omega = \{a, b, \ldots\}$ with a given ratio between two weights (i.e., a given Θ, η, k in the eight-vertex case). This is a one-parameter family of weights whose *overall* scale ρ is free. It is clear that in this family there is a member $\bar{\omega}$ such that

$$z(\bar{\omega}) = 1$$

This is because rescaling ω rescales z and we can make it 1. In fact

$$\bar{\omega} = \omega / z(\omega) \tag{1.9}$$

where ω is any member of the family and $z(\omega)$ the corresponding z; for

$$z(\bar{\omega}) = z(\omega/z(\omega)) = [1/z(\omega)]z(\omega) = 1 \tag{1.10}$$

Let us call $\bar{\omega}$ the *normalized weights*. For each given ratio (i.e., Θ, η, k), we will of course need a different $\bar{\omega}$, i.e., $\bar{\omega} = \bar{\omega}(\Theta, \eta, k)$. Clearly knowing $\bar{\omega}$ is knowing $z(\omega)$ [from Eq. (1.9)]. Now any other way to calculate $\bar{\omega}$ [other than solving for $z(\omega)$ and using Eq. (1.9)] is in effect another way to solve the problem. This is exactly what $z(S) = 1$ gives us—it says $\bar{\omega}$ is just S! Thus to find $\bar{\omega}$, we assemble the weights ω into a matrix $\omega_{\alpha\beta}^{\gamma\delta}$ (where α, β and γ, δ label the west–south and north–east bonds of the vertex, see Fig. 1) and choose the overall scale from the unitarity equation

$$\omega(\Theta)\omega^T(-\Theta) = I \tag{1.11}$$

The thrust of our derivation consists of showing that the "1" in the unitarity equation of the matrix of weights propagates all the way to the end and gives the "1" in $z(S) = 1$ in the PR. (Since z is a piecewise analytic function, rescaling the weights by an analytic function can make it unity only in some region, here the PR.). This is done by defining a function z_B which equals z in the PR (but not everywhere) and deriving certain functional equations for it. These then are used to show $z_B \equiv 1$ given some analytic properties A1 and A2. A2 states that $\tilde{\Lambda}_B(\Theta)$, a certain eigenvalue of $T(\Theta)$, is zero-free in a strip in the Θ-plane. In the eight-vertex case some progress has been made in this problem by showing that $\tilde{\Lambda}_B(\Theta)$ is itself the partition function of a one-dimensional Ising model in a magnetic field. For some special cases the Lee–Yang[8] and Suzuki–Fisher[9] theorems can be invoked to show that the zeros do not invade this strip. But the problem is unfinished. Thus the derivation lacks, at present, the rigor of the Bethe ansatz[4] or quantum inverse scattering methods.[10] It is more akin to the works of Straganov, Schultz, Perk or Baxter.[11–14] Indeed Baxter[14] solved the eight-vertex model (among others) this way, given similar assumptions. The present work is still useful because

(i) it clarifies the relation to the S-matrix problem;

(ii) it provides functional equations that are valid for other S-matrix-based models because they are derived using general principles like unitarity, positivity of weights, etc.;

(iii) it proves some of the assumptions made in earlier treatments.[10–14]

In the next section, the Z_4 eight-vertex problems will be discussed further and the way paved for Section 3, wherein the result $z(S) = 1$ (in the PR) is derived, given the assumptions A1 and A2. Partial results on the proof of A2 follow in Section 4. The generality of the result $z(S) = 1$ is then tested in Section 5 for the 19-vertex model based on the isospin-1 S matrix of Zamolodchikov and Fateev[15] which has the desired features like positivity of weights. The test is successful to the fourth order in a low-temperature expansion. Some concluding remarks and discussion of other models like the Potts model follow in the last section.

2. REVIEW OF THE BAXTER MODEL AND THE Z_4 S MATRIX

2.1. Baxter Model

Consider an $N \times N$ lattice on the bonds of which exist arrows which can point up or down or to the left or right. Demand that an even number of arrows point into each vertex. The eight allowed vertices shown in Fig. 1 are assigned energies $\epsilon_1, \ldots, \epsilon_8$. The partition function is

$$Z(N^2) = \sum \exp\left(-\beta \sum_{i=1}^{8} N_i \epsilon_i \right) \tag{2.1}$$

where N_i is the number of vertices of type i and the sum is over all allowed configurations. We also impose toroidal boundary conditions, i.e., periodicity in both directions. In the symmetric eight-vertex model that we consider, vertices related by reversal of all arrows are given the same Boltzmann weight. This leaves us with four independent weights—a, b, c, and d, associated with vertices 1, 3, 5, and 7 (or 2, 4, 6, 8), respectively.

In terms of

$$\omega_{1,2} = \frac{c \pm d}{2}, \qquad \omega_{3,4} = \frac{a \pm b}{2} \tag{2.2}$$

Fan and Wu[16] have shown that

$$Z(\omega_1, \omega_2, \omega_3, \omega_4) = Z(\pm\omega_i, \pm\omega_j, \pm\omega_k, \pm\omega_l) \tag{2.3}$$

where (i, j, k, l) is any permutation of $(1, 2, 3, 4)$. Thus it suffices to know Z in the region

$$\omega_1 > \omega_2 > \omega_3 > \omega_4 \geqslant 0 \tag{2.4}$$

Baxter evaluated[3]

$$z = \lim_{N \to \infty} \left[Z(N^2) \right]^{1/N^2} \tag{2.5}$$

in the *principal regime* (PR)

$$c > a + b + d \qquad \text{(all positive)} \tag{2.6}$$

which corresponds to replacing ω_4 by $|\omega_4|$ in Eq. (2.4).

The central entity in his approach was the transfer matrix T. Consider

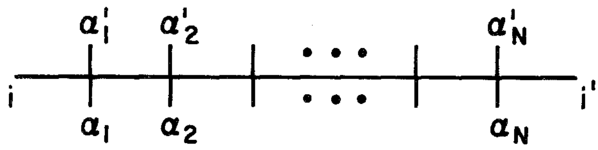

Fig. 2. Two successive rows of vertical bonds and the intervening row of horizontal bonds. In the S-matrix context the figure describes the matrix element $M_{i\alpha}^{i'\alpha'}$ for scattering between one right mover of rapidity Θ and N static targets.

two successive rows of vertical bonds, whose arrow states are labeled $\alpha = (\alpha_1, \ldots, \alpha_N)$ and $\alpha' = (\alpha'_1, \ldots, \alpha'_N)$ (see Fig. 2) and the intervening row of horizontal bonds. The elements of T are

$$T_{\alpha\alpha'} = \sum \exp\left(-\beta \sum_{j=1}^{8} n_j \epsilon_j \right) \tag{2.7}$$

where the sum is over allowed horizontal arrow configurations with periodic boundary conditions ($i = i'$ in Fig. 2) and n_j is the number of vertices of type j. It is easy to see that

$$Z(N^2) = \operatorname{Tr} T^N \tag{2.8}$$

where Tr is taken in the 2^N-dimensional space of vertical arrow states.

In Baxter's approach, as in all subsequent ones, one asks when

$$\left[T(a,b,c,d), T(a',b',c',d') \right] = 0$$

He found that if the ratios of weights are parametrized as follows

$$a : b : c : d = \frac{\operatorname{sn}\left[2\eta(1 + (i\Theta/\pi)), k\right]}{\operatorname{sn}(2\eta, k)} : \frac{-\operatorname{sn}\left[(2\eta i\Theta/\pi), k\right]}{\operatorname{sn}(2\eta, k)} :$$

$$: 1 : k \operatorname{sn}\left[2\eta(i\Theta/\pi), k\right] \cdot \operatorname{sn}\left[2\eta(1 + (i\Theta/\pi)), k\right] \tag{2.9}$$

then

$$\left[T(\Theta, \eta, k), T(\Theta', \eta, k) \right] = 0 \tag{2.10}$$

(I use weights which can be simultaneously all positive, i.e., as in Ref. 3b.)

In Eq. (2.10), $\operatorname{sn}(z,k)$ is the Jacobian elliptic function of argument z and modulus k[5]; Θ, η, and k are free parameters. This parametrization is in no way restrictive; to any given ratio of weights, there exists a choice of Θ, η, and k.[3] The PR corresponds to

$$0 < k < 1 \tag{2.11a}$$

$$\Theta \text{ imaginary}, \qquad 0 < \operatorname{Im} \Theta < \pi \tag{2.11b}$$

$$\eta \text{ imaginary}, \qquad 0 < \operatorname{Im} \eta < K'/2 \tag{2.11c}$$

where K' is the complete elliptic integral of modulus $k' = (1 - k^2)^{1/2}$. (For k in the above range, sn has a complex period $2iK'$ and a real period $4K$, K being the complete elliptic integral of modulus k.)

Baxter actually used a variable

$$v = \eta(1 + (2i\Theta/\pi)) \tag{2.12}$$

We have switched to Θ in view of things to come.

From the fact that

$$T(a,b,c,d,) = T^{\dagger}(a,b,d,c) \tag{2.13}$$

and that $c \leftrightarrow d$ does not alter η or k,[17] it follows that T is *normal*:

$$[T, T^{\dagger}] = 0 \tag{2.14}$$

Thus there exists a Θ-*independent* basis of eigenvectors $|\Lambda_i\rangle$ and the eigenvalues

$$\Lambda_i(\Theta) = \langle \Lambda_i | T(\Theta) | \Lambda_i \rangle \tag{2.15}$$

enjoy the same analyticity (*in* Θ) *as the weights*. This information was crucial to Baxter, whose weights were entire and will be so here when meromorphic weights are introduced.

From Eq. (2.8) we see that

$$Z(\Theta, N^2) = \sum_{i=1}^{2^N} \Lambda_i(\Theta)^N \tag{2.16}$$

and

$$z(\Theta) = \lim_{N\to\infty} \left[Z(\Theta, N^2)\right]^{1/N^2} \to \lim_{N\to\infty} \left[\Lambda_B(\Theta, N)\right]^{1/N} \tag{2.17}$$

where Λ_B is the dominant (in modulus) among the eigenvalues Λ_i of T.

