Journal of Statistical Physics, Vol. 35, Nos. 3/4, 1984

Eight-Vertex SOS Model and Generalized Rogers–Ramanujan-Type Identities

George E. Andrews,[1,2] R. J. Baxter,[3] and P. J. Forrester[3]

Received December 21, 1983

The eight-vertex model is equivalent to a "solid-on-solid" (SOS) model, in which an integer height l_i is associated with each site i of the square lattice. The Boltzmann weights of the model are expressed in terms of elliptic functions of period $2K$, and involve a variable parameter η. Here we begin by showing that the hard hexagon model is a special case of this eight-vertex SOS model, in which $\eta = K/5$ and the heights are restricted to the range $1 \leqslant l_i \leqslant 4$. We remark that the calculation of the sublattice densities of the hard hexagon model involves the Rogers–Ramanujan and related identities. We then go on to consider a more general eight-vertex SOS model, with $\eta = K/r$ (r an integer) and $1 \leqslant l_i \leqslant r-1$. We evaluate the local height probabilities (which are the analogs of the sublattice densities) of this model, and are automatically led to generalizations of the Rogers–Ramanujan and similar identities. The results are put into a form suitable for examining critical behavior, and exponents $\beta, \alpha, \bar{\alpha}$ are obtained.

KEY WORDS: Statistical mechanics; lattice statistics; number theory; eight-vertex model; solid-on-solid model; hard hexagon model; Rogers–Ramanujan identities.

1. THE LOCAL HEIGHT PROBABILITIES P_a

1.1. Introduction

Many of the exactly solved two-dimensional models in statistical mechanics are equivalent to special cases of the eight-vertex model.[1–3] For instance,

[1] Department of Mathematics, McAllister Building, Pennsylvania State University, University Park, Pennsylvania 16802.

[2] Supported by the Guggenheim Foundation and in part by the National Science Foundation, grant MCS 8201733.

[3] Research School of Physical Sciences, The Australian National University, G.P.O. Box 4, Canberra 2601, Australia.

0022-4715/84/0500-0193$03.50/0

the critical Potts model is equivalent to a six-vertex model,[4] which is the eight-vertex model with two-vertex configurations given zero weight. (One point that should be made is that the six-vertex model can be solved in an electric field,[5] which is not true of the eight-vertex model. Here we shall consider only the "zero-field" eight-vertex model.)

The triangular three-spin model was originally proposed and solved[6] quite independently of the eight-vertex model: three years went by before it was realized[7] that it can be transformed into a special eight-vertex model.

It seems that history must repeat itself: it is now a little over three years since the hard hexagon model was solved,[8,9] and until now it was believed to be distinct from the eight-vertex model (p. 453 of Ref. 10). However, one way of solving the eight-vertex model[11-14] is to convert it to a "solid-on-solid" (SOS) model (Section 3 of Ref. 12, and Section 9.1.2 of Ref. 14). We begin this paper by showing that the hard hexagon model is in fact a special case of this SOS model. For this case, the eight-vertex parameter η has the value

$$\eta = K/5 \tag{1.1.1}$$

where $2K$ and $2iK'$ are the periods of the elliptic functions that naturally occur.

One intriguing feature of the hard hexagon model is that the Rogers–Ramanujan and related identities occur very naturally in the calculation of the local densities.[9] These involve functions such as

$$\begin{aligned} G(x) &= \prod_{m=1}^{\infty} \left[(1 - x^{5m-4})(1 - x^{5m-1})\right]^{-1} \\ H(x) &= \prod_{m=1}^{\infty} \left[(1 - x^{5m-3})(1 - x^{5m-2})\right]^{-1} \end{aligned} \tag{1.1.2}$$

where the powers of x in the infinite products increase in intervals of 5.

In this paper we generalize the hard hexagon model by considering the corresponding eight-vertex SOS model with

$$\eta = K/r \tag{1.1.3}$$

r being a positive integer which is sometimes (but not always) restricted to odd values. We then use the corner-transfer-matrix technique to calculate the local height probabilities of this model (these are the analogs of the hard hexagon densities) and are led (as we hoped) to various generalizations of the Rogers–Ramanujan identities. These are almost certainly closely related to Gordon's generalization,[15,16] and involve the functions

$$G_{ri}(x) = \prod_{\substack{m=1 \\ m \not\equiv 0,\, \pm i \pmod r}}^{\infty} (1 - x^m)^{-1} \tag{1.1.4}$$

which are natural generalizations of $H(x), G(x)$, reducing to them when $r = 5$ and $i = 1, 2$.

1.2. Eight-Vertex SOS Model

Consider a square lattice $\mathscr{L}$, wound on a cylinder so that the last column is followed by the first. With each site i associate an integer "height" l_i. With each face associate a weight $W(l_i, l_j \mid l_m, l_n)$, where i, j, n, m are the four surrounding sites, ordered anticlockwise from the lower left, as in Fig. 1. Then the partition function is

$$Z = \sum \prod_{(i,j,n,m)} W(l_i, l_j \mid l_m, l_n) \tag{1.2.1}$$

where the sum is over all the allowed arrangements of heights on the lattice, and the product is over all faces (i, j, n, m) of the lattice.

Require that the heights of adjacent sites must differ by 1: this is equivalent to taking $W(l, m' \mid l', m)$ to be zero unless

$$|l - m'| = |m' - m| = |m - l'| = |l' - l| = 1 \tag{1.2.2}$$

To within a uniform additive shift of l, m', l', m, there are six possible ways of satisfying (1.2.2), as shown in Fig. 2.

Now consider the usual zero-field eight-vertex model, with weights a, b, c, d defined as in Eqs. (1)–(4) of Ref. 11. Define k, η, v, ρ by Eq. (8) of

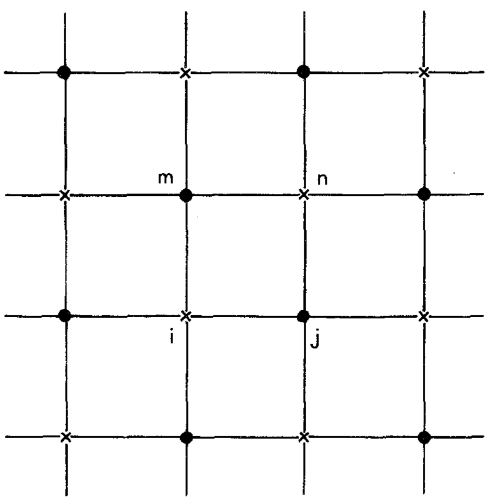

Fig. 1. The square lattice $\mathscr{L}$, showing a typical face (i, j, n, m) and the two sublattices X and Y (denoted, respectively, by crosses and circles).

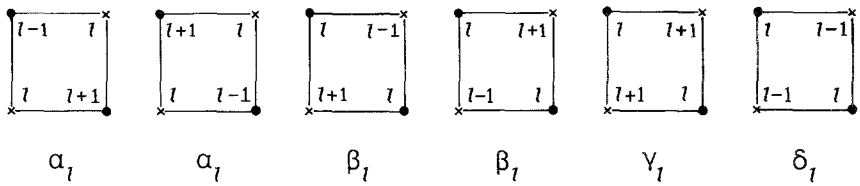

Fig. 2. The six possible arrangements of heights round a face of the lattice. The location of the sublattices is irrelevant here: crosses and circles may be interchanged.

Ref. 11, i.e., by

$$\begin{aligned} a &= \rho\Theta(2\eta)\Theta(v-\eta)H(v+\eta) \\ b &= \rho\Theta(2\eta)H(v-\eta)\Theta(v+\eta) \\ c &= \rho H(2\eta)\Theta(v-\eta)\Theta(v+\eta) \\ d &= \rho H(2\eta)H(v-\eta)H(v+\eta) \end{aligned} \tag{1.2.3}$$

where $H(u), \Theta(u)$ are the elliptic theta functions of argument u and modulus k. Let $h(u)$ be the function

$$h(u) = H(u)\Theta(u) \tag{1.2.4}$$

and let

$$\rho' = \rho\Theta(0), \qquad w_l = w_0 + 2l\eta \tag{1.2.5}$$

where w_0 is a constant, as yet arbitrary. Then it is shown in Refs. 11 and 12 that this eight-vertex model on $\mathscr{L}$ is equivalent to an SOS model on $\mathscr{L}$ with weights

$$\begin{aligned} &W(l,l+1\,|\,l-1,l) = \rho' h(v+\eta) \\ &W(l,l-1\,|\,l+1,l) = \rho' h(v+\eta)h(w_{l+1})/h(w_l) \\ &W(l+l,l\,|\,l,l-1) = \rho' h(v-\eta)h(w_{l-1})/h(w_l) \\ &W(l-1,l\,|\,l,l+1) = \rho' h(v-\eta) \\ &W(l+1,l\,|\,l,l+1) = \rho' h(2\eta)h(w_l+\eta-v)/\left[h(w_l)h(w_{l+1})\right] \\ &W(l-1,l\,|\,l,l-1) = \rho' h(2\eta)h(w_l-\eta+v) \end{aligned} \tag{1.2.6}$$

(taking $\mathscr{L}$ to be the dual of the lattice used in Ref. 12, where the heights are associated with faces rather than sites).

From the usual definitions[17,18] of the elliptic theta functions,

$$h(u) = 2p^{1/4}\sin\frac{\pi u}{2K}\prod_{n=1}^{\infty}\left(1-2p^n\cos\frac{\pi u}{K}+p^{2n}\right)(1-p^{2n})^2 \tag{1.2.7}$$

where K and K' are the complete elliptic integrals of the first kind, and p is the "nome"

$$p = e^{-\pi K'/K} \tag{1.2.8}$$

Noting that K is a function only of p, from (1.2.5)–(1.2.7) we see that we have five parameters at our disposal, which we can take to be p, $p^{1/4}\rho'$, η, w_0, and v. Here we shall regard the first four as real constants, and v as a variable (possibly complex). In particular, p must lie in the interval

$$-1 < p < 1 \tag{1.2.9}$$

which implies that K is real and positive.

In Refs. 11 and 12, particular attention is paid to the case when there exists an integer L such that $2L\eta$ is a period of the elliptic functions. This means that we can regard two heights l_i in the SOS model as being equal if they differ only by an integer multiple of L: it is then natural to interpret the adjacency restrictions (1.2.2) to modulo L, and this can be convenient in setting up a Bethe ansatz with which to solve the SOS model. [Incidentally, Pegg[19] has obtained the general conditions under which an SOS model satisfying (1.2.2) can be solved by a Bethe ansatz.]

Here we shall not adopt this "modulo L" viewpoint, but will interpret (1.2.2) strictly as written. This means that if we move from left to right along a row of $\mathscr{L}$, returning ultimately (because $\mathscr{L}$ is wound on a cylinder) to our starting point, then we must see as many increases in height (between adjacent sites) as decreases. (In Bethe ansatz terms, there are as many "up arrows" as "down arrows" per row.) It follows that we can in fact ignore the restriction (9) of Ref. 11, as well as the renormalization (10), and take $H(u)$, $\Theta(u)$ to be the usual elliptic theta functions.[17,18] In (1.2.3) we have used the fact that $\Theta(u)$ is an even function, while $H(u)$ is odd.

When we say, before (1.2.6), that the eight-vertex and SOS models are "equivalent," we mean that the eigenvalues of the transfer matrix of the former are also eigenvalues of the transfer matrix of the latter. If the SOS model has no eigenvalues larger than the maximum eight-vertex eigenvalues, then it follows that the two models have the same partition function per site in the limit when the number of rows of $\mathscr{L}$ becomes infinite.

The eigenvectors and eigenvalues of the SOS model transfer matrix are given by Eqs. (1.16)–(1.23) or Ref. 13, with $n = N/2$. [Actually these equations are over-restrictive: all the eight-vertex eigenvalues are certainly given by them[2] but the SOS model has other eigenvectors and eigenvalues obtained by multiplying the RH sides of (1.16) and (1.23) by z^l and z^2, respectively, while also multiplying the two additive terms on the right-hand side of (1.21) by z and z^{-1}. This z can be any complex number, and this extra degree of freedom reflects the fact that we are at present allowing the heights to take all integer values from $-\infty$ to ∞, with no "height boundary condition." When they are restricted to a finite set of values, boundary constraints are imposed on the eigenvectors which fix z. For the hard hexagon case discussed below, it appears that $z^{10}=1$. The eigenvectors are in general linear combinations of those given in Ref. 13 that correspond to degenerate eigenvalues.]

The eigenvectors of the SOS model are independent of v (reflecting the fact that eight-vertex transfer matrices with the same values of k and η, but different values of v, commute[2]). The eigenvalues are independent of the parameter w_0.

A trivial variation of the model is to associate extra weights $F(l_i,l_j)$, $F^{-1}(l_i,l_j)$ with the faces above and below each horizontal edge (i,j). Similarly, weights $G(l_i,l_m), G^{-1}(l_i,l_m)$ can be associated with the faces to the right and left of each vertical edge (i,m). Clearly such weights cancel out of the partition function, but they multiply the weight function $W(l,m'\mid l',m)$ by

$$F(l,m')G(l,l')/\left[F(l',m)G(m',m)\right] \tag{1.2.10}$$

One particular such transformation is to multiply W by $x(l)/x(m)$, for any function $x(l)$.

Performing both, taking

$$\begin{aligned} F(l,l\pm 1) = G(l+1,l) &= \left[G(l-1,l)\right]^{-1} \\ &= \left[h(w_l)\right]^{1/2} \end{aligned} \tag{1.2.11}$$

and $x(l)=i^l$, the weights (1.10) become

$$\begin{aligned} &W(l,l+1\mid l-1,l) = W(l,l-1\mid l+1,l) = \alpha_l \\ &W(l+1,l\mid l,l-1) = W(l-1,l\mid l,l+1) = \beta_l \\ &W(l+1,l\mid l,l+1) = \gamma_l, \qquad W(l-1,l\mid l,l-1) = \delta_l \end{aligned} \tag{1.2.12a}$$

where $\alpha_l, \alpha_l, \beta_l, \beta_l, \gamma_l, \delta_l$ are the weights of the six possible configurations, as shown in Fig. 2, and are given by

$$\begin{aligned} \alpha_l &= \rho' h(v+\eta) \\ \beta_l &= \rho' h(\eta - v)\left[h(w_{l-1})h(w_{l+1})\right]^{1/2}/h(w_l) \\ \gamma_l &= \rho' h(2\eta)h(w_l+\eta-v)/h(w_l) \\ \delta_l &= \rho' h(2\eta)h(w_l-\eta+v)/h(w_l) \end{aligned} \tag{1.2.12b}$$

From now on we shall take W to be given by (1.2.12) rather than (1.2.6). This is a more symmetric form, as it has the property that

$$W(l,m'\mid l',m) = W(l,l'\mid m',m) = W(m,m'\mid l',l) \tag{1.2.13}$$

The only further transformation of the type (1.2.10) that can be applied without violating these symmetries is to multiply $W(l,m' \mid l',m)$ by

$$x^{l-m} g_l g_m / (g_{l'} g_{m'}) \tag{1.2.14}$$

where $x^4 = 1$ and g_l is arbitrary. When $x = i$ and $g_l = 1$, this transformation merely negates β_l.

Note that if the height of one site is even, then the heights of all its neighbors must be odd. If we divide the square lattice into two sublattices X and Y (the crosses and circles of Fig. 1), such that any site of X has neighbors only in Y, and conversely, then it follows that Z is the sum of two individual partition functions Z_1 and Z_2. In Z_1, the only allowed configurations are those with odd heights on X sites, and even heights on Y sites. Conversely for Z_2. Thus in this sense the SOS model splits into two separate, but equivalent, models. We shall refer to them as "sub-SOS" models.

1.3. Hard Hexagon Model

The hard hexagon model[8] is a special case of the hard square model with diagonal interactions. With each site i of the square lattice $\mathscr{L}$ associate an occupation number σ_i, such that $\sigma_i = 0$ or 1 according to whether the site is empty or contains a particle. As with the SOS model, associate a weight $W_H(\sigma_i, \sigma_j \mid \sigma_m, \sigma_n)$ with each face (i, j, n, m). No two particles can be adjacent, which is equivalent to saying that $\sigma_i \sigma_j = 0$ for all edges (i, j) of $\mathscr{L}$.

This model can be solved provided a certain restriction (Eq. 24 of Ref. 8) is satisfied (the hard hexagon model is included in this solvable case). The weights W_H can then naturally be parametrized in terms of elliptic functions. Using the parametrization of Eq. (2.12) of Ref. 20, or equivalently Eq. (14.2.39) or Ref. 10, replacing u therein by $\pi(\eta - v)/(2K)$ and noting that $\theta(u)$ and $\theta_1(u)$ therein are proportional to $h(2Ku/\pi)$ (with appropriate elliptic moduli), we find that the nonzero values of W_H are

$$\begin{aligned}
&W_H(00 \mid 00) = h(5\eta - v)/h(4\eta) \\
&W_H(10 \mid 00) = W_H(00 \mid 01) = h(\eta - v)/\left[h(2\eta)h(4\eta)\right]^{1/2} \\
&W_H(01 \mid 00) = W_H(00 \mid 10) = h(\eta + v)/h(2\eta) \\
&W_H(10 \mid 01) = h(3\eta + v)/h(4\eta) \\
&W_H(01 \mid 10) = h(3\eta - v)/h(2\eta)
\end{aligned} \tag{1.3.1}$$

where

$$\eta = K/5 \tag{1.3.2}$$

We can convert this to an SOS model satisfying (1.2.2). For each occupation number σ_i define a height l_i by

$$\begin{aligned} l_i &= 3 - 2\sigma_i, && \text{if } i \text{ is on sublattice } X \\ &= 2\sigma_i + 2, && \text{if } i \text{ is on sublattice } Y \end{aligned} \tag{1.3.3}$$

It then follows that the restrictions $\sigma_i, \sigma_j = 0, 1$ and $\sigma_i\sigma_j = 0$ are equivalent to

$$1 \leqslant l_i, l_j \leqslant 4, \qquad |l_i - l_j| = 1 \tag{1.3.4}$$

for all pairs (i, j) of adjacent sites. Further, defining

$$W(l_i, l_j \mid l_m, l_n) = W_H(\sigma_i, \sigma_j \mid \sigma_m, \sigma_n) \tag{1.3.5}$$

and using the relations

$$h(-u) = -h(u) = h(2K + u) \tag{1.3.6}$$

we find that this function W is precisely the same as that given by (1.2.12), with

$$\rho' = [h(2\eta)]^{-1}, \qquad w_0 = 0 \tag{1.3.7}$$

The definitions (1.3.3) ensure that l_i is odd on an X site, even on a Y site, so we have a sub-SOS model. However, we could just as easily have interchanged the X and Y sublattices, and obtained the other SOS model [with the same weights, and still satisfying (1.3.4)]. Altogether, it follows that

$$Z_{\text{SOS}} = 2Z_{HH} \tag{1.3.8}$$

where Z_{HH} is the hard-hexagon partition function and Z_{SOS} is the eight-vertex SOS partition function, given by (1.2.1), (1.2.12) and (1.3.7), with each height l_i restricted to the values $1, 2, 3, 4$.

Restricted and Unrestricted SOS Models. Let us call this SOS model, where $1 \leqslant l_i \leqslant 4$, the "restricted" SOS model. It is not the same as the unrestricted model discussed in Section 1.2, where $-\infty < l_i < \infty$, but it is closely related.

Consider the unrestricted model. In particular, consider two successive rows R and R' of $\mathscr{L}$, and the product P of the weights of the intervening faces. Suppose that all the heights on R lie between 1 and 4. Then $P = 0$ unless the heights on R' also all lie between 1 and 4. To see this, note that if R' is the upper row, then it can only contain a height 0 or 5 via the configurations shown in Fig. 3. These contain weights $W(21 \mid 10)$, $W(34 \mid 45)$, respectively, and from (1.2.12) these vanish because

$$h(w_0) = h(0) = 0, \qquad h(w_5) = h(10\eta) = h(2K) = 0 \tag{1.3.9}$$

Similar considerations apply if R' is the lower row.

If all the weights were finite, this would mean that the transfer matrix of the unrestricted SOS model broke up into diagonal blocks, one of which

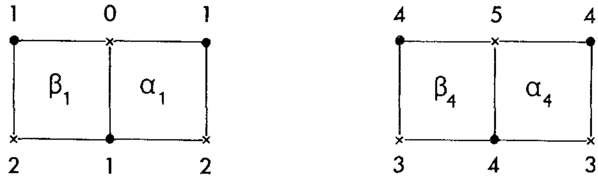

Fig. 3. Configurations containing heights 0 and 5 in the upper row, and only heights 1, 2, 3, 4 in the lower row.

would be the restricted SOS model transfer matrix. More precisely, this would happen if $\eta = K/5$ and w_0 is allowed to tend to zero.

In fact the situation is more complicated than this, since some of the other transfer matrix elements [e.g., those containing a factor $W(4,5\,|\,5,4)$] then become infinite. Even so, we have still been able to verify for a two-column lattice that the relevant eigenvalues and eigenvector elements of the unrestricted model become those of the restricted one. This suggests that the restricted and unrestricted SOS models may have at least some transfer matrix eigenvalues in common. This would be consistent with the fact that the hard hexagon model (in its regime I, III, and IV) has the same free energy per site as the corresponding eight-vertex model.[21]

Another argument in favor of this assertion is to compare the equations that determine the transfer matrix eigenvalues of the unrestricted SOS model and the hard hexagon model. Each eigenvalue Λ is an entire function of the variable v. For the SOS model, it follows from Eq. (1.23) of Ref. 13 that each $\Lambda(v)$ satisfies the functional equation

$$\Lambda(v)\,Q(v) = z\phi(v-\eta)\,Q(v+2\eta) + z^{-1}\phi(v+\eta)\,Q(v-2\eta) \qquad (1.3.10)$$

where z is the parameter discussed in Section 1.2,

$$\phi(v) = \left[\rho\Theta(0)h(v)\right]^{N} \qquad (1.3.11)$$

N is the (even) number of columns of the lattice $\mathscr{L}$, and

$$Q(v) = \prod_{j=1}^{N/2} h(v-u_j) \qquad (1.3.12)$$

$u_1, \ldots, u_{N/2}$ being constants.

Writing (1.3.10) with v replaced by $v + 2k\eta$, where $k = 0, 1, \ldots, 4$, noting that

$$Q(v+10\eta) = (-1)^{N/2}\,Q(v) \qquad (1.3.13)$$

we obtain five homogeneous linear equations for $Q(v+2k\eta)$, $k = 0, 1, \ldots, 4$. Eliminating these gives a 5th-degree equation for the function

$\Lambda(v)$, and it can be verified that this equation is automatically satisfied if

$$\Lambda(v)\Lambda(v-2\eta)=\phi(v+\eta)\phi(v-3\eta)+\phi(v-\eta)\Lambda(v+4\eta) \quad (1.3.14)$$

provided

$$z^5=-(-1)^{N/2} \quad (1.3.15)$$

(In fact the 5×5 matrix of coefficients then has rank of only 3.)

However, (1.3.14) is precisely the equation satisfied by the eigenvalues of the transfer matrix of the hard hexagon model [Eq. (3.3) of Ref. 20, using (1.3.7) and the transformations mentioned prior to Eq. (1.3.1) herein]. Thus there may be (and almost certainly are) eigenvalues satisfying both (1.3.10) and (1.3.14), provided z in the unrestricted SOS model is given by (1.3.15).

1.4. Restricted SOS Model with $\eta = K/r$

From now on we consider in this paper an SOS model satisfying (1.2.2), (1.2.5), and (1.2.12), with

$$\eta = K/r \quad (1.4.1)$$

$$w_0 = 0 \quad (1.4.2)$$

and each height l_i restricted to the integer values

$$l_i = 1, 2, \ldots, r-1 \quad (1.4.3)$$

Here r is an arbitrary integer not less than 3.

For $r=5$ we regain the restricted SOS model which we have shown to be equivalent to the hard hexagon model [more precisely, to the hard square model with diagonal interactions that has weights given by (1.3.1)]. Thus the present model is a natural generalization of the hard hexagon model. It has the property that $h(w_l)$ is nonzero for $l=1,\ldots,r-1$, while

$$h(w_0)=h(w_r)=0 \quad (1.4.4)$$

This means that the weights (1.2.12) are finite (and in general nonzero) provided the restriction (1.4.3) is satisfied.

As with the $r=5$ case, it is likely that this model is in some sense equivalent to the unrestricted model with $-\infty < l_i < \infty$, and hence to the eight-vertex model. However, we shall not need this equivalence. All we need is the fact that the restricted model, like the unrestricted one, satisfies the star–triangle relation [Eq. (13.3.6) of Ref. 10, (2.49) of Ref. 22]:

$$\sum_g W(b,c\,|\,a,g)W'(a,g\,|\,f,e)W''(g,c\,|\,e,d)$$
$$=\sum_g W''(a,b\,|\,f,g)W'(b,c\,|\,g,d)W(g,d\,|\,f,e) \quad (1.4.5)$$

Here W is the weight function defined by (1.2.2), (1.2.5), (1.2.12); W' and

W'' are defined similarly, but with v replaced by v' and v'', respectively, where

$$v' = v + v'' - \eta \tag{1.4.6}$$

[If we set $u = \eta - v$, then we regain the usual relation $u' = u + u''$: Eq. (13.3.10) of Ref. 10.]

