# q colourings of the triangular lattice



1986 J. Phys. A: Math. Gen. 19 2821

(http://iopscience.iop.org/0305-4470/19/14/019)

J. Phys. A: Math. Gen. 19 (1986) 2821-2839. Printed in Great Britain

# $q$ colourings of the triangular lattice

R J Baxter

Research School of Physical Sciences, The Australian National University, Canberra 2601, Australia

Received 9 December 1985

**Abstract.** Nienhuis has shown that the critical O($n$) model on the honeycomb lattice is equivalent to a zero-temperature antiferromagnetic Potts model on the triangular lattice, i.e. to the chromatic polynomial of the triangular lattice. Here the critical O($n$) model is solved by the Bethe ansatz method, thereby giving the large-lattice limit of the chromatic polynomial.

## 1. Introduction

Consider a honeycomb lattice $\mathscr{L}_H$ with $2N$ sites, or vertices. Place arrows on some (or none) of the edges so that at each vertex there are either no arrows, or just one arrow in and just one arrow out. The arrows therefore form oriented loops on $\mathscr{L}_H$, so let us call this the 'loop model'. For a given configuration $C$, let $L$ be the number of arrows, $l$ the number of vertices at which there are two arrows forming a left turn and $r$ the number where the arrows form a right turn. Let $t$ and $\alpha$ be given parameters and define the partition sum to be

$$Z_{\text{Loop}} = \sum_C t^{2N-L} \exp[\mathrm{i}\alpha(l-r)]. \tag{1.1}$$

Also, consider the Potts model on a triangular lattice $\mathscr{L}_T$ of $N$ sites. At each site $i$ there is a spin $\sigma_i$, with values $1, 2, \ldots, q$. Neighbouring spins interact, having energy $-kTJ$ if they are equal, 0 if different. The partition function is

$$Z_{\text{Potts}} = \sum_\sigma \exp\left(J \sum_{\langle ij \rangle} \delta(\sigma_i, \sigma_j)\right) \tag{1.2}$$

where the inner sum is over all $3N$ nearest-neighbour pairs $\langle i, j \rangle$ of sites and the outer sum is over all $q^N$ values of all the spins.

Nienhuis (1982, 1984) shows that the partition sum of the O($n$) model can be written as that of a loop model (his $x$ is here replaced by $t^{-1}$). Further, he shows that both the loop and Potts models are equivalent to six-vertex models on a Kagomé lattice of $3N$ sites. They are equivalent to the same six-vertex model (and hence to one another) when $t$, $\alpha$, $q$, $J$ satisfy the relations

$$2\cos 6\alpha = 2 - (2 - t^2)^2 \tag{1.3a}$$

$$q = 4 - t^2 = 2 - 2\sin 3\alpha \tag{1.3b}$$

$$\mathrm{e}^J = 0. \tag{1.3c}$$

0305-4470/86/142821+19$02.50  

In fact, when these are satisfied we have precisely (with appropriate boundary conditions)

$$Z_{\text{Loop}} = Z_{\text{Potts}}. \tag{1.4}$$

Nienhuis argues that these equations determine the critical point of the loop and O($n$) models (for $\alpha$ real and $0 \leqslant q \leqslant 4$). By taking $\sin 3\alpha = 2^{-1/2}$, $\cos 6\alpha = 0$, one can deduce that the connective constant for self-avoiding walks on the honeycomb lattice is $t = (2+\sqrt{2})^{1/2}$ (Guttmann 1984).

In early 1983, while we were both at a conference in Adelaide, A J Guttmann introduced me to Nienhuis' loop model and I was able to convince myself that it could be solved by a Bethe ansatz, provided the condition (1.3$a$) was satisfied. When I looked at Nienhuis' paper myself and saw the mapping to a Potts model, I assumed this had to be the ferromagnetic critical case, when $(e^J - 1)^2(e^J + 2) = q$. Since this had previously been solved (Baxter *et al* 1978), there seemed little point in doing so again by another route.

However, very recently A D Sokal made me look more carefully at Nienhuis' paper, making me realise that the associated triangular Potts model is not the usual ferromagnetic critical case, but satisfies (1.3$c$). This is the zero-temperature limit of the antiferromagnetic case, when two adjacent spins cannot be equal (i.e. this configuration has zero weight). This is a very interesting case, for we can now interpret the spins as 'colours': from (1.2), $Z_{\text{Potts}}$ is then the number of ways of colouring the sites of $\mathscr{L}_{\text{T}}$ with $q$ colours so that no two adjacent sites have the same colour. This is the chromatic polynomial of $\mathscr{L}_{\text{T}}$; chromatic polynomials of large planar graphs are extensively studied in graph theory and there are some interesting conjectures concerning the location of their zeros (Beraha *et al* 1980). Until now the chromatic polynomials of the large regular lattices have defied calculation. (For general values of $J$ the Potts model is equivalent to the dichromatic polynomial (see Kasteleyn and Fortuin 1969).)

Naturally I looked again at the loop model. It can indeed be solved by the Bethe ansatz when (1.3$a$) is satisfied, and the working is outlined in the following sections. Consequently one can evaluate the $N$th root of the chromatic polynomial:

$$W = Z_{\text{Loop}}^{1/N} = Z_{\text{Potts}}^{1/N} \tag{1.5}$$

in the large-lattice limit. For $q$ real, $q > 4$, one can define a parameter $x$ by

$$q = 2 - x - x^{-1} \qquad -1 < x < 0. \tag{1.6}$$

Then the result of the calculation is

$$W = -\frac{1}{x}\prod_{j=1}^{\infty}\frac{(1-x^{6j-3})(1-x^{6j-2})^2(1-x^{6j-1})}{(1-x^{6j-5})(1-x^{6j-4})(1-x^{6j})(1-x^{6j+1})}. \tag{1.7}$$

For $3 \leqslant q < 4$, we define $\theta$ by

$$q = 2 + 2\cos\theta \qquad 0 < \theta \leqslant \pi/3 \tag{1.8}$$

and find that

$$\ln W = \int_{-\infty}^{\infty} dk \frac{\sinh k\theta}{2k}\left(\frac{\sinh k(\pi/2-\theta)}{[\sinh(\pi k/2)](2\cosh k\theta - 1)} - \frac{\cosh k(\pi/2-\theta)}{[\cosh(\pi k/2)](2\cosh k\theta + 1)}\right). \tag{1.9}$$

There are some checks on the calculation. For $q$ large, Kim and Enting (1979) have obtained the first 13 terms of the expansion of $W$ in powers of $1/(q-1)$; they agree with the result (1.7). When $q = 3$ there are just 3! ways of colouring $\mathscr{L}_{\text{T}}$ (one

colour for each of the three sublattices); in this $\theta = \pi/3$ and the integrand of (1.9) vanishes, correctly giving $W = 1$. The case $q = 4$ can be obtained by either taking the $x \to -1$ limit in (1.7) or the $\theta \to 0$ limit in (1.9). Either way one obtains

$$W = \prod_{j=1}^{\infty} \frac{(3j-1)^2}{(3j-2)3j} = 3\gamma^3/4\pi^2 = 1.460\,99\ldots \tag{1.10}$$

where $\gamma = \Gamma(1/3) = 2.678\,9385\ldots$. This agrees with the known value (Baxter 1970).

If we require $q$ to be a positive integer, then the above results are completed by noting that $W = 0$ for $q = 1$ and 2. However, since $Z_{\text{Potts}}$ is just a polynomial in $q$, we can regard $q$ as a real or complex variable.