From now on we reserve the symbol $\Lambda_B(\Theta)$ *for the eigenvalue which dominates in the* PR. Baxter calculated it and obtained the following expression for z *in the* PR.

$$z(a,b,c,d)_{\Theta,\eta,k} = \frac{c}{S_c(\Theta, \eta, k)} \tag{2.18a}$$

where

$$S_c(\theta) = \exp\left\{4 \sum_{n=1}^{\infty} \frac{\sinh^2\left[\dfrac{2\pi n(\pi - \gamma)}{\gamma'}\right] \sin\left(\dfrac{2\pi n\theta}{\gamma'}\right) \sin\left[\dfrac{2\pi n(i\pi - \theta)}{\gamma'}\right]}{n \sinh\left(\dfrac{4\pi n\gamma}{\gamma'}\right) \cosh\left(\dfrac{2\pi^2 n}{\gamma'}\right)}\right\}, \tag{2.18b}$$

and

$$\gamma = \frac{i\pi K'}{2\eta}, \qquad \gamma' = \frac{2\pi i K}{\eta} \tag{2.18c}$$

2.2. The Z_4 S Matrix

Consider scattering in $1 + 1$ dimensions between particles and antiparticles whose charge is conserved modulo 4 so that a process like particle + particle $\to$ antiparticle + antiparticle is allowed. [If ψ is the field operator associated with the particle, invariance under $\psi \to U\psi$, where $U = \exp(2\pi i/N)$, corresponds to Z_n symmetry and charge conservation modulo N.] Assume that there exists a conserved tensor $T_{\mu\nu}$ which is the integral of some *local* density. (This is equivalent to assuming an infinite family of conservation laws.[18]) Consider a process

$$\alpha(p) + \beta(q) \to \gamma(r) + \delta(l) + \cdots$$

where α, β, ... label charge states and p, q, ... label the two-momenta. Translation invariance leads to the conservation of the operator $P^\mu = (H, P)$ (the energy-momentum vector) and the constraint

$$p^\mu + q^\mu = r^\mu + l^\mu + \cdots \tag{2.19a}$$

while conservation of $T^{\mu\nu}$ requires that

$$p^\mu p^\nu + q^\mu q^\nu = r^\mu r^\nu + l^\mu l^\nu + \cdots \tag{2.19b}$$

It is easily verified that in a $2 \to n$ process, with $n > 2$, these equations can be satisfied only at an isolated set of points. Analyticity of the S matrix then requires that $n = 2$.[19] In this case $p = r$ and $q = l$ and the S matrix is diagonal in the momenta and nontrivial in charge space:

$$S_{pq,\alpha\beta}^{rl,\gamma\delta} = 2p^0 2q^0 \delta(p^1 - r^1)\delta(q^1 - l^1) S_{\alpha\beta}^{\gamma\delta}(s, u) \tag{2.20}$$

where

$$s = (p_\mu + q_\mu)(p^\mu + q^\mu) \tag{2.21a}$$

$$u = (p_\mu - q_\mu)(p^\mu - q^\mu) \tag{2.21b}$$

are Mandelstam's invariants subject to the relation $s + u = 4m^2$, m being the particle mass. Hereafter we drop the δ functions and S shall mean $S_{\alpha\beta}^{\gamma\delta}$. There are eight nonzero elements which we can identify with the vertices in Fig. 1 in an obvious way. Imposing invariance under charge reversal, we get four amplitudes S_a, S_b, S_c, and S_d defined earlier [see Eq. (1.1)].

Now S has the usual decomposition

$$S = I + \frac{i}{\left[s(4m^2 - s)\right]^{1/2}} T \tag{2.22}$$

where I is the unit matrix in charge space, T is the reaction matrix, and the extra factor in front of T is due to rewriting $\delta^2(p^\mu + q^\mu - r^\mu - l^\mu)$ in terms of $2p^0 2q^0 \delta(p^1 - q^1)\delta(r^1 - s^1)$.

Zamolodchikov chooses to work with the rapidity Θ for reasons that will become apparent. Recall that the momentum p^μ of a particle of mass m can be written as

$$p^\mu = (E, p) = m(\cosh\Theta, \sinh\Theta) \tag{2.23a}$$

so that $p^\mu p_\mu = m^2$ is identically satisfied. For small values Θ becomes just the velocity, $v = p/E$, while in general $\Theta = \tanh v$, as a result of which Θ transforms *additively* under Lorentz boosts. A function of rapidity differences is thus Lorentz invariant and conversely.

For the two-body process

$$\alpha(p) + \beta(q) \to \gamma(p) + \delta(q) \tag{2.23b}$$

let us introduce the *CM* rapidities $\pm\Theta/2$ as follows:

$$\alpha(\Theta/2) + \beta(-\Theta/2) \to \gamma(\Theta/2) + \delta(-\Theta/2) \tag{2.23c}$$

In terms of Θ,

$$s = 4m^2 \cosh^2(\Theta/2) \tag{2.24a}$$

$$u = -4m^2 \sinh^2(\Theta/2) \tag{2.24b}$$

and

$$S_{\alpha\beta}^{\gamma\delta} = {}_{\text{out}}\langle \delta(-\Theta/2), \gamma(\Theta/2) \,|\, \alpha(\Theta/2), \beta(-\Theta/2)\rangle_{\text{in}} \tag{2.25}$$

is required to be a function of $(\Theta/2) - (-\Theta/2) = \Theta$.

The *crossing principle* tells us that if the *energy* and momentum of particles α and γ are reversed in Eq. (2.23), we get the amplitude for the *crossed reaction*

$$\bar{\gamma}(\Theta/2) + \beta(-\Theta/2) \to \bar{\alpha}(\Theta/2) + \delta(-\Theta/2)$$

where $\bar{\gamma}$ and $\bar{\alpha}$ denote antiparticles of γ and α. Since crossing corresponds [see Eqs. (2.21) and (2.24)] to

$$\Theta \leftrightarrow i\pi - \Theta \tag{2.26a}$$

or

$$s \leftrightarrow u \tag{2.26b}$$

we get

$$S_{\alpha\beta}^{\gamma\delta}(i\pi - \Theta) = S_{\bar{\gamma}\beta}^{\bar{\alpha}\delta}(\Theta) \tag{2.27}$$

or more explicitly

$$S_a(i\pi - \Theta) = S_b(\Theta) \tag{2.28a}$$

$$S_c(i\pi - \Theta) = S_c(\Theta) \tag{2.28b}$$

$$S_d(i\pi - \Theta) = S_d(\Theta) \tag{2.28c}$$

In the region for physical scattering, Θ real and positive or s real and $> 4m^2$, we have the *unitarity equation*

$$\sum_j S_{ij}(\Theta) S_{jk}^{\dagger}(\Theta) = \delta_{ik} \tag{2.29}$$

From this we may deduce the "optical theorem" for T, which tells us T has an imaginary part proportional to the "cross section." Since there is no cross section below the elastic threshold $s = 4m^2$, T is real below threshold down to $s = 0$, where it becomes complex again because the crossed reaction opens up here ($u = 4m^2$). If there is a bound state of mass M, it will produce an imaginary part proportional to $\delta(s - m^2)$. From Eq. (2.22) we see that S is then real between $s = 0$ and $s = 4m^2$ except at poles. The Schwartz reflection principle tells us that $S(s^*) = S^*(s)$ or in terms of Θ,

$$S(-\Theta^*) = S^*(\Theta) \tag{2.30}$$

which makes S real on the $\operatorname{Im}\Theta$ axis. Using this relation, we may write Eq. (2.29) as a relation between *analytic functions* that can be analytically continued for all Θ:

$$\sum_j S_{ij}(\Theta) S_{kj}(-\Theta) = \delta_{ik}$$

or

$$S(\Theta) S^T(-\Theta) = I \tag{2.31}$$

[Recall that $f^*(z)$ is generally not analytic if $f(z)$ is, for the Cauchy–Reimann conditions may be written as $\partial f/\partial z^* = 0$.]

Given $S^*(s) = S(s^*)$ and the optical theorem, we see that S has a discontinuity above $s = 4m^2$ and below $s = 0$, as we cross the real axis. Drawing cuts out to $\pm\infty$, we get the "physical sheet" shown in Fig. 3a. This sheet maps into the "physical strip," $0 < \operatorname{Im}\Theta < \pi$ in the Θ plane (Fig. 3b).

The branch points at $s = 0$ *and* $4m^2$ *are eliminated in going to* Θ.[7] [They arise because Θ and $-\Theta$, which carry complex conjugate values of S for Θ real, get mapped onto the same value of $s = 4m^2\cosh^2(\Theta/2)$.] *Thus S is expected to be meromorphic in* Θ, unless it has other singularities not

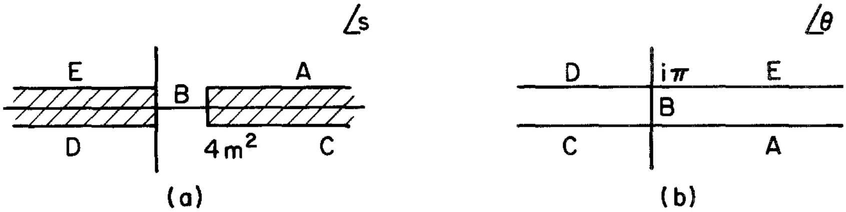

Fig. 3. The s and Θ planes. The physical sheet goes into the physical strip $0 < \text{Im}\,\Theta < \pi$. "B" is where the amplitudes are real and simultaneously positive. The mapping of several other points is also shown.

mandated by unitarity. One assumes it does not. (Had there been *inelasticity*, there would have been branch points in s at each threshold that could not be avoided in Θ.)

In addition to Eqs. (2.28), (2.30), (2.31) we need one more set of equations that will nail down $S(\Theta)$. These are called *Yang–Baxter equations*, encountered by Yang[20] in the study of the δ-function interaction and by Baxter in his quest for commuting transfer matrices. Their importance in relativistic scattering was realized by Karowski, Thun, Truong, and Weisz.[21] The intuitive derivation given below is due to Shankar and Witten.[22]

Consider a three-body collision depicted in Fig. 4a wherein the particles collide two at a time in three widely separated points in space-time. Here we expect the three-body S matrix to *factorize* into a product of

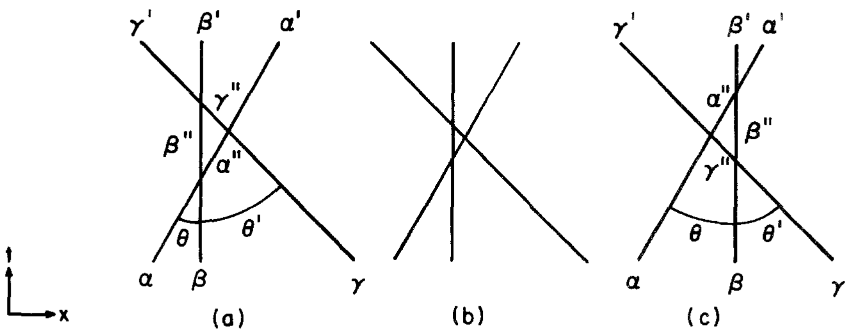

Fig. 4. (a) A scattering for which we expect factorization since the three two-body collisions are widely separated in space-time. (b) A scattering for which we do not expect factorization but obtain it, thanks to the symmetry generated by T_{11} which allows us to change the impact parameters. (c) Instead of shifting (b) into (a) we could shift it to (c). The Yang–Baxter equations are obtained by equating the amplitudes for the two cases.

on-shell two-body S matrices and to be diagonal in the momenta since each two-body collision is. But what if the impact parameters were as in Fig. 4b, with the particles headed for a collision where three-body forces are expected to be important? We shall see that elasticity and factorizability are valid even here. Consider the unitary operator $U = \exp iaT_{11}$, where T_{11} is the space–space component of the *conserved* tensor $T_{\mu\nu}$. Its action on a wave packet of mean momentum $\langle P \rangle = p_0$ is, by Lorentz invariance, that of $\exp iaP^2$, where P is the spatial component of the operator P^μ. We can show that this operator translates the packet by an amount proportional to *its mean momentum*. (A quick way is to replace one factor P by $\langle P \rangle$, i.e., $\exp iaP^2 \sim \exp ia\langle P \rangle P = \exp iap_0 P$, which translates by iap_0. A more careful stationary phase analysis gives the translation $2iap_0$.[22]) Using U, we can shift the particles in Fig. 4b *relative* to one another and get the impact parameters of Fig. 4a, without changing the amplitude. This proves elasticity and factorizability.