The equation (1.4.5) is to be true for all values of the six heights a, b, c, d, e, f. Since W, W', W'' each vanish unless their arguments satisfy (1.6), we must have $|a - b| = |b - c| = |c - d| = |d - e| = |e - f| = |f - a| = 1$. Apart from a uniform additive shift of all the heights, this means that there are 20 cases to consider. These occur in pairs, one being obtained from the other by interchanging a with d, b with e, and c with f, which merely interchanges the two sides of the equation (1.4.5).

It follows that (1.4.5) consists of ten distinct equations. Taking $\alpha_l', \ldots, \delta_l'$ to be the values of $\alpha_l, \ldots, \delta_l$ in (1.2.12) when W, v are replaced by W', v'; and similarly for $\alpha_l'', \ldots, \delta_l''$; seven of the equations are

$$\begin{aligned}
\beta_l\alpha_l'\beta_l'' + \gamma_l\delta_{l+1}'\gamma_l'' &= \delta_{l+1}\gamma_l'\delta_{l+1}'' + \beta_{l+1}\alpha_{l+1}'\beta_{l+1}'' \\
\beta_l\gamma_{l-1}'\beta_l'' + \gamma_l\alpha_l'\gamma_l'' &= \alpha_l\gamma_l'\alpha_l'' \\
\beta_l\alpha_l'\delta_l'' + \gamma_l\delta_{l+1}'\beta_l'' &= \alpha_l\beta_l'\delta_{l+1}'' \\
\delta_l\alpha_l'\delta_l'' + \beta_l\delta_{l+1}'\beta_l'' &= \alpha_l\delta_l'\alpha_l'' \\
\delta_l\gamma_{l-1}'\beta_l'' + \beta_l\alpha_l'\gamma_l'' &= \alpha_l\beta_l'\gamma_{l-1}'' \\
\alpha_{l+1}\alpha_l'\alpha_{l+1}'' &= \alpha_l\alpha_{l+1}'\alpha_l'' \\
\alpha_{l+1}\beta_{l+1}'\beta_l'' &= \alpha_l\beta_l'\beta_{l+1}''
\end{aligned} \tag{1.4.7}$$

The other three equations can be obtained by interchanging the unprimed and double-primed symbols.

For the unrestricted model, l in (1.4.7) takes all integer values and we can verify directly, using (1.2.12b), that the equations are satisfied.

For the restricted model, each of $a, b, \ldots, f$ must lie in the interval $1, \ldots, r-1$. This means that l takes the values $1, \ldots, r-2$ in the first equation in (1.4.7), the values of $2, \ldots, r-2$ in the next four equations, and the values $2, \ldots, r-3$ in the last two. Further, the height g in (1.4.5) must also lie in the interval $1, \ldots, r-1$: this means that in the first equation the terms

$$\beta_1\alpha_1'\beta_1'', \qquad \beta_{r-1}\alpha_{r-1}'\beta_{r-1}'' \tag{1.4.8}$$

should be deleted. (They occur for $l = 1$ and $l = r-2$, and correspond to $g = 0, r$, respectively.)

However, from (1.2.12b) and (1.4.4) it is clear that the terms (1.4.8) vanish for the unrestricted model. It follows that (1.4.7), and hence the star–triangle relation (1.4.5), is indeed satisfied by the restricted SOS model.

The Case When r Is Odd. The model has some special features when r is odd, i.e., when there exists an integer n such that

$$r = 2n + 1 \tag{1.4.9}$$

(Note that the hard hexagon model is such a case, with $n = 2$.) In this case, there are as many even heights in the allowed range (1.4.3) as there are odd heights. Each of the two sub-SOS models can then naturally be expressed as a "lattice gas" generalization of the hard hexagon model. For the one with odd heights on X sites and even heights on Y sites, define σ_i by

$$\begin{aligned} \sigma_i &= \tfrac{1}{2}(2n - 1 - l_i), \qquad &&\text{if} \quad i \in X \\ &= \tfrac{1}{2}(l_i - 2), &&\text{if} \quad i \in Y \end{aligned} \tag{1.4.10}$$

Then the σ_i are integers with permitted values $0, 1, \ldots, n-1$, satisfying

$$\sigma_i + \sigma_j = n - 2 \quad \text{or} \quad n - 1 \tag{1.4.11}$$

for all adjacent pairs (i, j) of sites. We can regard these σ_i as "occupation numbers" and use these to define the states of the system.

The SOS model weight function W is the same for all faces of the lattice $\mathscr{L}$, but the mapping (1.4.10) is not: it depends on whether the face has an X or a Y site at its lower-left corner (if $\mathscr{L}$ were a checkerboard it would depend on whether the face were black or white). It follows that W, when regarded as a function of the occupation numbers σ_i, may similarly depend on the face under consideration. In fact this does not occur, because the function W defined by (1.2.12), (1.4.1), and (1.4.2) has the symmetry property

$$W(r - l, r - m' \mid r - l', r - m) = W(l, m' \mid l', m) \tag{1.4.12}$$

Thus when (1.4.9) is satisfied, each sub-SOS model is equivalent to a uniform lattice gas in which there must be either $n - 2$ or $n - 1$ particles on every pair of adjacent sites.

1.5. Expressions for the Local Height Probabilities

The Rogers–Ramanujan identities naturally enter the calculation of the sublattice densities of the hard hexagon model.[9] For the more general case of our restricted SOS model, the analogs of these densities are the local height probabilities:

$$P_a = Z^{-1} \sum \delta(l_1, a) \prod_{(i,j,n,m)} W(l_i, l_j \mid l_m, l_n) \tag{1.5.1}$$

where Z is the partition function defined by (1.2.1) and the sum and product have the same meanings as therein, l_1 is the height of the center site

1 of the lattice, and a is an integer between 1 and $r-1$. Thus P_a is the probability that site 1 has height a, and it follows from (1.2.1) and (1.5.1) that

$$\sum_{a=1}^{r-1} P_a = 1 \tag{1.5.2}$$

In evaluating the probability P_a, we suppose initially that the lattice $\mathscr{L}$ is planar (i.e., *not* wound on a cylinder or torus) and finite, and fix the boundary heights to have the values they would assume in a particular ground state configuration [i.e., a set of values of the heights that maximizes the summand in (1.2.1)]. Finally, we shall take the limit when the lattice becomes infinitely large, all boundary sites being infinitely far from the center site 1.

For definiteness, let us suppose that site 1 lies on the X sublattice, and refer to sub-SOS models, states, and boundary conditions as "even" if they have even heights on X sites, odd heights on Y sites. If the converse, then "odd." For every even ground state, an odd one can be formed by interchanging the X and Y sublattices. Thus the ground states of the total SOS model can be divided into equal numbers of even and odd ground states.

P_a will depend on the particular ground state that is chosen for the boundary conditions. One significant but trivial dependence is that if the ground state is even, clearly

$$P_a = 0, \qquad \text{unless } a \text{ is even} \tag{1.5.3}$$

Thus in this case the summation in (1.5.2) can be restricted to even values of a. (In the other case it can be restricted to odd values.)

If there is only one even (and one odd) ground state, then we expect the system to be "disordered." This means that in the limit of a large lattice we expect P_a to be independent of the boundary conditions, so long as they are even. If there is more than one even ground state, then the system is "ordered" and P_a depends on the boundary conditions.

As with the hard hexagon model, we can calculate P_a by using corner transfer matrices.[10,23] This is done in Appendix A, attention being focused on the cases when v is real and $-\eta < v < 3\eta$. These are the cases when the weight function W is positive, or can be made so by an appropriate choice of ρ' and by applying a transformation of the type (1.2.14) with $x = i$, i.e., by negating β_l.

We have to distinguish the cases when v is greater or less than η, and when p in (1.2.7) is positive or negative. Altogether this gives us four "regimes" to consider. Labeling them analogously to those of the hard hexagon model, they are

$$\begin{array}{lll} \text{regime I:} & -1 < p < 0, & \eta < v < 3\eta \\ \text{regime II:} & 0 < p < 1, & \eta < v < 3\eta \\ \text{regime III:} & 0 < p < 1, & -\eta < v < \eta \\ \text{regime IV:} & -1 < p < 0, & -\eta < v < \eta \end{array} \tag{1.5.4}$$

It is convenient to define an integer t such that

$$\begin{aligned} t &= 2 - r, && \text{in regimes I and II} \\ &= 2, && \text{in regimes III and IV} \end{aligned} \tag{1.5.5}$$

and to define a function $E(z,x)$ by

$$E(z,x) = \prod_{n=1}^{\infty} (1 - x^{n-1}z)(1 - x^n z^{-1})(1 - x^n) \tag{1.5.6}$$

for all complex numbers z, x such that $|x| < 1$. This is basically an elliptic theta function, and by Jacobi's triple product identity (Theorem 2.8 of Ref. 16), it has the series expansion:

$$E(z,x) = \sum_{n=-\infty}^{\infty} (-1)^n x^{n(n-1)/2} z^n \tag{1.5.7}$$

Regimes II and III. The parameter p is positive in regimes II and III, so we can define ϵ and x (both real and positive) so that

$$p = e^{-\epsilon}, \qquad x = \exp\left[-4\pi^2/(r\epsilon)\right] \tag{1.5.8}$$

We find in Appendix A that

$$P_a = S^{-1} u_a X_m(a,b,c;x^t) \tag{1.5.9}$$

for $a = 1, \ldots, r-1$, where

$$u_a = x^{(2-t)(2a-r)^2/(16r)} E(x^a, x^r) \tag{1.5.10}$$

$$X_m(a,b,c;q) = \sum_{l_2, \ldots, l_m} q^{\phi(\mathbf{l})} \tag{1.5.11}$$

$$\phi(\mathbf{l}) = \sum_{j=1}^{m} j \,| l_{j+2} - l_j | /4 \tag{1.5.12}$$

$$S = \sum_{1 \leqslant a < r} u_a X_m(a,b,c;x^t) \tag{1.5.13}$$

Here $\mathbf{l} = \{l_1, l_2, \ldots, l_{m+2}\}$ is a set of integer heights satisfying the restrictions

$$1 \leqslant l_j \leqslant r-1, \qquad |l_{j+1} - l_j| = 1 \tag{1.5.14a}$$

for $j \geqslant 1$. The summation in (1.5.11) is over all allowed values of

$l_2, \ldots, l_m$, the end heights l_1, l_{m+1}, l_{m+2} being fixed at the values

$$l_1 = a, \qquad l_{m+1} = b, \qquad l_{m+2} = c \tag{1.5.14b}$$

Thus a, b, c must all lie in the range $1, 2, \ldots, r-1$, and must satisfy

$$|b-c| = 1, \qquad m + a - b = \text{even} \tag{1.5.15}$$

We can verify the symmetry properties

$$u_{r-a} = u_a, \qquad X_m(r-a, r-b, r-c; q) = X_m(a, b, c; q) \tag{1.5.16}$$

which are consequences of the symmetry (1.4.12) of W.

The heights $l_1, \ldots, l_{m+2}$ correspond to the heights of sites on the center row of $\mathscr{L}$, starting at the center site 1 and moving rightwards to the boundary. Thus l_{m+1} and l_{m+2}, i.e., b and c, are boundary heights, and are to be fixed according to the ground state chosen for the boundary conditions.

In a ground state, the values of $l_1, \ldots, l_{m+2}$ are such as to maximize the summand in (1.5.11). In regime III, t is positive, so $|q| < 1$ and the ground states are obtained by minimizing $\phi(\mathbf{l})$. Plainly this is achieved by taking

$$\begin{aligned} l_j &= l_X, && \text{if } j \text{ is odd} \\ &= l_Y, && \text{if } j \text{ is even} \end{aligned} \tag{1.5.17}$$

where l_X, l_Y are any two integers satisfying $|l_X - l_Y| = i$, $1 \leqslant l_X, l_Y \leqslant r-1$. There are $2r-4$ such values of l_X and l_Y, so there are $2r-4$ ground states, of which $r-2$ are even and $r-2$ are odd. (For the hard hexagon model, with $r=5$, this gives $5-2=3$ ground states, in agreement with the observation of Huse[24].)

We remark in Appendix A that all ground states (for all regimes) are invariant under uniform shifts in the southwest to northeast diagonal direction. We can use this to obtain the ground state heights on all rows of $\mathscr{L}$, given the heights $l_1, \ldots, l_{m+2}$ on the center row. From (1.5.17) it follows that in regime III a ground state has

$$\begin{aligned} l_j &= L_X, && \text{if } j \in X \\ &= l_Y, && \text{if } j \in Y \end{aligned} \tag{1.5.18}$$

for all sites j of the lattice.

In regime II, t is negative and the ground states maximize $\phi(\mathbf{l})$. This is achieved by taking the l_j (for the linear sequence $l_1, \ldots, l_{m+2}$) to increase in unit steps to $r-1$, then decrease to 1, then increase again to $r-1$, and so on. More precisely, let j_0 be any integer from 1 to $2r-4$ and, for given j, let

$$k = j - j_0, \qquad \operatorname{mod}(2r-4) \tag{1.5.19a}$$

(thus $0 \leqslant k < 2r - 4$). Then

$$l_j = 1 + |k - r + 2| \tag{1.5.19b}$$

There are $2r - 4$ such sequences, depending on the value of j_0, so there are $2r - 4$ ground states, as in regime III.

In these regimes II and III there are as many ground states as there are permitted values of b and c, with $|b - c| = 1$. Thus we can either select a ground state and then evaluate b and c from (1.5.14b), or we can choose b and c and then determine the corresponding ground state.

As remarked above, we ultimately want to calculate P_a in the limit of an infinitely large lattice. This corresponds to letting $m \to \infty$ in (1.5.9)–(1.5.15), while keeping the ground state [i.e., l_X and l_Y in (1.5.17), or j_0 in (1.5.19)] fixed.

Regimes I and IV. In these regimes p is negative. We now define ϵ and x (both real and positive) by

$$p = -e^{-\epsilon}, \qquad x = \exp(-2\pi^2/r\epsilon) \tag{1.5.20}$$

(For the hard hexagon case, this x is the square of that used in Refs. 8, 9, and 20.) We find that

$$P_a = T^{-1} v_a Y_m(a,b,c;x^{-t}) \tag{1.5.21}$$

for $a = 1, \ldots, r-1$, where

$$v_a = x^{[(t-1)r-(t-2)a]a/2r} E(x^a, -x^{r/2}) \tag{1.5.22}$$

$$Y_m(a,b,c;q) = \sum_{l_2,\ldots,l_m} q^{\psi(\mathbf{l})} \tag{1.5.23}$$

$$\psi(\mathbf{l}) = \sum_{j=1}^{m} jH(l_j, l_{j+1}, l_{j+2}) \tag{1.5.24}$$

$$T = \sum_{1 \leqslant a < r} v_a Y_m(a,b,c;x^{-t}) \tag{1.5.25}$$

Here $\mathbf{l}$ and the $l_2, \ldots, l_m$ summation have the same meaning as in (1.5.14). The function $H(l,l',l'')$ is defined by

$$H(l,l',l'') = 0, \qquad \text{if} \quad l \neq l'' \tag{1.5.26a}$$

$$\begin{aligned} H(l,l+1,l) &= 0, && \text{if} \quad l \leqslant r/2 \\ &= 1, && \text{if} \quad l > r/2 \end{aligned} \tag{1.5.26b}$$

$$\begin{aligned} H(l,l-1,l) &= 1, && \text{if} \quad l \leqslant r/2 \\ &= 0, && \text{if} \quad l > r/2 \end{aligned} \tag{1.5.26c}$$

These formulas (1.5.20)–(1.5.26) apply regardless of whether r is even or odd, but only when r is odd do we have both the symmetries

$$v_{r-a} = v_a, \qquad Y_m(r-a, r-b, r-c; q) = Y_m(a,b,c;q) \quad (1.5.27)$$

[which are analogous to (1.5.16) and consequences of (1.4.12)]. Further, there are complications in discussing the ground states when r is even. (Both these difficulties arise because the corner transfer matrices are not diagonal in the limit $x \to 0$ when r is even: the necessary diagonalization means that if $l_{j-1} = l_{j+1} = r/2$, then l_j is merely an eigenvalue label and does not correspond to a height on the lattice.)

For r odd, the ground states are then of the form given by (1.5.17) and (1.5.18). For regime I we want to minimize $\psi(\mathbf{l})$, which corresponds to taking

$$(l_X, l_Y) = (n, n+1) \quad \text{or} \quad (n+1, n), \qquad \text{where} \quad n = (r-1)/2 \quad (1.5.28)$$

Thus there are just two ground states, one for each sub-SOS model, and we expect the system to be disordered.

For regime IV we want to maximize $\psi(\mathbf{l})$, which corresponds to taking l_X, l_Y to be any pair of integers in the range $1, \ldots, r-1$ satisfying $|l_X - l_Y| = 1$, except for the values (1.5.29). Thus there are $2r-6$ ground states in regime IV.

2. EVALUATION OF THE PROBABILITIES P_a

2.1. General Comments

We have obtained the expressions (1.5.9), (1.5.21) for the local height probabilities P_a, involving the m-fold sums X_m and Y_m defined by (1.5.11) and (1.5.23). Our object in this part is to obtain more tractable expressions for X_m and Y_m, considered as functions of their argument q; and to study the limit $m \to \infty$. In this limit we find that X_m and Y_m are actually modular forms (p. 114 of Ref. 25). Indeed the results in regimes I, III, and IV are sums of at most two simple quotients of elliptic theta functions (Ref. 18, Chap. 21). Regime II yields multidimensional theta series.

Our approach is to first keep m finite, and show that X_m and Y_m can be written as sums of Gaussian polynomials. This treatment has its genesis in Ref. 26, which in turn was suggested by I. Schur's original treatment of the Rogers–Ramanujan–Schur identities.[27] While this approach may be more cumbersome than the one used for regimes I, III, and IV of the original hard hexagon model (Ref. 10, pp. 432–443), it does seem essential in this more general setting. In particular the classical theory of q-difference equations and q series[28,29] appears barely adequate (see Appendix B) to handle regime II via polynomials. Unfortunately we do not know what the

Rogers–Ramanujan-type "series product" identities are for either regime II or IV except when $r = 5$ [Eqs. (14.5.22) and (14.5.50) of Ref. 10]. We hope to discuss the Rogers–Ramanujan aspects of regimes I and III in a subsequent paper; indeed the appropriate identities for regime I are given in Ref. 30 and in Eq. (7.3.7) of Ref. 16.

2.2. Gaussian Polynomials

The Gaussian polynomials (which we shall use for calculating X_m, Y_m) are defined as

$$\begin{aligned} \begin{bmatrix} N \\ M \end{bmatrix} \equiv \begin{bmatrix} N \\ M \end{bmatrix}_q &= \prod_{j=1}^{M} \frac{1-q^{N-M+j}}{1-q^j}, \qquad 0 \leqslant M \leqslant N \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{2.2.1}$$

We shall use the subscript q only when confusion might otherwise arise.

It is easy to prove (by mathematical induction on N) that these are indeed polynomials in q. One merely utilizes either of the fundamental recurrences (p. 35 of Ref. 16)

$$\begin{bmatrix} N \\ M \end{bmatrix} - \begin{bmatrix} N-1 \\ M \end{bmatrix} = q^{N-M} \begin{bmatrix} N-1 \\ M-1 \end{bmatrix} \tag{2.2.2}$$

$$\begin{bmatrix} N \\ M \end{bmatrix} - \begin{bmatrix} N-1 \\ M-1 \end{bmatrix} = q^{M} \begin{bmatrix} N-1 \\ M \end{bmatrix} \tag{2.2.3}$$

In addition to the recurrences, we need that

$$\begin{bmatrix} N \\ M \end{bmatrix} = \begin{bmatrix} N \\ N-M \end{bmatrix} \tag{2.2.4}$$

$$\begin{bmatrix} N \\ M \end{bmatrix}_{q^{-1}} = q^{-M(N-M)} \begin{bmatrix} N \\ M \end{bmatrix}_q \tag{2.2.5}$$

and for $|q| < 1$,

$$\lim_{M,N \to \infty} \begin{bmatrix} N+M \\ M \end{bmatrix} = Q(q)^{-1} \tag{2.2.6}$$

where

$$Q(q) = \prod_{n=1}^{\infty} (1-q^n) \tag{2.2.7}$$

We shall also use the notations

$$(A)_n = (A;q)_n = \prod_{j=0}^{n-1} (1-q^j A). \tag{2.2.8}$$

Thus in particular

$$(q)_\infty = \prod_{j=0}^{\infty} (1 - q^{j+1}) = Q(q) \tag{2.2.9}$$

Two simple identities that we shall use are

$$(A)_{m+n} = (A)_m (q^m A)_n \tag{2.2.10}$$

$$(A)_n = (-A)^n q^{(1/2)n(n-1)} (q^{1-n} A^{-1})_n \tag{2.2.11}$$

2.3. Regime III

We shall consider the regimes in order of difficulty: regime III is the easiest to handle and provides a good prototype for subsequent work.

The function $X_m(a,b,c;q)$, or simply $X_m(a,b,c)$, is defined by (1.5.11), (1.5.12), and (1.5.14). By examining admissible sequences of $\{l_2, \ldots, l_m\}$ and splitting them into two classes according to whether $l_m = b - 1$ or $l_m = b + 1$, we see that the $X_m(a,b,c)$ are totally defined by the following recurrences and initial values ($m > 0$, $1 \leqslant a \leqslant r - 1$):

$$X_m(a,b,b+1) = X_{m-1}(a,b+1,b) + q^{m/2} X_{m-1}(a,b-1,b), \qquad 1 \leqslant b \leqslant r-2 \tag{2.3.1}$$

$$X_m(a,b,b-1) = X_{m-1}(a,b-1,b) + q^{m/2} X_{m-1}(a,b+1,b), \qquad 2 \leqslant b \leqslant r-1 \tag{2.3.2}$$

$$X_m(a,0,1) = X_m(a,r,r-1) = 0 \tag{2.3.3}$$

$$\begin{aligned} X_0(a,b,c) &= 1, \qquad \text{if} \quad a = b \quad \text{and} \quad c = b \pm 1 \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{2.3.4}$$

Our main object is to show that as $m \to \infty$, X_m converges to a difference of two theta series divided by $Q(q)$.

Theorem 2.3.1. For $m \geqslant 0$, $1 \leqslant a,b,c < r$, $c = b \pm 1$, $m + a - b$ an even integer,

$$X_m(a,b,c) = q^{(a^2-a)/4} \{ F_m(a,b,c) - F_m(-a,b,c) \} \tag{2.3.5}$$

where

$$F_m(a,b,c) = \sum_{\lambda=-\infty}^{\infty} q^{(r-1)\lambda(r\lambda - a) + [bc + (2r\lambda - a)(b+c-1)]/4} \begin{bmatrix} m \\ \frac{1}{2}(m+a-b) - r\lambda \end{bmatrix} \tag{2.3.6}$$

[Because of the definition (2.2.1), for given m, a, b there are only a finite number of nonzero terms in the λ summation.]

Proof. Since (2.3.1)–(2.3.4) uniquely define the $X_m(a,b,c)$, we need only show that the right-hand side of (2.3.5) satisfies the same conditions. Now

$$F_m(a,b,b+1) - F_{m-1}(a,b+1,b)$$
$$= \sum_{\lambda=-\infty}^{\infty} q^{(r-1)\lambda(r\lambda-a)+b(b+1+4r\lambda-2a)/4}$$
$$\times \left\{ \begin{bmatrix} m \\ \frac{1}{2}(m+a-b)-r\lambda \end{bmatrix} - \begin{bmatrix} m-1 \\ \frac{1}{2}(m+a-b)-1-r\lambda \end{bmatrix} \right\} \tag{2.3.7}$$

Using (2.2.3) to simplify the bracketted difference of Gaussian polynomials, it follows at once that

$$F_m(a,b,b+1) - F_{m-1}(a,b+1,b) = q^{m/2} F_{m-1}(a,b-1,b) \tag{2.3.8}$$

Similarly,

$$F_m(a,b,b-1) - F_{m-1}(a,b-1,b)$$
$$= \sum_{\lambda=-\infty}^{\infty} q^{(r-1)\lambda(r\lambda-a)+(b-1)(b+4r\lambda-2a)/4}$$
$$\times \left\{ \begin{bmatrix} m \\ \frac{1}{2}(m+a-b)-r\lambda \end{bmatrix} - \begin{bmatrix} m-1 \\ \frac{1}{2}(m+a-b)-r\lambda \end{bmatrix} \right\} \tag{2.3.9}$$

and on using (2.2.2) we obtain

$$F_m(a,b,b-1) - F_{m-1}(a,b-1,b) = q^{m/2} F_{m-1}(a,b+1,b) \tag{2.3.10}$$

Thus $F_m(a,b,c)$ satisfies the linear recurrence relations (2.3.1) and (2.3.2). Since these are independent of a, so does $F_m(-a,b,c)$, and hence the expression (2.3.5) for $X_m(a,b,c)$.