The transfer matrix of the loop model is defined in § 2. In § 3 it is shown how its eigenvalues can be obtained by adapting the Bethe ansatz method employed by Lieb for planar models (Lieb 1967, Lieb and Wu 1972). This gives each eigenvalue in terms of $n$ complex numbers $z_1, \ldots, z_n$, which are in turn defined by $n$ equations. There are many solutions, corresponding to the different eigenvalues, and one has to find the solution corresponding to the eigenvalue of largest modulus. This is done in § 4 by using continuity from the easily handled $q = \infty$ case. We assume that the eigenvalue thus selected remains the largest contributing eigenvalue. From the independent checks mentioned above, it seems that this assumption is justified, at least for $q$ real and $q \geqslant 3$.

For other values of $q$, in particular for $0 < q < 3$, one has to be very careful. There are two problems: one is that it is possible that the previously largest eigenvalue may be surpassed by another. In fact this is necessary if the partition function is to have zeros, and we know from the graph theorists (Tutte 1973, 1982, Beraha *et al* 1978, Beraha and Kahane 1979) that the chromatic polynomial does have zeros on and close to the segment (0, 4) of the real axis.

A more fundamental difficulty is that the equivalence (1.4) should be treated with caution. For instance, in (2.1) it is shown that $Z_{\text{loop}}$ can be written as a sum weighted by $\cos 6\alpha$ to the power of the number of loops on $\mathscr{L}_{\text{H}}$. If we take $2\cos 6\alpha = 1$ the model is equivalent to the triangular Ising model (put Ising spins on the faces of $\mathscr{L}_{\text{H}}$ so that adjacent spins are different if a loop lies on the intervening edge, the same if no loop lies on it), with $\exp(2J/kT) = t$. From (1.3*a*), $t^2 = 1$ or 3, from which it follows that the Ising model is either non-interacting or critical. Either way one obtains $Z_{\text{Loop}} \geqslant 2^N$. On the other hand, from (1.3*b*) $q$ is either 3 or 1, so $Z_{\text{Potts}}$ is 6 or 0.

We can resolve this apparent contradiction by repeating Nienhuis' argument for a finite lattice. We start by writing the triangular Potts model as a Kagomé lattice six-vertex model, as has been done by Baxter *et al* (1976). We interpret this as an SOS model by placing spins on faces of the Kagomé lattice, adjacent spins differing by $\frac{1}{2}$, the greater being to the left of the intervening arrow. Choosing the hexagonal (triangular) faces to have integer (half-integer) spins and taking $J = -\infty$, we can sum over the half-integer spins to obtain Nienhuis' SOS form of the loop model, but with special boundary weights. For $0 < q < 4$ these boundary weights are complex, so it is possible for them to modify the bulk behaviour. Obviously this is happening in the $q = 1$ and 3 cases mentioned above.

In terms of the loop model transfer matrices, these boundary conditions may mean that not all eigenvalues contribute to the partition function. (If an eigenvector is 'orthogonal to the boundary vector', then it never enters the calculation of $Z_{\text{Loop}}$.) This must be happening in the $q = 1$ and 3 cases, so is presumably a general phenomenon. (Similar problems with using the six-vertex form of the Potts model have previous been observed (see Baxter 1982b).)

Thus not only is care necessary to select the largest eigenvalue of the loop model transfer matrix but one should also verify that it is a contributing eigenvalue. Even so, it should be possible to determine $W$ throughout the complex $q$ plane. As is remarked in § 5, this would provide interesting information on the limiting distribution of the zeros of the chromatic polynomial. I hope to do this.

For sufficiently large negative values of $q$, the result (1.7) must be valid (with $0<x<1$). It may well be valid for $-\infty<q<0$, with limiting value

$$|W|=\prod_{j=1}^{\infty}\frac{(6j-3)(6j-2)^2(6j-1)}{(6j-5)(6j-4)(6j)(6j+1)}=3^{3/2}\gamma^9/(2\pi)^5=3.770\,92\ldots \quad (1.11)$$

for $q=0$. The contribution of the corresponding eigenvalue through $q=0$ is given in § 5. If it is still the largest contributing eigenvalue, then ln $W$ is equal to the RHS of (5.3). This may be so for $0<q<1$, but as yet we have no independent checks, so this result is more speculative than (1.7) and (1.9).

The Bethe ansatz calculation is similar to that for the regular zero-field six-vertex model, and hence to that for the ferromagnetic critical Potts model (though not sufficiently so as to suggest a simple mapping from one to the other). In both cases one uses Fourier series for $q>4$ and Fourier integrals for $0<q<4$. This in itself supports Nienhuis' assumption that the zero-temperature antiferromagnetic triangular Potts model is critical for $0<q<4$. The $q=2$ Ising case is known to be critical (Wannier 1950, Stephenson 1964, Nienhuis *et al* 1984, Domany and Schaub 1984).

## 2. Transfer matrix eigenvalue equations

For a given loop on $\mathscr{L}_H$, the arrows either make six more turns to the left than the right, or vice versa. Thus (1.1) is equivalent to

$$Z_{\text{Loop}}=\sum t^{2N-L}(2\cos 6\alpha)^P \quad (2.1)$$

where the sum is now over all ways of putting non-intersecting loops on $\mathscr{L}_H$, $P$ being the number of loops and $L$ the number of edges covered by a loop. This is Nienhuis' original form for the loop model partition function, but we shall use the form (1.1), which has the advantage that the summand is a product of local vertex weights.

We can construct the summand in (1.1) by assigning to each vertex of $\mathscr{L}_H$ a weight $t$ if no arrows are incident to it, a weight $e^{i\alpha}$ if thre are two arrows forming a left turn and a weight $e^{-i\alpha}$ if they form a right turn. Draw $\mathscr{L}_H$ as in figure 1. It is bi-partite, with two types of site A and B, as indicated. The seven possible configurations at each type of site are shown in figure 2. Let $\omega_1,\ldots,\omega_7'$ be the corresponding weights.

There are three types of edge of $\mathscr{L}_H$, as indicated in figure 1. Consider the b-type edges, with an A site on the left and a B site on the right. The partition function is unchanged if for every right-pointing arrow on such an edge we associate a weight $b$ with its tip, $b^{-1}$ with its tail, and incorporate these into $\omega_1,\ldots,\omega_7'$. Similarly for left-pointing arrows, but with $b$ replaced by $b'$; and for the other two types of edge with tip weights $a$, $a'$, $c$, $c'$. Setting $\exp(i\alpha)=e$, the vertex weights are then

$$\begin{aligned}\omega_1,\ldots,\omega_7&=t,\frac{c'e}{a},\frac{a'e}{b},\frac{b'e}{c},\frac{a'}{ce},\frac{b'}{ae},\frac{c'}{be}\\ \omega_1',\ldots,\omega_7'&=t,\frac{a}{c'e},\frac{b}{a'e},\frac{c}{b'e},\frac{ce}{a'},\frac{ae}{b'},\frac{be}{c'}.\end{aligned} \quad (2.2)$$

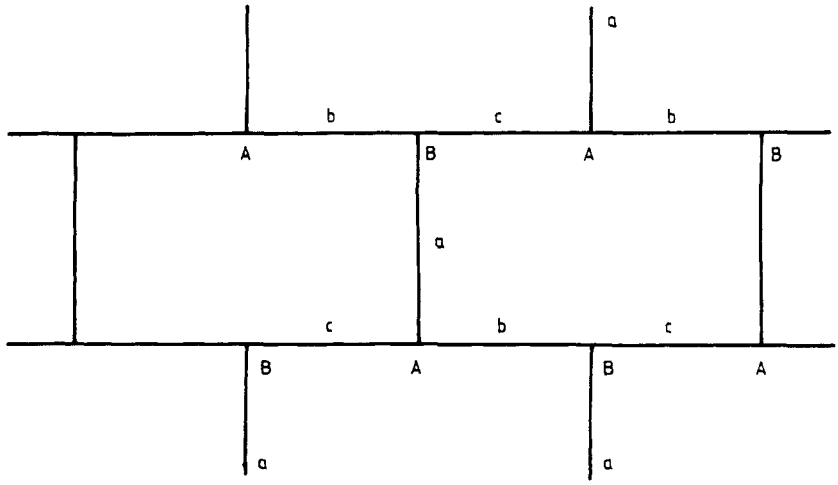


**Figure 1.** The honeycomb lattice $\mathscr{L}_H$, showing its two types of sites (A and B), and three types of edges (a, b, c).