Suppose we moved them around instead to get the configuration in Fig. 4c? We must get the same answer, of course, and this imposes the constraint

$$S_{\beta''\gamma''}^{\beta'\gamma'}(\Theta') S_{\alpha''\gamma}^{\alpha'\gamma''}(\Theta + \Theta') S_{\alpha\beta}^{\alpha''\beta''}(\Theta) = S_{\alpha''\beta''}^{\alpha'\beta'}(\Theta) S_{\alpha\gamma''}^{\alpha''\gamma'}(\Theta + \Theta') S_{\beta\gamma}^{\beta''\gamma''}(\Theta') \tag{2.32}$$

Zamolodchikov found the solution to these overdeterminate equations that are also consistent with crossing and real analyticity. The result, amazingly, was that the amplitudes S_i be in the ratio

$$S_a : S_b : S_c : S_d = a : b : c : d \tag{2.33}$$

where $a : b : c : d$ is given in Eq. (2.9)! In the present context, η and k are free parameters associated with the underlying field theory ($\gamma - \pi$, where $\gamma = i\pi K'/2\eta$ plays the role of the coupling constant). One chooses $0 < k < 1$ and η to be purely imaginary (to satisfy real analyticity) and requires $0 < -i\eta < \frac{1}{2}K'$ (or $\gamma > \pi$) to keep complex poles off the physical strip, for these violate casuality. (My notation differs from Zamolodchikov's.[2] I have chosen ξ in his Eq. (3.1) to be the purely imaginary η, set his $l = k$, and lastly relabeled his S, S_t, S_r, and $-S_a$ as S_a, S_b, S_c, and S_d, respectively.)

The significance of Eq. (2.29) did not escape Zamolodchikov: it implied that the S_i could be used as vertex weights to define an eight-vertex model with commuting transfer matrices. Of course, the physical region for scattering (Θ real and positive) and the physical region for the statistical problem (Θ imaginary, and $0 < \operatorname{Im}\Theta < \pi$) are different. *Notice, however, that unlike the continuation off-mass shell (which is ambiguous since there is only one value of m for which "experiments" are possible), the continuation of*

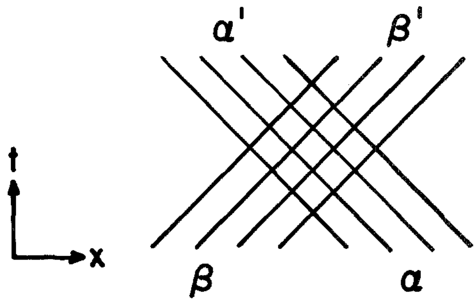

Fig. 5. The space-time lattice defined by the collision between N particles of rapidity $\Theta/2$ and N of rapidity—$\Theta/2$. The toroidally periodic partition function Z is the trace of the S matrix for this process.

the on-shell scattering function, "measurable" on the positive real Θ axis to all Θ, is unique.

Indeed, Zamolodchikov succeeded in showing[5] that *any* factorizable S matrix obeying Eq. (2.32) generates likewise a vertex model with $[T(\Theta), \mathrm{T}(\Theta')]=0$.

Another way to relate the two problems, which also appear in a more general form in Zamolodchikov's more recent publication,[23] is the following: Consider a collision (Fig. 5) between N right movers of rapidity $\Theta/2$ with N left movers of rapidity $-\Theta/2$, in charge states β and α respectively. Let (β', α') be the final state. From Fig. 5 it is apparent that

$$Z = \operatorname{Tr} \mathbf{S} \equiv \sum_{\alpha,\, \beta} \mathbf{S}^{\alpha\beta}_{\alpha\beta}(\Theta, 2N) \tag{2.34}$$

the Tr being taken in the 4^N-dimensional internal space of $\mathbf{S}$, the $2N$-body S matrix. Thus, *even though the particles move in a space-time continuum, a lattice emerges thanks to elasticity and the conservation laws in each collision.* This connection between an S-matrix and a statistical problem on a lattice is very different from the usual connection[25] between *Euclidean* field theories and *critical* statistical systems. While the present connection is interesting and serves to establish some not so obvious properties, like Baxter's Z-invariance,[24] it does not tell us how to *solve* for Z since computing the trace of $\mathbf{S}^{\alpha'\beta'}_{\alpha\beta}$, the giant $2N$-body S matrix, given the elements S_i of the two-body S matrix, amounts to solving for Z given the matrix of weights. To solve for Z, we must use a trick described in the next section.

Let us continue with the saga of two-body S matrix. Given the ratios of S_i, we need find just one, say S_c. Eliminating the others in the unitarity equation, one finds

$$S_c(\Theta)S_c(-\Theta) = \frac{\operatorname{sn}^2(2\eta, k)}{\operatorname{sn}^2(2\eta, k) - \operatorname{sn}^2(2\eta i\Theta/\pi, k)} \tag{2.35a}$$

We also have the crossing equations

$$S_c(\Theta) = S_c(i\pi - \Theta) \tag{2.35b}$$

The solution to these equations is not unique, even in the realm of meromorphic, real analytic solutions. One can multiply any given S_c by a function $\Phi(\Theta)$ of the same genre obeying

$$\begin{aligned} \Phi(\Theta)\Phi(-\Theta) &= 1 \\ \Phi(\Theta) &= \Phi(i\pi - \Theta) \end{aligned} \tag{2.36}$$

If, however, we demand that S_c be *minimal*, i.e., free of poles or zeros in the physical strip and be positive on the $\operatorname{Im}\Theta$ axis, we get $\Phi \equiv 1$. To see this, note that

$$\Phi(-\Theta) = \frac{1}{\Phi(\Theta)} = \frac{1}{\Phi(i\pi - \Theta)}$$

implies $\Phi(\Theta) = \Phi(\Theta + 2\pi i)$. Given that Φ is zero free in the physical strip implies it is pole free for $0 > \operatorname{Im}\Theta > -\pi$. On the lines $\operatorname{Im}\Theta = 0$ and π unitarity and crossing assure us Φ is unimodular. A meromorphic function free of poles in a full period is a constant by Liouville's theorem. At $\Theta = 0$ we see $\Phi = 1$. Zamolodchikov found that *the unique minimal solution, obeying all of the above conditions was just the S_c that occurred in Eq.* (2.18)! *This implied, given Eq.* (2.18), *that if $S_a, \ldots, S_d$ were used as weights,* $z = S_c/S_c = 1$ in the PR! Zamolodchikov concluded his paper with this dazzling coincidence and said an explanation ought to be found. The aim of this paper is to provide the same, i.e., to derive the result $z(S) = 1$ in the PR, to which we now turn our attention in the next section. It will be seen that much of the derivation, based on general principles like unitarity, applies to all S-matrix based models.

3. THE S-MATRIX SOLUTION

Perhaps it is best to state at the very outset that the derivation of the result $z(s) = 1$ in the PR is based on the following two assumptions.

Assumption A1. *A single eigenvalue $\Lambda_B(\Theta, \eta, k)$ dominates in the* PR. Because of the conservation of Q mod 4, T will break up into two blocks with $Q = 0$ or 2 (for N even). While Perron's theorem[26] assures us that each block, being irreducible and positive, has a dominant eigenvalue, it does not rule out a cross-over between the two. At low temperatures, $S_c \gg S_a + S_b + S_d$, the lattice is antiferroelectrically ordered (look at vertex c in Fig. 1) and the $Q=0$ sector provides the dominant eigenvalue $\Lambda_B(\Theta, \eta, k)$. We assume this is so down to $S_c = S_a + S_b + S_d$.

Assumption A2. Corresponding to the dominant eigenvalue $\Lambda_B(\Theta, \eta, k)$, there must be a nondegenerate eigenket with real positive components,[26] which we denote by $|\Lambda_B(\eta, k)\rangle$ since the eigenkets are Θ-independent. Let us define for *all* Θ the function

$$\Lambda_B(\Theta, \eta, k) = \langle \Lambda_B(\eta, k)|T(\Theta, \eta, k)|\Lambda_B(\eta, k)\rangle \tag{3.1}$$

Clearly Λ_B is meromorphic in Θ, with no poles in the strip PS′, $0 \leqslant \operatorname{Im}\Theta < \pi$. (PS′ differs slightly from the physical strip PS: $0 < \operatorname{Im}\Theta < \pi$). *We assume that* $\Lambda_B \neq 0$ *in* PS′. Some progress in proving this assumption is reported in Section 3. This constitutes a very interesting problem by itself.

Consider now a sequence of functions

$$z_B(\Theta, N) = \Lambda_B(\Theta, N)^{1/N} \tag{3.2}$$

It is clearly analytic and bounded for all N, for Θ in PS′ (and η and k in the PR). Assuming the thermodynamic limit for any dense set of points in the PR, Vitali's convergence theorem[27] assures us that

$$z_B(\Theta) = \lim_{N \to \infty} z_B(\Theta, N) \tag{3.3}$$

exists throughout the PS′, and is analytic therein. Further $z_B(\Theta) \neq 0$ in the PS′ by Hurwitz's theorem.[27]

We will now derive two functional equations for z_B:

$$z_B(\Theta) = z_B(i\pi - \Theta) \tag{3.4}$$

$$z_B(\Theta) z_B(-\Theta) = 1 \tag{3.5}$$

η and k being fixed everywhere at some value in the PR.

It now follows, as in the case of the function $\Phi(\Theta)$ encountered in Eq. (2.37), that

$$z_B(\Theta) \equiv 1 \tag{3.6}$$

[Recall that Eqs. (3.4) and (3.5) imply $z_B(\Theta) = z_B(\Theta + 2\pi i)$. Crossing tells us z_B is analytic and nonzero on $0 \leqslant \operatorname{Im}\Theta \leqslant \pi$, given the same on PS′: $0 \leqslant \operatorname{Im}\Theta < \pi$. Inversion gives us the same for $0 \geqslant \operatorname{Im}\Theta \geqslant -\pi$. Since z_B is analytic and bounded in a full period, it is constant. At $\Theta = 0$, Eq. (3.5) tells us $z_B = 1$.] Since Λ_B dominates in the PR, $z_B = z$ there, and the result $z(S) = 1$ follows in the PR.

In Ref. 1 I had attempted to derive $z(S) = 1$ without recourse to A2. I am now cogniscent of the fact that the arguments presented therein can be invalidated by the peculiarities of the $N = \infty$ limit. I am obliged to Professor Baxter for suggesting the route based on Vitali's theorem. Naturally, I am responsible for any misrepresentation.