It remains to verify that (2.3.3) and (2.3.4) are satisfied. From (2.3.6)

$$F_m(a,0,1) = \sum_{\lambda=-\infty}^{\infty} q^{(r-1)\lambda(r\lambda-a)} \begin{bmatrix} m \\ \frac{1}{2}(m+a)-r\lambda \end{bmatrix} \tag{2.3.11}$$

$$F_m(a,r,r-1) = \sum_{\lambda=-\infty}^{\infty} q^{(r-1)(\lambda+1/2)[r(\lambda+1/2)-a]} \begin{bmatrix} m \\ \frac{1}{2}(m+a)-r(\lambda+1/2) \end{bmatrix} \tag{2.3.12}$$

Using (2.2.4) and replacing λ by $-\lambda$, $-1-\lambda$, respectively, in these two equations, we find that

$$F_m(a,0,1) = F_m(-a,0,1), \qquad F_m(a,r,r-1) = F_m(-a,r,r-1) \tag{2.3.13}$$

From (2.3.5) it follows immediately that (2.3.3) is satisfied.

Finally, from (2.2.1) and (2.3.6), $F_0(a,b,c)$ will be zero unless $a \equiv b \pmod{2r}$. Remembering that $1 \leqslant a, b < r$, from (2.3.5) we obtain

$$X_0(a,b,c) = q^{(a^2-a)/4} F_0(a,a,c)\delta_{a,b} \tag{2.3.14}$$

where, as in part I,

$$\begin{aligned} \delta_{a,b} \equiv \delta(a,b) &= 1, \qquad \text{if} \quad a = b \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{2.3.15}$$

For $F_0(a,b,c)$, the only nonzero term in the summation in (2.3.6) is that with $\lambda = 0$; hence

$$F_0(a,a,c) = q^{(a-a^2)/4} \tag{2.3.16}$$

It follows that (2.3.4) is also satisfied, which completes the proof of Theorem 2.3.1. ■

We now go on to consider the limit $m \to \infty$. From (1.5.9), the argument q of the function $X_m(a,b,c;q)$ is $q = x^t$. From (1.5.5), $t = 2$ in regime III, so here we have

$$|q| < 1 \tag{2.3.17}$$

We shall need the function $\Delta(a,d;q)$, defined for $1 \leqslant a \leqslant r-1$, $1 \leqslant d \leqslant r-2$ by

$$\begin{aligned} \Delta(a,d;q) = \sum_{\lambda=-\infty}^{\infty} & q^{r(r-1)\lambda^2 + rd\lambda + a(a-1)/4} \\ & \times \left\{ q^{-(r-1)a\lambda - (1/2)ad} - q^{(r-1)a\lambda + (1/2)ad} \right\} \end{aligned} \tag{2.3.18}$$

i.e., using (1.5.7),

$$\begin{aligned} \Delta(a,d;q) = q^{a(a-1)/4} \Big\{ & q^{-(1/2)ad} E\left[-q^{(r-a)(r-1)+rd}, q^{2r(r-1)} \right] \\ & - q^{(1/2)ad} E\left[-q^{(r+a)(r-1)+rd}, q^{2r(r-1)} \right] \Big\} \end{aligned} \tag{2.3.19}$$

Theorem 2.3.2. For $1 \leqslant a, b, c < r$, $c = b \pm 1$,

$$\lim_{m\to\infty} X_m(a,b,c) = Q(q)^{-1} q^{bc/4} \Delta(a, \tfrac{1}{2}(b+c-1); q) \tag{2.3.20}$$

where the limit is taken through values of m with same parity as $a - b$.

Proof. This is a direct consequence of Theorem 2.3.1. One merely applies (2.2.6) to the Gaussian polynomial in (2.3.6), then uses (2.3.5) and (1.5.7). The passage to a limit inside the summation in (2.3.6) is easily justified due to the uniform boundedness of the Gaussian polynomial inside $|q| \leqslant 1 - \epsilon$ and the rapid absolute convergence of the λ series caused by the exponent on q being quadratic in λ.

We must distinguish the case in which the large-m limit is taken through even values of $m - b$ from that in which it is taken through odd values. In the former case a must be an even integer (we have the "even sub-SOS model" discussed in Section 1); in the latter case a is an odd integer (the "odd sub-SOS model").

Substituting (2.3.18) into (1.5.9) and (1.5.13), using (1.5.5), we find that the local height probability P_a tends to a limit as $m \to \infty$, being given by

$$P_a = u_a \Delta(a, d; x^2) \Big/ \sum_{1 \leqslant a' \leqslant r-1} u_{a'} \Delta(a', d; x^2) \tag{2.3.21}$$

where

$$d = \tfrac{1}{2}(b + c - 1), \qquad 1 \leqslant d \leqslant r - 2 \tag{2.3.22}$$

and a, a' are either both even integers satisfying $1 \leqslant a, a' \leqslant r - 1$ [this is the "even" case (1.5.3)], or are both odd integers satisfying $1 \leqslant a, a' \leqslant r - 1$.

Allowing for these even and odd cases, and noting that the integer d takes $r - 2$ values, we see that there are $2r - 4$ different functions P_a given by (2.3.21). These correspond to the $2r - 4$ ground states (1.5.17). ■

2.4. Regime I

The results for this regime closely parallel those of regime III. Our work is slightly more complicated owing to the intricacy of the summands in the definition (1.5.23)–(1.5.26) of the relevant polynomials $Y_m(a, b, c; q)$.

As before, we examine admissible sequences of $l_2, \ldots, l_m$ satisfying (1.5.14), and split them into two classes according to whether $l_m = b - 1$ or $l_m = b + 1$. Writing the polynomials $Y_m(a, b, c; q)$ simply as $Y_m(a, b, c)$ we find that they are completely defined by the following recurrences and initial values ($m > 0$, $1 \leqslant a < r$):

$$Y_m(a, b, b+1) = Y_{m-1}(a, b-1, b) + q^{m\chi(b+1)} Y_{m-1}(a, b+1, b), \qquad 1 \leqslant b \leqslant r - 2 \tag{2.4.1}$$

$$Y_m(a, b, b-1) = Y_{m-1}(a, b+1, b) + q^{m[1-\chi(b-1)]} Y_{m-1}(a, b-1, b), \qquad 2 \leqslant b \leqslant r - 1 \tag{2.4.2}$$

$$Y_m(a, 0, 1) = Y_m(a, r, r-1) = 0 \tag{2.4.3}$$

$$Y_0(a, b, b \pm 1) = \delta_{a,b} \tag{2.4.4}$$

where $\chi(i)$ is the characteristic function of all numbers $\leqslant n$:

$$\begin{aligned} \chi(i) &= 1, \qquad \text{if} \quad i \leqslant n \\ &= 0, \qquad \text{if} \quad i > n \end{aligned} \tag{2.4.5}$$

and n is the integer part of $r/2$, i.e.,

$$\begin{aligned} n &= (r-1)/2, && \text{if } r \text{ odd} \\ &= r/2, && \text{if } r \text{ even} \end{aligned} \tag{2.4.6}$$

We shall in fact find that we can solve (2.4.1)–(2.4.5) without using (2.4.6); so in this section and the next we shall ignore (2.4.6) and allow n to be any integer in the interval $1 \leqslant n \leqslant r-2$.

Note that $\chi(b+1) = 1 - \chi(b-1) = 0$ for $b = n$ and for $b = n+1$. For these values of b it follows that the right-hand sides of (2.4.1) and (2.4.2) are the same and hence

$$Y_m(a,b,b-1) = Y_m(a,b,b+1), \qquad b = n, n+1 \tag{2.4.7}$$

We shall need the following two functions:

$$\alpha(a,b,c) = -2a + \tfrac{1}{2}r(b-c+3) - r\chi(c) \tag{2.4.8}$$

$$\beta(m;a,b,c) = \tfrac{1}{4}(b-c+1)(m+b-a) - \tfrac{1}{2}a + \tfrac{1}{2}\chi(c)\left[m(c-b)+a-b\right] \tag{2.4.9}$$

Lemma 2.4.1. For a, b integers,

$$\alpha(a,b,b+1) = \alpha(a,b-1,b), \qquad b \neq n \tag{2.4.10}$$

$$\beta(m;a,b,b+1) = \beta(m-1;a,b-1,b), \qquad b \neq n \tag{2.4.11}$$

$$\alpha(a,b,b-1) = \alpha(a,b+1,b), \qquad b \neq n+1 \tag{2.4.12}$$

$$\beta(m;a,b,b-1) = \beta(m-1;a,b+1,b), \qquad b \neq n+1 \tag{2.4.13}$$

$$\alpha(a,b-1,b) = \alpha(a,b+1,b), -r \tag{2.4.14}$$

$$\beta(m-1;a,b-1,b) = \beta(m-1;a,b+1,b), + \tfrac{1}{2}(a-b-m) + m\chi(b) \tag{2.4.15}$$

Also, for a an integer and $b = n$ or $n+1$,

$$\alpha(a,b,b-1) = \alpha(a,b,b+1) \tag{2.4.16}$$

$$\beta(m;a,b,b-1) = \beta(m;a,b,b+1) \tag{2.4.17}$$

These lemmas follow directly from the definitions (2.4.8) and (2.4.9).

The following five lemmas contain the results necessary to provide expressions for the $Y_m(a,b,c)$; they each concern the polynomial

$G_m(a,b,c)$, defined by ($m - a - b$ an even integer)

$$G_m(a,b,c) = \sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2+\alpha(a,b,c)\lambda+\beta(m;a,b,c)} \begin{bmatrix} m \\ \frac{1}{2}(m+a-b) - r\lambda \end{bmatrix} \tag{2.4.18}$$

Lemma 2.4.2. For a and b integers, with $b \neq n$,

$$G_m(a,b,b+1) = G_{m-1}(a,b-1,b) + q^{m\chi(b+1)} G_{m-1}(a,b+1,b) \tag{2.4.19}$$

Proof. Using (2.4.10) and (2.4.11),

$$\begin{aligned} &G_m(a,b,b+1) - G_{m-1}(a,b-1,b) \\ &\quad = \sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2+\alpha(a,b-1,b)\lambda+\beta(m-1;a,b-1,b)} \\ &\quad \times \left\{ \begin{bmatrix} m \\ \frac{1}{2}(m+a-b) - r\lambda \end{bmatrix} - \begin{bmatrix} m-1 \\ \frac{1}{2}(m+a-b) - r\lambda \end{bmatrix} \right\} \end{aligned} \tag{2.4.20}$$

Using (2.2.2) to simplify the bracketted difference of Gaussian polynomials, then using (2.4.14) and (2.4.15) and comparing the result with (2.4.18), we find that

$$G_m(a,b,b+1) - G_{m-1}(a,b-1,b) = q^{m\chi(b)} G_{m-1}(a,b+1,b) \tag{2.4.21}$$

Noting that $\chi(b) = \chi(b+1)$ for $b \neq n$, we obtain (2.4.19). ■

Lemma 2.4.3. For a and b integers, with $b \neq n+1$,

$$G_m(a,b,b-1) - G_{m-1}(a,b+1,b) = q^{m[1-\chi(b-1)]} G_{m-1}(a,b-1,b) \tag{2.4.22}$$

Proof. The proof parallels that of the preceding lemma. We use (2.4.12) and (2.4.13) to write the left-hand side of (2.4.22) as a sum involving the difference of two Gaussian polynomials then use (2.2.3) to express this difference as a single polynomial, and finally use (2.4.14), (2.4.15) and the fact that $\chi(b) = \chi(b-1)$ for $b \neq n+1$. ■

Lemma 2.4.4. For $b = n$ and $b = n+1$,

$$G_m(a,b,b-1) = G_m(a,b,b+1) \tag{2.4.23}$$

This follows at once from (2.4.16) and (2.4.17).

Lemma 2.4.5. For any integer a,

$$G_m(a,0,1) = G_m(-a,0,1) \tag{2.4.24}$$

$$G_m(a,r,r-1) = G_m(-a,r,r-1) \tag{2.4.25}$$

Proof. From (2.4.8), (2.4.9), and (2.4.18),

$$G_m(a,0,1) = \sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2-2a\lambda+\frac{1}{2}m}\begin{bmatrix} m \\ \frac{1}{2}(m+a)-r\lambda \end{bmatrix} \tag{2.4.26}$$

$$G_m(a,r,r-1) = \sum_{\lambda=-\infty}^{\infty} q^{2r(\lambda+1/2)^2-2a(\lambda+1/2)+(1/2)m}\begin{bmatrix} m \\ \frac{1}{2}(m+a)-r(\lambda+\frac{1}{2}) \end{bmatrix} \tag{2.4.27}$$

Using (2.2.4), we observe that the right-hand side of (2.4.26) is unchanged by negating a and λ. Similarly, the right-hand side of (2.4.27) is unchanged by negating a and replacing λ by $-\lambda-1$. The symmetries (2.4.24) and (2.4.25) immediately follow. ■

Lemma 2.4.6. For $1 \leqslant |a|, b < r$, and $c = b \pm 1$,

$$G_0(a,b,c) = q^{-a/2}\delta_{a,b} \tag{2.4.28}$$

Proof. From (2.2.1), when $m = 0$ the Gaussian polynomial in (2.4.18) vanishes unless $a - b = 2r\lambda$, which can happen only when $a = b$ and $\lambda = 0$. Using (2.4.8) and (2.4.9), the summand in (2.4.18) then has value $q^{-a/2}$. The result (2.4.28) follows. ■

We are now in a position to state and establish our result for the function $Y_m(a,b,c)$.

Theorem 2.4.7. For $m \geqslant 0$, $1 \leqslant a,b,c < r$, $c = b \pm 1$, and $b - a - m$ an even integer,

$$Y_m(a,b,c) = q^{a/2}\left[G_m(a,b,c) - G_m(-a,b,c)\right] \tag{2.4.29}$$

Proof. Since $Y_m(a,b,c)$ is completely defined by the recurrences and boundary conditions (2.4.1)–(2.4.4), we need only show that they are satisfied by (2.4.29). From Lemma 2.4.2, the recurrence relation (2.4.1) is satisfied (for $b \neq n$) by $Y_m(a,b,c) = G_m(a,b,c)$. Since the relation is unchanged by negating a, it is also satisfied by $Y_m(a,b,c) = G_m(-a,b,c)$. Since it is linear and homogeneous, it is also satisfied by (2.4.29).

Similarly, Lemma 2.4.3 implies that Eq. (2.4.2) is satisfied by (2.4.29) for $b \neq n+1$, while Lemma 2.4.4 implies that Eq. (2.4.7) is satisfied for $b = n, n+1$.

Thus when $b = n$ we have proved that (2.4.2) and (2.4.7) are satisfied. Since (2.4.7) was obtained by eliminating the right-hand sides of (2.4.1) and (2.4.2), it follows that (2.4.1) is also satisfied for $b = n$. Similarly, (2.4.2) is satisfied for $b = n+1$. (In fact, Lemmas 2.4.2 and 2.4.3 are valid for $b = n$ and $b = n+1$, but their proofs are then different from those given above. We have introduced (2.4.7) to avoid the need to give the proofs for these special cases.)

Thus the expression (2.4.29) for $Y_m(a,b,c)$ satisfies the recurrences (2.4.1) and (2.4.2) for all allowed values of b. From Lemmas 2.4.5 and 2.4.6 it is easily seen that the boundary conditions (2.4.3) and (2.4.4) are also satisfied. This completes the proof of Theorem 2.4.7. ■

The Limit $m \to \infty$**.** We now want to take the limit $m \to \infty$. From (1.5.5) we see that $t < 0$ in regime I, so from (1.5.21) the argument q of our function $Y_m(a,b,c;q)$ is numerically less than 1:

$$|q| < 1 \tag{2.4.30}$$

We have essentially four cases to consider: $c = b \pm 1$, $\chi(c) = 0$, or 1. In order to avoid digressions during the proofs of the main results, we first prove the following lemma, using the definitions (2.2.1), (2.2.8), and (2.2.9).

Lemma 2.4.8.

$$\begin{aligned} &\lim_{m\to\infty} q^{-m/2}\left\{\begin{bmatrix} m \\ \frac{1}{2}m + B \end{bmatrix} - \begin{bmatrix} m \\ \frac{1}{2}m + B + b \end{bmatrix}\right\} \\ &\qquad = (1-q^b)(q^{-B-b+1} - q^{B+1})/\left[(1-q)(q)_\infty\right] \end{aligned} \tag{2.4.31}$$

Proof. Set $m = 2k$ and note that

$$\begin{aligned} q^{-k}&\left\{\begin{bmatrix} 2k \\ k+B \end{bmatrix} - \begin{bmatrix} 2k \\ k+B-b \end{bmatrix}\right\} \\ &= q^{-k}\sum_{j=0}^{b-1}\left\{\begin{bmatrix} 2k \\ k+B+j \end{bmatrix} - \begin{bmatrix} 2k \\ k+B+j+1 \end{bmatrix}\right\} \\ &= \sum_{j=0}^{b-1} \frac{(q)_{2k}(q^{-B-j} - q^{B+j+1})}{(q)_{k+B+j+1}(q)_{k-B-j}} \end{aligned} \tag{2.4.32}$$

using the definitions (2.2.1) and (2.2.8). Taking the limit $k \to \infty$, the right-hand side becomes

$$\sum_{j=0}^{b-1} (q^{-B-j} - q^{B+j+1})/(q)_\infty \tag{2.4.33}$$

Performing the j summation, we obtain (2.4.31), as desired.

We shall save space in the subsequent developments if we define "$\lim_{m\to\infty}$" to mean the limit restricted to those values of m of the same parity as $a-b$.

Theorem 2.4.9. Let a and b be integers, $1 \leqslant a < r$, $1 \leqslant b < n$. Then

$$\begin{aligned}\lim_{m\to\infty} q^{-m} Y_m(a,b,b+1) &= \frac{q^{1-b}(1-q^b)}{(1-q)(q)_\infty} \sum_{\mu=-\infty}^{\infty} (-1)^\mu q^{(1/2)r\mu(\mu+1)-a\mu} \\ &= q^{1-b}(1-q^b)E(q^a,q^r)/[(1-q)(q)_\infty]\end{aligned} \tag{2.4.34}$$

where the sum is over all integers μ, and the function $E(x,z)$ is given by (1.5.7).

Proof. We begin by noting that for $1 \leqslant b < n$,

$$\alpha(a,b,b+1) = -2a, \qquad \beta(m;a,b,b+1) = \tfrac{1}{2}(m-b) \tag{2.4.35}$$

Using these expressions in (2.4.18) and (2.4.29), negating λ in $G_m(-a,b,b+1)$ and using (2.2.4), we obtain

$$\begin{aligned}Y_m(a,b,b+1) = \sum_{\lambda=-\infty}^{\infty} & q^{2r\lambda^2-2a\lambda+\frac{1}{2}(m-b+a)} \\ & \times \left\{ \begin{bmatrix} m \\ \frac{1}{2}(m+a-b)-r\lambda \end{bmatrix} - \begin{bmatrix} m \\ \frac{1}{2}(m+a+b)-r\lambda \end{bmatrix} \right\}\end{aligned} \tag{2.4.36}$$

Taking the limit $m \to \infty$, using (2.4.31), we obtain

$$\begin{aligned}&\lim_{m\to\infty} q^{-m} Y_m(a,b,b+1) \\ &\quad = \frac{q^{1-b}(1-q^b)}{(1-q)(q)_\infty} \sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2}\left[q^{(r-2a)\lambda} - q^{a-(r+2a)\lambda}\right]\end{aligned} \tag{2.4.37}$$

Splitting the summation into two series, taking $\lambda = \mu/2$, (μ even) in the first, and taking $\lambda = (\mu+1)/2$ (μ odd) in the second, we obtain the desired result (2.4.34). ■

Theorem 2.4.10. Let a and b be integers, $1 \leqslant a < r$, $n \leqslant b \leqslant r-2$. Then

$$\lim_{m\to\infty} Y_m(a,b,b+1) = \left[(q)_\infty\right]^{-1} \sum_{\mu=-\infty}^{\infty} (-1)^\mu q^{(1/2)r\mu(\mu+1)-a\mu} = E(q^a,a^r)/(q)_\infty \tag{2.4.38}$$

Proof. This case is simpler than the previous, in that we can use the elementary formula (2.2.6), rather than the more sophisticated Lemma 2.4.8.

Since $n \leqslant b \leqslant r-2$, we have

$$\alpha(a,b,b+1) = r-2a, \qquad \beta(m;a,b,b+1) = -\tfrac{1}{2}a \tag{2.4.39}$$

From (2.4.18) and (2.4.29), letting $m \to \infty$ and using (2.2.6), we obtain

$$\lim_{m\to\infty} Y_m(a,b,b+1) = \left[(q)_\infty\right]^{-1} \sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2+r\lambda}\left(q^{-2a\lambda} - q^{2a\lambda+a}\right) \tag{2.4.40}$$

Splitting this series into two, setting $\lambda = \mu/2$ in the first, and $(-\mu-1)/2$ in the second, we obtain the desired result (2.4.38).

Theorem 2.4.11. Let a and b be integers, $1 \leqslant a < r$, $2 \leqslant b \leqslant n+1$. Then

$$\lim_{m\to\infty} Y_m(a,b,b-1) = E(q^a,q^r)/(q)_\infty \tag{2.4.41}$$

Proof. We merely note from (2.4.8) and (2.4.9) that

$$\alpha(a,b,b-1) = r-2a, \qquad \beta(m;a,b,b-1) = -\tfrac{1}{2}a \tag{2.4.42}$$

These are the same expressions as those in (2.4.39), so the proof proceeds exactly as for Theorem 2.4.10. ■

Theorem 2.4.12. Let a and b be integers, $1 \leqslant a < r$, $n+2 \leqslant b < r$. Then

$$\lim_{m\to\infty} q^{-m} Y_m(a,b,b-1) = q^{1-r+b}(1-q^{r-b})E(q^a,q^r)/\left[(1-q)(q)_\infty\right] \tag{2.4.43}$$

Proof. In this case

$$\alpha(a,b,b-1) = 2r-2a, \qquad \beta(m;a,b,b-1) = \tfrac{1}{2}(m+b-2a) \tag{2.4.44}$$

Using (2.4.18) and (2.4.29); negating λ in $G_m(a,b,b-1)$ and using (2.2.4), while replacing λ in $G_m(-a,b,b-1)$ by $\lambda-1$, we find that

$$Y_m(a,b,b-1)=\sum_{\lambda=-\infty}^{\infty} q^{2r\lambda^2-2(r-a)\lambda+\frac{1}{2}(m+b-a)} \times\left\{\begin{bmatrix} m \\ \frac{1}{2}(m-a+b)-r\lambda\end{bmatrix} -\begin{bmatrix} m \\ \frac{1}{2}(m+2r-a-b)-r\lambda\end{bmatrix}\right\} \tag{2.4.45}$$

The right-hand side of this equation is the same as that of (2.4.36), but with a,b replaced by $r-a, r-b$. Making these replacements throughout Theorem 2.4.9, and using the identity

$$E(z,x)=E(x/z,x) \tag{2.4.46}$$

we at once obtain our desired result (2.4.43). ■

The last four theorems give the limiting behavior of $Y_m(a,b,c)$ as m becomes large, for all integers a,b,c such that $1\leqslant a,b,c\leqslant r-1$, $c=b\pm 1$. Note that in every case the appropriate limit factors into the form

$$E(q^a,q^r)\times[\text{function of } b,c,q] \tag{2.4.47}$$

Substituting the appropriate limiting forms of $Y_m(a,b,c)$ into (1.5.21) and (1.5.25), it follows that P_a tends to a limit as $m\to\infty$, and that the function of b,c,q in (2.4.47) cancels out of the final expression for P_a, leaving

$$P_a=v_aE(q^a,q^r)\Big/\sum_{1\leqslant a'\leqslant r-1} v_{a'}E(q^{a'},q^r), \tag{2.4.48}$$

where, as in regime III (and indeed in all regimes), a and a' are either both even integers, or are both odd. From (1.5.21) and (1.5.5), the argument q of $Y_m(a,b,c;q)$ is here related to the parameter x by

$$q=x^{r-2} \tag{2.4.49}$$

We see that the limit (2.4.48) is indepedent of the values b,c of the boundary heights l_{m+1}, l_{m+2}. [It is even independent of the integer n, which as we mentioned after (2.4.6) can be regarded in this section as an independent integer.] This indicates that the model is disordered in regime I, and is consistent with the fact that there are then just the two ground states (1.5.29).