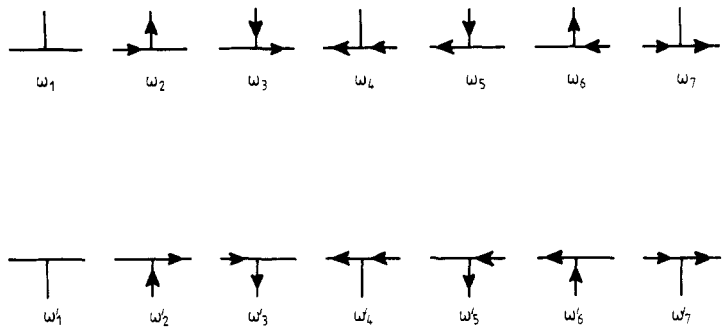


**Figure 2.** The seven allowed arrow and non-arrow configurations at each type of site.

Choosing $a$, $b$, $c$, $a'$, $b'$, $c' = e^3$, $e$, $e^5$, 1, $e^4$, $e^2$, the expressions simplify to

$$\begin{aligned} &\omega_1, \ldots, \omega_7 = t, 1, 1, 1, s^{-1}, 1, 1 \\ &\omega'_1, \ldots, \omega'_7 = t, 1, 1, 1, s, 1, 1 \end{aligned} \tag{2.3}$$

where

$$s = e^6 = \exp(6ia). \tag{2.4}$$

Now instead of drawing arrows on edges, represent the edges by dotted, full, double-full or wiggly lines, according to figure 3. For instance, if a b-type edge contains no arrow, draw it as a dotted line. If it contains a left-pointing (right-pointing) arrow, draw it as a full (wiggly) line. Then in this picture, figure 2 is replaced by figure 4. We shall regard an edge as empty (occupied, doubly occupied) if it contains a dotted (full or wiggly, double-full) line.

Notice that the full/wiggly lines do not terminate. If one follows a full line from below into a B site, it either becomes a wiggly line (in which case it moves left to the next A site and then goes upwards again as a full line), or it goes to the right (in which case it stays as a full line and can move some distance to the right before it exists

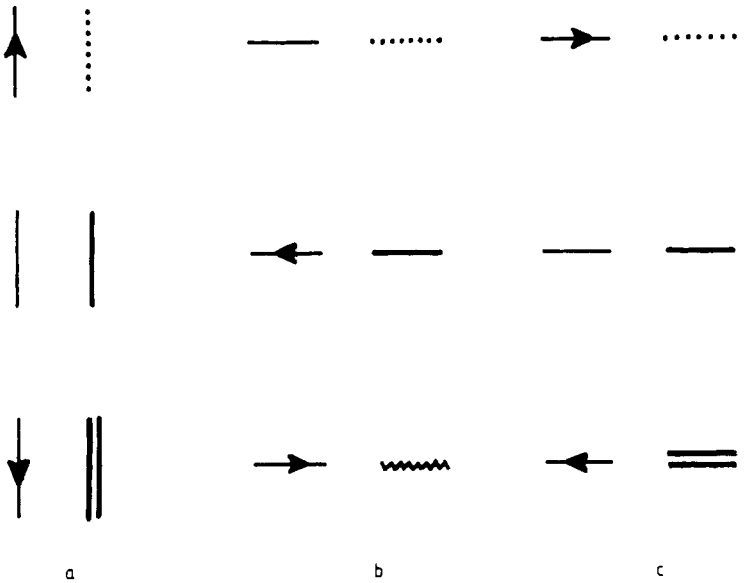


**Figure 3.** Conversion (on edge types a, b, c) from arrows and non-arrows to dotted, full, and double-full or wiggly lines.

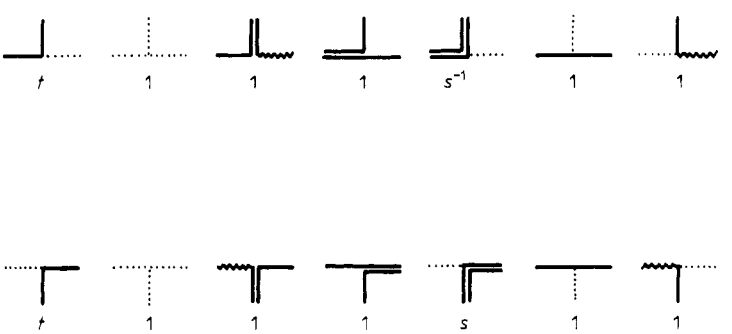


**Figure 4.** Translation of figure 2 into the line representation, showing the weights (2.3).

upwards at an A site). In either case it is conserved: if there are $n$ full lines in one row of vertical bonds, there are $n$ in every row.

It is permissible for two full lines to occupy the same a-type or c-type (but not b-type) edge. A typical arrangement of lines (with $n = 4$) between two successive rows is shown in figure 5.

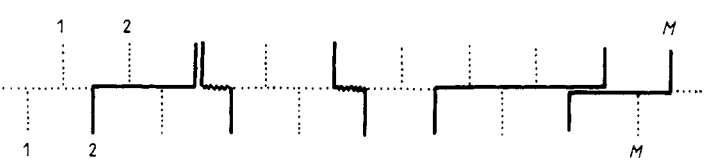


**Figure 5.** A typical line arrangement on successive rows of $\mathscr{L}_H$. The numbering of the vertical edges is indicated.

Consider two successive rows of vertical edges, such as in figure 5. Let $\phi$ denote the configuration of vertical lines in the lower row, $\phi'$ that in the upper row. Then the row-to-row transfer matrix $V$ has entries

$$V(\phi, \phi') = \sum \prod (\text{vertex weights}) \tag{2.5}$$

where the sum is over all allowed arrangements of full and wiggly lines on the intervening horizontal edges, and for each such arrangement the product is over the weights of the vertices in the intervening horizontal row. (There are at most two allowed arrangements of the intervening lines.) If $\mathscr{L}_H$ has $L$ rows and $M$ columns, and we use cyclic (toroidal) boundary conditions, then in the usual way (ch 2 and 7 of Baxter 1982a)

$$Z_{\text{Loop}} = \text{Tr } V^L. \tag{2.6}$$

Let $\Lambda$ be an eigenvalue of $V$ and $g$ the corresponding eigenvector. Then

$$\Lambda g = Vg. \tag{2.7}$$

If $\Lambda_{\max}$ is the eigenvalue of largest modulus, then from (2.6), it follows that for large $L$

$$Z_{\text{Loop}} \sim \Lambda_{\max}^L. \tag{2.8}$$

## 3. Bethe ansatz

Because the number of vertical lines $n$ is the same for each row, the transfer matrix breaks up into diagonal blocks, in which $\phi$ and $\phi'$ have the same value of $n$. This means we can try to use the Bethe ansatz to solve the eigenvalue equation (2.7), rather as one does for the six-vertex model (Lieb and Wu 1972, Baxter 1982a).