To prove Eq. (3.4) we recall that[17] in the PR,

$$T(a,b,c,d) = T^{\dagger}(a,b,d,c) \tag{3.7}$$

Now $a \leftrightarrow b$, $c \leftrightarrow d$ does not affect T since this amounts to reversal of horizontal spins which are summed over anyway. Applying this symmetry to Eq. (3.7), we get

$$T(a,b,c,d) = T^{\dagger}(b,a,c,d) \tag{3.8}$$

Sandwiching this equation between $\langle \Lambda_B |$ and $| \Lambda_B \rangle$, which have real components, we get (for the case $a = S_a$, $b = S_b$, etc.)

$$\Lambda_B(S_a, S_b, S_c, S_d) = \Lambda_B(S_b, S_a, S_c, S_d) \tag{3.9}$$

or, in view of the crossing equations (2.28),

$$\Lambda_B(\Theta, N) = \Lambda_B(i\pi - \Theta, N) \tag{3.10}$$

This relation, established on the $\operatorname{Im}\Theta$ axis, of course holds for all Θ. Taking the Nth root of the above and letting $N \to \infty$, we get Eq. (3.4). It is important in all this to know that Λ_B has no zeros [i.e., $z_B(\Theta, N)$ is analytic] in the vicinity of $\Theta = i\pi/2$, the fixed point of the transformation $\Theta \to i\pi - \Theta$. To see what can otherwise go wrong, consider a model function

$$f(\tilde{\Theta}, N) = [2\cos N\tilde{\Theta}]^{1/N}$$

where $\tilde{\Theta} = \Theta - i\pi/2$. At every N, $f(\tilde{\Theta}, N) = f(-\tilde{\Theta}, N)$. As $N \to \infty$ however, $f(\tilde{\Theta}) = \lim_{N\to\infty} f(\tilde{\Theta}, N)$ is only *piecewise analytic*:

$$f(\tilde{\Theta}) = f_{\pm}(\tilde{\Theta}) = \exp(\mp\tilde{\Theta}) \qquad \text{for} \quad \operatorname{Im}\tilde{\Theta} \gtrless 0$$

and we have only the relation

$$f_{+}(\tilde{\Theta}) = f_{-}(-\tilde{\Theta})$$

between *two different* functions and not a functional equation characterizing a single function. We can of course concentrate instead on either of the analytic functions $f_{\pm}(\tilde{\Theta})$, but these do not obey $f_{\pm}(-\tilde{\Theta}) = f_{\pm}(\tilde{\Theta})$. Instead they obey something not true at finite N: $f_{\pm}(\tilde{\Theta}) f_{\pm}(-\tilde{\Theta}) = 1$. Of course Vitali's theorem does not apply here since $\cos N\tilde{\Theta}$ has a line of zeros on $\operatorname{Im}\tilde{\Theta} = 0$.

In the case of Λ_B, however, A2 together with Vital's theorem tells us Z_B is a single analytic function in the region considered. More generally if we derive many functional relations of this type, with fixed points Θ_1, $\Theta_2, \ldots$, etc, and want these all to represent statements about a *single* analytic function, we must be given a connected region D of nonzero thickness containing all these points. Indeed this is what A2 gives for Eqs. (3.4) and (3.5).

Another way to derive Eq. (3.10) is to anticipate the result of Section 4, namely, that Λ_B depends on S_a and S_b only through some combination which is crossing symmetric.

Yet another way is to use the Fan and Wu identity

$$Z(S_a, S_b, S_c, S_d) = Z(S_b, S_a, S_c, S_d)$$

to infer that

$$z(\Theta) = z(i\pi - \Theta) \tag{3.11}$$

which implies Eq. (3.4) since $z = z_B$ in the PR. However, implicit in Eq. (3.11) is the assumption that z is analytic at $\Theta = i\pi/2$, i.e., Z has no zeros nearby for all N.

Now, we turn to Eq. (3.5). Consider a projectile of rapidity Θ and in internal state i that collides with N targets at rest and in internal state α. Let (i', α') be the final state. This collision can be represented by Fig. 2 if we append to it x and t axes as in Fig. 6. Let us denote by M the corresponding $(N+1)$-body S matrix (to remind us it is the monodromy matrix of Faddeev and Takhtadzhan[10]). Clearly,

$$\mathrm{Tr}_2 M \equiv \sum_i \sum_{i'} \delta_{ii'} M_{i\alpha}^{i'\alpha'}(\Theta) = T_{\alpha\alpha'}(\Theta) \tag{3.12}$$

Since M rather than T is the natural entity in S-matrix theory, we bring it into the picture by considering the following skewed partition functions Z_s: the spins are periodic in the vertical direction, but skewed as follows in the horizontal direction: i'_n, the rightmost spin of row n equals i_{n+1} the leftmost of row $n+1$, with $i'_{N+1} = i_1$. (See Fig. 6.) Since the final

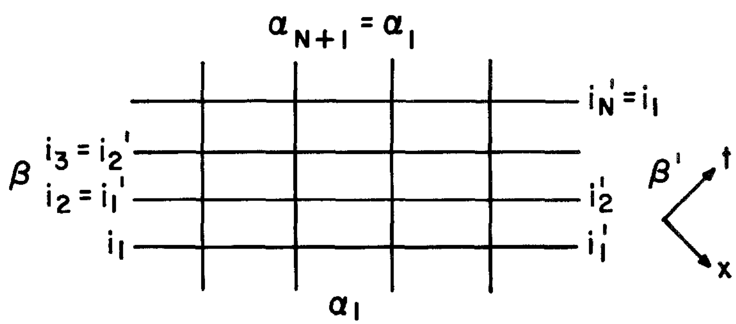

Fig. 6. The lattice with skewed boundary conditions. The vertical bonds are periodic, the horizontal ones are skewed: the rightmost state in each now equals the leftmost in the next.

state (i', α') of each row is the initial state for the next

$$Z_s = \operatorname{Tr} M^N(\Theta) \tag{3.13}$$

the trace being taken in the 2^{N+1}-dimensional space of M.

Let us now evaluate Z_s column by column using $T^c(\Theta)$, the column transfer matrix which acts on the horizontal spins, the vertical spins being summed over with periodic conditions (PBC) because $\alpha_{N+1} = \alpha_1$ in Fig. 6. It is readily seen that

$$T^c(a,b,c,d) = T(b,a,c,d)$$

i.e.,

$$T^c(\Theta) = T(i\pi - \Theta) \tag{3.14}$$

In terms of T we have then

$$Z_s(\Theta) = \operatorname{Tr}\left[T^N(i\pi - \Theta)T(0)\right] \tag{3.15}$$

if we note that (i) the final state $|\beta'\rangle$ is related to the initial state $|\beta\rangle$ by a cyclic shift (see Fig. 6); (ii) $T(0) \equiv T(\Theta = 0)$ is a cyclic shift operator, a result due to Baxter.[28] [To be precise, we must supplement his result with our normalization, $S_a(0) = S_c(0) = 1$; $S_b(0) = S_d(0) = 0$.] Comparing Eqs. (3.14) and (3.15), we get for all Θ,

$$\operatorname{Tr} M^N(\Theta) = \operatorname{Tr}\left[T^N(i\pi - \Theta)T(0)\right] \tag{3.16}$$

Consider this equation in the PR as $N \to \infty$. Perron's theorem applies to M also and we get, in terms of $\Lambda_s(\Theta)$, its biggest eigenvalue,

$$\Lambda_s(\Theta)(1 + \text{negligible terms}) = \Lambda_B(\Theta)(1 + \text{negligible terms}) \tag{3.17}$$

i.e., as $N \to \infty$

$$z_s(\Theta) = z_B(\Theta) \tag{3.18}$$

where

$$z_s = \lim_{N \to \infty} z_s(\Theta, N) \tag{3.19a}$$

and

$$z_s(\Theta, N) = \Lambda_B(\Theta, N)^{1/N} \tag{3.19b}$$

In passing from Eq. (3.16) to Eq. (3.17) I have used the fact that $\Lambda_B(\Theta) = \Lambda_B(i\pi - \Theta)$ and that

$$T(0)|\Lambda_B\rangle = \Lambda_B(0)|\Lambda_B\rangle = 1 \cdot |\Lambda_B\rangle \tag{3.20}$$

which follows because Λ_B is real on the $\operatorname{Im}\Theta$ axis, and unimodular at $\Theta = 0$ since the shift operator is unimodular: $T(0)^N = 1$ since N shifts = no shift. There are some subtleties associated with asymptotic degeneracy discussed in the Appendix.

Our motive for introducing z_s and proving $z_s = z_B$ in the PR is that z_s, when analytically continued to $\operatorname{Im}\Theta < 0$, obeys

$$z_s(\Theta)z_s(-\Theta) = 1 \tag{3.21}$$

It then follows that $z_B(\Theta)$ has an analytic continuation $z_B^{ac}(\Theta)$ that obeys the same equation. However, we shall see $z_B(\Theta)$ itself is analytic at $\Theta = 0$, and hence Eq. (3.5).

To prove Eq. (3.21) we shall establish that at every N,

$$\Lambda_S(\Theta, N)\Lambda_S(-\Theta, N) = 1 \tag{3.22}$$

Taking the Nth root we get

$$z_s(\Theta, N) \cdot z_s(-\Theta, N) = 1 \tag{3.23}$$

Upon showing (as we will) that $N \to \infty$ does not produce singularities at the reflection (or fixed) point $\Theta = 0$, Eq. (3.21) follows.

So consider Eq. (3.22). First, what do we mean by the function Λ_s for Θ not in the PR? (This problem does not arise for $\Lambda_B(\Theta)$ defined for all Θ by $\Lambda_B(\Theta) = \langle \Lambda_B | T(\Theta) | \Lambda_B \rangle$.) In the PR, the characteristic equation for M has a real nondegenerate root $\Lambda_s(\Theta, N)$. Because it is nondegenerate, it is analytic.[29] Given this germ, an analytic function (possibly multivalued) is defined for all Θ by analytic continuation. The claim is that the continuation down the $\operatorname{Im}\Theta$ axis to $\operatorname{Im}\Theta < 0$ obeys Eq. (3.22). The proof is as follows. Since $S(\Theta)$ obeys the unitarity equation $S(\Theta)S^T(-\Theta) = I$, so will M:

$$M(\Theta)M^T(-\Theta) = I \tag{3.24}$$

a result that is readily checked and also expected in a factorizable theory. So $M^T(-\Theta)$ is just $M^{-1}(\Theta)$. Since M^T and M have the same eigenvalues, we learn that the eigenvalues of M at $\Theta = -i|\Theta|$ are inverses of the eigenvalues at $\Theta = i|\Theta|$. Consequently, there will be a nondegenerate, real, smallest eigenvalue $\lambda_s(\Theta)$ such that

$$\Lambda_s(\Theta)\lambda_s(-\Theta) = 1 \tag{3.25}$$

for $|\Theta| < \pi$. If we can show that λ_s, the smallest eigenvalue for $\operatorname{Im}\Theta < 0$, is just the analytic continuation of Λ_s, the largest for $\operatorname{Im}\Theta > 0$, we are done. Evidently we can explore the cross-over which takes place at $\Theta = 0$, by first-order perturbation theory in Θ. Let

$$M(\Theta) = M(0) + \Theta M^1 + \cdots \tag{3.26}$$

Since Θ real corresponds to physical scattering, $M(0)$ is unitary. In fact it is a cyclic shift operator on the $N + 1$ spins (i, α). By standard perturbation theory, the eigenvalues will be (in obvious notation)

$$\Lambda_i(\Theta) = \Lambda_i(0) + \Theta\langle\Lambda_i^0|M^1|\Lambda_i^0\rangle \tag{3.27}$$

Clearly $|\Lambda_i(0)| = 1$. But we require also that $\Lambda_i(\Theta)$ be unimodular for Θ real. Consequently, the first order correction must be orthogonal to $\Lambda_i(0)$ in the complex plane, i.e.,

$$\begin{aligned}\Lambda_i(\Theta) &= \Lambda_i(0) - i\Theta\Delta_i\Lambda_i(0)\\ &\simeq \Lambda_i(0)e^{-i\Theta\Delta_i} \qquad \text{(to order } \Theta)\end{aligned} \tag{3.28}$$

where Δ_i is some real number. Now Perron's theorem assures us that there is a $\Delta_s > \Delta_i$ for all $i \neq s$, in order that the corresponding Λ_s dominate in the PR. But for the same reason, if we continue Eq. (3.28) down to $\operatorname{Im}\Theta < 0$ we see that Λ_s evolves smoothly into the *smallest* eigenvalue, i.e., $\lambda_s(\Theta)$ is just the same function $\Lambda_s(\Theta)$.