2.5. Regime IV

We now come to our first exploitation of the duality principle that was utilized for treating the Rogers–Ramanujan-type identities occurring in the original hard hexagon model.[26] In this regime we see from (1.5.5) that $t > 0$, so from (1.5.21) the argument q of the polynomial $Y_m(a,b,c;q)$ is now greater than 1. Our interest therefore now centers on the polynomials reciprocal to those of regime I.

Throughout this section we shall write $y_m(a,b,c)$ for $Y_m(a,b,c)$ with q replaced by q^{-1}, and $g_m(a,b,c)$ for $G_m(a,b,c)$ with q replaced by q^{-1}. As before $b-a-m$ is even and $c = b \pm 1$.

It is convenient to define

$$\gamma(m;a,b,c) = \tfrac{1}{4}(a-b)^2 - \beta(m;a,b,c) \tag{2.5.1}$$

Lemma 2.5.1.

$$q^{m^2/4} g_m(a,b,c) = q^{\gamma(m;a,b,c)} \sum_{\lambda=-\infty}^{\infty} q^{r(r-2)\lambda^2 - [\alpha(a,b,c)+ra-rb]\lambda} \times \begin{bmatrix} m \\ \tfrac{1}{2}(m+a-b) - r\lambda \end{bmatrix} \tag{2.5.2}$$

Proof. This follows immediately by applying (2.2.5) to (2.4.18). ■

As in section 2.4, we shall denote by $\lim_{m\to\infty}$ the limit taken through those integers m of the same parity as $a-b$.

Lemma 2.5.2.

$$\begin{aligned}\lim_{m\to\infty} q^{m^2/4} y_m(a,b,c) &= Q(q)^{-1} \sum_{\lambda=-\infty}^{\infty} q^{r(r-2)\lambda^2 + rb\lambda - \frac{1}{2}a} \\ &\quad \times \left\{ q^{\gamma(m;a,b,c) - [\alpha(a,b,c)+ra]\lambda} - q^{\gamma(m;-a,b,c) - [\alpha(-a,b,c)-ra]\lambda} \right\}\end{aligned} \tag{2.5.3}$$

Proof. This follows from applying (2.2.6) to (2.5.2), and using (2.4.29) with q replaced by q^{-1}. ■

The result (2.5.3) can be simplified by defining two further functions:

$$\tau(b,c) = \tfrac{1}{2}(c-b-1) + \chi(c) \tag{2.5.4}$$

for $1 \leqslant b,c \leqslant r-1$, $c = b \pm 1$; then

$$\Gamma(a,d;q) = \sum_{\lambda=-\infty}^{\infty} q^{r(r-2)\lambda^2 + rd\lambda + (a-d-1)^2/4} \left\{ q^{-(r-2)a\lambda} - q^{(r-2)a\lambda + ad} \right\} \tag{2.5.5}$$

for $1 \leqslant a \leqslant r-1$ and $1 \leqslant d \leqslant r-3$. Using (1.5.7), we can write (2.5.5) as

$$\Gamma(a,d;q) = q^{(a-d-1)^2/4}\left\{E\left[-q^{r(r-2+d)-(r-2)a}, q^{2r(r-2)}\right] - q^{ad}E\left[-q^{r(r-2+d)+(r-2)a}, q^{2r(r-2)}\right]\right\} \quad (2.5.6)$$

The function $\tau(b,c)$ is an integer, with values $-1, 0, 1$.

Lemma 2.5.3.

$$\lim_{m\to\infty} q^{[m+|\tau(b,c)|]^2/4} y_m(a,b,c) = \Gamma\left[a, b+\tau(b,c)-1; q\right]/Q(q) \quad (2.5.7)$$

Proof. This is just a restatement of Lemma 2.5.2, as is most easily verified by explicitly considering the four cases $c = b \pm 1$, $c \leqslant n$ or $c > n$. ■

From (1.5.21), (1.5.25), (1.5.5), it follows that P_a tends to a limit as $m \to \infty$, given by

$$P_a = v_a \Gamma(a,d;x^2) \Big/ \sum_{1 \leqslant a' \leqslant r-1} v_{a'} \Gamma(a',d;x^2) \quad (2.5.8)$$

where

$$d = b + \tau(b,c) - 1 \quad (2.5.9)$$

and a, a' are either both even integers (the "even" sub-SOS model), or are both odd integers (the "odd" sub-SOS model).

We note that P_a depends on d. From (2.5.5) and (2.4.5),

$$\begin{aligned} d &= b, & &\text{if} \quad b+1 = c \leqslant n \\ &= b-1, & &\text{if} \quad b+1 = c > n \\ &= b-1, & &\text{if} \quad b-1 = c \leqslant n \\ &= b-2, & &\text{if} \quad b-1 = c > n \end{aligned} \quad (2.5.10)$$

Since $1 \leqslant b, c \leqslant r-1$, we see that d is an integer satisfying

$$1 \leqslant d \leqslant r-3 \quad (2.5.11)$$

Thus d takes only $r-3$ values. Allowing for the two possible sets of values of a (even or odd), it follows that there are $2(r-3)$ functions P_a, corresponding to the $2r-6$ ground states in regime IV.

2.6. Regime II

This is by far the most complicated of the regimes. From (1.5.5) we see that $t < 0$, so from (1.5.9) the argument q of the function $X_m(a,b,c;q)$ is greater than 1. We are therefore interested in the polynomials (in $q^{1/2}$) that are reciprocal to those of regime III, which we can define as

$$x_m(a,b,c) = x_m(a,b,c;q) = q^{m(m+1)/4} X_m(a,b,c;q^{-1}) \tag{2.6.1}$$

From (2.3.1) and (2.3.2) we see that the $x_m(a,b,c)$ satisfy the recurrences ($m > 0$, $1 \leqslant a < r$):

$$x_m(a,b,b+1) = q^{m/2} x_{m-1}(a,b+1,b) + x_{m-1}(a,b-1,b), \qquad 1 \leqslant b \leqslant r-2 \tag{2.6.2}$$

$$x_m(a,b,b-1) = q^{m/2} x_{m-1}(a,b-1,b) + x_{m-1}(a,b+1,b), \qquad 2 \leqslant b \leqslant r-1 \tag{2.6.3}$$

Lemma 2.6.1. For $m \geqslant 0$, $1 \leqslant a,b,c < r$, $c = b \pm 1$, $m+a-b$ an even integer

$$x_m(a,b,c) = q^{(m+a-b)/4}\{f_m(a,b,c) - f_m(-a,b,c)\} \tag{2.6.4}$$

where

$$f_m(a,b,c) = \sum_{\lambda=-\infty}^{\infty} q^{r\lambda^2 - a\lambda + (b+1-c)(2r\lambda+b-a)/4} \begin{bmatrix} m \\ \frac{1}{2}(m+a-b) - r\lambda \end{bmatrix} \tag{2.6.5}$$

Proof. This follows immediately on replacing q by q^{-1} in Theorem 2.3.1 and using (2.6.1). ■

Remarks on Procedure. Using (2.2.6), we can easily verify that $f_m(a,b,c)$ tends to a limit as $m \to \infty$. Unfortunately this limit is an even function of a (this follows by replacing λ by $-\lambda$, or by $-\lambda-1$), so the expression $\{f_m(a,b,c) - f_m(-a,b,c)\}$ in (2.6.4) vanishes.

This is still true even if we expand the Gaussian polynomial in (2.6.5) in increasing powers of $q^{m/2}$ and keep a finite number of terms in the expansion, which is basically the method we employed in regime I. (We shall in fact find for large m that $\{f_m(a,b,c) - f_m(-a,b,c)\}$ is q raised to the power of a *quadratic* form in m, which accounts for the failure of this method.)

We therefore need a more sophisticated procedure to determine the large-m behavior of $x_m(a,b,c)$, and it is this that makes regime II so difficult. We shall begin by showing that $x_m(a,2,1)$ can be expressed in terms of the function

$$\rho_N(q, w, s; d_1, \ldots, d_2) = \sum_{j=0}^{N} q^{j^2+j-Nj} w^{-j}$$

$$\times (1 - q^{N-2j} w) \prod_{i=1}^{s} \left[(d_i)_j (w d_i)_{N-j} \right]^{-1} \quad (2.6.6)$$

where N, s are nonnegative integers and we are using the notation (2.2.8).

We shall further show that one of $d_1, \ldots, d_s$ is equal to the argument q, while another is equal to q/w. As is shown in Appendix B, the function ρ_N is then proportional to a special case of a "well-poised q-hypergeometric series".(28,29,32) Its behavior in the limit $N \to \infty$ is determined in Appendix B, using a standard theorem for such series (Theorem 4 of Ref. 29).

Using this result (Theorem B5), we can evaluate $x_m(a, 2, 1)$ in the limit of m large. We then obtain the limits of all the $x_m(a, b, c)$ by sequentially solving the recurrence relations (2.6.2), (2.6.3) for $x_m(a, 1, 2)$, $x_m(a, 3, 2)$, $x_m(a, 2, 3), \ldots, x_m(a, r-2, r-1)$. Finally we identify our results when $r = 5$ with those of the original hard hexagon model.(9,10)

Evaluation of $x_m(a, 2, 1)$ in the Limit of m Large. We begin by expressing $x_m(a, 2, 1)$, for $m + a$ an even integer, in terms of the function ρ_N. We shall need the integers $u, v, \lambda_0, \lambda_1, N$, defined by

$$\tfrac{1}{2}(m - a) + 1 = r\lambda_0 + u, \qquad 0 \leqslant u < r \quad (2.6.7)$$

$$\tfrac{1}{2}(m + a) + 1 = r\lambda_1 + v, \qquad 0 \leqslant v < r \quad (2.6.8)$$

$$N = \lambda_0 + \lambda_1 \quad (2.6.9)$$

Adding (2.6.7) and (2.6.8), we see that $m, N, u + v$ are related by

$$m + 2 = rN + u + v \quad (2.6.10)$$

Theorem 2.6.2. For $1 \leqslant a < r$, $m \geqslant 0$, $m + a$ even,

$$x_m(a, 2, 1) = L_m(a) \rho_N(q^r, q^{v-u}, r; q^{u+1}, q^{u+2}, \ldots, q^{u+r}) \quad (2.6.11)$$

where

$$L_m(a) = \frac{q^{r\lambda_0^2 + (a-r)\lambda_0 + (m+2-a)/4} (q)_{m+1}}{(q)_u (q)_v} \quad (2.6.12)$$

Proof. Using Lemma 2.6.1, negating λ in $f_m(-a,2,1,)$,

$$x_m(a,2,1) = q^{(m+2-a)/4} \sum_{\lambda=-\infty}^{\infty} q^{r\lambda^2+(r-a)\lambda} \times \left\{ \begin{bmatrix} m \\ \frac{1}{2}(m+a)-1-r\lambda \end{bmatrix} - q^{a-2\lambda} \begin{bmatrix} m \\ \frac{1}{2}(m+a)+1-r\lambda \end{bmatrix} \right\} \tag{2.6.13}$$

The summand in (2.6.13) is zero unless

$$-\tfrac{1}{2}(m-a)-1 \leqslant r\lambda \leqslant \tfrac{1}{2}(m+a)+1 \tag{2.6.14}$$

The definitions (2.6.7) and (2.6.8) are arranged so that $-\lambda_0$ is the smallest value of λ in the range (2.6.14), while λ_1 is the largest. Thus we can restrict the summation in (2.6.13) to

$$-\lambda_0 \leqslant \lambda \leqslant \lambda_1 \tag{2.6.15}$$

From (2.2.1) it is readily verified, for $-1 \leqslant k \leqslant m+1$, that

$$\begin{bmatrix} m \\ k-1 \end{bmatrix} - q^{2k-m} \begin{bmatrix} m \\ k+1 \end{bmatrix} = \frac{(1-q^{2k-m})(q)_{m+1}}{(q)_{k+1}(q)_{m-k+l}} \tag{2.6.16}$$

Using this identity, (2.6.13) can be written as

$$x_m(a,2,1) = q^{(m+2-a)/4} \sum_{\lambda=-\lambda_0}^{\lambda_1} q^{r\lambda^2+(r-a)\lambda} \times \frac{(1-q^{a-2r\lambda})(q)_{m+1}}{(q)_{(1/2)(m+a)+1-r\lambda}(q)_{(1/2)(m-a)+1+r\lambda}} \tag{2.6.17}$$

Now we replace λ by $j-\lambda_0$, and use (2.6.7)–(2.6.9), (2.6.12), and (2.2.10) to obtain

$$x_m(a,2,1) = L_m(a) \sum_{j=0}^{N} q^{rj^2+(r-rN+u-v)j} \frac{1-q^{rN-2rj+v-u}}{(q^{u+1})_{rj}(q^{v+1})_{rN-rj}} \tag{2.6.18}$$

Now we note that

$$(q^{u+1})_{rj} = (q^{u+1};q^r)_j(q^{u+2};q^r)_j \cdots (q^{u+r};q^r)_j \tag{2.6.19}$$

and similarly for $(q^{v+1})_{r(N-j)}$. Comparing the summation in (2.6.18) with that in (2.6.6), we obtain the desired result (2.6.11). ■

Now we want to let $m \to \infty$. We shall find it convenient, both here and in Appendix B, to adopt the convention that by

$$\lim_{n\to\infty} g_n = f_n \tag{2.6.20a}$$

we mean

$$\lim_{n\to\infty} (g_n/f_n) = 1 \tag{2.6.20b}$$

We shall need the function

$$\Phi_a(z) = \frac{E(q^a,q^r)E(-z,q)[Q(q^r)]^3}{E(-z,q^r)E(-q^a z,q^r)[Q(q)]^2} \tag{2.6.21}$$

This is an analytic function of z in the domain $0<|z|<\infty$, so has a Laurent expansion:

$$\Phi_a(z) = \sum_{j=-\infty}^{\infty} \eta_{a,j} z^j \tag{2.6.22}$$

We shall find that we can express the $x_m(a,b,c)$ in terms of these coefficients $\eta_{a,j}$.

Theorem 2.6.3. For $1 \leqslant a < r$, $m+a$ even,

$$\lim_{m\to\infty} x_m(a,2,1) = (-1)^N q^{r\lambda_0^2+(a-r)\lambda_0+(m+2-a)/4-(1/2)u(u+1)}\eta_{v-u,-N-u} \tag{2.6.23}$$

Proof. When $m\to\infty$ we see from (2.6.7)–(2.6.10) that u and v remain between 0 and $r-1$, while $N\to\infty$. Also, one of the arguments $q^{u+1},\ldots,q^{u+r}$ of ρ_N in Theorem 2.6.2 must be equal to q^r, while another must be equal to q^{r+u-v}. These two special arguments must be distinct, since if $u=v$ then (2.6.7) and (2.6.8) give $a=r(\lambda_1-\lambda_0)$, i.e., $a\equiv 0 \pmod r$, which are not allowed values of a.

We can therefore use Theorem B5 to obtain the limiting behavior of ρ_N in Theorem 2.6.2. This gives that $x_m(a,2,1)$ is the coefficient of z^{-N} in the Laurent expansion of

$$\frac{L_m(a)(q^r;q^r)_\infty^3 E(q^{v-u},q^r)}{E(z,q^r)E(q^{v-u}z,q^r)} \prod_{i=1}^{r} \frac{E(q^{u+i}z^{-1},q^r)}{(q^r;q^r)_\infty (q^{u+i};q^r)_\infty (q^{v+i};q^r)_\infty} \tag{2.6.24}$$

Now we use the triple product form (1.5.6) of $E(z,x)$ to note that

$$E(q^{u+i}z^{-1},q^r)/(q^r;q^r)_\infty = (q^{u+i}z^{-1};q^r)_\infty (q^{r-u-i}z;q^r)_\infty \tag{2.6.25}$$

Using the identity (2.6.19), it follows that the product over i in (2.6.24) is simply

$$(q^{u+1}z^{-1})_\infty (q^{-u}z)_\infty / [(q^{u+1})_\infty (q^{v+1})_\infty] \tag{2.6.26}$$

Substituting this, and using the definition (2.6.12) of $L_m(a)$, with $(q)_{m+1}$ replaced by $(q)_\infty$, the function (2.6.24) becomes

$$\frac{q^{r\lambda_0^2+(a-r)\lambda_0+(m+2-a)/4}(q^r;q^r)_\infty^3 E(q^{v-u},q^r)E(q^{-u}z,q)}{E(z,q^r)E(q^{v-u}z,q^r)(q)_\infty^2} \tag{2.6.27}$$

From (1.5.6) we can readily verify that for all integers u,

$$E(q^{-u}z,q) = (-z)^u q^{-(1/2)u(u+1)}E(z,q) \tag{2.6.28}$$

so from (2.6.21) and (2.2.9) we see that (2.6.27) is equal to

$$q^{r\lambda_0^2+(a-r)\lambda_0+(m+2-a)/4-(1/2)u(u+1)}(-z)^u\Phi_{v-u}(-z) \tag{2.6.29}$$

Recalling that $x_m(a,2,1)$ is the coefficient of z^{-N} in the Laurent expansion of this function (2.6.29), we obtain the desired result (2.6.23). ■

The form for $x_m(a,2,1)$ given by the preceding theorem involves the integers λ_0, v, u, N defined by (2.6.7)–(2.6.9). It turns out that we can eliminate them in favor of our original variables a and m. To do this we need certain recurrence and symmetry properties of $\Phi_a(z)$ and $\eta_{a,j}$.

Lemma 2.6.4. For all integers a, and nonzero complex numbers

$$\Phi_{a+r}(z) = -z\Phi_a(z) \tag{2.6.30}$$

$$\Phi_a(q^r z) = z^{2-r}q^{a-(1/2)r(r-1)}\Phi_a(z) \tag{2.6.31}$$

$$\Phi_a(q^{-a}z^{-1}) = z^{1-a}q^{(1/2)a(1-a)}\Phi_a(z) \tag{2.6.32}$$

Proof. These three properties all follow directly from the definition (2.6.21), the recurrence formula (2.6.28), and the identity [readily verifiable from (1.5.6)]:

$$E(z^{-1},q) = -z^{-1}E(z,q) \quad ■ \tag{2.6.33}$$

Lemma 2.6.5. For all integers a and j,

$$\eta_{a+r,j} = -\eta_{a,j-1} \tag{2.6.34}$$

$$\eta_{a,j+r-2} = q^{rj-a+(1/2)r(r-1)}\eta_{a,j} \tag{2.6.35}$$

$$\eta_{a,a-1-j} = q^{(1/2)a(a-2j-1)}\eta_{a,j} \tag{2.6.36}$$

Proof. Substituting the series form (2.6.22) of $\Phi_a(z)$ into Lemma 2.6.4 and equating coefficients in the resulting expansions, we obtain (2.6.34)–(2.6.36). Repeated application of (2.6.34) and (2.6.35) yields the corollaries:

$$\eta_{a+rk,j} = (-1)^k \eta_{a,j-k} \tag{2.6.37}$$

$$\eta_{a,j+(r-2)k} = q^{k[rj-a+(1/2)r+(1/2)r(r-2)k]}\eta_{a,j} \tag{2.6.38}$$

for all integers a, j, k. ■

Theorem 2.6.6. For $1 \leqslant a < r$, $m + a$ even,

$$\lim_{m\to\infty} x_m(a, 2, 1) = q^{-(m+2-a)^2/8}\eta_{a,(1/2)(a-m-2)} \tag{2.6.39}$$

Proof. From (2.6.7) and (2.6.8),

$$v - u = a + r(\lambda_0 - \lambda_1) \tag{2.6.40}$$

From (2.6.37), with $k = \lambda_0 - \lambda_1$, it follows that

$$\begin{aligned}(-1)^N \eta_{v-u,-N-u} &= (-1)^{N+\lambda_0-\lambda_1}\eta_{a,-N-u-\lambda_0+\lambda_1}\\ &= \eta_{a,-u-2\lambda_0}\end{aligned} \tag{2.6.41}$$

after using (2.6.9).

Now we use (2.6.38), with $j = -u - r\lambda_0$ and $k = \lambda_0$, to obtain

$$\eta_{a,-u-2\lambda_0} = q^{\lambda_0[(1/2)r-a-ru-(1/2)r(r+2)\lambda_0]}\eta_{a,-u-r\lambda_0} \tag{2.6.42}$$

so from this and (2.6.41), Theorem 2.6.3 becomes

$$\lim_{m\to\infty} x_m(a, 2, 1) = q^{-(1/2)(u+r\lambda_0)(u+r\lambda_0+1)+(m+2-a)/4}\eta_{a,-u-r\lambda_0} \tag{2.6.43}$$

Note that u, λ_0 enter this expression only in the combination $u + r\lambda_0$, which from (2.6.7) is $\frac{1}{2}(m - a + 2)$. The result (2.6.39) follows. ■

Evaluation of $x_m(a, b, c)$ in the Limit of m Large. Now that we have $x_m(a, 2, 1)$, we can use the recurrence relations (2.6.2) and (2.6.3) to obtain $x_m(a, b, c)$ in general. We first need a lemma concerning the large-j behavior of $\eta_{a,j}$.

Lemma 2.6.7. For all integers a, j, define $\hat{\eta}_{a,j}$ by

$$\eta_{a,j} = q^{[(1/2)rj(j+1)-aj]/(r-2)}\hat{\eta}_{a,j} \tag{2.6.44}$$

Then

$$\hat{\eta}_{a,j+r-2} = \hat{\eta}_{a,j} \tag{2.6.45}$$

Proof. This follows immediately from (2.6.35). Since $\hat{\eta}_{a,j}$ is periodic in j, it is bounded, so this lemma gives the large-j behavior of $\eta_{a,j}$.

Theorem 2.6.8. For $1 \leqslant a < r$, $m + a - b$ even,

$$\lim_{m\to\infty} x_m(a,b,b+1) = q^{-(m+a-b)^2/8}\eta_{a,(1/2)(m+a-b)}, \qquad 1 \leqslant b \leqslant r-2 \tag{2.6.46a}$$

$$\lim_{m\to\infty} x_m(a,b,b-1) = q^{-(m-a+b)^2/8}\eta_{a,(1/2)(a-m-b)}, \qquad 2 \leqslant b \leqslant r-1 \tag{2.6.46b}$$

Proof. From (2.3.3) and (2.6.1), $x_m(a,0,1)$ in the recurrence relation (2.6.2) is to be interpreted as zero. Setting $b = 1$ therein, we therefore obtain

$$x_m(a,1,2) = x_{m-1}(a,2,1) \tag{2.6.47}$$

Using Theorem 2.6.6 and Eq. (2.6.36), we can at once verify that (2.6.46a) is satisfied for $b = 1$, and (2.6.46b) is satisfied for $b = 2$.

We now proceed by induction. Suppose that (2.6.46a) is satisfied for $1 \leqslant b \leqslant b_0 - 1$, and (2.6.46b) is satisfied for $2 \leqslant b \leqslant b_0$, where $2 \leqslant b_0 \leqslant r-2$. (We have just shown this is so for $b_0 = 2$.) Setting $b = b_0$ in the recurrence relation (2.6.3) and rearranging, we obtain

$$x_{m-1}(a,b_0+1,b_0) = x_m(a,b_0,b_0-1) - q^{m/2}x_{m-1}(a,b_0-1,b_0) \tag{2.6.48}$$

In the limit $m \to \infty$, both terms on the right-hand side are known to be given by (2.6.46). Thus their ratio is, using Lemma 2.6.7 and temporarily dropping the suffix on b_0,

$$\frac{q^{m/2}x_{m-1}(a,b-1,b)}{x_m(a,b,b-1)} = \frac{q^{(1/2)m(b-a+1)}\eta_{a,(1/2)(m+a-b)}}{\eta_{a,(1/2)(a-b-m)}} = \frac{q^{m(r-1-b)/(r-2)}\hat{\eta}_{a,(1/2)(m+a-b)}}{\hat{\eta}_{a,(1/2)(a-b-m)}} \tag{2.6.49}$$

Since $\hat{\eta}_{a,j}$ is bounded and $b < r-1$, this ratio tends to zero as $m \to \infty$. Thus we can ignore the last term in (2.6.48). Using (2.6.46b) to evaluate $x_m(a, b_0, b_0 - 1)$, it follows at once that $x_{m-1}(a, b_0 + 1, b_0)$ is also given, for m large, by (2.6.46b). Thus (2.6.46b) is true for $b = b_0 + 1$.