First consider the case $n = 0$, when all vertical edges are empty (dotted). Then on a given horizontal row, either all horizontal edges are empty or they are full. From figure 4 (columns 2 and 6) all the contributing vertex weights are unity, so

$$\Lambda = V(\phi_0, \phi_0) = 1 + 1 = 2. \tag{3.1}$$

### *3.1. $n = 1$*

When $n = 1$ there is just one full line per row of vertical edges. Let $x$ be the position of the line in one row and $y$ the position in the row above, labelling the vertical edges from 1 to $M$ as in figure 5. Then the four possible arrangements of $x$ and $y$ are shown in figure 6. Using figure 4, we find that the product of the weights of the $2M$ vertices in the row is 1, $t^2$, $t^2$, 1 for the four cases, respectively. If $g(x)$ is the entry of the

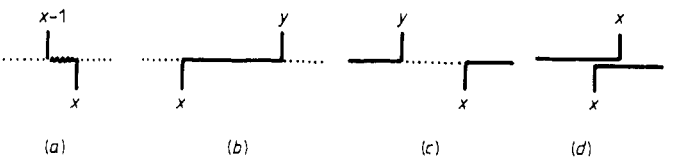


**Figure 6.** The four cases that arise when $n = 1$.

eigenvector $g$ corresponding to the state with a line in position $x$, then from (2.5) the eigenvalue equation (2.7) is

$$\Lambda g(x) = g(x-1) + t^2 \sum_{y=x}^{M} g(y) + t^2 \sum_{y=1}^{x-1} g(y) + g(x) \tag{3.2}$$

for $1 \leqslant x \leqslant M$.

The four terms on the RHS of (3.2) correspond respectively to the four diagrams in figure 6. We see that ($a$) and ($b$) are similar in that they have no full horizontal at the boundary; we call the corresponding first two terms on the RHS of (3.2) the 'left' terms. Similarly the last two terms (corresponding to diagrams ($c$) and ($d$), with a full line at the boundary), we call 'right' terms.

We now try as a solution

$$g(x) = z^x \tag{3.3}$$

and define

$$\begin{aligned} \lambda(z) &= z^{-1} + t^2/(1-z) \\ \mu(z) &= 1 - t^2/(1-z). \end{aligned} \tag{3.4}$$

Then the left terms in (3.2) give

$$\lambda(z) z^x - t^2 z^{M+1}/(1-z) \tag{3.5a}$$

while the right terms give

$$\mu(z) z^x + t^2 z/(1-z). \tag{3.5b}$$

The expressions proportional to $z^x$ are 'wanted', in that they are of the same type as the LHS of (3.2); their contribution gives

$$\Lambda = \lambda(z) + \mu(z). \tag{3.6}$$

The other expressions in (3.5) (the ones independent of $x$) are 'unwanted boundary terms'. They cancel if

$$z^M = 1. \tag{3.7}$$

This equation has $M$ solutions for $z$. (3.6) then gives the $M$ eigenvalues of $V$ in the $n = 1$ block.

### *3.2. $n = 2$*

When $n = 2$, let $x_1, x_2$ be the positions of the two full lines in one row and $y_1, y_2$ the positions in the row above. Then the general situation is that $y_1, y_2$ interlace $x_1, x_2$. More precisely, if we choose $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$, then for the 'left' terms $x_1 - 1 \geqslant y_1 \leqslant x_2$, $x_2 - 1 \leqslant y_2 \leqslant M$. For the 'right' terms $1 \leqslant y_1 \leqslant x_1$, $x_1 - 1 \leqslant y_2 \leqslant x_2$.

As is usual in the Bethe ansatz (Lieb 1967, Lieb and Wu 1972, Baxter 1982a, pp 131–8) it is easiest to regard the $y_1$ and $y_2$ summations as independent and then introduce correction terms to compensate for errors when $y_1 \geqslant y_2$ in the summations.

Let $g(x_1, x_2)$ be the entry of $g$ corresponding to lines in positions $x_1, x_2$, provided $x_1 < x_2$. Let $h(x)$ be the entry when the lines are both in position $x$ (i.e. there is a

double-full line on edge $x$). Define, for $x-1 \leqslant y \leqslant x'$,

$$\begin{aligned} D(x, y, x') &= t^2 \qquad \text{if } x \leqslant y < x' \\ &= 1 \qquad \text{if } y = x-1 \text{ or } y = x'. \end{aligned} \tag{3.8}$$

Then the eigenvalue equation (2.7) is, for $1 \leqslant x_1 < x_2 \leqslant M$,

$$\begin{aligned} \Lambda g(x_1, x_2) = &\sum_{y_1=x_1-1}^{x_2} D(x_1, y_1, x_2) \sum_{y_2=x_2-1}^{M} D(x_2, y_2, M+1) g(y_1, y_2) + C(x_2) \\ &+ \sum_{y_1=1}^{x_1} D(0, y_1, x_1) \sum_{y_2=x_1-1}^{x_2} D(x_1, y_2, x_2) g(y_1, y_2) + C(x_1) \end{aligned} \tag{3.9}$$

and, for $1 \leqslant x \leqslant M$,

$$\begin{aligned} \Lambda h(x) = &t \sum_{y=x}^{M} g(x-1, y) + st \sum_{y=x+1}^{M} g(x, y) + h(x) \\ &+ t \sum_{y=1}^{x-2} g(y, x-1) + st \sum_{y=1}^{x-1} g(y, x) + h(x-1). \end{aligned} \tag{3.10}$$

As for $n = 1$, the terms on the RHS of (3.9) and (3.10) can be grouped into 'left' and 'right' terms, depending on whether the horizontal boundary edge is empty or full. The terms up to and including $+C(x_2)$ in (3.9), and $+h(x)$ in (3.10), are 'left' terms. The remainder are 'right' terms.

The function $C(x)$ incorporates the corrections to the double sums in (3.9) and is given by

$$C(x) = -g(x, x-1) - t^2 g(x-1, x-1) - t^2 g(x, x) + th(x-1) + ts^{-1} h(x). \tag{3.11}$$

Note that the double sums necessitate extending the function $g(x_1, x_2)$ to the argument values $x_2 = x_1$ and $x_2 = x_1 - 1$. However, these are then subtracted off again via $C(x)$. It follows that we can give $g(x, x)$ and $g(x, x-1)$ any values we like, so long as we use the same ones in both (3.9) and (3.11). We need not (and shall not) require $g(x, x) = h(x)$.

First we try, for $x_2 \geqslant x_1 - 1$,

$$g(x_1, x_2) = z_1^{x_1} z_2^{x_2}. \tag{3.12}$$

Then the first double sum in (3.9) becomes

$$[\lambda(z_1) z_1^{x_1} + \mu(z_1) z_1^{x_2}][\lambda(z_2) z_2^{x_2} - t^2 z_2^{M+1}/(1-z_2)]. \tag{3.13}$$

Expanding this gives one term proportional to $z_1^{x_1} z_2^{x_2}$. The second double sum gives a similar contribution. Altogether these 'wanted' terms cancel out of (3.9) if

$$\Lambda = \lambda(z_1)\lambda(z_2) + \mu(z_1)\mu(z_2). \tag{3.14}$$

This is unaltered by interchanging $z_1$ with $z_2$, so we can more generally try

$$\begin{aligned} g(x_1, x_2) &= A_{12} z_1^{x_1} z_2^{x_2} + A_{21} z_2^{x_1} z_1^{x_2} \\ h(x) &= B(z_1 z_2)^x \end{aligned} \tag{3.15}$$

the parameters $A_{12}$, $A_{21}$, $B$, $z_1$, $z_2$ being at our disposal.