Our analysis, based on analyticity at $\Theta = 0$, needs to be justified. Now $M(\Theta)$ is of course analytic at $\Theta = 0$ since it is meromorphic and the poles are away from $\Theta = 0$. However, the eigenvalues can be nonanalytic if they are degenerate. At each point of degeneracy Θ_0, the eigenvalues will be given by a Pusieux series [Taylor series in $(\Theta - \Theta_0)^{1/m}$][29]

$$\Lambda_i(\Theta) = \sum_{n=0}^{\infty} c_n(\Theta - \Theta_0)^{n/m} \tag{3.29}$$

As we go around Θ_0, m degenerate roots will get permuted among themselves. However, unitarity of $M(\Theta)$ for Θ real precludes such a singularity. To see this, note that for Θ real, we can write

$$M(\Theta) = e^{iH(\Theta)} \tag{3.30}$$

where $H(\Theta)$ is Hermitian. At any real Θ let

$$H(\Theta) = H^0 + \Theta H^1 + \cdots \tag{3.31}$$

In general, we expect the eigenvalues to be of the form

$$h_i(\Theta) = \sum_{n=0}^{\infty} d_n(\Theta - \Theta_0)^{n/m} \tag{3.32}$$

However, the requirement that $h_i(\Theta)$ be real for $\Theta - \Theta_0$ positive and real, and *also* negative and real, tells us that $d_n = 0$ unless n is a multiple of m, i.e., $h_i(\Theta)$ has a Taylor series in $\Theta - \Theta_0$. By exponentiation, the same goes for $\Lambda_i(\Theta)$ for all real Θ and in particular $\Theta = 0$. [The above result for $H(\Theta)$, called Rellich's theorem, is familiar to students of quantum mechanics in restricted form: if $H = H^0 + \Theta H^1$, where H^0 and H^1 are Hermitian, the eigenvalues have a perturbative expansion in Θ even if H^0 is degenerate.]

So we have Eq. (3.22). Since $\Lambda_s(\Theta)$ is analytic and nonzero near $\Theta = 0$, we take the Nth root and obtain $z_s(\Theta, N) z_s(-\Theta, N) = 1$. $z_s(\Theta, N)$ will be analytic at $\Theta = 0$ for the above reasons, i.e., have a Taylor expansion with a nonzero radius of convergence $\delta(N)$. But what if $\delta(N) \to 0$ as $N \to \infty$? This may happen because H^i defined in Eq. (3.31) will grow in norm with N and the expansion may have a zero radius of convergence as $N \to \infty$. We do not, however, expect this to happen at $\Theta = 0$. This is because at $\Theta = 0$, the logarithmic derivatives of $M(\Theta)$, i.e., H^0, H^1, etc, are *local* operators, i.e., a sum over N sites of two-spin, three-spin, etc. coupling. [This result is well known for T but is also true for M thanks to the fact that $M(0)$ is also a shift operator.] Consequently, $\|H^i/N\|$ will be finite as $N \to \infty$ and we expect $h_i(\Theta)/N$ to have a Taylor series with a nonzero radius of convergence even as $N \to \infty$. But $z_S(\Theta, N) = \exp[h_i(\Theta)/N]$, and so the same goes for $z_S(\Theta, N)$ as $N \to \infty$. Some readers may worry that the nth logarithmic derivative, H^n, involves n-spin couplings. However, one can show that the radius of convergence for $h_i(\Theta)/N$ defined as $\lim_{n \to \infty} a_n^{-1/n}$, a_n being the nth coefficient, is nonzero once the extensive growth has been neutralized by the factor $1/N$. Similar arguments show that $z_B(\Theta)$ is also analytic at $\Theta = 0$.

This concludes the derivation. For readers who did not follow the connection with the S-matrix problem in detail, I add that it is not absolutely necessary to do so. The main point is that if the weight matrix (called S here and R in Baxter's paper) obeys $S(\Theta) S^T(-\Theta) = I$, then $M(\Theta) M^T(-\Theta) = I$ follows. (This does not require use of the factorizability equation. But S-matrix theory does allow us to anticipate it since a factorizable theory unitarized at the two-body level must remain so at all levels.) From this we get $z_s(\Theta) z_s(-\Theta) = 1$. Equation (3.16) applied in the PR as $N \to \infty$, then allows us to equate z_B to z_s and get $z_B(\Theta) z_B(-\Theta) = 1$ (once z_s is shown to be analytic at the inversion point $\Theta = 0$).

4. ON THE ZEROS OF $\Lambda_B(\Theta)$

Since the seminal work of Lee and Yang[8] we know the importance of locating the zeros of the partition function Z in some variable in which it is entire for all finite N, such as $\exp -\beta H_z$ in the case of the Ising model in a magnetic field H_z. Since Z is entire, the only singularities in Z^{1/N^2} arise from zeros of Z and hence the interest in the latter. In the present case we are interested in the zeros of the eigenvalue Λ_B in the Θ plane. The commutativity of $T(\Theta)$ and $T(\Theta')$ leads to nice analytic properties for Λ_B (same as the weights) and hence the interest in its zeros. Here, however, to prove that $\Lambda_B(\Theta) \neq 0$ in the PS′ is to completely solve the problems, given the functional equations of Section 3 and A1.

In this section it will be shown that $\Lambda_B(\Theta, N)$ itself can be related to the partition function of a one-dimensional Ising model. For some range of parameters the zeros can be located by the Lee–Yang or Suzuki–Fisher theorems and shown to lie outside the PS′. The general problem is, however, unsolved and posed to the reader.

For our discussions it is convenient to work, not with $\Lambda_B(\Theta) \equiv \Lambda_B(S_a, S_b, S_c, S_d)$ *where S_i are minimal*, but with

$$\tilde{\Lambda}_B(\Theta) = \Lambda_B(a, b, 1, d) \tag{4.1}$$

Since

$$(S_a, S_b, S_c, S_d) = S_c \cdot (a, b, 1, d)$$

and rescaling all weights by a factor S_c rescales every matrix element of T by S_c^N,

$$\Lambda_B(\Theta) = S_c^N(\Theta)\tilde{\Lambda}_B(\Theta) \tag{4.2}$$

Since $S_c \neq 0$ or ∞ in the PS′ (for η and k in the PR) Λ_B has these features if $\tilde{\Lambda}_B$ does. Let $\tilde{T}$ denote $T\ (a, b, 1, d)$.

Now $|\tilde{\Lambda}_B|^{1/N} < \infty$ in the PS′ because $(a, b, 1, d)$ is bounded therein, the nearest poles being at

$$\Theta_p = i(\pi - \gamma) \pm m\gamma^{1/2} \tag{4.3a}$$

and $i\pi - \Theta_p$, where m is an integer and

$$\gamma = \frac{i\pi K'}{2\eta} \quad \text{and} \quad \gamma' = \frac{2\pi i K}{\eta} \tag{4.3b}$$

In the PR, γ is greater than π and PS′ is pole free. So the problem is to prove that $|\tilde{\Lambda}_B| > 0$ in the PS.

Here are some properties of $\tilde{\Lambda}_B$ that should help in proving $\tilde{\Lambda}_B \neq 0$ in the PS′ (see Fig. 7a).

Property P1. $\tilde{\Lambda}_B$ is doubly periodic with periods $\gamma'/2$ and $2i\gamma$.

Proof. Under $\Theta \to \Theta + 2i\gamma$, $2\eta i\Theta/\pi \to 2\eta i\Theta/\pi + 2iK'$ and $2iK'$ is a period of a, b, 1, and d. Under $\Theta \to \Theta + (-\gamma'/2)$, $2\eta i\Theta/\pi \to 2\eta i\Theta/\pi - 2K$ and only a and b get negated $[\mathrm{sn}(z \pm 2K) = -\mathrm{sn}(z)]$. $\tilde{\Lambda}_B$ is, however, unaffected since $n_a + n_b = N - n_c - n_d$ is even since $n_c + n_d$ is even to ensure horizontally PBC. (Here n_i is the number of i-type vertices in a row.) Thus we need just examine $\tilde{\Lambda}_B$ in one full period. ■

Property P2. $\tilde{\Lambda}_B \to \tilde{\Lambda}_B^*$ if we reflect on the $\operatorname{Im}\Theta$ axis or the line $\operatorname{Im}\Theta = \pi/2$.

Proof. Since $|\Lambda_B\rangle$ is real, $\tilde{\Lambda}_B$ is real on the $\operatorname{Im}\Theta$ axis and $\tilde{\Lambda}_B \to \tilde{\Lambda}_B^*$ when we reflect on the $\operatorname{Im}\Theta$ axis. P2 then follows using $\tilde{\Lambda}_B(\Theta) = \tilde{\Lambda}_B(i\pi -$

Θ). This property is true for any function (like d or S_c) real on the $\operatorname{Im}\Theta$ axis and invariant under $\Theta \to i\pi - \Theta$. ■

Property P3. In the rectangle bounded by $\operatorname{Re}\Theta = \pm\frac{1}{4}\gamma'$, $\operatorname{Im}\Theta = \pi/2$ and $\operatorname{Im}\Theta = -(\gamma - \pi)/2$, $\tilde{\Lambda}_B$ has $(1/2)N$ zeros and $\tilde{Z} = Z(a, b, 1, d)$ has $(1/2)N^2$ zeros.