The second step of the inductive proof is to set $b = b_0$ in (2.6.2). Both terms on the right-hand side are now known to be given by (2.6.46), so (temporarily dropping the suffix on b_0)

$$\frac{q^{m/2}x_{m-1}(a,b+1,b)}{x_{m-1}(a,b-1,b)} = \frac{q^{(1/2)m(a-b+1)}\eta_{a,(1/2)(a-m-b)}}{\eta_{a,(1/2)(m+a-b)}} = \frac{q^{m(b-1)/(r-2)}\hat{\eta}_{a,(1/2)(a-m-b)}}{\hat{\eta}_{a,(1/2)(m+a-b)}} \tag{2.6.50}$$

Since $b > 1$, this ratio tends to zero as $m \to \infty$, so we can ignore the first term on the right-hand side of (2.6.2). Using (2.6.46a) to evaluate the second term, we at once find that $x_m(a, b_0, b_0 + 1)$ on the left-hand side is also given by (2.6.46a). Thus (2.6.46a) is true for $b = b_0$.

This completes the inductive loop. Taking $b_0 = 2, 3, \ldots, r-2$, we establish the results (2.6.46). ■

Throughout the rest of this section, let

$$j = -b + (c - b)m \tag{2.6.51}$$

Since $c = b \pm 1$, j is either $m - b$ or $-m - b$. It is an integer, and has the same parity as a.

We shall use the function $\hat{x}_m(a,b,c)$, defined in terms of $x_m(a,b,c)$ by

$$x_m(a,b,c) = q^{j(j+r)/(4r-8)} \hat{x}_m(a,b,c) \tag{2.6.52}$$

This function $\hat{x}_m$ contains fractional powers of q, but its large-m behavior is simpler to discuss than that of x_m itself, as is evident from the following theorem.

Theorem 2.6.9. For $1 \leqslant a < b$, $c < r$, $c = b \pm 1$, $m + a - b$ even,

$$\lim_{m \to \infty} \hat{x}_m(a,b,c) = q^{a(r-a)/(4r-8)} \hat{\eta}_{a,(1/2)(a+j)} \tag{2.6.53}$$

Proof. This is just a restatement of Theorem 2.6.8, using the definitions (2.6.44), (2.6.51), (2.6.52). ■

Note that m, b, c enter the right-hand side of (2.6.53) only via the integer j, and that $a + j$ must be even. Further, from (2.6.45), $\hat{\eta}_{a,i}$ is a periodic function of i, of period $r - 2$. Thus as m is increased (in increments of two), $\hat{x}_m(a,b,c)$ sequentially takes $r - 2$ different values, returning after $r - 2$ increments to its original value. Varying b and c merely changes the starting point of this cycle.

From (1.5.5.), (1.5.9), (1.5.13), (2.6.1), and (2.6.52), the local height probabilities are given by

$$P_a = u_a \hat{x}_m(a,b,c;q) \Big/ \sum_{1 \leqslant a' \leqslant r-1} u_{a'} \hat{x}_m(a',b,c;q) \tag{2.6.54}$$

where a, a' are either confined to the set of even integers (the "even" sub-SOS model), or to the set of odd integers (the "odd" model). The argument q is related to the parameter x of (1.5.8) (which is not to be confused with the functions $x_m(a,b,c;q)$ and $\hat{x}_m(a,b,c;q)$ by

$$q = x^{r-2} \tag{2.6.55}$$

For m large and a fixed, there are just $r - 2$ different values of $\hat{x}_m(a,b,c;q)$ that can be obtained by varying m, b, and c, i.e., by varying the boundary conditions on the lattice. From (2.6.54) it follows that for a given parity of a, there are just $r-2$ different functions P_a that can be formed by varying the boundary conditions. Altogether this gives us $2r-4$ functions, corresponding to the $2r-4$ ground states (1.5.19).

The Case $r = 5$. As was shown in Sections 1.3 and 1.4, when $r = 5$ we regain the origin hard hexagon model. Thus the results (2.6.46)–(2.6.55) of this section should then reduce to those for regime II of the hard hexagon model.[9,10,26] From their definitions, the quantities $F_k(0), F_k(1)$ in Refs. 9, 10, and 26 are related to our functions $\hat{x}_m(a,b,c)$ by

$$\begin{aligned} F_k(\sigma) &= q^{(1/2)(\sigma-1)-(2-k)(3-k)/6}\hat{x}_{6i-2k+b-2}(2+2\sigma,b,b+1), \qquad 1 \leqslant b \leqslant 3 \\ &= q^{(1/2)(\sigma-1)-(2-k)(3-k)/6}\hat{x}_{6i-2k-b}(2+2\sigma,b,b-1), \qquad 2 \leqslant b \leqslant 4 \end{aligned} \tag{2.6.56}$$

where $k = 1,2,3$, $\sigma = 0,1$, and the right-hand side is to be evaluated in the limit $i \to \infty$.

We can take the limit $i \to \infty$ by using Theorem 2.6.9, which gives

$$\begin{aligned} F_k(\sigma) &= q^{[2\sigma-(2-k)(3-k)]/6}\hat{\eta}_{2+2\sigma,\sigma-k+3i} \\ &= q^{[2\sigma-(2-k)(3-k)]/6}\hat{\eta}_{2+2\sigma,\sigma+k-3i+1} \end{aligned} \tag{2.6.57}$$

We see that both expressions in (2.6.57) are independent of b, so it is irrelevant what value we took in (2.6.56). Also, from (2.6.36) and (2.6.44), $\hat{\eta}_{a,a-1-j} = \hat{\eta}_{a,j}$ for all integers a, j. It follows that the two expressions in (2.6.57), and hence in (2.6.56), are equivalent.

The integer i still appears in (2.6.57), but from the periodicity relation (2.6.45) it plays no role, the expressions in (2.6.57) being independent of i. We can therefore now set $i = 0$ and use the first expression. Then from (2.6.44) we obtain

$$F_k(\sigma) = q^{\sigma k-1+k-k^2}\eta_{2+2\sigma,\sigma-k} \tag{2.6.58}$$

for $k = 1,2,3$ and $\sigma = 0,1$.

If we now use the definitions (2.6.21), (2.6.22) of $\eta_{a,j}$, the resulting expressions for the $F_k(\sigma)$ are not of the same form as those obtained originally.[9,10] To establish the equivalence we need the following theorem. As throughout this paper, $E(z,x)$ is the elliptic theta function defined by (1.5.6), or equivalently (1.5.7).

Theorem 2.6.10. For $|q| < 1$, and all nonzero complex numbers z, w,

$$\begin{aligned} &E(z,q)E(w,q)E(zw,q)E(w/z,q) \\ &\quad = [Q(q)]^2\{E(w^3,q^3)[E(qz^3,q^3) - zE(qz^{-3},q^3)] \\ &\qquad - (w/z)E(z^3,q^3)[E(qw^3,q^3) - wE(qw^{-3},q^3)]\} \end{aligned} \tag{2.6.59}$$

Proof. Like all elementary theta function identities this can be proved using Liouville's theorem (Section 15.3 of Ref. 10). Regard q, w as constants and z as a variable. Let

$$f(z) = \frac{\text{right-hand side of (2.6.59)}}{\text{left-hand side of (2.6.59)}} \tag{2.6.60}$$

Then by using (2.6.28) it is readily verified that $f(z)$ has the periodicity property $f(qz) = f(z)$. Since $E(z,q)$ is analytic for $0 < |z| < \infty$, the only possible singularities of $f(z)$ in this domain are simple poles, occurring when the left-hand side of (2.6.59) vanishes, in particular when $z = 1$, w, or w^{-1}. However, using the obvious identities $E(1,q) = 0$, $E(w^{-3},q) = -w^{-3}E(w^3,q)$, we can immediately verify that the right-hand side of (2.6.59) then also vanishes, so the poles have zero residue and $f(z)$ is in fact analytic at $z = 1, w, w^{-1}$. Since $f(qz) = f(z)$, it is therefore analytic at $z = q^n, q^n w, q^n w^{-1}$ for all integers n, which exhausts all the zeros of the left-hand side of (2.6.59).

It follows that $f(z)$ is analytic in $0 < |z| < \infty$, so is bounded in $q \leqslant |z| \leqslant 1$. Again using the periodicity property $f(qz) = f(z)$, the function $f(z)$ must be bounded everywhere. This means that $f(z)$ must be analytic at $z = 0$ and ∞, so by Liouville's theorem it must be a constant.

From the infinite product expansions (1.5.6) and (2.2.7), it is readily established that if $\omega = e^{2\pi i/3}$, then (for all complex numbers z)

$$E(z,q)E(\omega z,q)E(\omega^{-1}z,q) = \left[\,Q(q)\right]^3 E(z^3,q^3)/Q(q^3) \tag{2.6.61}$$

$$E(q,q^3) = Q(q) \tag{2.6.62}$$

$$E(\omega,q^3) = (1-\omega)Q(q^3) \tag{2.6.63}$$

Setting $z = \omega$ in (2.6.59) and (2.6.60), and using these identities, it follows that

$$f(\omega) = 1 \tag{2.6.64}$$

Since $f(z)$ is a constant, this implies that $f(z) \equiv 1$, which establishes the theorem. ■

Lemma 2.6.11. For $|q| < 1$ and all complex numbers z,

$$E(z,q) = Q(q) \prod_{i=1}^{r} \left[E(q^{i-1}z, q^r)/Q(q^r) \right] \tag{2.6.65}$$

Proof. Like (2.6.19), this follows directly from the infinite product expansions (1.5.6) and (2.2.7). ■

We can use the preceding lemma and theorem to write $\Phi_a(z)$ in a form from which the original hard hexagon results can be regained. First we define

$$\zeta_a = q^{(1/2)(a-1)(3a-4)}\left[E(q^{3a+5},q^{15}) - q^a E(q^{5-3a},q^{15})\right]/Q(q) \tag{2.6.66}$$

$$\zeta_a' = q^{(1/2)(a-1)(3a-4)}E(q^{3a},q^{15})/Q(q) \tag{2.6.67}$$

Theorem 2.6.12. For $r = 5$, $1 \leqslant a \leqslant 4$ and all complex numbers z,

$$\Phi_a(z) = \zeta_a z^{1-a}E(-q^{6a}z^{-3},q^{15}) + \zeta_a' z^{-a}\left[q^a E(-q^{6a+5}z^{-3},q^{15}) + q^{3a}z^{-1}E(-q^{5-6a}z^3,q^{15})\right] \tag{2.6.68}$$

Proof. Negate z in Lemma 2.6.11 and substitute the resulting expression for $E(-z,q)$ into (2.6.21). This gives

$$\Phi_a(z) = \frac{E(q^a,q^5)E(-q^iz,q^5)E(-q^jz,q^5)E(-q^kz,q^5)}{Q(q)Q(q^5)^2} \tag{2.6.69}$$

where i,j,k are the integers $1,2,3,4$ excluding a. Using the quasi-periodic property (2.6.68) of the E function, together with the simple identity $E(z,x) = E(x/z,x)$, the result (2.6.69) can be written as

$$\Phi_a(z) = q^{(1/2)(a-1)(3a-4)}z^{1-a}E(q^a,q^5)E(-q^{3a}z^{-1},q^5)E(-q^{2a}z^{-1},q^5) \times E(-q^a z^{-1},q^5)/\left[Q(q)Q(q^5)^2\right] \tag{2.6.70}$$

This expression contains a product of four E functions, which is the same as that on the left-hand side of Theorem 2.6.10, provided that z,w,q in that theorem are replaced by q^a, $-q^{2a}z^{-1},q^5$. We can therefore apply Theorem 2.6.10 to (2.6.70). Doing so, we obtain the desired result (2.6.68). ■

From (2.6.22) and (1.5.7), an immediate *corollary* of this theorem is that, for $1 \leqslant a \leqslant 4$ and all integers i,

$$\begin{aligned} \eta_{a,3i+1-a} &= \zeta_a q^{15i(i+1)/2-6ai} \\ \eta_{a,3i-a} &= \zeta_a' q^{15i(i+1)/2-(6a+5)i+a} \\ \eta_{a,3i-a-1} &= \zeta_a' q^{15i(i-1)/2-(6a-5)i+3a} \end{aligned} \tag{2.6.71}$$

[These expressions satisfy the periodicity and symmetry relations (2.6.34)–(2.6.36).]

Using these expressions (2.6.71) for $\eta_{a,j}$, from (2.6.58) it follows that

$$\begin{aligned}
F_1(0) &= q^{-1}\zeta_2 = \left[E(q^4,q^{15}) + qE(q,q^{15})\right]/Q(q) \\
F_1(1) &= q^{-9}\zeta_4 = \left[-qE(q^2,q^{15}) + E(q^7,q^{15})\right]/Q(q) \\
F_2(0) &= F_3(0) = q^{-1}\zeta_2' = E(q^6,q^{15})/Q(q) \\
F_2(1) &= F_3(1) = q^{-11}\zeta_4' = qE(q^{12},q^{15})/Q(q)
\end{aligned} \tag{2.6.72}$$

which are the results originally obtained for the hard hexagon model [Eq. 64 of Ref. 9, (14.5.22) of Ref. 10, Theorems 1–6 of Ref. 26].

3. NORMALIZATION OF THE PROBABILITIES P_a

3.1. The Normalization Factor M

The probabilities P_a are given, for the four regimes, by (2.3.21), (2.4.48), (2.5.8), and (2.6.54). Each of these equations is of the form

$$P_a = \mu_a/M \tag{3.1.1}$$

where

$$M = \sum_{1 \leqslant a \leqslant r}^{*} \mu_a \tag{3.1.2}$$

the $*$ indicating that the sum is either confined to the set of even integers or to the set of odd integers. It is sometimes convenient to write μ_a in the form

$$\mu_a = \mu_a' + \mu_{-a}' \tag{3.1.3}$$

Using (2.3.19), (2.5.6), (2.6.53), and the definitions (1.5.5), (1.5.10), (1.5.22) of u_a, v_a, we find that for the four regimes, μ_a or μ_a' is given by

$$\text{I} \qquad \mu_a = x^{(1/2)a(a+1-r)}E(x^a, -x^{r/2})E\left[x^{(r-2)a}, x^{r(r-2)}\right] \tag{3.1.4a}$$

$$\text{II} \qquad \mu_a = E(x^a, x^r)\hat{\eta}_{a,(1/2)(a+j)} \tag{3.1.4b}$$

$$\text{III} \qquad \mu_a' = x^{(1/2)a(a-1-2d)}E(x^a, x^r)E\left[-x^{2(r-a)(r-1)+2rd}, x^{4r(r-1)}\right] \tag{3.1.4c}$$

$$\text{IV} \qquad \mu_a' = x^{(1/2)a(a-1-2d)}E(x^a, x^{r/2})E\left[-x^{2(r-a)(r-2)+2rd}, x^{4r(r-2)}\right] \tag{3.1.4d}$$

We have omitted factors that are independent of a, and hence cancel out of (3.1.1). The function $\hat{\eta}_{a,i}$ is defined by (2.6.21), (2.6.22), and (2.6.44) with $q = x^{r-2}$. It satisfies the relations

$$\hat{\eta}_{a,i} = \hat{\eta}_{a,i+r-2} = x^{-2a-r}\hat{\eta}_{a+2r,i+r} = -x^{-a}\hat{\eta}_{-a,i-a} \tag{3.1.5}$$

for all integers a, i.

In (3.1.4), a, j, d are integers satisfying the restrictions

$$a + j = \text{even}, \qquad 1 \leqslant d \leqslant d_{\max} \tag{3.1.6}$$

where $d_{\max} = r - 2, r - 3$ for regimes III, IV, respectively. The integers j and d are to be regarded as fixed (i.e., independent of a): they determine the phase of the system in the ordered regimes.

The function μ_a is needed only for $1 \leqslant a < r$, and μ'_a for $1 \leqslant |a| < r$, but it can be convenient to extend the definitions (3.1.3) and (3.1.4) to all integers a. Using (2.6.28), (2.6.33), and (3.1.5), we can then verify that

$$\mu_{-a} = \mu_a = \mu_{a+2r}, \qquad \mu_0 = \mu_{\pm r} = 0 \tag{3.1.7}$$

Thus (3.1.2) can be put into the form

$$M = \frac{1}{2} \sum_{-r \leqslant a < r}^{*} \mu_a \tag{3.1.8}$$

It is clear from (3.1.1) and (3.1.2) that M is a normalization factor whose presence in (3.1.1) ensures that the local height probabilities satisfy the condition

$$\sum_{1 \leqslant a < r}^{*} P_a = 1 \tag{3.1.9}$$

This means that M plays a similar role to the partition function: in fact if we trace back its origin we see that M is proportional to S and T in (1.5.13) and (1.5.25), to the denominators in (A26) and (A1), and hence to the partition function (1.2.1) of the eight-vertex SOS model.

We also see from (3.1.7) and (3.1.8) that M has a simple mathematical significance: it is the sum of $\frac{1}{2}\mu_a$ over all equal-parity values of a within a complete period $2r$. For those cases (at present regimes III and IV) where μ'_a is defined, it is also true that

$$\mu'_a = \mu'_{a+2r}, \qquad \mu'_0 = \mu'_{\pm r} = 0 \tag{3.1.10}$$

so we can, using (3.1.3), write (3.1.8) as

$$M = \sum_{-r \leqslant a < r}^{*} \mu'_a \tag{3.1.11}$$

3.2. Theta-Function "Sums-of-Products" Identities

The Case $r = 5$. For the original hard hexagon model, which is when $r = 5$, it has been noted that the normalization factor M always turns out to be a simple product of theta functions. (There are nine different possibilities, depending on the regime and the phase.) Further, this simplification is always a corollary of one or more of the 40 "sums-of-products" identities[34] listed by Ramanujan and subsequently proved by Rogers,[35] Darling,[36] Mordell,[37] and Watson.[38] These identities involve the function $Q(x)$ defined by (2.2.7), as well as the function $P(x), G(x), H(x)$ defined by

$$P(x) = \prod_{n=1}^{\infty} (1 - x^{2n-1}) \tag{3.2.1}$$

$$G(x) = \prod_{n=1}^{\infty} \left[(1 - x^{5n-4})(1 - x^{5n-1})\right]^{-1} \tag{3.2.2}$$

$$H(x) = \prod_{n=1}^{\infty} \left[(1 - x^{5n-3})(1 - x^{5n-2})\right]^{-1} \tag{3.2.3}$$

For instance, consider regime I with $r = 5$ and a even. Then from (3.1.2) and (3.1.4a), setting $x_1 = -x^{1/2}$.

$$\begin{aligned} M &= \mu_2 + \mu_4 \\ &= x^{-2}E(x^2, -x^{2.5})E(x^6, x^{15}) + E(x^4, -x^{2.5})E(x^{12}, x^{15}) \\ &= x_1^{-4}Q(x_1)\,Q(x_1^6)\left\{H(x_1)G(x_1^6) - x_1G(x_1)H(x_1^6)\right\} \end{aligned} \tag{3.2.4}$$

However, Eq. 8 in Birch's list[34] of Ramanujan's identities is

$$G(x^6)H(x) - xG(x)H(x^6) = P(x)/P(x^3) \tag{3.2.5}$$

so we deduce at once that

$$\begin{aligned} M &= x_1^{-4}Q(x_1)\,Q(x_1^6)P(x_1)/P(x_1^3) \\ &= x_1^{-4}E(x_1, x_1^2)E(-x_1^3, x_1^{12}) \end{aligned} \tag{3.2.6}$$

The other $r = 5$ cases involve Eqs. 2, 5, 6, and 23 of Birch's[34] list, and are given in Refs. 9, 10, and 20.

Do these identities generalize to arbitrary values of r, so that M is always a simple product of theta functions? It turns out that the answer is yes, as is shown in the remainder of this section. As usual, regime II has to be treated separately; for regimes I, III, and IV there is one general identity (Theorem 3.2.1) that covers all cases.

Regimes I, III, and IV. The general identity that we shall need is the following.

Theorem 3.2.1. Let x, y be real numbers such that $|x| < 1$ and

$$y^{m/2} = -\epsilon x^r, \qquad \epsilon = \pm 1 \tag{3.2.7}$$

where m, r are positive integers and $1 \leqslant m < 2r$; then for all complex numbers z

$$\sum_{-r<a<r}^{*} x^{(1/2)a(a-1)} z^{a} E(x^{a}, y) E\left[\epsilon^{m-1} x^{(2r-m)(r+a)} z^{2r}, x^{2r(2r-m)}\right]$$
$$= \tfrac{1}{2}\left\{E(-z,x)E(z^{-1}, y/x) \pm E(z,x)E(-z^{-1}, y/x)\right\} \quad (3.2.8)$$

where the upper (plus) choice of the $\pm$ sign in (3.2.8) is to be made if the summation is restricted to even values of a; the lower (minus) sign applies if the sum is restricted to odd values.

Proof. First consider the quantity

$$J = \sum_{-\infty<a<\infty}^{*} \sigma(a) \quad (3.2.9)$$

where

$$\sigma(a) = x^{(1/2)a(a-1)} z^{a} E(x^{a}, y) \quad (3.2.10)$$

From (3.2.7) and the quasiperiodicity property (2.6.28) of $E(z,q)$, we can verify that (for all integers k)

$$\sigma(a+2kr) = x^{2kra+2k^2r^2-kr} z^{2kr} (-x^{a})^{-km} y^{-km(km-1)/2} \sigma(a)$$
$$= (-\epsilon^{m-1})^{k} x^{(2r-m)k(a+kr)} z^{2kr} \sigma(a) \quad (3.2.11)$$

For any integer a, there are unique integers a_0, k such that

$$a = a_0 + 2kr, \qquad -r \leqslant a_0 < r \quad (3.2.12)$$

and a_0 has the same parity as a. Thus (3.2.9) can be written

$$J = \sum_{-r\leqslant a_0<r}^{*} \sum_{k=-\infty}^{\infty} \sigma(a_0+2kr) \quad (3.2.13)$$

Dropping the suffix on a_0 and substituting the expression (3.2.11) for $\sigma(a+2kr)$ into (3.2.13), the k summation can be performed by using (1.5.7). This gives

$$J = \sum_{-r\leqslant a<r}^{*} \sigma(a) E\left[\epsilon^{m-1} x^{(2r-m)(r+a)} z^{2r}, x^{2r(2r-m)}\right] \quad (3.2.14)$$

Using (3.2.10) and noting that $\sigma(-r)=0$, we see that J is equal to the left-hand side of (3.2.8).

Now we go back to the definition (3.2.9)–(3.2.10) of J. Let J_e and J_0 be the values of J when the summation in (3.2.9) is restricted, respectively, to even and odd values of a. Then

$$J_e + \tau J_0 = \sum_{a=-\infty}^{\infty} \tau^{a} \sigma(a) \quad (3.2.15)$$

where $\tau = \pm 1$ and the summation is now over *all* integers a. Using (3.2.10) and (1.5.7), it follows that

$$J_e + \tau J_0 = \sum_{a=-\infty}^{\infty} \tau^a x^{(1/2)a(a-1)} z^a \sum_{k=-\infty}^{\infty} (-1)^k x^{ak} y^{(1/2)k(k-1)}$$

$$= \sum_{k=-\infty}^{\infty} (-\tau)^k z^{-k} (y/x)^{(1/2)k(k-1)} \sum_{j=-\infty}^{\infty} \tau^j z^j x^{(1/2)j(j-1)} \tag{3.2.16}$$

where we have interchanged the a, k summations and set $a = j - k$. Since the summand in (3.2.16) factors into a function of k times a function of j, the sum is a product of two E functions. Using (1.5.7), we obtain

$$J_e + \tau J_0 = E(-\tau z, x) E(\tau z^{-1}, y/x) \tag{3.2.17}$$

Taking sums and differences of this equation for $\tau = +1$ and $\tau = -1$, we can evaluate J_e and J_0. Remembering that the left-hand side of (3.2.8) is either J_e or J_0, we obtain (3.2.8), as desired. ■

Simple special cases of Theorem 3.2.1 occur when

$$z = x^l \quad \text{or} \quad z = -(y/x)^l \tag{3.2.18}$$

where l is an integer. The last term in (3.2.8) then vanishes, so the right-hand side reduces to the single product

$$\tfrac{1}{2} E(-z, x) E(z^{-1}, y/x) \tag{3.2.19}$$

and is independent of whether the summation is restricted to even or to odd values of a.

The particular identities that we shall need are all such simple special cases of Theorem 3.2.1. We shall also need the symmetry and periodicity properties (2.4.46) and (2.6.28) of $E(z, q)$, as well as the elementary identities (true for all integers d and complex numbers z, q with $|q| < 1$)

$$E(z, q) = E(-qz^2, q^4) - zE(-qz^{-2}, q^4) \tag{3.2.20}$$

$$E(-1, q) = 2E(-q, q^4) \tag{3.2.21}$$

$$E(-q^{-d}, q) = 2q^{-(1/2)d(d+1)} E(-q, q^4) \tag{3.2.22}$$

[The first follows from (1.5.7) by splitting the sum therein into two parts, one with n even and the other with n odd; the second is obtained from the first by taking $z = -1$; the third follows from the second and from (2.6.28).]