Substituting (3.15) into (3.9), all terms proportional to $z_1^{x_1}z_2^{x_2}$ or $z_2^{x_1}z_1^{x_2}$ cancel because of (3.14). The 'left' terms include expressions proportional to $(z_1z_2)^{x_2}$, while the 'right' terms include ones proportional to $(z_1z_2)^{x_1}$. Both these 'unwanted internal' terms cancel if

$$a_{12}A_{12}+a_{21}A_{21}+bB=0 \tag{3.16}$$

where

$$\begin{aligned} a_{ij} &= \mu(z_i)\lambda(z_j)-\frac{1}{z_j}-\frac{t^2}{z_1z_2}-t^2 \\ b &= \frac{t(s+z_1z_2)}{sz_1z_2}. \end{aligned} \tag{3.17}$$

Other terms of this type, proportional to $(z_1z_2)^x$, occur in (3.10). We regard $\Lambda$ as given by (3.14), and then all terms in (3.10) can be naturally classified as 'left' or 'right'. To be able to continue to apply this method to large values of $n$, we need the internal unwanted terms to cancel separately in both the left and right groupings. Thus we require

$$\begin{aligned} a'_{12}A_{12}+a'_{21}A_{21}+b'B&=0 \\ a''_{12}A_{12}+a''_{21}A_{21}+b''B&=0 \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} a'_{ij} &= \frac{t(1+sz_1z_2)}{z_i(1-z_j)} \\ a''_{ij} &= \frac{-t(1+sz_1z_2)}{z_1z_2(1-z_i)} \\ b' &= 1-\lambda(z_1)\lambda(z_2) \\ b'' &= \frac{1}{z_1z_2}-\mu(z_1)\mu(z_2). \end{aligned} \tag{3.19}$$

All other terms in (3.9) and (3.10) arise from the $y_2$, $y=M$ or $y_1$, $y=1$ limits of the summations. They are 'boundary' terms. The ones in (3.9) cancel if

$$A_{21}=z_2^MA_{12} \qquad A_{12}=z_1^MA_{21}. \tag{3.20}$$

Those in (3.10) cancel if $z_1^Mz_2^M=1$, which can be seen to be a corollary of (3.20).

We therefore want to satisfy (3.16), (3.18) and (3.20). The first two form three homogeneous linear equations for $A_{12}$, $A_{21}$, $B$, so will have non-zero solutions only if the determinant

$$\Delta=\begin{vmatrix} a_{12} & a_{21} & b \\ a'_{12} & a'_{21} & b' \\ a''_{12} & a''_{21} & b'' \end{vmatrix} \tag{3.21}$$

vanishes. After some tedious algebra, we find that

$$\Delta=\frac{t^2(z_1-z_2)(1-z_1z_2)(s+s^{-1}+2-4t^2+t^4)}{z_1^2z_2^2(1-z_1)^2(1-z_2)^2}. \tag{3.22}$$

Thus $\Delta$ does indeed vanish (for all $z_1, z_2$) provided the parameters $s$ and $t$ satisfy the relation

$$s+s^{-1}=4t^2-2-t^4. \tag{3.23}$$

From (2.4), $s=\exp(6i\alpha)$, so (3.23) is the condition (1.3$a$). Because our $t$ is the inverse of Niehuis' $x$, it is also Nienhuis' criticality condition, equation (20) of Nienhuis (1982).

We can now solve (3.16) and (3.18) for $A_{12}: A_{21}: B$. Define

$$S(z, w)=(1-z-w+zw+t^2z)(1-2w+zw+t^2w). \tag{3.24}$$

then by eliminating $B$ we obtain

$$S(z_1, z_2)A_{12}+S(z_2, z_1)A_{21}=0. \tag{3.25}$$

If we define $A_{12}=S(z_2, z_1)$, then $A_{12}$, $A_{21}$, $B$ are given by

$$A_{12}=S(z_2, z_1) \qquad A_{21}=-S(z_1, z_2) \qquad B=t(z_2-z_1)(1+sz_1z_2). \tag{3.26}$$

Substituting these results into the boundary conditions (3.20), we obtain the two equations

$$z_1^M=-\frac{S(z_2, z_1)}{S(z_1, z_2)} \qquad z_2^M=-\frac{S(z_1, z_2)}{S(z_2, z_1)}. \tag{3.27}$$

We can in principle solve these for $z_1$ and $z_2$. (There will be many solutions, corresponding to the different eigenvalues of $V$.) Then $\Lambda$ is given by (3.14).

### *3.3. General values of n*

Provided the condition (3.23) is satisfied, the above working can be generalised to arbitrarily large values of $n$. The notation can become very complicated, but basically the calculation proceeds as for the square lattice six-vertex models (Lieb and Wu 1972, Baxter 1982a).

For $1 \leqslant x_1 < x_2 < \ldots < x_n \leqslant M$ the elements of the eigenvector are

$$g(x_1, \ldots, x_n)=\sum_P A_{P1,\ldots,Pn} z_{P1}^{x_1} \cdots z_{Pn}^{x_n} \tag{3.28}$$

where the sum is over all permutations $P=\{P1, \ldots, Pn\}$ of the integers $1, \ldots, n$. If $\varepsilon_P$ is the sign (+1 or −1) of the permutation, then

$$A_{P1,\ldots,Pn}=\varepsilon_P \prod_{1 \leqslant i<j \leqslant n} S(z_{Pj}, z_{Pi}) \tag{3.29}$$

where $S(z, w)$ is defined by (3.24).

If two adjacent $x$ are equal, say if $x_3=x_4$, then the terms in (3.28) can be grouped in pairs, differing only in the interchange of $P3$ with $P4$. The sum of each pair contains a factor $S(z_{P4}, z_{P3})-S(z_{P3}, z_{P4})$. This should be replaced by $t(z_{P4}-z_{P3})(1+sz_{P3}z_{P4})$ to obtain the correct eigenvector element. Since no more than two $x$ can equal one another, this covers all cases.

The parameters $z_1, \ldots, z_n$ are given by

$$z_j^M=(-1)^{M-1} \prod_{l=1}^{n} \frac{S(z_l, z_j)}{S(z_j, z_l)} \tag{3.30}$$

for $j = 1, \ldots, n$. Then the eigenvalue $\Lambda$ is given by

$$\Lambda = \lambda(z_1) \cdots \lambda(z_n) + \mu(z_1) \cdots \mu(z_n) \tag{3.31}$$

$\lambda(z)$ and $\mu(z)$ being defined in (3.4).

Ultimately we want to calculate the 'partition function (or chromatic polynomial) per site' $W$, defined by (1.5). Since $\mathscr{L}_H$ has twice as many sites as vertical edges,

$$2N = 2LM \tag{3.32}$$

so, using (2.8),

$$\ln W = M^{-1} \ln \Lambda_{\max}. \tag{3.33}$$

## 4. Maximum eigenvalue

We want to solve the equations (3.30), $S(z, w)$ being defined by (3.24), $M$ the number of vertical bonds per row of $\mathscr{L}_H$, and $n$ the number of full lines per row. In particular, we want to obtain the solution corresponding to the largest eigenvalue $\Lambda$ of the transfer matrix $V$. We require that $z_1, \ldots, z_n$ be distinct, as otherwise all elements of the eigenvector $g$ are given by (3.28) to be zero.