Proof. Under $\Theta \to \Theta + i\gamma$

$$(a, b, 1, d) \to \frac{1}{d}(-b, -a, d, 1) \tag{4.4}$$

since $\operatorname{sn}(z + K', k) = k^{-1}\operatorname{sn}(z, k)$. Using

$$\Lambda_B(a, b, c, d) = \Lambda_B(b, a, d, c) = \Lambda_B(-b, -a, d, c)$$

we get

$$\tilde{\Lambda}_B(\Theta + i\gamma) = d^{-N}(\Theta)\tilde{\Lambda}_B(\Theta) \tag{4.5}$$

Now let $\Theta = \Theta_r + i(\pi - \gamma)/2$:

$$\begin{aligned}\tilde{\Lambda}_B(\Theta_r + i(\pi + \gamma)/2) &= d^{-N}\tilde{\Lambda}_B(\Theta_r + i(\pi - \gamma)/2)\\ &= \tilde{\Lambda}_B^*(\Theta_r + i(\pi - \gamma)/2)\end{aligned} \tag{4.6}$$

using P2. Thus

$$\frac{\tilde{\Lambda}_B}{\tilde{\Lambda}_B^*} = d^N \qquad \text{on} \quad \operatorname{Im}\Theta = \frac{(\pi - \gamma)}{2} \tag{4.7a}$$

$$\Delta\phi_{\tilde{\Lambda}} = \tfrac{1}{2}N\Delta\phi_d \tag{4.7b}$$

where $\Delta\phi_f$ is the change in phase of the function f over a distance $\Delta\Theta_r$. Consider now the closed curve BCDEFGHI and apply the argument principle[27]

$$\frac{1}{2\pi}\oint_c \Delta\phi_f = Z_f$$

where Z_f is the number of zeros of the function f, analytic in and on the contour C.

Since $\tilde{\Lambda}_B$ and d are real on CDE and have real period $(1/2)\gamma'$ (so that IBC cancels EFG), we get from Eq. (4.7) that

$$\frac{1}{2\pi}\oint \Delta\phi_{\tilde{\Lambda}_B} = \frac{1}{2\pi}\cdot\frac{N}{2}\cdot\oint \Delta\phi_d = \frac{N}{2}$$

since d has a simple zero at $\Theta = 0$ and no others within the contour BC . . . I.

A similar argument applied to $\tilde{Z}$, which obeys $\tilde{Z}(\Theta + i\gamma) = d^{-N^2}\tilde{Z}(\Theta)$, gives the number of zeros to be $(1/2)N^2$.

Property P4. On GHI and its reflection on $\operatorname{Im}\Theta = \pi/2$, $|d| = 1$.

Proof. On going from GHI to its reflection, $\Theta \to \Theta + i\gamma$ so that $d \to d^{-1}$. But also $d \to d^*$ (see P2). Note also, for later use that $\Delta\phi_d = 2\pi$ as we go from G to H to I (argument principle applied to d on GHIBCDEF)

Property P5. *With η and k fixed, $\tilde{\Lambda}_B$ is a polynomial in d of the form*

$$\tilde{\Lambda}_B(d, \eta, k) = \sum_{n=0}^{N} \gamma^n d^n, \qquad \gamma_n = \gamma_n^* = \gamma_{N-n} > 0 \tag{4.8}$$

Proof. Consider a row of the transfer matrix corresponding to the final state $\tilde{\alpha} = (+ - + - + \cdots) \equiv |\text{afe}\rangle$, where afe mean antiferroelectric. Let us write the equation $\tilde{T}|\Lambda_B\rangle \equiv \tilde{\Lambda}_B|\Lambda_B\rangle$ more explicitly:

$$\begin{bmatrix} & & \\ & & \\ & & \\ \tilde{T}_{\tilde{\alpha}\alpha_1} & \tilde{T}_{\tilde{\alpha}\alpha_2} \cdots & \tilde{T}_{\tilde{\alpha}\alpha_2 N} \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \Lambda_B^1 \\ \Lambda_B^2 \\ \vdots \\ \Lambda_B^{\text{afe}} \\ \vdots \\ \Lambda_B^{2^N} \end{bmatrix} = \tilde{\Lambda}_B \begin{bmatrix} \Lambda_B^1 \\ \Lambda_B^2 \\ \vdots \\ \Lambda_B^{\text{afe}} \\ \vdots \\ \Lambda_B^{2^N} \end{bmatrix} \tag{4.9}$$

When the indicated row is dotted with the column vector $|\Lambda_B(\eta, k)\rangle$, we get $\tilde{\Lambda}_B$ times Λ_B^{afe}, where $\Lambda_B^{\text{afe}} = \langle \text{afe}|\tilde{\Lambda}_B\rangle$. Let us choose $\Lambda_B^{\text{afe}} = 1$ for convenience. Now, it is easy to verify that in the elements of the row $\tilde{T}_{\tilde{\alpha}\alpha}$, the weights a and b occur only in the combination ab (more on this later). Using Baxter's relation[3b]

$$\frac{1}{x^2} \equiv -k\,\text{sn}^2 2\eta = \frac{cd}{ab} = \frac{d}{ab} \qquad (\text{recall } c = 1) \tag{4.10}$$

we see that at fixed η and k, $\tilde{\Lambda}_B = \tilde{\Lambda}_B(d, \eta, k)$. It is evidently a polynomial of Nth order. Since x^2 and the components Λ_B^i are real and positive, so are the coefficients γ_n in Eq. (4.8). Finally $\gamma_n = \gamma_{N-n}$ follows from Eq. (4.5) which may be written as

$$\tilde{\Lambda}_B(1/d) = d^{-N}\tilde{\Lambda}_B(d) \tag{4.11a}$$

since

$$d \to 1/d \qquad \text{under} \quad \Theta \to \Theta + i\gamma \quad \blacksquare \tag{4.11b}$$

Thus we should really study $\tilde{\Lambda}_B$ in the d plane. Thanks to Eq. (4.10) we need consider just the unit disk $|d| \leqslant 1$, which corresponds to the rectangle BC . . . I, (see Fig. 7), while the shaded region within corresponds to ABCDEF. (If $\tilde{\Lambda}_B \neq 0$ in ABCDEF, $\tilde{\Lambda}_B \neq 0$ in the PS′ by use of periodicity,

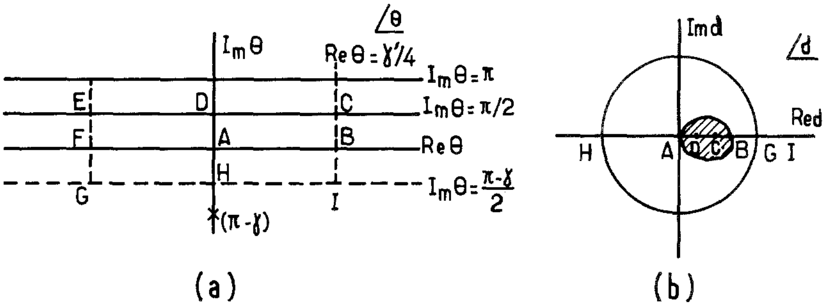

Fig. 7. (a) Given the periods and symmetries of $\tilde{\Lambda}_B$ we need consider just the rectangle BCDEFGHI in the Θ plane. The corresponding part of the PS′ is ABCDEF. (b) The same regions in the d plane: BCDEFGHI becomes the unit disk $|d| \leqslant 1$, while the shaded region corresponds to PS′.

crossing, etc. . . .) As for the map Fig. 7a$\rightarrow$Fig. 7b, we have already seen that GHI$\rightarrow |d| = 1$ with $\Delta\phi_d = 2\pi$, going from G to I. Next, d is real negative/positive on AH/AD. Finally $d = ab/x^2$ is real positive on DC because $a = b^*$ here (crossing = reflection on Im Θ axis) and on CB because $a = i|a|$ and $b = -i|b|$ here.

Property P6. *$\tilde{\Lambda}_B(d)$ may be interpreted as the partition function of a one-dimensional Ising model in a magnetic field H_z given by $d = \exp - \beta H_z$.*

Proof. Consider the totally *disordered* case: $\Lambda_B^i \equiv 1$ in Eq. (4.9), i.e., consider the row sum. Instead of viewing this as a sum over α (with α' fixed at $\tilde{\alpha}$) let us view it as a sum over $\{\tau\}$, the horizontal spins. These may be chosen freely with $\tau_{N+1} = \tau_1$ i.e., for a given $\alpha' = \tilde{\alpha}$ and any $\{\tau\}$, we can find a value of α in this eight-vertex model. In fact $\{\tau\}$ and $\{-\tau\}$ give the same α, i.e., each element $\tilde{T}_{\tilde{\alpha}\alpha}$ is a sum of two elements, due to $\{\tau\}$ and $\{-\tau\}$. Let us call $\{\tau\} = (+ - + - + - + \cdots)$ the *standard configuration*, with respect to which deviations are made. To emphasize this, let us introduce $\{\tau'\}$ such that $\{\tau'\} = (+ + + + + + \cdots)$ in the standard configuration. The corresponding Boltzmann factor is $c^N = 1$.

Let us now flip one spin, say at site $n = 2$. This state is $\{\tau\} = (+ + + - + - \cdots)$ and $\{\tau'\} = (+ - + + + + \cdots)$. What is its Boltzmann weight? Evidently if we flip τ_n or τ'_n, we must flip α_{n-1} and α_n, since only a pair of arrows can be flipped at each vertex. Under this change $c \rightarrow a$ at the site $n-1$ and $c \rightarrow b$ at site n and so the weight is $c^{N-2}ab = ab$. If we flip m spins in a row, we get a factor ab from the ends and a factor d^{m-2} from the interior since $c \rightarrow d$ under reversal of horizontal arrows. Writing abd^{m-2} as x^2d^m, we see that we can associate the factor d^m with m flipped spins in an external magnetic field and the factor x with each nearest-neighbor transition from flip $\leftrightarrow$ nonflip. Thus the row sum obeys the Lee–Yang theorem if $x \leqslant 1$, and the zeros lie on $|d| = 1$, i.e., outside the PS′. The dividing line $x = 1$ has special significance in the S-matrix problem ($\gamma = 2\pi$ and there are no poles in the second strip $0 > \operatorname{Im}\Theta > -\pi$) and in Baxter's solution as the point $\tau = 2\lambda$ at which the location of zeros changes. Indeed for $x \leqslant 1$, i.e., $\tau < 2\lambda$, Baxter finds the zeros do lie very close to $|d| = 1$. We have, however, not managed to rederive this result yet since we are considering the row sum, unweighted by the components Λ_B^i. If we put them back we get, in general (in contrast to the totally disordered case),

$$\tilde{\Lambda}_B(d) = \sum_{\{\tau'\}} d^{\sum_{i=1}^{N}(1-\tau_i')/2} \cdot \exp\left[-\frac{1}{2} \ln x \sum_{i=1}^{N} (\tau_i' \tau_{i+1}' - 1) \right] \cdot \langle \{\tau'\} \,|\, \Lambda_B \rangle \tag{4.12}$$

Since $\langle \{\tau'\} \,|\, \Lambda_B \rangle$ is

(i) real, positive for all $\{\tau'\}$

(ii) translationally invariant ($T(0) \,|\, \Lambda_B\rangle = |\Lambda_B\rangle$)

(iii) reflection invariant ($\langle \{\tau'\} \,|\, \Lambda_B \rangle = \langle \{-\tau'\} \,|\, \Lambda_B \rangle = \langle \alpha \,|\, \Lambda_B \rangle$

we can view it as a physical interaction with properties (ii) and (iii), appended to our Ising model. ■

What about the zeros now? This will be decided by the form of the "function" $\langle \{\tau'\} \,|\, \Lambda_B \rangle$ expanded in terms of two-spin, four-spin, six-spin, etc. couplings. If we can "fit" $\langle \{\tau'\} \,|\, \Lambda_B \rangle$ with just ferromagnetic two-body interactions (not necessarily nearest neighbor) we can still use the Lee–Yang theorem (for $x \leqslant 1$). If not, we must turn to the Suzuki–Fisher theorem which tells us that if r-spin terms are present, zeros lie on $|d| = 1$ for a range of parameters that shrinks like $1/\ln r$. If it happens (and I have other signals which suggests this may be so for $x < 1$) that $r \simeq N$, then there is no circle theorem as $N \to \infty$. But there is, however, hope since *we do not really want a circle theorem, we just want no zeros in the* PS′. For example at $N = 4$ an elementary analysis shows that if the zeros are not on $|d| = 1$ they are in the left-half plane, which is enough since PS′ is in the right-half plane. Further, even if the exact $\langle \{\tau'\} \,|\, \Lambda_B \rangle$ is not reproducible by r-spin terms with r finite even as $N \to \infty$, it may be well approximated by such terms. Since the zeros will depend smoothly on the components Λ_B^i, if an approximation has zeros on $|d| = 1$ the exact one will have them nearby. The fact that $|\Lambda_B\rangle$ is also the eigenket of a local operator like $H_{xyz} \propto d \ln T/d\Theta$ at $\Theta = 0$, suggests that a local approximation should be possible. Note also that as $x \to 0$, $\tilde{\Lambda}_B \simeq 1 + d^N$ and the zeros are again on $|d| = 1$, for this totally *ordered* case.