Taking $m = 2$, $\epsilon = -1$, $y = x^r$, and $z = x^{-d}$ (d an integer) in Theorem 3.2.1, using (2.4.46) and (3.2.22), we obtain

$$\sum_{-r<a<r}^{*} x^{(1/2)a(a-1-2d)} E(x^a, x^r) E\left[-x^{2(r-1)(r-a)+2rd}, x^{4r(r-1)}\right]$$

$$= x^{-(1/2)d(d+1)} E(-x, x^4) E(x^d, x^{r-1}) \tag{3.2.23}$$

Similarly, taking $m = 4$, $\epsilon = -1$, $y = -x^{r/2}$, and $z = x^{-d}$ (d an integer), we obtain

$$\sum_{-r<a<r}^{*} x^{(1/2)a(a-1-2d)}E(x^{a}, -x^{r/2})E\left[-x^{2(r-2)(r-a)+2rd}, x^{4r(r-2)}\right]$$
$$= x^{-(1/2)d(d+1)}E(-x, x^{4})E\left[x^{d}, -x^{(r-2)/2}\right] \qquad (3.2.24)$$

Identities (3.2.32) and (3.2.24) are relevant to regimes III and IV, respectively. For regime I we take $m=4$, $\epsilon = -1$, $y = -x^{r/2}$, and $z = w^{-(r-2)/2}$: this ensures that $z = -x/y$, so the second possibility in (3.2.18) is satisfied (for r even, so is the first). We then group the $(a, -a)$ terms together in (3.2.8) and use (2.6.33), (3.2.20), (2.6.28), and (3.2.21) to obtain

$$\sum_{1\leqslant a<r}^{*} x^{(1/2)a(a+1-r)}E(x^{a}, -x^{r/2})E\left[x^{(r-2)a}, x^{r(r-2)}\right]$$
$$= x^{(2r-r^{2}-\nu)/8}E(-x^{\nu/2}, x)E(x^{(r-2)/2}, x^{2r-4}) \qquad (3.2.25)$$

where

$$\begin{aligned} \nu &= 0, && \text{if } r \text{ is even} \\ &= 1, && \text{if } r \text{ is odd} \end{aligned} \qquad (3.2.26)$$

[For r even, (3.2.25) is a special case of (3.2.24), with $d = (r-2)/2$.]

Regime II. For regime II, the identity we want to establish is

$$\sum_{1\leqslant a<r}^{*} E(x^{a}, x^{r})\hat{\eta}_{a,(1/2)(a+j)} = Q(q) \qquad (3.2.27)$$

for all integers j, the sum being over integers a with the same parity as j. The function $\hat{\eta}_{a,i}$ is defined by (2.6.21), (2.6.22), and (2.6.44) with $q = x^{r-2}$. From (3.1.2) and (3.1.4b), M is equal to the left-hand side of (3.2.27), so it follows that

$$M = Q(q) \qquad (3.2.28)$$

One way to prove (3.2.27) is to first establish the following theorem.

Theorem 3.2.2. Let $\Phi_{a}(z)$ be defined by (2.6.21), with $q = x^{r-2}$. Then for all complex numbers z,

$$\sum_{1\leqslant a<r}^{*} x^{ra(a-2)/8}z^{-a/2}E(x^{a}, x^{r})\Phi_{a}(q^{-a/2}z)$$
$$= x^{-r\alpha/8}z^{(1/2)\alpha-1}Q(q)E(-x^{(1/2)r\alpha}z, x^{r}) \qquad (3.2.29)$$

where $\alpha = 0$ if the sum is restricted to even values of a, $\alpha = 1$ if the sum is restricted to odd values.

Proof. We regard z as a complex variable and use Liouville's theorem, rather as we did in Theorem 2.6.10. First we verify from (2.6.21) and

(1.5.6) that

$$\lim_{z\to -1} \Phi_a(z) = Q(q)$$
$$\lim_{z\to -1} \Phi_a(q^{-a}z) = (-1)^{a-1} q^{-(1/2)a(a-1)} Q(q) \tag{3.2.30}$$

Then we consider (3.2.29) when $z = -q^{b/2}$, where b is an integer of the same parity as a and α. If $1 \leqslant |b| < r$, then all terms in the sum on the left-hand side vanish except the one with $a = |b|$. Using (3.2.30) and (2.6.28), we can verify that (3.2.29) is satisfied. If $b = 0$ or r (and has the same parity as a and α), then all terms on the left-hand side and right-hand side vanish, so (3.2.29) is trivially satisfied.

Now let

$$f(z) = (\text{left-hand side} - \text{right-hand side}) \text{ of } (3.2.29) \tag{3.2.31}$$

This function is analytic in the domain $0 < |z| < \infty$, and by using (2.6.28) we can verify that

$$f(q^r z) = z^{2-r} q^{-(1/2)r(r-1)} f(z) \tag{3.2.32}$$

(This quasiperiodicity relation is satisfied by each term in the a summation.) From the above remarks it follows that

$$f(-q^{b/2}) = 0 \tag{3.2.33}$$

for all integers b of the same parity as a and α. This means that the function

$$g(z) = f(z)/E(-q^{-\alpha/2}z, q) \tag{3.2.34}$$

is also analytic in $0 < |z| < \infty$ [the denominator has only simple zeros, at which $f(z)$ vanishes].

Finally, consider the function

$$G(z) = z^{\alpha/2} g(z) E(-q^{\alpha/2}z, q^r) E(-q^{-\alpha/2}z, q^r) \tag{3.2.35}$$

[which is proportional to $f(z)$ divided by the $a = \alpha$ term of the sum in (3.2.29)]. This is analytic in $0 < |z| < \infty$ and is periodic:

$$G(q^r z) = G(z) \tag{3.2.36}$$

It is therefore bounded in $q^r \leqslant |z| \leqslant 1$, and hence in $0 < |z| < \infty$. This means that it must be analytic at $z = 0$ and ∞, so by Liouville's theorem it must be a constant. It vanishes when $z = -q^{\alpha/2}$, so the constant must be zero. Hence $G(z)$, $g(z)$, and $f(z)$ vanish identically, which proves the theorem. ■

The required identity (3.2.27) can now be established by Laurent expanding both sides of (3.2.29) in powers of z and equating coefficients, using (2.6.22), (2.6.44), and (1.5.7).

There is an alternative way to establish (3.2.28). Note that $\hat{\eta}$ enters

(3.1.4b) only because of the result (2.6.53) of Theorem 2.6.9. If we go back to the original functions $\hat{x}_m$ and, using (2.6.52) to x_m, we see that (3.1.2) is equivalent to

$$M = \lim_{m\to\infty} M_m(b,c) \tag{3.2.37}$$

where, for $1 \leqslant b < c < r$ and $c = b \pm 1$,

$$M_m(b,c) = \sum_{1\leqslant a<r}^{*} x^{(a+j)(a-j-r)/4} E(x^a, x^r) x_m(a,b,c) \tag{3.2.38}$$

the integer j being defined by (2.6.51); the summation is over integers a with the same parity as j.

In Section 2.6 we evaluated the large-m limit of $x_m(a,b,c)$. The derivation was very complicated, involving (in Appendix B) the theory of very well-poised q-hypergeometric series. By contrast, the large-m limit of $M_m(b,c)$ can be obtained quite simply and directly: the key is the following theorem.

Theorem 3.2.3. Let $M_n(b,c)$ be defined by (3.2.38), (2.6.51) and (2.6.4)–(2.6.5), with $q = x^{r-2}$. Define ξ (an integer or half an integer) by

$$\xi = (c-b)(\tfrac{1}{2}r - b) \tag{3.2.39}$$

Then for $|x| < 1$, $1 \leqslant b$, $c < r$ and $c = b \pm 1$,

$$M_m(b,c) = \sum_{t=-\infty}^{\infty} (-1)^t x^{(1/2)rt^2-(m+\xi)t} (q^{1-t}; q)_m \tag{3.2.40}$$

where we are using the notation (2.2.8).

Proof. From (2.6.4) and (2.6.33) we can verify that the summand in (3.2.38) is an even function of a, and that

$$M_m(b,c) = \sum_{-r\leqslant a<r}^{*} E(x^a, x^r)\sigma(a) \tag{3.2.41}$$

where

$$\sigma(a) = x^{(a+j)(a-j-r)/4} q^{(m+a-b)/4} f_m(a,b,c) \tag{3.2.42}$$

and we have used the fact that $\sigma(0) = \sigma(-r) = 0$.

We now proceed similarly to the first part of the proof of Theorem 3.2.1. Incrementing a by $2r$, and replacing λ in (2.6.5) by $\lambda + 1$, we find that

$$\sigma(a+2r) = x^{2a+r}\sigma(a) \tag{3.2.43}$$

and hence, by recurrence,

$$\sigma(a+2kr) = x^{2ka+rk(2k-1)}\sigma(a) \tag{3.2.44}$$

for all integers k.

If we define

$$J = \sum\nolimits^*_{-\infty < a < \infty} \sigma(a)(1 - x^a) \tag{3.2.45}$$

then by proceeding as in (3.2.12) and (3.2.13) we obtain, using (1.5.7) and (3.2.41),

$$\begin{aligned} J &= \sum\nolimits^*_{-r \leqslant a_0 < r} \sigma(a_0) \sum_{k=-\infty}^{\infty} \left[x^{2ka_0 + rk(2k-1)} - x^{(2k+1)a_0 + rk(2k+1)} \right] \\ &= \sum\nolimits^*_{-r \leqslant a < r} \sigma(a) \sum_{l=-\infty}^{\infty} (-1)^l x^{la + (1/2)rl(l-1)} \\ &= \sum\nolimits^*_{-r \leqslant a < r} \sigma(a) E(x^a, x^r) \\ &= M_m(b,c) \end{aligned} \tag{3.2.46}$$

Thus J is equal to $M_m(b,c)$. On the other hand, substituting (2.6.5) into (3.2.42) and (3.2.45), interchanging the a, λ summations and setting

$$a = -j + 2(r-2)\lambda + 2t \tag{3.2.47}$$

(t must be an integer), we obtain

$$\begin{aligned} J = \sum_{\lambda=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} & (1 - q^{2\lambda} x^{2t-j}) q^{\lambda(2\lambda + b - c) - (1/2)(b+1-c)t} \\ & \times x^{t(t-j-1)} \begin{bmatrix} m \\ \frac{1}{2}(1+b-c)m + t - 2\lambda \end{bmatrix} \end{aligned} \tag{3.2.48}$$

Let $F(\lambda, t)$ be the summand in (3.2.48) with the factor $(1 - q^{2\lambda} x^{2t-j})$ omitted. Then one can verify that

$$F(\lambda + \tfrac{1}{2}, t+1) = q^{2\lambda} x^{2t-j} F(\lambda, t) \tag{3.2.49}$$

so (3.2.48) can be written as

$$J = \sum_{\lambda=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \left[F(\lambda, t) - F(\lambda + \tfrac{1}{2}, t+1) \right] \tag{3.2.50}$$

The two terms in (3.2.50) can be grouped together by replacing (λ, t) by $(\frac{1}{2}\mu, t)$ in the first, and by $(\frac{1}{2}\mu - \frac{1}{2}, t-1)$ in the second. Then μ is even for the first term, odd for the second: combining them we get

$$J = \sum_{\mu=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} (-1)^{\mu} F(\tfrac{1}{2}\mu, t) \tag{3.2.51}$$

the sums being over all integers μ, t.

Next we interchange the μ, t summations (such interchanges are justified by the rapid convergence of the series due to the quadratic exponents in the summand), and set

$$\mu = t + (b - c)k \tag{3.2.52}$$

(Since $|b - c| = 1$, k is an integer.) Then, after substituting the explicit form of the summand, (3.2.51) becomes

$$J = \sum_{t=-\infty}^{\infty} (-1)^t x^{(1/2)rt(t-1)-jt} \sum_{k=0}^{m} (-1)^k q^{(1/2)k(k+1)+(b-c)kt} \begin{bmatrix} m \\ k \end{bmatrix} \tag{3.2.53}$$

where we have used (2.2.4), and the fact that $|b - c| = 1$. The k summation can be restricted to $0 \leqslant k \leqslant m$ because of (2.2.1).

The k summation can now be performed by using the identity

$$\sum_{k=0}^{m} (-1)^k q^{(1/2)k(k-1)} z^k \begin{bmatrix} m \\ k \end{bmatrix} = (z)_m \tag{3.2.54}$$

(Theorem 3.3, p. 36 of Ref. 16). This gives

$$j = \sum_{t=-\infty}^{\infty} (-1)^t x^{(1/2)rt(t-1)-jt} (q^{1+(b-c)t}; q)_m \tag{3.2.55}$$

Replacing t by $(c - b)t$, using (2.6.51) and remembering that $J = M_m(b, c)$, we obtain the desired result (3.2.40). This completes the proof of Theorem 3.2.3. ■

It is instructive to explicitly expand the product in the summand of (3.2.40):

$$(q^{1-t}; q)_m = (1 - q^{1-t})(1 - q^{2-t}) \cdots (1 - q^{m-t}) \tag{3.2.56}$$

We see that this vanishes if $t = 1, 2, \ldots, m$, so we can restrict the t summation to the two regions $t \leqslant 0$ and $t > m$. Setting $t = -s$ in the first, $t = s + m + 1$ in the second, and using (2.2.11), we can write (3.2.40) as

$$M_m(b, c) = \sum_{s=0}^{\infty} (-1)^s x^{(1/2)rs^2 + ms} \left[x^{\xi s} - x^{-\xi s + [(1/2)r - \xi](m+1)} \right] (q^{s+1}; q)_m \tag{3.2.57}$$

It is now easy to take the limit $m \to \infty$. From (3.2.39), $|\xi| < \frac{1}{2}r$, so as $m \to \infty$ all terms in the summation in (3.2.57) tend uniformly to zero, except for the first part of the $s = 0$ term. Thus, using (3.2.37),

$$M = \lim_{m \to \infty} M_m(b, c) = (q; q)_\infty = Q(q) \tag{3.2.58}$$

We have thus rederived (3.2.28), without using the working of Appendix B and Section 2.6, and without introducing the functions η and $\hat{\eta}$ (which are multidimensional theta series).

3.3. Final Results

We can now write down explicit expressions for the local height probabilities. They are given by (3.1.1)–(3.1.4), and for regimes I, II, III, IV, we can write M as a simple product of theta functions by using (3.2.25), (3.2.28), (3.2.23), (3.2.24), respectively. We obtain

$$\text{I:}\quad P_a = x^{(1/2)a(a+1-r)+(r^2-2r+\nu)/8}E(x^a, -x^{r/2}) \times \frac{E[x^{(r-2)a}, x^{r(r-2)}]}{\{E(-x^{\nu/2},x)E(x^{(r-2)/2},x^{2r-4})\}} \tag{3.3.1a}$$

$$\text{II:}\quad P_a = E(x^a, x^r)\hat{\eta}_{a,(1/2)(a+j)}/Q(q) \tag{3.3.1b}$$

III and IV:

$$P_a = x^{(1/2)a(a-1)+(1/2)d(d+1)}E(x^a, y) \times \frac{\{x^{-ad}E[-x^{2l(r-a)+2rd}, x^{4rl}] - x^{ad}E[-x^{2l(r+a)+2rd}, x^{4rl}]\}}{[E(-x,x^4)E(x^d, y/x)]} \tag{3.3.1c}$$

where ν is defined by (3.2.26); $y = x^r$, $l = r-1$ in regime III; while in regime IV $y = -x^{r/2}$, $l = r-2$. In all cases x lies in the interval $(0, 1)$.

The function $\hat{\eta}$ is defined by (2.6.21), (2.6.22), and (2.6.44), with $q = x^{r-2}$. Setting $j = j_0 + (r-2)k$ in (2.6.22), where j_0 and k are integers such that $0 \leqslant j_0 \leqslant r-3$, we can perform the k summation by using the periodicity property (2.6.45). [This procedure parallels that in (3.2.9)–(3.2.14).] We obtain the identity

$$\Phi_a(z) = \sum_{j=0}^{r-3} x^{(1/2)rj(j+1)-aj}\hat{\eta}_{a,j}z^j E[-q^{(1/2)r(r-1)+rj-a}z^{r-2}, q^{r(r-2)}] \tag{3.3.2}$$

where $q = x^{r-2}$. We can regard the $\hat{\eta}_{a,j}$ as being defined by this identity and the periodicity relation (2.6.45).

We can readily use (1.5.6) or (1.5.7) to expand all these results in increasing powers of x, so these forms are convenient for examining the behavior when x is small, which is when the eight-vertex SOS model is completely disordered (regime I) or completely ordered (regime II, III, and IV): the local height probabilities P_a take their ground state values.

Critical Behavior. The other extreme is when $x \to 1$. From (1.5.8) and (1.5.20), ϵ is then large and p in (1.2.3)–(1.2.9) is numerically small. In the limit when $x = 1$ and $p = 0$ the eight-vertex weights a, b, c, d in (1.2.3) satisfy

$$d = 0, \qquad |(a^2 + b^2 - c^2)/(2ab)| < 1 \tag{3.3.3}$$

The model is then critical,[10] so it is of particular interest to consider this case and to be able to develop perturbation expansions in powers of p.[24] This can be done by transforming to elliptic functions of conjugate modulus. [Examples of such transformations have already occurred in (A30) and (A51).] The results are conventionally expressed in terms of the standard theta functions,[17,18] defined for all complex numbers u, q, with $|q| < 1$, by

$$\theta_1(u, q^2) = 2|q|^{1/4}\sin u \prod_{n=1}^{\infty} (1 - 2q^{2n}\cos 2u + q^{4n})(1 - q^{2n}) \quad (3.3.4a)$$

$$\theta_4(u, q^2) = \prod_{n=1}^{\infty} (1 - 2q^{2n-1}\cos 2u + q^{4n-2})(1 - q^{2n}) \quad (3.3.4b)$$

$$\theta_2(u, q^2) = \theta_1\left(\frac{\pi}{2} + u, q^2\right), \qquad \theta_3(u, q^2) = \theta_4\left(\frac{\pi}{2} + u, q^2\right) \quad (3.3.4c)$$

In regimes I and IV we encounter θ_1 functions with q^2 negative, which is why we take q^2 rather than q to be the second argument. Note that θ_1 remains real even when q^2 is negative. Sometimes we shall find it convenient to remove the $2|q|^{1/4}$ factor and to work with the functions

$$\tilde{\theta}_i(u, q^2) = \tfrac{1}{2}|q|^{-1/4}\theta_i(u, q^2), \qquad i = 1, 2 \quad (3.3.5)$$

When $q = 0$, these functions $\tilde{\theta}_1$ and $\tilde{\theta}_2$ reduce to $\sin u$ and $\cos u$.

The "conjugate modulus" identities that we shall need are [for all complex numbers u, ϵ such that $\mathrm{Re}(\epsilon) > 0$]

$$\theta_1(u, e^{-\epsilon}) = \rho(u, \epsilon) E(e^{-4\pi u/\epsilon}, e^{-4\pi^2/\epsilon}) \quad (3.3.6a)$$

$$\theta_4(u, e^{-\epsilon}) = \rho(u, \epsilon) E(-e^{-4\pi u/\epsilon}, e^{-4\pi^2/\epsilon}) \quad (3.3.6b)$$

$$\theta_1\left(\tfrac{1}{2}u, -e^{-\epsilon/4}\right) = 2^{1/2}\rho(u, \epsilon) E(e^{-4\pi u/\epsilon}, -e^{-4\pi^2/\epsilon}) \quad (3.3.6c)$$

where

$$\rho(u, \epsilon) = (2\pi/\epsilon)^{1/2}\exp\left[\left(2\pi u - 2u^2 - \tfrac{1}{2}\pi^2\right)/\epsilon\right] \quad (3.3.6d)$$

These are Eqs. (14.2.42) and (14.6.1) of Ref. 10, after allowing for differences of notation. Note that if ϵ is large then the second argument of the function E is numerically close to 1, which is what happens near criticality in our results (3.3.1). The product or series expansions (1.5.6) and (1.5.7) then converge very slowly. On the other hand, the second argument of the functions θ_1 and θ_4 is then small, so the product expansions (3.3.4) converge rapidly. This is why it is desirable near criticality to use the transformation (3.3.6).

First apply the transformation to the definition (3.3.2) of the function $\hat{\eta}_{a,j}$ that occurs in regime II. The parameters x, p, and ϵ are then related by

(1.5.8). Using (2.6.21) and setting

$$\lambda_{a,j} = \left[(r-2)/r\right]^{1/2} \exp\left[-\frac{\epsilon}{8} + \frac{r\epsilon}{24(r-2)} - \frac{(r-2)\pi^2}{6r\epsilon} + \frac{(r-a)^2\pi^2}{2r^2\epsilon} \right] \hat{\eta}_{a,j} \tag{3.3.7}$$

$$z = \exp\left[-4\pi(r-2)u/\epsilon\right] \tag{3.3.8}$$

$$t = p^{1/(r-2)} = e^{-\epsilon/(r-2)} \tag{3.3.9}$$

$$F_a(u) = \frac{Q(t)^3 \tilde{\theta}_1(\pi a/r, t)\theta_4(ru, t^r)}{Q(t^r)^2 \theta_4(u,t)\theta_4(u+\pi a/r, t)} \tag{3.3.10}$$

the identity (3.3.2) transforms to

$$F_a(u) = \frac{r}{4(r-2)} \sum_{j=0}^{r-3} \lambda_{a,j} \theta_4\left[u + \frac{\pi(r-1)}{2(r-2)} + \frac{\pi j}{r-2} - \frac{\pi a}{r(r-2)}, t^{1/(r-2)} \right] \tag{3.3.11}$$

true for all complex numbers u. This $F_a(u)$ is an entire function of u, periodic of period π.

From (2.6.45) and (3.3.7)

$$\lambda_{a,j+r-2} = \lambda_{a,j} \tag{3.3.12}$$

We can regard the $\lambda_{a,j}$ as defined by (3.3.8)–(3.3.12). We can make this more explicit by introducing the Fourier coefficients $f_{a,n}$ of $F_a(u)$:

$$F_a(u) = \sum_{n=-\infty}^{\infty} f_{a,n} e^{2inu} \tag{3.3.13}$$

which satisfy the quasiperiodicity relation

$$f_{a,n+r-2} = -t^{n+(1/2)(r-2)} e^{-2\pi i a/r} f_{a,n} \tag{3.3.14}$$

Using the series expansion of $\theta_4(u)$ (Eq. 8.192.1 of Ref. 17), we find that the $\lambda_{a,j}$ are given by

$$\lambda_{a,j} = \frac{4}{r} \sum_{n=0}^{r-3} \exp\left[\frac{-2\pi i n(j+\frac{1}{2}-a/r)}{r-2} \right] t^{-n^2/(2r-4)} f_{a,n} \tag{3.3.15}$$

for all integers a, j.

Now we apply the transformations (3.3.6) to our results (3.3.1), using the relations (1.5.8) (between x, ϵ, and p) in regimes II and III, and the relations (1.5.20) in regimes I and IV. It is convenient to define

$$s = p^{1/(4r-r)}, \qquad t = |p|^{1/(r-2)} \tag{3.3.16}$$

$$R_a = r/\theta_1(\pi a/r, p) \tag{3.3.17}$$

[This is consistent with (3.3.9).] After a fair bit of cumbersome manipulation we obtain, for the four regimes I–IV,

$$\text{I:}\quad P_a = \frac{2\theta_1(\pi a/r, t^2)}{R_a\theta_4(\pi\nu/2, p^{2r})\theta_2(0, -t^r)} \tag{3.3.18a}$$

$$\text{II:}\quad P_a = \lambda_{a,(1/2)(a+j)}\tilde{\theta}_1(\pi a/r, p)/Q(t^r) \tag{3.3.18b}$$

$$\text{III:}\quad P_a = \frac{\theta_3(\pi a/2r - \pi d/(2r-2), s) - \theta_3(\pi a/2r + \pi d/(2r-2), s)}{R_a\theta_4(0, p^r)\theta_1(\pi d/(r-1), s^{4r})} \tag{3.3.18c}$$

$$\text{IV:}\quad P_a = \frac{\theta_3(\pi a/2r - \pi d/(2r-4), t^{1/2}) - \theta_3(\pi a/2r + \pi d/(2r-4), t^{1/2})}{R_a\theta_4(0, p^{2r})\theta_1(\pi d/(r-2), -t^r)} \tag{3.3.18d}$$

As before, ν is defined by (3.2.26), $Q(q)$ by (2.2.7). The integers j and d determine the phase of the system in the ordered regimes. They must satisfy (3.1.6) and are defined in terms of the boundary conditions by (2.6.51), (2.3.18)–(2.3.20), and (2.5.10), for regimes II, III, IV, respectively. [Thus $d = \frac{1}{2}(b + c - 1)$ in regime III.]