In the limit when $|t|$ becomes large, $|z_1|, \ldots, |z_n|$ remaining finite, (3.30) simplifies to

$$z_j^M = (-1)^{M-1} \qquad j = 1, \ldots, n. \tag{4.1}$$

From (3.4) and (3.31), the solution of this type that maximises $|\Lambda|$ is the one with $n = M$, $z_1, \ldots, z_n$ being uniformly distributed around the unit circle. Further, this gives (for $M$ even) $\Lambda = t^{2M}$, which is indeed the largest eigenvalue of $V$ in the large $t$ limit, corresponding to the limiting behaviour $Z_{\text{Potts}} \sim q^N$ of the colouring polynomial for large $q$.

In all previous Bethe ansatz calculations there has been a 'transformation to a difference kernel' (Lieb and Wu 1972) in which the ratio $S(z, w)/S(w, z)$ is transformed to a function only of the difference of its two arguments. This case appears to be more complicated, as $S(z, w)$ is the product of two bilinear factors (rather than one, as in the six-vertex model (Baxter 1982a)). Even so, such a transformation does exist. Define

$$r = \mathrm{i}\, \mathrm{e}^{3\mathrm{i}\alpha} \tag{4.2}$$

then from (2.4), $s = -r^2$, so from (3.23) we can choose

$$t^2 = 2 - r - r^{-1}. \tag{4.3}$$

This is consistent with (1.3$b$), from which we see that the argument $q$ of the chromatic polynomial is related to $r$ by

$$q = 2 + r + r^{-1}. \tag{4.4}$$

Now define $w_1, \ldots, w_n$ by

$$z_j = (1 + rw_j)/(r + w_j) \qquad j = 1, \ldots, n. \tag{4.5}$$

Then from (3.24)

$$\frac{S(z_l, z_j)}{S(z_j, z_l)} = \frac{(w_j + rw_l)(w_l - r^2 w_j)}{(w_l + rw_j)(w_j - r^2 w_l)}. \tag{4.6}$$

This depends on $w_l$ and $w_j$ only via their ratio: by taking logarithms we can readily convert this to a difference.

From (3.4),

$$\lambda(z_j) = \frac{(1+rw_j)(r^{-1}-rw_j)}{(r+w_j)(1-w_j)}$$
$$\mu(z_j) = \frac{r-r^{-1}w_j}{1-w_j}. \tag{4.7}$$

For $q$ real and positive, we see from (4.4) that $r$ is real if $q>4$, unimodular if $q<4$. We must distinguish these cases.

*4.1. $q>4$*

In this case we can choose $0<r<1$ and define $\eta$ so that

$$r = e^{-\eta} \qquad \eta > 0. \tag{4.8}$$

Guided by the large-$t$ limit, we expect $z_1, \ldots, z_n$, and hence $w_1, \ldots, w_n$, to be distributed round the unit circle. Set

$$w_j = e^{-iu_j} \qquad -\pi < u_j \leqslant \pi. \tag{4.9}$$

Using (4.5) and (4.6), we can write $z_j$ and $S(z_l, z_j)/S(z_j, z_l)$ as functions of $u_j$ and $u_j - u_l$, respectively. Let us write them as

$$z_j = \exp[ik(u_j)]$$
$$S(z_l, z_j)/S(z_j, z_l) = = \exp[i\Theta(u_j - u_l)]. \tag{4.10}$$

Then we readily find that

$$k'(u) = \sinh\eta/(\cosh\eta + \cos u) = 1 + 2\sum_{m=1}^{\infty}(-r)^m \cos mu \tag{4.11}$$

$$\Theta'(u) = 2\sum_{m=1}^{\infty}[r^{2m} - (-r)^m]\cos mu. \tag{4.12}$$

Note that $k(u)$ is monotonic increasing: as $u$ increases from $-\pi$ to $\pi$, so does $k(u)$.

Taking logarithms of both sides of (3.30), using (4.10) and dividing by $iM$, we obtain for $j = 1, \ldots, n$

$$k(u_j) = (2I_j + 1 - M)\pi/M + M^{-1}\sum_{l=1}^{n}\Theta(u_j - u_l) \tag{4.13}$$

where $I_j$ is an integer. For $q$ and $t$ large we know that the largest eigenvalue is given by choosing

$$n = M \qquad I_j = j - 1. \tag{4.14}$$

There is no evidence to suppose that the chromatic polynomial is not analytic for $q>4$, so let us keep these choices of $n$ and $I_1, \ldots, I_n$.

In the limit of $M, n$ large we expect $u_1, \ldots, u_n$ to be in increasing order and to form a dense distribution in the interval $(-\pi, \pi)$. Let $M\rho(u)\,du$ be the number of $u_j$ between $u$ and $u+du$. Since there are $M$ $u_j$ in the $(-\pi, \pi)$, $\rho(u)$ must satisfy the relation

$$\int_{-\pi}^{\pi} \rho(u)\,du = 1. \tag{4.15}$$

Equation (4.13) now becomes

$$k(u) = -\pi + 2\pi \int_{-\pi}^{u} \rho(v)\,dv + \int_{-\pi}^{\pi} \Theta(u-v)\rho(v)\,dv \tag{4.16}$$

for $-\pi < u < \pi$. This is a linear integral equation for $\rho(u)$. It can readily be solved by differentiating with respect to $u$ and using Fourier series, together with (4.11) and (4.12). We obtain

$$2\pi\rho(u) = 1 + 2\sum_{m=1}^{\infty} \frac{(-r)^m}{1+r^{2m}-(-r)^m} \cos mu \tag{4.17}$$

which does indeed satisfy the condition (4.15).

Using (4.7)-(4.9) and (4.17):

$$\begin{aligned} M^{-1}\sum_{j=1}^{M} \ln \mu(z_j) &= \int_{-\pi}^{\pi} \left[\eta + \ln\left(\frac{1-r^2\,e^{iu}}{1-e^{iu}}\right)\right]\rho(u)\,du \\ &= \eta + \sum_{m=1}^{\infty} \frac{1-r^{2m}}{m} \int_{-\pi}^{\pi} e^{imu}\rho(u)\,du \\ &= \eta + \sum_{m=1}^{\infty} \frac{(1-r^{2m})(-r)^m}{m[1-(-r)^m-r^{2m}]}. \end{aligned} \tag{4.18}$$

(We have taken some liberties in expanding $\ln(1-e^{iu})$, but they are justified by the convergence of the series in (4.17).) One obtains the same answer if one calculates $M^{-1}\Sigma \ln \lambda(z_j)$, so the two products in (3.11) are equal. (This is true for finite even values of $M$, being a consequence of the fact that $z_1, \ldots, z_n$ occur in mutually inverse pairs.) If we now define $x$ so that

$$x = -r = -e^{-\eta} \qquad q = q - x - x^{-1} \qquad -1 < x < 0 \tag{4.19}$$

then from (4.18), (3.31) and (3.33),

$$\ln W = \eta + \sum_{m=1}^{\infty} \frac{x^m - x^{3m}}{m(1-x^m+x^{2m})}. \tag{4.20}$$

Taylor expanding the summand in powers of $x^m$ and summing term by term, we obtain the product form (1.7) for $W$.

### *4.2. $3 < q < 4$*

If $0 < q < 4$, then from (4.4), $r$ is unimodular. We can define $\theta$ so that

$$r = e^{i\theta} \qquad 0 < \theta < \pi \tag{4.21}$$

which is consistent with (1.8).

We still expect $n = M$, and that $z_1, \ldots, z_n$ will be unimodular. From (4.5) it follows that $w_1, \ldots, w_n$ will be real. In the six-vertex model (Lieb and Wu 1972, Baxter 1982a) they are all positive, but this is not true here. We have to be careful to correctly apportion the $w_j$ between the positive and negative parts of the real axis.