But these are all special cases and one must prove $\tilde{\Lambda}_B \neq 0$ on the PS′ using the above-mentioned properties of $|\Lambda_B\rangle$ as well as some I have not listed above but which can likewise be derived from general considerations. (Baxter's work[3b] suggests that the zeros obey a circle theorem for all x in the variable $\exp i\Theta$. For $0 < x < 1$, the radius of the circle grows in such a way that it is a circle in d also.)

5. ANOTHER TEST CASE–THE 19-VERTEX MODEL

It will be recalled that much of the derivation in Section 3 was based on general S-matrix principles (like unitarity) and therefore should apply to a wide class of S-matrix based models. I consider the 19-vertex model of Zamolodchikov and Fateev[15] to illustrate what features are needed before the inversion and crossing relations, Eqs. (3.4), (3.5) can be derived. These must of course be supplemented with assumptions A1 and A2 to derive $z(s) = 1$ in the PR. I have not been able to prove the assumptions. I have verified however, to fourth order in a perturbation expansion, that $z(s)$ indeed equals 1 in the PR (i.e., $0 < \operatorname{Im}\Theta < \pi$, other parameters to be specified later).

In this model each bond can be in one of the three states with $Q = 1, 0, -1$, which we can denote by arrows pointing up or down (or left or right) or no arrow at all. Of the total of $3^4 = 81$ vertices only 19 are allowed and these correspond to absolute charge conservation. The vertex weights and S-matrix elements depend on a parameter λ (which we take to be real positive) and the variable $\alpha = -i\Theta$. In S-matrix notation, i.e. (see Fig. 1),

$$
\begin{aligned}
S_{\alpha\beta}^{\gamma\delta}(\Theta) &= \langle \delta(-\Theta/2)\gamma(\Theta/2) \,|\, \alpha(\Theta/2)\beta(-\Theta/2)\rangle \\
S_{++}^{++} &= s = \frac{\operatorname{sh}\lambda(\pi-\alpha)}{\operatorname{sh}\lambda(\pi+\alpha)} \\
S_{+0}^{+0} &= t = \frac{\operatorname{sh}\lambda\alpha\,\operatorname{sh}\lambda(\pi-\alpha)}{\operatorname{sh}\lambda(2\pi-\alpha)\operatorname{sh}\lambda(\pi+\alpha)} \\
S_{+0}^{0+} &= r = \frac{\operatorname{sh}2\pi\lambda\,\operatorname{sh}\lambda(\pi-\alpha)}{\operatorname{sh}\lambda(2\pi-\alpha)\operatorname{sh}\lambda(\pi+\alpha)} \\
S_{00}^{+-} &= a = r(\pi-\alpha) \\
S_{+-}^{+-} &= T = s(\pi-\alpha) \\
S_{+-}^{-+} &= R = \frac{\operatorname{sh}\pi\lambda\,\operatorname{sh}2\pi\lambda}{\operatorname{sh}\lambda(\pi+\alpha)\operatorname{sh}\lambda(2\pi-\alpha)} \\
S_{00}^{00} &= \sigma = \frac{\operatorname{sh}\pi\lambda\,\operatorname{sh}2\pi\lambda - \operatorname{sh}\lambda\alpha\,\operatorname{sh}\lambda(\pi-\alpha)}{\operatorname{sh}\lambda(\pi+\alpha)\operatorname{sh}\lambda(2\pi-\alpha)}
\end{aligned}
\tag{5.1}
$$

Elements not given above are obtained from the following symmetries:

$$\begin{aligned} S_{\alpha\beta}^{\gamma\delta} &= S_{\bar{\alpha}\bar{\beta}}^{\bar{\gamma}\bar{\delta}} && \text{charge reversal} \\ &= S_{\beta\alpha}^{\delta\gamma} && \text{parity} \\ &= S_{\delta\gamma}^{\beta\alpha} && \text{time reversal} \end{aligned} \tag{5.2}$$

Here are the relevant facts.

(i) For α real, $0 < \alpha < \pi$, all weights are real positive. Hence Perron's theorem applies to T and M. Here there are blocks for each Q. In the low-temperature region $\lambda \to \infty$, weights σ and R dominate, i.e., $Q = 0$. This is assumed to hold down to some $\lambda_{\min}$, probably 0. This defines the PR, along with α real, $0 < \alpha < \pi$.

(ii) Since $[T(\alpha), T(\alpha')] = 0$ and $H = d \ln T(\alpha)/d\alpha$ at $\alpha = 0$ is Hermitian, there exists an α-independent basis and $\Lambda_B(\alpha)$ is meromorphic.

(iii) Consider a row of T in which $Q = 0$ for all bonds in the final state α'. Let the initial state α also be such. The possible vertices are σ and t (which are crossing symmetric). If we now vary α, staying within the sector with $Q = 0$, it will be seen that besides σ and t, the vertices r and a occur but only in the crossing symmetric combination ra. Thus $\Lambda_B(\alpha) = \Lambda_B(\pi - \alpha)$ and Eq. (3.5) follows given A2.

We can also get this result another way. Since every vertex is accompanied by one which is related to it by a 90° rotation, Z is invariant under the exchange of all vertices related by a 90° rotation (i.e., in the sum Z, each configuration of the lattice is accompanied by another obtained by a 90° rotation). Now a 90° rotation corresponds to crossing plus parity

$$\begin{tikzcd} \langle \delta\gamma \mid \alpha\beta \rangle \arrow[r, "\text{crossing}"'] \arrow[dr, "90^\circ\text{ rot.}"'] & \langle \delta\bar{\alpha} \mid \bar{\gamma}\beta \rangle \arrow[r, "\text{parity}"'] & \langle \bar{\alpha}\delta \mid \beta\bar{\gamma} \rangle \\ & \langle \bar{\alpha}\delta \mid \beta\bar{\gamma} \rangle & \end{tikzcd}$$

Thus in this model, which is parity invariant,

$$z(\alpha) = z(\pi - \alpha)$$

Using $z = z_B$ in the PR, and assuming z is analytic at $\alpha = \pi/2$, the result follows.

(iv) The inversion formula (3.5) follows from unitarity, and the existence of a dominant Λ_s, and the locality of $d^n \ln M/d^n\alpha$ at $\alpha = \Theta = 0$.[15] [This will be true whenever $S_{\alpha\beta}^{\gamma\delta}(0) \propto \delta_{\alpha\delta}\delta_{\beta\gamma}$. There are many such S matrices.[6]]

No attempt is made to prove A1 and A2 here. Assuming these, $z(s) = 1$ follows for $0 < \operatorname{Re} \alpha < \pi$, α real and $\lambda > \lambda_{\min}$ (not precisely known). The result $z(S) = 1$ can be verified in the low-temperature limit $\lambda \to \infty$. Here σ and R dominate and equal unity, $Z = 3$, and $z = 1$. For λ

large but finite, we can develop a series in $x = e^{-\alpha\lambda}$ and $y = e^{-\lambda(\pi-\alpha)}$ for $0 < \alpha < \pi$. I find $z(S) = 1$ to fourth order in the double power series, and have run into enough miraculous cancellation to become a believer.

It must be pointed out that unlike in the eight-vertex case (i) there are not enough symmetries to map every region into the PR; (ii) due to the dearth of parameters, we cannot assume every weight is of the form $\exp -\beta\epsilon$, where ϵ is β independent. However, as $\lambda \to \infty$ all weights are indeed of this form with $\beta = \lambda$.

6. CONCLUDING REMARKS

In this paper an attempt was made to show that the knowledge of a factorizable S matrix not only allows us to define a solvable vertex model *à la* Zamolodchikov, but also to solve it at once, the answer being $z(S) = 1$ in the PR. [The reader is reminded $z(S) = 1$ does not imply a trivial problem; in this parlance $z(s) = 1$ even for the Baxter model.] We made assumptions A1 and A2 and derived two functional equations for z_B. These implied $z_B \equiv 1$. Since $z = z_B$ in the PR, $Z(S) = 1$ followed in the PR.

The work of Section 4 suggests that it may be possible to prove A2, concerning the zeros of $\tilde{\Lambda}_B$, the eigenvalue that dominates in the PR. We saw $\tilde{\Lambda}_B$ is itself the partition function of an Ising model, the interaction being decided by the ket $|\Lambda_B\rangle$. It is the latter that prevents us from saying anything definite about the zeros except in special cases with complete order or disorder. It is of course possible that there is some general feature of $|\Lambda_B\rangle$ (i.e., a feature like translational invariance that can be known without explicitly solving for it) that I have not considered, which will prove that the zeros do not lie on PS′. This I pose as a problem to the readers, not just in the eight-vertex case but for all S-matrix based vertex models. Besides following the approach of Section 4, namely, viewing the corresponding $\tilde{\Lambda}_B$ as the partition function of a spin problem in one dimension, one can try to use factorization in the following sense. Consider the zeros of Λ_s, which seems an equally hard problem, but is not, because $\Lambda_s = 0 \to M^{-1}$ does not exist $\to S^{-1}$ does not exist. The points where $\det S = 0$ are instantly located: $a^2 = d^2$ or $b^2 = c^2$ [see Eq. (1.1)]. If we can locate the zeros of Λ_s so easily (and further, they do not lie in PS′) why not consider $z_S = \Lambda_S^{1/N}$ which finally equals z_B in the PR? The problem is that unlike $\tilde{\Lambda}_B$, Λ_s *is not meromorphic*: $[M(\Theta), M(\Theta')] \neq 0$ [though $\mathrm{Tr}_2 M(\Theta)M(\Theta') = \mathrm{Tr}_2 M(\Theta')M(\Theta)$ the significance of which eludes me]. If we can show that Λ_s dominates in the PS (on $\mathrm{Im}\,\Theta = 0$ it is analytic and unimodular), i.e., is analytic therein, we are done because $z_S(\Theta) = z_s(i\pi - \Theta)$ follows from $z_S = z_B$ assuming z_B is analytic at $\Theta = i\pi/2$.