These results are expressed in terms of the original parameter p that entered (1.2.3)–(1.2.9), which is negative for regimes I and IV, positive for II and III. For fixed values of w_0/K, v/K, and η/K [from (1.4.1), $\eta/K = r^{-1}$], the SOS model weights W are real functions of p, analytic for $|p| < 1$. At $p = 0$ we are on the interface between regimes I and II (for $\eta < v < 3\eta$), or between III and IV (for $-\eta < v < \eta$). The model is then critical, so we can naturally regard p as a "deviation from criticality" variable.

At $p = 0$ we find from (3.3.18) that for all regimes

$$P_a = P_a^{(c)} = 4r^{-1}\sin^2(\pi a/r) \tag{3.3.19}$$

The (c) denoting the critical value of P_a. Note that this result is independent of j and d, i.e., is independent of boundary conditions and is the same for all phases. This is what we expect to happen at criticality: long-range order has disappeared.

We can expand the P_a in increasing powers of p (or s, or t). For regime II we first need to note from (3.3.10) and (3.3.13) that

$$f_{a,n} = t^{n/2}e^{in\pi a/r}\sin\frac{(n+1)\pi a}{r}\left\{1 + 0(t^2)\right\}, \qquad 0 \leqslant n \leqslant r-2 \tag{3.3.20}$$

$$f_{a,n} = \sin(\pi a/r) + t^2\sin(3\pi a/r) + \cdots \tag{3.3.21}$$

We then find that

$$\text{I:}\quad P_a = P_a^{(c)}\{1 - (-p)^{2/(r-2)}[1 + 2\cos(2\pi a/r)] + O[p, p^{6/(r-2)}]\} \tag{3.3.22a}$$

$$\text{II:}\quad P_a = P_a^{(c)}\left\{1 + 4p^{(1/2)(r-3)/(r-2)^2}\cos\frac{\pi a}{r}\cos\frac{\pi(j+1)}{r-2} + O\left[p, p^{(r-4)/(r-2)^2}\right]\right\} \tag{3.3.22b}$$

$$\text{III:}\quad P_a = P_a^{(c)}\left\{1 + 4p^{3/(8r-8)}\cos\frac{\pi a}{r}\cos\frac{\pi d}{r-1} + O[p, p^{1/(r-1)}]\right\} \tag{3.3.22c}$$

$$\text{IV:}\quad P_a = P_a^{(c)}\left\{1 + 4(-p)^{3/(4r-8)}\cos\frac{\pi a}{r}\cos\frac{\pi d}{r-2} + O[p, p^{2/(r-2)}]\right\} \tag{3.3.22d}$$

[The terms of order $p^{(r-4)/(r-2)^2}$ in (3.3.22b) occur only for $r \geqslant 6$.]

For the ordered regimes II, III, IV, the leading deviations from criticality are proportional to the differences between the P_a for different phases, and hence to the long-range-order parameter. Their exponents are therefore the critical exponent β,[10] so

$$\beta = \frac{r-3}{2(r-2)^2}, \qquad \frac{3}{8(r-1)}, \qquad \frac{3}{4(r-2)} \tag{3.3.23}$$

for regimes II, III, IV, respectively.

The probabilities P_a are in general local properties of the lattice: they depend on the site to which the height l_1 in (1.5.1) refers. For definiteness we have up to now taken this to be the center site, but it could be any site deep within the lattice.

In regime I the P_a depend only on whether the site lies on the X or the Y sublattice. (We can think of these sublattices as the centers of the black and white faces of a checkerboard.) For a given phase, the same is true in regimes III and IV. For regime II, there are $r-2$ different functions P_a, depending on the position of the site, corresponding to the $r-2$ distinct (modulo $2r-4$) equal-parity values of j that are allowed in (3.3.1b) and (3.3.18b). Thus in regime II we can define an average function $\tilde{P}_a$ obtained by averaging over these values of j. From (3.3.18b), (3.3.15), and (3.3.21), this is given by

$$\begin{aligned}\bar{P}_a &= 4r^{-1}f_{a,0}\tilde{\theta}_1(\pi a/r, p)/Q(t^r)\\ &= P_a^{(c)}\{1 + p^{2/(r-2)}[1 + 2\cos(2\pi a/r)] + \cdots\}\end{aligned} \tag{3.3.24}$$

In regime I, the corresponding average $\overline{P}_a$ is equal to P_a. For both regimes, $\overline{P}_a$ is a nonlocal property of the model, independent of boundary conditions and phase. (For the hard hexagon model $\overline{P}_a$ corresponds to the mean density.) From (3.3.22a) and (3.3.24) we see that for regimes I and II,

$$\overline{P}_a = P_a^{(c)}\left\{1 + \operatorname{sgn}(p)|p|^{1-\bar{\alpha}}\left[1 + 2\cos(2\pi a/r)\right] + \cdots\right\} \quad (3.3.25a)$$

where the critical exponent $\bar{\alpha}$ is given by

$$1 - \bar{\alpha} = 2/(r-2) \quad (3.3.25b)$$

Free Energy. To complete the discussion of the critical behavior, we should consider the free energy of the SOS model, with heights restricted by (1.4.3). This can be obtained by the "inversion relation" method,[21] using (A11), (A12), and (A13). It is the same as the free energy of the unrestricted model, and of the original eight-vertex model with weights given by (1.2.3), except that in regime II the free energy is that of the "SDP-like" eight-vertex model.[21]

Let ρ', η, v, K, p be the parameters entering the definitions (1.2.3)–(1.2.9) of the Boltzmann weights of the model, and $h(u)$ the function defined by (1.2.4) and (1.2.7). In regimes II and III define

$$\begin{aligned} \rho_0 &= \rho' h(2\eta)\exp\left[-\pi(v^2-\eta^2)/(2KK')\right] \\ \tau &= 2\pi K/K', \qquad \lambda = 2\pi\eta/K' \\ u &= \pi(\eta+v)/K', \qquad q = e^{-\tau} \end{aligned} \quad (3.3.26)$$

Then we can verify that the definition (1.2.3) of the eight-vertex model weights a, b, c, d is equivalent to that of Eqs. (6.1)–(6.5) and (6.11) of Ref. 21 (apart from negating and permuting some of $w_1, \ldots, w_4$ therein: such transformations leave the partition function unchanged[2]). It follows that κ, the partition function per site, is given by Eqs. (6.30) and (6.32) of Ref. 21. Using (1.5.4), we see that $-\eta < v < \eta$, $0 < u < \lambda$ in regime III, so Eq. (6.32a) is applicable; in regime II, $\eta < v < \kappa - \eta$, $\lambda < u < \frac{1}{2}\tau$, so (6.32b) applies.

For regime III we can use the original eight-vertex model results.[2] The $q, \pi/\mu$ of Eq. (E10) of Ref. 2 correspond to $p^{1/2}, K/\eta$ herein, so we find that the dominant critical singularity in κ is

$$\text{III:} \quad \kappa_{\text{sing}} = p^{K/(2\eta)} \quad (3.3.27)$$

This is multiplied by a factor that in general is a nonzero analytic function of p at $p = 0$. Exceptions occur when $r = K/\eta$ is an integer, which from (1.4.1) is precisely the case we are considering. If r is even (3.3.27) should be multiplied by a factor $\ln p$; if r is odd the singular part of κ disappears altogether, as has been observed for the hard hexagon model.[8,10]

For regime II, a similar calculation to that of Appendix E of Ref. 2 yields

$$\text{II:}\quad \kappa_{\text{sing}} = p^{K/(K-2\eta)} \tag{3.3.28}$$

In regimes I and IV the parameter p is negative. Define L'/K by (A50), i.e.,

$$p = -e^{-\pi L'/K} \tag{3.3.29}$$

and define

$$\begin{gathered} \rho_0 = \rho' h(2\eta)\exp\left[-\pi(v^2 - \eta^2)/2KL'\right] \\ q^2 = -e^{-\pi K/L'}, \qquad \lambda = \pi\eta/L' \\ u = \pi(\eta + v)/2L' \end{gathered} \tag{3.3.30}$$

These definitions ensure that (1.2.3) is equivalent to Eqs. (6.1)–(6.5) of Ref. 21. It is now Eqs. (6.9a) and (6.9b) that are appropriate, the former applying in regime IV ($-\eta < v < \eta$), the latter in regime I, ($\eta < v < K - \eta$). From them we find that the singular part of κ is

$$\text{IV:}\quad \kappa_{\text{sing}} = (-p)^{K/2\eta} \tag{3.3.31}$$

$$\text{I:}\quad \kappa_{\text{sing}} = (-p)^{K/(K-2\eta)} \tag{3.3.32}$$

These free energy results (3.3.26)–(3.3.32) apply for all values of η in the interval $(0, K/2)$. Setting $r = K/\eta$ as in (1.4.1), we see that

$$\kappa_{\text{sing}} = |p|^{2-\alpha} \tag{3.3.33}$$

where

$$\begin{aligned} 2 - \alpha &= r/(r-2), && \text{in regimes I and II} \\ &= r/2, && \text{in regimes III and IV} \end{aligned} \tag{3.3.34}$$

For the hard hexagon case, when $r = 5$, these values of α agree with those previously obtained (Eq. 14.6.10 of Ref. 10), with the proviso, noted after (3.3.27), that κ then has no singular part in regime III.

We note also that in regimes I and II the free energy exponent α is the same as the average height probability exponent $\tilde{\alpha}$ given by (3.3.25). In general it is not obvious why this should be so, since differentiating the logarithm of the partition function with respect to p gives not only terms proportional to $\overline{P}_a$, but also terms proportional to the correlation of four heights round a face of the lattice. (For the hard hexagon case, with $r = 5$

and $v = 3\eta$, such multiple-height correlations do not occur, so then we must have $\bar{\alpha} = \alpha$.)

4. SUMMARY

We have considered the SOS model with weights given by (1.2.6), or equivalently (1.2.12b), which is known[11,12] to have the same partition function as the eight-vertex model. We have shown that the hard hexagon model is equivalent to a special case of this SOS model, in which $\eta = K/5$ and each height l_i is restricted to the range $1 \leqslant l_i \leqslant 4$.

One intriguing feature of the hard hexagon model is that the Rogers–Ramanujan[39,40] identities, and many related identities, occur naturally in the calculation of the sublattice densities.[9,20,26] In the SOS model these densities correspond to the probability that the height at a particular site has a given value. We have therefore considered a more general SOS model, in which $\eta = K/r$ (r an integer), each height being restricted to the range $1 \leqslant l_i \leqslant r-1$. Sure enough, when we calculate the local height probabilities we are led to generalizations of the Rogers–Ramanujan identities.

There are two main sorts of such identities. In Section 2 we evaluate the $(m-1)$-fold sums (1.5.11) and (1.5.23), and show that in the limit $m \to \infty$ they are modular forms. (In regimes I, III, and IV they are sums of at most two simple quotients of elliptic theta functions.) These identities are generalizations of the Rogers–Ramanujan identities, as well as of some similar identities listed by Slater.[31] They are presumably related to Gordon's generalization.[15,16]

The other sort of identity occurs in Section 3, where we calculate the normalization factors M occurring in the definition of the local height probabilities P_a. [These factors are proportional to (1.5.13) or (1.5.25), and ensure that the necessary condition (1.5.2) is satisfied.] For regimes I, III, and IV these identities involve sums of products of theta functions of different nomes. They are generalizations of some of the 40 such identities listed by Ramanujan,[34] and are derivable from Theorem 3.2.1.

Regime II is much more difficult than the other regimes. Its solution (in the limit $m \to \infty$) involves the theory of very well-poised q – hypergeometric series (see Appendix B), and introduces the multi-dimensional theta series $\eta_{a,j}$ and $\hat{\eta}_{a,j}$. (Curiously, the normalization factor M can be evaluated without these complications, as is shown in Theorem 3.2.3.)

In Section 3.3. we present our final results for the P_a, put them into a form suitable for examining critical behavior, and obtain some of the critical exponents $\beta, \alpha, \bar{\alpha}$ of the restricted SOS model.

For definiteness, we have focused our attention on the case $\eta = K/r$ (r an integer), which is the obvious immediate generalization of the hard hexagon case $\eta = K/5$. However, the working can be further generalized to the case $\eta = sK/r$ (r and s integers, $1 \leqslant s < r$), as is being done by one of us (PJF).

APPENDIX A: CORNER TRANSFER MATRICES

Let A, B, C, D be the corner transfer matrices (CTMs) corresponding to the lower-right, upper-right, upper-left, and lower-left quadrants of the lattice, as on p. 366 of Ref. 10. Each will break up into $r-1$ diagonal blocks, one for each value of the center spin l_1. Let S_a be the corresponding diagonal matrix whose diagonal entries are unity for the block with $l_1 = a$, all other elements being zero. Then (1.5.1) is equivalent to

$$P_a = \mathrm{Trace}\, S_a ABCD/\mathrm{Trace}\, ABCD \tag{A1}$$

The matrices A, B, C, D are products of local "face transfer matrices" U_j and V_j that merely add one face at a time to the lattice [Eq. (13.2.4) of Ref. 10, Eq. (9) of Ref. 23]. These matrices have elements

$$(U_j)_{\mathbf{l}\mathbf{l}'} = W(l_j, l_{j+1} \,|\, l_{j-1}, l_j') \prod_{\substack{k=1\\ \neq j}}^{m} \delta(l_k, l_k') \tag{A2}$$

$$(V_j)_{\mathbf{l}\mathbf{l}'} = W(l_{j-1}, l_j \,|\, l_j', l_{j+1}) \prod_{\substack{k=1\\ \neq j}}^{m} \delta(l_k, l_k') \tag{A3}$$

Here $\mathbf{l}$ denotes the set of heights $\{l_1, l_2, \ldots, l_m\}$, $\mathbf{l}'$ denotes $\{l_1', l_2', \ldots, l_m'\}$, and $j = 2, 3, \ldots, m+1$. The heights $l_1, l_2, \ldots$ correspond to those an observer would see if he started at the center of the lattice and walked outwards along the edges. Thus they (and $l_1', l_2', \ldots$) must satisfy the adjacency condition (1.2.2), i.e.,

$$|l_{j+1} - l_j| = |l_{j+1}' - l_j'| = 1, \qquad j \geqslant 1 \tag{A4}$$

The boundary heights of the lattice are to be fixed at their ground state values (there may be more than one ground state, in which case we have to select a particular one). The heights $l_{m+1}, l_{m+1}', l_{m+2}$ that occur in the definitions of $U_m, V_m, U_{m+1}, V_{m+1}$ are to be given these boundary values, which will in general mean that these boundary face transfer matrices depend on the face to which they refer.

Special Properties of the U_j, V_j. From (1.2.12), all the matrices U_j, V_j, A, B, C, D are functions of v, so we can for instance write U_j as

$U_j(v)$. When $v = \eta$, we see from (1.2.12) that, for l, m, n, p satisfying (1.2.2),

$$W(l,m \mid n,p) = \rho' h(2\eta)\delta(l,p) \tag{A5}$$

By substituting this result into (A2), it follows that

$$U_j(\eta) = \rho' h(2\eta) I \tag{A6}$$

where I is the identity matrix.

Similarly, when $v = -\eta$ we find from (1.2.12) that

$$W(l,m \mid n,p) = \rho' h(2\eta)\left[\frac{h(w_l)h(w_p)}{h(w_m)h(w_n)}\right]^{1/2} \delta(m,n) \tag{A7}$$

Using (A3), it follows that

$$V_j(-\eta) = \rho' h(2\eta) R_{j-1} R_{j+1} R_j^{-2} \tag{A8}$$

where R_j is the diagonal matrix with entries

$$(R_j)_{ll'} = \left[h(w_{l_j})\right]^{1/2} \prod_{k=1}^{m} \delta(l_k, l'_k) \tag{A9}$$

For general values of v, the matrices U_j and V_j are very sparse, breaking up into one-by-one and two-by-two blocks. They can therefore easily be inverted. Defining

$$\xi(v) = \left[\rho'^2 h(2\eta + v) h(2\eta - v)\right]^{-1} \tag{A10}$$

we find that the inverse of U_j is also given by (A2), but with W multiplied by $\xi(v-\eta)$ and v in (1.2.12) replaced by $2\eta - v$. Thus

$$U_j^{-1}(v) = \xi(v-\eta) U_j(2\eta - v) \tag{A11}$$

Similarly, the inverse of V_j is given by (A3), but with $W(l,m \mid n,p)$ multiplied by $\xi(v+\eta)h(w_m)h(w_n)/[h(w_l)h(w_p)]$, and v replaced by $-2\eta - v$. It follows that

$$V_j^{-1}(v) = \xi(v+\eta) R_{j-1}^{-2} R_j^2 V_j(-2\eta - v) R_j^2 R_{j+1}^{-2} \tag{A12}$$

These inversion relations (A11), (A12) are consistent with (A6) and (A8), being satisfied by them when $v = \eta, -\eta$, respectively.

Actually, we shall use neither (A11) nor (A12), but a simple corollary of (A12) that follows from (1.4.1) and the fact that W is negated if v is increased by $2K$:

$$V_j^{-1}(v) = -\xi(v+\eta) R_{j-1}^{-2} R_j^2 V_j\left[(2r-2)\eta - v\right] R_j^2 R_{j+1}^{-2} \tag{A13}$$

Properties of A, B, C, D. The matrix A, i.e., $A(v)$, is defined as

$$A(v) = F_2 F_3 \dots F_{m+1} \tag{A14a}$$

where

$$F_j = U_{m+1}^{(j)} U_m^{(j)} U_{m-1} \cdots U_j \tag{A14b}$$

the superscripts (j) on U_{m+1} and U_m denoting the fact that these matrices (but not $U_2, \ldots, U_{m-1}$) depend on the boundary heights and hence on the value of j in (A14b).

The matrices B, C, D are defined similarly, but with each U_j replaced by V_j, U_j^T, V_j^T, respectively. In fact U_j and V_j are symmetric matrices, but it does not necessarily follow that $C = A$ or $D = B$, since the boundary heights for C (D) may differ from those of A (B).

The ground states [i.e., the values of the heights of all lattice sites that maximize the summand in (1.2.1)], do themselves depend upon v, but only in the sense that they have fixed values in one domain of the complex v plane, then change discontinuously to other fixed values as v crosses from this domain to another. We find that we have two domains to consider:

$$\begin{aligned} &\mathscr{D}_1: \quad -1 < \operatorname{Re}(v/\eta) < 1 \\ &\mathscr{D}_2: \quad 1 < \operatorname{Re}(v/\eta) < r-1 \end{aligned} \tag{A15}$$

We can allow v to reach, or even briefly cross, the domain boundaries, so long as we interpret our results as the appropriate analytic continuations of the intradomain values.

In both cases we can therefore let $v = \eta$. From (A6) and (1.14) it follows that (to within irrelevant scalar factors)

$$A(\eta) = C(\eta) = I \tag{A16}$$

This result is consistent with the fact that the ground states are unchanged by uniform shifts in the SW–NE direction [i.e., the height on site (x, y) is the same as that on site $(x+1, y+1)$]. Since U_j, V_j are symmetric, it follows that

$$\begin{gathered} A^T(v) = A(v), \qquad C^T(v) = C(v) \\ D^T(v) = B(v) \end{gathered} \tag{A17}$$

In the domain $\mathscr{D}_1$, we can also let $v = -\eta$. Substituting the form (A8) of $V_j(-\eta)$ into the appropriate analogs of (A14), we find that $R_2, R_3 \ldots$ all cancel out, leaving (to within irrelevant scalar factors)

$$B(-\eta) = D(-\eta) = R_1 \tag{A18}$$

Thus in domain $\mathscr{D}_1$

$$A(\eta)B(-\eta)C(\eta)D(-\eta) = R_1^2 \tag{A19}$$

In the domain $\mathscr{D}_2$ we can allow v to be close to the value $(r-1)\eta$ (this is a "virtual inversion point"[21]), and use (A13). Together with (A14) and the symmetry property of the ground states, this implies that

$$\left[B(v)\right]^{-1} = R_1^{-2}D\left[(2r-2)\eta - v\right] \tag{A20}$$

[There is a problem in deriving this, in that (A13) is not necessarily satisfied for $j = m + 1$. However, it seems that we can ignore this boundary difficulty in the limit of m large.]

The matrix R_1, like $S_1, \ldots, S_{r-1}$, commutes with all the corner transfer matrices, so in domain $\mathscr{D}_2$ we have the relation

$$A(\eta)B(v)C(\eta)D\left[(2r-2)\eta - v\right] = R_1^2 \tag{A21}$$

provided v is sufficiently close to $(r-1)\eta$.

A, B, C, D **as Exponentials in** v. The star–triangle relation (1.33), (1.34) implies[10,22] that in the limit of m large the corner transfer matrices have the form (to within irrelevant scalar factors)

$$\begin{aligned} A(v) &= Q_1 M_1 e^{v\mathscr{H}} Q_2^{-1} \\ B(v) &= Q_2 M_2 e^{-v\mathscr{H}} Q_3^{-1} \\ C(v) &= Q_3 M_3 e^{v\mathscr{H}} Q_4^{-1} \\ D(v) &= Q_4 M_4 e^{-v\mathscr{H}} Q_1^{-1} \end{aligned} \tag{A22}$$

where $\mathscr{H}$, $Q_1, \ldots, Q_4$, $M_1, \ldots, M_4$ are matrices that are independent of v and commute with $R_1, S_1, \ldots, S_{r-1}$, and $\mathscr{H}, M_1, \ldots, M_4$ are diagonal. Substituting these forms into (A1), we obtain

$$P_a = \operatorname{Trace} S_a M_1 M_2 M_3 M_4 / \operatorname{Trace} M_1 M_2 M_3 M_4 \tag{A23}$$

Thus we need to calculate the product $M_1 M_2 M_3 M_4$. First substitute the forms (A22) into (A19) or (A21) depending on whether v is in domain $\mathscr{D}_1$ or $\mathscr{D}_2$. We obtain

$$M_1 M_2 M_3 M_4 e^{2t\eta\mathscr{H}} = R_1^2 \tag{A24}$$

where

$$t = 2, 2 - r, \qquad \text{in domains } \mathscr{D}_1, \mathscr{D}_2 \tag{A25}$$

respectively. Thus (A23) becomes

$$P_a = \operatorname{Trace} S_a R_1^2 e^{-2t\eta\mathscr{H}} / \operatorname{Trace} R_1^2 e^{-2t\eta\mathscr{H}} \tag{A26}$$

This simplifies the problem to the calculation of $\mathscr{H}$. However, setting $v = \eta$ in the first of the equations (A22) and using (A16), we find that

$$Q_1 M_1 = Q_2 e^{-\eta\mathscr{H}} \tag{A27}$$

and hence

$$A(v) = Q_2 e^{(v-\eta)\mathscr{H}} Q_2^{-1} \tag{A28}$$

Thus $\exp[(v-\eta)\mathscr{H}]$ is simply the diagonal form of the single corner transfer matrix $A(v)$. We can calculate this by using certain quasiperiodic properties of $A(v)$ (corresponding to incrementing v by iK', or $2iK'$): these imply that the elements of $\mathscr{H}$ are integer multiples of π/K', and we can calculate these integers from special limiting cases.

To do this, we have to distinguish whether the nome p in the definition (1.2.7) of the elliptic function $h(u)$ is positive or negative. Together with the division (A15) of the complex v plane, this gives us four cases to consider. If we restrict v to be real and less than 3η (which ensures that the Boltzmann weights can be chosen to be real and nonnegative), then these four cases are the four "regimes" in (1.5.4). However, here we shall still find it convenient to regard v as a complex variable, and extend the allowed values of v to

$$\begin{aligned} v &\in \mathscr{D}_2, \qquad \text{in regimes I and II} \\ &\in \mathscr{D}_1, \qquad \text{in regimes III and IV} \end{aligned} \tag{A29}$$

From (A25), we see that the values of t in the four regimes are given by (1.5.5).

Regimes II and III. When $0 < p < 1$ (regimes II and III), we can make a conjugate modulus transformation[10,17] so as to write the definition (1.2.7) of the function $h(u)$ as

$$h(u) = \tau \exp\left[-\pi(u-K)^2/(2KK')\right] E(e^{-2\pi u/K'}, y) \tag{A30}$$

where

$$y = \exp(-4\pi K/K') \tag{A31}$$

$$\tau = \frac{K}{K'} \prod_{n=1}^{\infty} \frac{1-y^{n/2}}{1+y^{n/2}} \tag{A32}$$

and the function $E(z, y)$ is defined by (1.5.6). Note that y, like p, lies in the interval (0, 1), but is close to one when p is small, and vice-versa.