To do this, consider what happens when $q$ is just bigger than four in the previous working. Then $r$ is just less than one. In this limit (4.17) becomes

$$\rho(u)=\tfrac{1}{3}\delta(u)+\tfrac{2}{3}\delta(u-\pi) \tag{4.22}$$

where $\delta(u)=(2\pi)^{-1}\sum e^{imu}$ is the usual $2\pi$ periodic delta function. Thus $\frac{1}{3}$ of $u_1,\ldots,u_n$ lie close to 0, the other $\frac{2}{3}$ close to $\pi$. This means that $\frac{1}{3}$ of $z_1,\ldots,z_n$ are close to one (in fact, very close to one, the deviation being proportional to $\eta^2$ rather than $\eta$), while $\frac{2}{3}$ are distributed over the unit circle.

Assuming (as seems reasonable) that $z_1,\ldots,z_n$ do not discontinuously shift at $q=4$, this general picture must still be true for $q$ just less than 4, implying that $\frac{1}{3}$ of $w_1,\ldots,w_n$ are close to 1 and $\frac{2}{3}$ are close to $-1$.

For $0<q<4$, if $w_j$ is positive (negative) then $z_j$ lies on the right (left) segment of the unit circle between $e^{-i\theta}$ and $e^{i\theta}$. Define

$$\begin{aligned} w_j&=-e^{-\alpha_j} \qquad & z_j&=\exp[ik_-(\alpha_j)] \qquad & &\text{for } w_j<0\\ w_j&=e^{-\alpha_j} \qquad & z_j&=\exp[ik_+(\alpha_j)] \qquad & &\text{for } w_j>0. \end{aligned} \tag{4.23}$$

Then we can choose the functions $k_+$, $k_-$ so as to lie in the intervals $(-\theta,\theta)$, $(\theta-2\pi,-\theta)$, respectively. Logarithmically differentiating (4.5) with respect to $\alpha$, we obtain, for $\gamma=\pm$,

$$k'_\gamma(\alpha)=-\frac{\gamma\sin\theta}{\cosh\alpha+\gamma\cos\theta}. \tag{4.24}$$

Similarly, we can define functions $\Theta_{++}$, $\Theta_{+-}$, $\Theta_{-+}$, $\Theta_{--}$ so that (4.6) is equivalent to

$$S(z_l,z_j)/S(z_j,z_l)=\exp[i\Theta_{\gamma\delta}(\alpha_j-\alpha_l)] \tag{4.25}$$

where $\gamma(=\pm)$ is the sign of $w_j$, $\delta$ is the sign of $w_l$. Provided $0<\theta<\pi/2$, we can choose these functions so that they are continuous and $|\Theta_{\gamma\delta}(\alpha)|<|\gamma\theta-\delta\pi|$. Logarithmically differentiating (4.6), we obtain

$$\Theta'_{\gamma\delta}(\alpha)=\frac{\gamma\sin\theta}{\delta\cosh\alpha+\gamma\cos\theta}+\frac{\gamma\sin 2\theta}{\delta\cosh\alpha-\gamma\cos 2\theta}. \tag{4.26}$$

Let $Y_+(Y_-)$ be the set of values of $j$ for which $w_j$ is positive (negative); let $n_+(n_-)$ be the number of such values, so $n=n_++n_-$. Then the analogue of (4.13) is, for $\gamma=\pm$ and $j\in Y_\gamma$,

$$Mk_\gamma(\alpha_j)=(2I_j+1-M)\pi+\sum_{l\in Y_+}\Theta_{\gamma+}(\alpha_j-\alpha_l)+\sum_{l\in Y_-}\Theta_{\gamma-}(\alpha_j-\alpha_l). \tag{4.27}$$

The integers $I_j$ must be distinct (modulo $M$), but are no longer necessarily given by (4.14). We determine them so as to match $z_1,\ldots,z_n$ for $q$ just less than 4 with the values already obtained for $q>4$.

For $q>4$, $I_1,\ldots,I_n$ increase uniformly (in unit intervals), while $z_1,\ldots,z_n$ are ordered anticlockwise round the unit circle. However, as $q\to 4$, $\frac{1}{3}$ of $z_1,\ldots,,z_n$ cluster round unity, their deviation from it being proportional to $q-4$. It follows that as $q$ passes through 4, these $z_j$ reverse their ordering. These are the ones that correspond to $w_j$ being positive, i.e. to $\arg(z_j)$ lying in the interval $(-\theta,\theta)$.

This suggests that for $q<4$ we can still take the $I_j$ to be given by (4.14), provided we arrange the $z_j$ so that those with arguments between $-\pi$ and $-\theta$ come first, the arguments increasing with $j$; then come those in the interval $(-\theta,\theta)$, but with the arguments *decreasing* with $j$; finally come those in $(\theta,\pi)$, with increasing arguments.

It is convenient to combine the first and last sets, and to take $z_1, \ldots, z_{n_-}$ to have arguments in the interval $(\theta - 2\pi, -\theta)$, in increasing order, while $z_{n_-+1}, \ldots, z_n$ have arguments in $(-\theta, \theta)$, in *decreasing* order. Thus $Y_-$ is the set $\{1, \ldots, n_-\}$ and $Y_+$ is $\{n_-+1, \ldots, n\}$. Provided $n_-$ is even, the effect of this is simply to replace (4.14) by

$$I_j = j - 1 - \tfrac{1}{2}n_-. \tag{4.28}$$

From (4.23), it follows that the sequences $\{\alpha_1, \ldots, \alpha_{n_-}\}$ and $\{\alpha_{n_-+1}, \ldots, \alpha_n\}$ are both increasing, in the interval $(-\infty, \infty)$. In the limit when $n_-$, $n_+$, $n$, $M$ become large, we expect these $\alpha_j$ to form two continuous distributions. Let $M\rho_\gamma(\alpha)\,\mathrm{d}\alpha$ be the number of $\alpha_j$ (with $j \in Y_\gamma$) in the interval $(\alpha, \alpha + \mathrm{d}\alpha)$. Then

$$n_\gamma / M = \int_{-\infty}^{\infty} \rho_\gamma(\alpha)\,\mathrm{d}\alpha \tag{4.29}$$

and (4.27) becomes the coupled pair of linear integral equations

$$k_\gamma(\alpha) = 2\pi \int_{-\infty}^{\alpha} \rho_\gamma(\beta)\,\mathrm{d}\beta + c_\gamma + \sum_{\delta=\pm} \int_{-\infty}^{\infty} \Theta_{\gamma\delta}(\alpha - \beta)\rho_\delta(\beta)\,\mathrm{d}\beta \tag{4.30}$$

where $\gamma = \pm$ and $c_+ = -n_+\pi/M$, $c_- = -(M + n_-)\pi/M$.