At present the necessity of having to make A1 and A2 renders this work in the same league as that of Straganov, Schultz, or Baxter. It differs, however, in that the central inversion formula, based on unitarity, is universally true for all S-matrix based models and in that some other assumptions on analyticity are proven here.

Instead of starting with a known S matrix and solving the problem it defines, we could also take a known problem and try to solve it this way. The case in point is the q-state Pott's model, which can be viewed, thanks to the Temperley–Lieb equivalence,[30] as a staggered six-vertex model with different weights ω_A and ω_B on sublattices A and B. At the critical point, $\omega_A = \omega_B$, and Baxter solved for the free energy.[31] What about the general case? We can take as the basic block a 2×2 lattice containing two A and two B sites, which repeats simply. (I am obliged to Daniel Fisher for this suggestion.) We can view it as an unnormalized S matrix for objects in the $\frac{1}{2} \otimes \frac{1}{2}$ representation and try to find the unitarizing factor, i.e., which will finally give $z(S) = 1$. To do this of course one needs the parameter one can identify with Θ (or $\alpha = -i\Theta$), and proof of $[T(\Theta), T(\Theta')] = 0$, which is presently unavailable.

Meanwhile Jaekel and Maillard[32] have verified (to fifth order in a power series) that z satisfies two functional equations in *two* complex variables x and y. They report, however, that the minimal solution to these equations (minimal in a sense they define) coincides with the series expansion for the exact answer only at the critical (self-dual) point. We then have the option of trying to redefine minimality or changing variables so that the problem essentially involves just one complex variable. (The latter may in fact be done in the self-dual case as well as in the antiferromagnetic case recently solved by Baxter.[33] Here the equations will take the form of unitarity and crossing equations and minimality will have the same meaning: no poles or zeros in a certain strip.)

ACKNOWLEDGMENTS

I welcome this opportunity to thank Professor R. J. Baxter for so readily sharing his insights in his correspondence and Professor J. L. Gervais of the Laboratoire de Physique Théorique de l'Ecole Normale Supérieure for its generous hospitality. I thank also my colleagues M. Jaekel, B. Julia, J. M. Maillard, and especially A. Neveu for many fruitful discussions. An NSF grant INT-8117361, which supported this research in part, is gratefully acknowledged.

APPENDIX

Here we reexamine the passage from Eq. (3.16)

$$\operatorname{Tr} M^N(\Theta, N) = \operatorname{Tr}\left[T^N(i\pi - \Theta) T(0) \right] \tag{A.1}$$

to Eq. (3.17)

$$\Lambda_s^N(\Theta, N)(1 + \text{negligible terms}) = \Lambda_B^N(\Theta, N)(1 + \text{negligible terms}) \quad \text{(A.2)}$$

as $N \to \infty$. The argument was based on Perron's theorem, which tells us that Λ_s and Λ_B dominate the respective traces. Perron's theorem is, however, valid only for finite N and it is possible as $N \to \infty$ that other eigenvalues, Λ_s' and Λ_B', asymptotically degenerate in modulus with Λ_s and Λ_B emerge. We will see that this can happen here, with $\Lambda_B' = -\Lambda_B$; $\Lambda_s' = -\Lambda_s$. For even N, this poses no real problem on the left-hand side, whereas in the right-hand side there will be a cancellation between $\Lambda_B^N \cdot \langle \Lambda_B | T(0) | \Lambda_B \rangle \equiv \Lambda_B^N C_B$ and $\Lambda_B' \cdot \langle \Lambda_B' | T(0) | \Lambda_B' \rangle \equiv \Lambda_B' C_B'$ because $C_B' = -C_B$. [Note that C_B' is just $\Lambda_B'(0)$.] Thus the leading eigenvalue of T drops out! (Incidentally this provides us with an inversion formula for the next eigenvalue Λ_B''. We cannot, however, go further than that.)

The origin of this problem and its cure are as follows. Consider very low temperatures when $S_c \gg S_a$, S_b, or S_d. Here the lattice is antiferroelectrically (afe) ordered. Imagine a 4×4 lattice covered with c type vertices. (It might help to have a sketch.) Clearly

$$T(\Theta)|+-+-\rangle = S_c^4|-+-+\rangle$$
$$T(\Theta)|-+-+\rangle = S_c^4|+-+-\rangle$$

Thus we can form eigenstates

$$|\pm\rangle = \frac{|+-+-\rangle \pm |-+-+\rangle}{\sqrt{2}}$$

with eigenvalues $\Lambda_B = S_c^4$, $\Lambda_B' = -S_c^4$ and C_B or $C_B' = \pm 1$, respectively.

Unfortunately, the matrix M rejects the afe ordered state. This can be seen in two ways: (i) for $N = 4$ (or any even number) the rightmost horizontal bond of row n will not equal the leftmost bond of row $n+1$, as required to form M^4; (ii) the action of M on an afe ordered state is not simple:

$$M(\Theta)|-+-+-\rangle = |--+-+\rangle$$
$$M(\Theta)|+-+-+\rangle = |++-+-\rangle$$

where the first label is just i (the leftmost horizontal bond) and the next four are α (see Fig. 2). Thus we cannot make an afe ordered eigenstate of M, which cyclically shifts the spins (i, α).

The cure is simple: consider a lattice with 5 columns and 4 rows (or generally $N+1$ columns and N rows, N being even). Now we have

$$M(\Theta)|-+-+-+\rangle = |+-+-+-\rangle$$

etc. and we can form $|+\rangle$ and $|-\rangle$ states that are afe ordered and have Λ_s

and Λ_s' equal to $\pm S_c^5$. Equation (1) now becomes

$$\operatorname{Tr} M^4(\Theta, 5) = \operatorname{Tr} T^5(\Theta, 4)$$

and Eq. (2) becomes

$$(\Lambda_s^4 + \Lambda_s'^4)(1 + \cdots) = [\Lambda_B^5 \cdot 1 + \Lambda_B'^5 \cdot (-1)][1 + \cdots]$$

and more generally, as $N \to \infty$,

$$\Lambda_s^N(\Theta, N+1) = \Lambda_B^{N+1}(\Theta, N)$$

Taking the $N(N+1)$th root of both sides we get $z_s = z_B$.

These arguments were made at low temperatures where the scenario is simple. As we raise the temperature, other weights will come in, $|\Lambda_B'|$ and $|\Lambda_s'|$ will lie below Λ_s and Λ_B (by Perron's theorem, since M and T now have irreducible positive blocks), and only asymptotically approach Λ_s and Λ_B. The constants C_B and C_B' are expected to be the same throughout, namely, ± 1.

NOTE ADDED IN PROOF

Recently some asymmetrix 8-vertex models have been solved this way and will be reported upon soon.

REFERENCES

1. R. Shankar, *Phys. Rev. Lett.* **47**:1177 (1981).
2. A. B. Zamolodchikov, *Commun. Math. Phys.* **69**:165 (1979).
3. (a) R. J. Baxter, *Ann. Phys. (N.Y.)* **70**:193 (1970); (b) R. J. Baxter, *J. Stat. Phys.* **8**:25 (1973).
4. E. H. Lieb and F. Y. Wu, Two-dimensional Ferroelectric Models, in *Phase Transition and Critical Phenomena*, C. Domb and M. S. Green, eds. (Academic Press, New York, 1976); P. W. Kastelyn, Exactly Solvable Lattice Models, in *Fundamental Problems in Statistical Mechanics*, E. D. G. Cohen, ed. (North Holland, Amsterdam, 1975), Vol. 3.
5. I. S. Gradstein and I. M. Ryzik, *Tables of Integrals, Series and Products*, (Academic Press, New York, 1965).
6. A. B. Zamolodchikov and Al. B. Zamolodchikov, *Ann. Phys. (N.Y.)* **120**:253 (1979); for a more simple-minded review see Exact S-Matrices in Two-Dimensional Field Theories—A Review," R. Shankar, Yale preprint No. C00 3075-199 (1979).
7. A. B. Zamolodchikov, *Sov. Sci. Rev.* **2** (1980).
8. T. D. Lee and C. N. Yang, *Phys. Rev.* **87**:410 (1952).
9. M. Suzuki and M. E. Fisher, *J. Math. Phys.* **12**:235 (1971).
10. L. A. Takhtadzhan and L. D. Faddeev, *Russ. Math. Survey* **345**:11 (1979).
11. Y. A. Stroganov, *Phys. Lett.* **74A**:116 (1979).
12. C. L. Schultz, *Phys. Rev. Lett.* **46**:629 (1981).
13. J. H. H. Perk and C. L. Schultz, New Families of Commuting Transfer Matrices in q-State Vertex Models, Stony Brook preprint No. ITP-SB-81-25.

14. R. J. Baxter, Exactly Solved Models, in *Fundamental Problems in Statistical Mechanics*, E. D. G. Cohen, ed. (North Holland, Amsterdam, 1980); also *J. Stat. Phys.* **28**:1 (1982).
15. A. B. Zamolodchikov and V. A. Fateev, *Sov. J. Nucl. Phys.* **32**:298 (1981).
16. C. Fan and F. Y. Wu, *Phys. Rev.* **82**:723 (1970).
17. J. D. Johnson, S. Krinsky, and B. M. McCoy, *Phys. Rev. A* **8**:2526 (1973).
18. R. Shankar, *Phys. Lett.* **92B**:333 (1980).
19. A. M. Polyakov, *Phys. Lett.* **72B**:224 (1977); S. Parke, Harvard preprint, HUTP 80/A021 (1980).
20. C. N. Yang, *Phys. Rev.* **168**:1920 (1968).
21. M. Karowski, H. J. Thun, J. T. Troung, and P. Weisz, *Phys. Lett.* **62B**:321 (1977).
22. R. Shankar and E. Witten, *Phys. Rev. D* **17**:213 (1978).
23. A. B. Zamolodchikov, *Commun. Math. Phys.* **29**:489 (1981).
24. R. J. Baxter, *Phil. Trans. R. Soc.* **289**:315 (1978).
25. See for example, J. Kogut, *Rev. Mod. Phys.* **51**:659 (1979).
26. F. R. Gantmacher, *Application of the Theory of Matrices* (Interscience Publishers, New York, 1959), pp. 64, 101, 107, and 111.
27. E. C. Titchmarsh, *Theory of Functions* (Oxford University Press, London, 1939); for application in the Ising model or hardsphere gas see Ref. 8.
28. R. J. Baxter, *Ann. Phys. (N.Y.)* **70**:323 (1972).
29. T. Kato, *Perturbation Theory for Linear Operators* (Springer Verlag, Berlin, 1976); M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic Press, London, 1978), Vol. 4, Chap. XII.
30. H. N. V. Temperley and E. H. Lieb, *Proc. R. Soc. London* **A322**:251 (1971).
31. R. J. Baxter, *J. Phys. C* **6**:L445 (1973).
32. M. T. Jaekel and J. M. Maillard, Inverse Functional Relation on the Potts Model, Saclay preprint DPhT-G/PSRM/81-1930.
33. R. J. Baxter, *Proc. R. Soc. London.* **A383**:43 (1982).