Using the form (A30) for $h(u)$, and defining

$$x = e^{-4\pi\eta/K'}, \qquad w = e^{-2\pi(\eta - v)/K'} \tag{A33}$$

$$g_l = \exp\left[\pi(v-\eta)(w_l - K)^2/8\eta KK'\right] \tag{A34}$$

$$\nu = \rho' \tau E(x, y) \exp\left[\pi(4K\eta - K^2 - v^2 - 3\eta^2)/2KK'\right] \tag{A35}$$

$$\mu_l = \exp(-2\pi w_l/K'), \qquad E_l = E(\mu_l, y) \tag{A36}$$

then, using (1.2.5) but not (1.4.1) nor (1.4.2), we can write (1.2.12b) as

$$\begin{aligned}
\alpha_l &= \nu\left[g_l^2/(g_{l-1}g_{l+1})\right] w^{1/2} E(xw^{-1})/E(x) \\
\beta_l &= \nu(g_{l-1}g_{l+1}/g_l^2)\left[xE_{l-1}E_{l+1}/(wE_l^2)\right]^{1/2} E(w)/E(x) \\
\gamma_l &= \nu(g_{l+1}/g_l)^2 E(\mu_l w)/E(\mu_l) \\
\delta_l &= \nu(g_{l-1}/g_l)^2 E(\mu_l w^{-1})/E(\mu_l)
\end{aligned} \tag{A37}$$

where here we have written $E(z, y)$ simply as $E(z)$.

[These reduce to the hard hexagon weights in the case $\eta = K/5$, provided that here we replace ν, g_l by unity, and take r in Eq. (28b) of Ref. 8 to be $w^{1/2}$.] Now using (1.4.1) and (1.4.2), we can verify that the x defined by (A33) is the same as that defined by (1.5.8), and that

$$y = x^r, \qquad \mu_l = x^l \tag{A38}$$

When calculating A, we can ignore the parameter ν in (A37), since it merely contributes an irrelevant scalar factor. The terms involving g_l have the effect of multiplying $W(l,m' \mid l',m)$ by $g_l g_m / g_{l'} g_{m'}$ and this in turn simply divides $A_{\mathbf{l}\mathbf{l}'}$ by g_{l_1}. The remaining terms in (A37) involve v only via integer or half-integer powers of w, so are periodic functions of v, of period $2iK'$. Within the appropriate domain $\mathscr{D}_1$ or $\mathscr{D}_2$, defined in (A15), it appears that both A and its diagonal form are analytic functions of v, so from (A28) it follows that the elements of $\exp[(v-\eta)\mathscr{H}]$ are

$$\left[e^{(v-\eta)\mathscr{H}}\right]_{\mathbf{l}\mathbf{l}'} = g_{l_1}^{-1} w^{N(\mathbf{l})/2} \delta(\mathbf{l},\mathbf{l}') \tag{A39}$$

where $N(\mathbf{l})$ is an integer function of $\mathbf{l}$, and we use the notation

$$\delta(\mathbf{l},\mathbf{l}') \equiv \prod_{k=1}^{m} \delta(l_k, l_k') \tag{A40}$$

Assuming (as seems perfectly reasonable) that $\mathscr{H}$ does not change discontinuously with p, the integers $N(\mathbf{l})$ must be independent of p. We can therefore obtain them by studying the simple case when p and w are close to one, i.e., when η/K' and K/K' are large, $-1 < \mathrm{Re}(v/\eta) < 3$ and

$$x \ll |x| \ll x^{-1} \tag{A41}$$

It then follows from (A37) that

$$|\beta_l^2/(\gamma_l\delta_l)| \ll 1 \tag{A42}$$

for $l = 2, \ldots, r-2$, which implies that the matrices U_j, and hence A, are near diagonal (in the sense that their off-diagonal elements give a negligible contribution to their eigenvalues). In the limit when x is small, we can take

β_l to be zero: from (1.2.12a) and (A37) we then obtain

$$W(l, m' \mid l', m) = v\left[g_l g_m / (g_{l'} g_{m'}) \right] w^{|l'-m'|/4} \delta_{lm} \tag{A43}$$

The corner transfer matrix A is then diagonal, with entries

$$\begin{aligned} A_{\mathbf{ll}'} &= \prod_{j=1}^{m} \left[W(l_{j+1}, l_{j+2} \mid l_j, l_{j+1}) \right]^j \delta(\mathbf{l}, \mathbf{l}') \\ &= g_{l_1}^{-1} w^{\phi(\mathbf{l})} \delta(\mathbf{l}, \mathbf{l}') \end{aligned} \tag{A44}$$

where $\phi(\mathbf{l})$ is defined by (1.5.12) and we have ignored some $\mathbf{l}$-independent (i.e., scalar) factors in A. The boundary heights l_{m+1}, l_{m+2} that occur here must be fixed at their values for the particular ground state under consideration.

From (A28), $\exp[(v-\eta)\mathscr{H}]$ is the diagonal form of A, so in this small-x limit it is given by (A44). From (A39) it follows that

$$N(\mathbf{l}) = 2\phi(\mathbf{l}) \tag{A45}$$

This is indeed an integer function; using the continuity argument mentioned above, $N(\mathbf{l})$ must be given by (A45) for all p in the interval $(0, 1)$. From (A39), (A33), and (A34) (with $\eta - v$ replaced by $2t\eta$), it follows that

$$\left[e^{-2t\eta\mathscr{H}} \right]_{\mathbf{ll}'} = x^{[-t(2l_1 - r)^2/16r + t\phi(\mathbf{l})]} \delta(\mathbf{l}, \mathbf{l}') \tag{A46}$$

Also, from (A9) and (A30),

$$\left(R_1^2 \right)_{\mathbf{ll}'} = \tau x^{(2l_1 - r)^2/8r} E(x^{l_1}, y) \delta(\mathbf{l}, \mathbf{l}') \tag{A47}$$

Substituting these results into (A26), cancelling the scalar factor τ, and remembering that

$$(S_a)_{\mathbf{ll}'} = \delta(l_1, a) \delta(\mathbf{l}, \mathbf{l}') \tag{A48}$$

we obtain the result (1.5.9)–(1.5.14).

Regimes I and IV. When $-1 < p < 0$ (regimes I and IV), the elliptic integral K is real, but K' is complex, being of the form

$$K' = L' + iK \tag{A49}$$

where L' is real and positive, and

$$p = -e^{-\pi L'/K} \tag{A50}$$

We can write (1.2.7) in the "conjugate modulus" form[10]

$$h(u) = \tau \exp\left[\pi u(K - u)/2KL' \right] E(e^{-\pi u/L'}, y) \tag{A51}$$

where the function $E(z, y)$ is again defined by (1.5.6), and in these regimes

I and IV we define y and τ by

$$y = -\exp(-\pi K/L') \tag{A52}$$

$$\tau = y^{1/4}\frac{K}{L'}\prod_{n=1}^{\infty}\frac{1-y^{2n}}{1-y^{2n-1}} \tag{A53}$$

Since y is negative, the constant τ is complex. This is just a trivial complication due to the standard definitions of the elliptic theta functions. It is readily compensated by an appropriate definition of ρ' in (1.2.12b). In any case, τ will cancel out of our final result for P_a.

Substituting this form for $h(u)$ into (1.2.12b), defining

$$x = e^{-2\pi\eta/L'}, \qquad w = e^{-\pi(\eta-v)/L'} \tag{A54}$$

$$g_l = \exp\left[\pi(v-\eta)(w_l - K)^2/8\eta KL'\right] \tag{A55}$$

$$\nu = \rho'\tau E(x,y)\exp\left[\pi(2K\eta - 3\eta^2 - v^2)/2KL'\right] \tag{A56}$$

$$\mu_l = \exp(-\pi w_l/L'), \qquad E_l = E(\mu_l, y) \tag{A57}$$

we find that α_l, β_l are again given by (A37), while the expressions for γ_l, δ_l now contain extra factors $w^{1/2}, w^{-1/2}$, respectively. Thus

$$\begin{aligned}\gamma_l &= \nu(g_{l+1}/g_l)^2 w^{1/2} E(\mu_l w)/E(\mu_l)\\ \delta_l &= \nu(g_{l-1}/g_l)^2 w^{-1/2} E(\mu_l w^{-1})/E(\mu_l)\end{aligned} \tag{A58}$$

[For the case $\eta = K/5$, the hard hexagon weights, Eq. (28a) of Ref. 8, can be obtained by replacing x, w, g_l, ν herein by $x^2, w^{-1}, w^{-l/2}, w^{1/2}$.]

From (1.4.1) and (1.4.2),

$$y = -x^{r/2}, \qquad \mu_l = x^l \tag{A59}$$

As in regimes II and III, we expect the elements of the diagonal form of A to be integer or half-integer powers of w, divided by g_{l_1}, i.e., we expect (A39) to be valid. To obtain these powers, we focus attention on the case when x is small and

$$x^{1/2} \ll |w| \ll x^{-1/2} \tag{A60}$$

Provided $r \neq 2l$, we find that (A42) is again satisfied, so in the limit of x small we can take $\beta_l = 0$. Provided l' and m' are not both equal to $r/2$, it follows that

$$W(l, m' \mid l', m) = \nu w^{1/2}(g_l g_m / g_{l'} g_{m'}) w^{-H(l', l, m')}\delta_{lm} \tag{A61}$$

where the function $H(l, l', l'')$ is defined by (1.5.26).

For odd values of r, it follows immediately that A is diagonal in the limit $x \to 0$, with elements

$$A_{\mathbf{l}\mathbf{l}'} = g_{l_1}^{-1} w^{-\psi(\mathbf{l})} \delta(\mathbf{l}, \mathbf{l}') \tag{A62}$$

where $\psi(\mathbf{l})$ is defined by (1.5.24) and we have ignored l-independent factors. As in regimes II and III, the boundary heights l_{m+1}, l_{m+2} are to be fixed at their ground state values.

The argument is a little more complicated if r is even, since (A42) then fails for $l = r/2$. This means that in the small-x limit most of the off-diagonal elements $A_{\mathbf{l}\mathbf{l}'}$ of A are zero, but nonzero elements occur when

$$l_j \neq l_j' \quad \text{and} \quad l_{j-1} = l_{j+1} = l_{j-1}' = l_{j+1}' = r/2 \tag{A63}$$

It follows that A is block diagonal, with blocks that are direct products of one-by-one or two-by-two matrices. It can therefore readily be diagonalized: the eigenvectors are found to be independent of w [as (A28) implies], while the diagonal form is again given by (A62), (1.5.24), and (1.5.26).

From (A28), (A39), and (A62), we see that in the small-x limit

$$N(\mathbf{l}) = -2\psi(\mathbf{l}) \tag{A64}$$

This is indeed an integer function, so should be valid for all x in the interval $0 < x < 1$, i.e., for all p in $(-1, 0)$. Replacing $\eta - v$ by $2t\eta$ in (A39), (A54), and (A55), it follows that

$$\left[e^{-2t\eta\mathscr{H}} \right]_{\mathbf{l}\mathbf{l}'} = x^{[-t(2l_1 - r)^2/8r - t\psi(\mathbf{l})]} \delta(\mathbf{l}, \mathbf{l}') \tag{A65}$$

while from (A9) and (A51),

$$\left(R_1^2 \right)_{\mathbf{l}\mathbf{l}'} = \tau x^{l_1(2l_1 - r)/2r} E(x^{l_1}, y) \delta(\mathbf{l}, \mathbf{l}') \tag{A66}$$

Substituting these expressions into (A26), using (A48) and (A59), and canceling a factor $\tau x^{-tr/8}$, we obtain the result (1.5.21)–(1.5.26).

APPENDIX B: q-HYPERGEOMETRIC SERIES

The study of q-hypergeometric series has an extensive literature.[28,29,32] We shall consider just those developments necessary for the treatment of regime II. We shall use the notation (2.2.8) and its generalization

$$(A_1, A_2, \ldots, A_r; q)_n = (A_1)_n (A_2)_n \cdots (A_r)_n \tag{B1}$$

The *ordinary q-hypergeometric series* is defined by

$${}_n\Phi_s\begin{pmatrix} \alpha_1, \ldots, \alpha_n; q, t \\ \beta_1, \ldots, \beta_s \end{pmatrix} = \sum_{j=0}^{\infty} \frac{(\alpha_1, \ldots, \alpha_n; q)_j t^j}{(q, \beta_1, \ldots, \beta_s; q)_j} \tag{B2}$$

This series terminates after finitely many terms if one of the α_i is a nonpositive integer power of q: this is the case we shall be considering.

The series is said to be "well poised" if $s = n - 1$ and

$$\beta_{j-1}\alpha_j = qa, \qquad j = 1, \ldots, n \tag{B3}$$

where $a = \alpha_1$ and we take β_0 to be the fixed entry q in (B2). It is "very well poised" if it is also true that

$$\alpha_2, \alpha_3, \beta_1, \beta_2 = qa^{1/2}, -qa^{1/2}, a^{1/2}, -a^{1/2} \tag{B4}$$

One can readily verify that in this case the entries $\alpha_2, \alpha_3, \beta_1, \beta_2$ in (B2) merely contribute a combined factor

$$\frac{(qa^{1/2})_j(-qa^{1/2})}{(a^{1/2})_j(-a^{1/2})_j} = \frac{1 - aq^{2j}}{1 - a} \tag{B5}$$

to the summand.

We shall restrict our attention to such very well poised q-hypergeometric series, and shall take n to be even, $n \geqslant 6$, and α_n to be the nonpositive integer power of q. Then we can write $\alpha_1, \ldots, \alpha_n$ as

$$\alpha_1, \ldots, \alpha_n = a, qa^{1/2}, -qa^{1/2}, b_1, c_1, b_2, c_2, \ldots, b_k, c_k, q^{-N} \tag{B6}$$

where

$$n = 2k + 4, \qquad s = 2k + 3 \tag{B7}$$

and N is a nonnegative integer. We take t to be

$$t = a^k q^{k+N}/(b_1 \ldots b_k c_1 \ldots c_k) \tag{B8}$$

The series (B2) can then be written as a $(k-1)$-dimensional summation (Theorem 4 of Ref. 29). At first sight this may not appear to be progress, but for $k = 2$ the identity yields Watson's[33] proof of the Rogers–Ramanujan identities. We shall in fact find that it enables us to handle the tricky $m \to \infty$ limit for regime II.

We do not need the full theorem, but can restrict our attention to the limit when

$$c_1, \ldots, c_k \to 0 \tag{B9}$$

(Since the series is a rational function of each c_i, there is no difficulty in letting them become zero.) Then Theorem 4 of Ref. 29 becomes the following.

Theorem B1. For N, k integers, $N \geqslant 0$, $k \geqslant 1$,

$$\sum_{j=0}^{N} q^{(1/2)kj(1-j)} a^j \frac{1 - q^{2j}a}{1 - a} \prod_{i=1}^{k+2} \frac{(-b_i)^{-j}(b_i)_j}{(aq/b_i)_j}$$
$$= q^{(1/2)N(N-1)}(aq)_N \sum_{\{m\}} \prod_{i=1}^{k} \frac{(b_{i+1})_{M_i} q^{(1/2)(m_i + M_i)(m_i - M_i + 1)}}{(-b_i)^{M_i}(q)_{m_i}(aq/b_i)_{M_i}} \tag{B10}$$

where

$$b_{k+1} = q^{-N}, \qquad b_{k+2} = a \tag{B11}$$

$$M_i = m_1 + m_2 + \cdots + m_i \tag{B12}$$

and the $\{m\}$ summation is over all integer values of $m_1, \ldots, m_k$ such that

$$m_i \geqslant 0, \quad i = 1, \ldots, k; \qquad m_1 + m_2 + \cdots + m_k = N \tag{B13}$$

Remarks. The left-hand side of (B10) is the q-hypergeometric series (B2) with the substitutions (B3)–(B9). This theorem is precisely the same as Theorem 4 of Ref. 29 with $c_1, \ldots, c_k$ set equal to zero: we have merely condensed the notation somewhat by using the definitions (B11)–(B13) and the properties (2.2.10), (2.2.11).

Theorem B2. Let the function ρ_N be defined by (2.6.6), and set

$$d_{k+1} = q/w, \qquad d_{k+2} = q \tag{B14}$$

Then for N, k integers, $N \geqslant 0$, $k \geqslant 1$:

$$\rho_N(q, w, k+2; d_1, d_2, \ldots, d_{k+2}) = (1-w) \sum_{\{m\}} \prod_{i=1}^{k} \frac{q^{(1/2)m_i(m_i-1)}(-d_i)^{m_i}}{(q)_{m_i}(d_i)_{M_i}(wd_i)_{N-M_{i-1}}} \tag{B15}$$

where the M_i are defined by (B12) (wht $M_0 = 0$) and the summation is over all integers $m_1, \ldots, m_k$ satisfying (B13).

Proof. This is merely a restatement of Theorem B1, in which we have set

$$a = q^{-N}w^{-1}, \qquad b_i = aq/d_i, \qquad i = 1, \ldots, k+2 \tag{B16}$$

Doing this, using (2.2.10) and (2.2.11), and multiplying both sides of (B10) by

$$(1 - q^N w)/\left[(wd_1)_N (wd_2)_N \cdots (wd_{k+2})_N\right] \tag{B17}$$

we obtain the desired result (B15). It also follows that the function ρ_N in (B15) is proportional to the very well-poised series in (B10).

We shall be interested in the limit when $N \to \infty$, while $q, w, k, d_1, \ldots, d_k$ remain fixed. We must exclude the cases when the function ρ_N is infinite, so for the rest of this Appendix it is to be understood that, for all integers i and all integers n such that $n \geqslant 0$,

$$d_i \neq q^{-n}, \qquad wd_i \neq q^{-n} \tag{B18}$$

We also restrict our attention to the case when

$$|q| < 1 \tag{B19}$$

Theorem B3. For $k \geqslant 1$,

$$\lim_{N\to\infty} \rho_N(q, w, k+2; d_1, \ldots, d_k, qw^{-1}, q)$$

$$= (1-w) \sum_{\{m\}} \prod_{i=1}^{k} \frac{q^{(1/2)m_i(m_i-1)}(-d_i)^{m_i}}{(q)_\infty (d_i)_\infty (wd_i)_\infty} \tag{B20}$$

where now the $\{m\}$ summation is over all integers m_i such that

$$-\infty < m_i < \infty, \quad i = 1, \ldots, k; \qquad m_1 + m_2 + \cdots + m_k = N \tag{B21}$$

we are using the convention (2.6.20).

Proof. The denominator in (B15) is uniformly bounded. The numerator is maximized when $m_i = k^{-1}N + O(1)$, $i = 1, \ldots, k$. Let f be some number which is large compared with unity but much less than N/k, e.g., $N/2k$. Split the $\{m\}$ summation in Theorem B2 into two parts A and B, A having all m_i greater than f, B containing all other values of $m_1, \ldots, m_k$. Then the power $\frac{1}{2}\sum m_i(m_i - 1)$ of q ensures that for N large B is negligible compared with A. In part A the suffixes $m_i, M_i, N - M_{i-1}$ (for $i = 1, \ldots, k$) all tend to infinity with N, so the terms in the denominator of (B15) can all be replaced by their respective limits. The $\{m\}$ summation can then be extended to include negative values of the m_i, since the extra terms thus introduced are also negligible compared with A.

Theorem B4. For $k \geqslant 1$,

$$\lim_{N\to\infty} \rho_N(q, w, k+2; d_1, \ldots, d_k, qw^{-1}, q)$$

is equal to the coefficient of z^{-N} in the Laurent expansion in powers of z of

$$(1-w) \prod_{i=1}^{k} \frac{E(d_i/z, q)}{(q)_\infty (d_i)_\infty (wd_i)_\infty} \tag{B22}$$

where $E(z, x)$ is the elliptic function given by (1.5.7).

Proof. Substituting the series form (1.5.7) of the function $E(z, x)$ into (B22) and collecting terms in the k-fold product that are proportional to z^{-N}, we obtain the right-hand side of (B20).

Theorem B5. For $s \geqslant 3$, one of $d_1, \ldots, d_s$ equal to q/w and another equal to q,

$$\lim_{N\to\infty} \rho_N(q, w, s; d_1, \ldots, d_s)$$

is equal to the coefficient of z^{-N} in the Laurent expansion in powers of z of

$$\frac{(q)_\infty^3 E(w, q)}{E(z, q) E(zw, q)} \prod_{i=1}^{s} \frac{E(d_i/z, q)}{(q)_\infty (d_i)_\infty (wd_i)_\infty} \tag{B23}$$

Proof. Using the definition (2.6.6), both sides of (B23) are symmetric functions of $d_1, \ldots, d_s$, so it is sufficient to prove the relation for some particular ordering of $d_1, \ldots, d_s$. Choose

$$d_{s-1} = qw^{-1}, \qquad d_s = q \tag{B24}$$

Substituting these values into (B23), setting $s = k + 2$ and noting from (1.5.6) that $E(q/z, q) = E(z, q)$ and

$$E(w, q) = (1 - w)(q)_\infty (qw)_\infty (qw^{-1})_\infty \tag{B25}$$

NOTE ADDED IN PROOF

Many of the identities of Secs. 2 and 3 were suggested and/or checked by partially expanding them in powers of the variable q or x, using a computer and (in Sec. 2) IBM's symbolic manipulation language "Scratchpad." For r odd, the identity (3.2.25) has been obtained by Bressoud (Eq. 5.3 of Ref. 41, with $k = (r-1)/2$, and $a - x^n$ therein corrected to $1 + x^n$).

REFERENCES

1. C. Fan and F. Y. Wu, *Phys. Rev. B* **2**:723 (1970).
2. R. J. Baxter, *Ann Phys. (N.Y.)* **70**:193 (1972).
3. F. Y. Wu, *Phys. Rev. B* **4**:2312 (1971); L. P. Kadanoff and F. J. Wegner, *Phys. Rev. B* **4**:3989 (1971).
4. R. J. Baxter, *J. Phys. C* **6**:L445 (1973).
5. E. H. Lieb and F. Y. Wu, in *Phase Transitions and Critical Phenomena*, Vol. 1, C. Domb and M. S. Green, eds. (Academic, London, 1972), pp. 321–490.
6. R. J. Baxter and F. Y. Wu, *Austral. J. Phys.* **27**:357 and 369 (1974).
7. R. J. Baxter and I. G. Enting, *J. Phys. A* **9**:L149 (1976).
8. R. J. Baxter, *J. Phys. A*, **13**:L61 (1980).
9. R. J. Baxter, *J. Stat. Phys.* **26**:427 (1981).
10. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
11. R. J. Baxter, *Ann. Phys. (N.Y.)*, **76**:1 (1973).
12. R. J. Baxter, *Ann. Phys. (N.Y.)*, **76**:25 (1973).
13. R. J. Baxter, *Ann. Phys. (N.Y.)*, **76**:48 (1973).
14. M. Gaudin, *La fonction d'onde de Bethe* (Masson, Paris, 1983).
15. B. Gordon, *Am. J. Math.* **83**:393 (1961).
16. G. E. Andrews, *The Theory of Partitions* (Addison-Wesley, Reading, Massachusetts, 1976).
17. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
18. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, 1950).
19. N. E. Pegg, *J. Phys. A* **15**:L549 (1982).
20. R. J. Baxter and P. A. Pearce, *J. Phys. A* **15**:897 (1982); **16**:2239 (1983).
21. R. J. Baxter, *J. Stat. Phys.* **28**:1 (1982).

22. R. J. Baxter, in *Fundamental Problems in Statistical Mechanics V*, E. G. D. Cohen, ed. (North-Holland, Amsterdam, 1980).
23. R. J. Baxter, *Physica* **106A**:18 (1981).
24. D. A. Huse, *Phys. Rev. Lett.* **49**:1121 (1982).
25. T. M. Apostol, *Modular Functions and Dirichlet Series in Number Theory* (Springer, New York, 1976).
26. G. E. Andrews, *Proc. Natl. Acad. Sci. (U.S.A.)* **78**:5290 (1981).
27. I. Schur, *Sitzungsber. Akad. Wissensch. Berlin Phys.-Math. Kl.* p. 488 (1926).
28. C. R. Adams, *Bull. Am. Math. Soc.* **37**:361 (1931).
29. G. E. Andrews, in *Theory and Application of Special Functions*, R. Askey, ed. (Academic, New York, 1975).
30. G. E. Andrews, *Proc. Natl. Acad. Sci. (U.S.A.)* **71**:4082 (1974).
31. L. J. Slater, *Proc. London Math. Soc. (2)* **54**:147 (1952).
32. L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge University Press, Cambridge, 1966).
33. G. N. Watson, *J. London Math. Soc.* **4**:4 (1929).
34. B. J. Birch, *Math. Proc. Cambridge Philos. Soc.* **78**:73 (1975).
35. L. J. Rogers, *Proc. London Math. Soc. (2)*, **19**:387 (1921).
36. H. B. C. Darling, *Proc. London Math. Soc. (2)* **19**:350 (1921).
37. L. J. Mordell, *Proc. London Math. Soc. (2)* **20**:408 (1922).
38. G. N. Watson, *J. Indian Math. Soc.* **20**:57 (1933).
39. L. J. Rogers, *Proc. London Math. Soc.* **25**:318 (1894).
40. S. Ramanujan, *Proc. Cambridge Philos. Soc.* **19**:214 (1919).
41. D. M. Bressoud, *J. London Math. Soc. (2)* **16**:9 (1977).