As is usual with Bethe ansatz calculations, we can solve these equations by differentiating and using Fourier transforms (in this case Fourier integrals). Let

$$\begin{aligned} 2\pi\rho_\gamma(\alpha) &= \int_{-\infty}^{\infty} \hat{\rho}_\gamma(x)\,\mathrm{e}^{\mathrm{i}\alpha x}\,\mathrm{d}x \\ k'_\gamma(\alpha) &= \int_{-\infty}^{\infty} \hat{k}_\gamma(x)\,\mathrm{e}^{\mathrm{i}\alpha x}\,\mathrm{d}x \\ \Theta'_{\gamma\delta}(\alpha) &= \int_{-\infty}^{\infty} \hat{\Theta}_{\gamma\delta}(x)\,\mathrm{e}^{\mathrm{i}\alpha x}\,\mathrm{d}x. \end{aligned} \tag{4.31}$$

Then we obtain, for $\gamma = \pm$,

$$\hat{k}_\gamma(x) = \hat{\rho}_\gamma(x) + \sum_{\delta=\pm} \hat{\Theta}_{\gamma\delta}(x)\hat{\rho}_\delta(x). \tag{4.32}$$

From (4.24) and (4.26), provided $0 < \theta < \pi/2$,

$$\hat{k}_+(x) = -\frac{\sinh \theta x}{\sinh \pi x} \qquad \hat{k}_-(x) = \frac{\sinh(\pi - \theta)x}{\sinh \pi x} \tag{4.33}$$

$$\begin{aligned} \hat{\Theta}_{++}(x) = \hat{\Theta}_{--}(x) &= \frac{\sinh \theta x + \sinh(\pi - 2\theta)x}{\sinh \pi x} \\ \hat{\Theta}_{+-}(x) = \hat{\Theta}_{-+}(x) &= -\frac{\sinh(\pi - \theta)x + \sinh 2\theta x}{\sinh \pi x}. \end{aligned} \tag{4.34}$$

Equations (4.32) are easily solved by summing and differencing. After some considerable simplifications, we obtain

$$\begin{aligned} \hat{\rho}_+(x) + \hat{\rho}_-(x) &= \frac{1}{2\cosh \theta x - 1} \\ \hat{\rho}_-(x) - \hat{\rho}_+(x) &= \frac{1}{2\cosh \theta x + 1}. \end{aligned} \tag{4.35}$$

It follows at once from (4.29) that

$$n_+ + n_- = M \qquad n_- - n_+ = M/3 \tag{4.36}$$

so $n_+ = M/3$, $n_- = 2M/3$. Thus we still have $n = M$, and $\frac{1}{3}$ of $w_1, \ldots, w_n$ are positive and $\frac{2}{3}$ are negative, in agreement with the behaviour we expected near $q = 4$.

Using (4.7) and the evenness of $\rho_\pm(\alpha)$,

$$M^{-1} \sum_{j=1}^{M} \ln \mu(z_j) = \tfrac{1}{2} \int_{-\infty}^{\infty} \ln\left(\frac{\cosh \alpha - \cos 2\theta}{\cosh \alpha - 1}\right) \rho_+(\alpha)\, d\alpha + \tfrac{1}{2} \int_{\infty}^{\infty} \ln\left(\frac{\cosh \alpha + \cos 2\theta}{\cosh \alpha + 1}\right) \rho_-(\alpha)\, d\alpha. \tag{4.37}$$

Substituting the Fourier integrals (4.31) for $\rho_\pm(\alpha)$, the RHS becomes

$$\tfrac{1}{2} \int_{-\infty}^{\infty} \frac{\cosh \pi x - \cosh(\pi - 2\theta)x}{x \sinh \pi x} \hat{\rho}_+(x)\, dx - \tfrac{1}{2} \int_{-\infty}^{\infty} \frac{\cosh 2\theta x - 1}{x \sinh \pi x} \hat{\rho}_-(x)\, dx. \tag{4.38}$$

Using the expressions given by (4.35) for $\hat{\rho}_\pm(x)$, (4.38) in turn becomes the same as the RHS of (1.9) (with $x$ instead of $k$ as the integration variable). Since the products in (3.31) are equal, for $M$ large the LHS of (4.37) is the same as $M^{-1} \ln \Lambda$. From (3.33) this is $\ln W$, so we have obtained the result (1.9).

We have only quoted this result in § 1 for $3 \leqslant q < 4$, i.e. for $0 < \theta \leqslant \pi/3$, where we expect $W$ to be a continuous function of $q$. It is presumably valid for smaller values of $q$, though not smaller than 2, since then $\theta > \pi/2$ and the second formula in (4.34) ceases to be valid.

## 5. Other real values of $q$

We expect $W$ to have singularities only at limit points (for $N$ large) of zeros of the chromatic polynomial $Z_{\text{Potts}}$. The above calculations are certainly correct for sufficiently large $q$. If (as seems likely) there are no limit points of zeros on the real axis for $q > 4$, then they are valid for $q > 4$.

It is slightly surprising that our results have a singularity at $q = 4$. This seems to imply that $q = 4$ is a limit point of zeros, even though $Z_{\text{Potts}}$ is positive (and grows exponentially with $N$) for $q = 4$.

Our result (1.9) for $q < 4$ has been obtained by extending the $q > 4$ result in a natural way. If one analytically continues it back to $q > 4$ and compares it with the correct result, the difference is very small, proportional to

$$\exp[-\pi^2/3(q-4)^{1/2}]. \tag{5.1}$$

This type of essential singularity occurs in the six-vertex antiferroelectric $F$ model (Lieb 1967, Lieb and Wu 1972) and is very weak. It may be that although there are zeros of $Z_{\text{Potts}}$ near $q = 4$, their distribution is very sparse, so that one would have to consider large values of $N$ to see them in a numerical calcuation.

The $q < 4$ result is analytic for $2 < q < 4$ and gives the correct trivial result for $q = 3$. This suggests that there are no limit points of zeros for $3 \leqslant q < 4$. If we accept the Beraha conjecture (Bevaha *et al* 1980) that real limit points occur only for $q = 2 + 2\cos(2\pi/n)$, $n$ being an integer, then the only remaining possible singular points on the real axis are $q = 0, 1, 2$ and $(3+\sqrt{5})/2 = 2.618\ldots$. Thus the result (1.9) is probably correct for $q > 2.618\ldots$ (i.e. for $\theta < 2\pi/5$) and may be correct for $q > 2$.

*5.1. $q<0$*

The '$q>4$' result (1.7) is correct for sufficiently large $q$, positive, negative or complex. The Beraha conjecture (together with the analyticity of the result) suggests that it may be true for all negative real values of $q$. Certainly there is no problem in applying the above $q>4$ working to the $q<0$ case: $r$ is now negative (between $-1$ and 0) and $z_j$ is given by (4.5) and (4.9), where $u_1, \ldots, u_n$ are distributed over the interval $(-\pi, \pi)$, with the distribution function $\rho(u)$ given by (4.17).

*5.2. $0<q<1$*

We can extend these negative-$q$ results through $q=0$ in much the same way as we extended the $q>4$ results through $q=4$. As $r\to-1$, (4.17) becomes $\rho(u)=\delta(u)$, so $u_1, \ldots, u_n$ are clustered round $u=0$. The distribution has half-width proportional to $\ln(-r)$, from which it follows that $z_1, \ldots, z_n$ are distributed right round the unit circle. For $q$ just greater than zero it follows that $w_1, \ldots, w_n$ are all positive, so now we repeat the working of (4.21)–(4.38), but taking all the $w_j$ to be positive and the arguments of $z_1, \ldots, z_n$ to be increasing.

This case is simpler than the $3\leqslant q<4$ case discussed above and is analogous to the usual $-1<\Delta<1$ case of the six-vertex model. We replace (4.23) by $w_j=\exp(\alpha_j)$, so as to ensure that $\{\alpha_1, \ldots, \alpha_n\}$ is an increasing sequence. The effect of this is to negate $k'_+(\alpha)$ and $\Theta'_{++}(\alpha)$. We only take $\gamma=\delta=+$ in (4.30) and (4.32), obtaining

$$\hat{\rho}_+(x)=\frac{1}{2\cosh(\pi-\theta)x-1}. \tag{5.2}$$

Equation (4.38) is still valid, except that the second integral (involving $\hat{\rho}_-$) does not occur, so

$$M^{-1}\ln\Lambda=\int_{-\infty}^{\infty}\frac{\sinh\theta x\sinh(\pi-\theta)x}{x\sinh\pi x[2\cosh(\pi-\theta)x-1]}\,\mathrm{d}x. \tag{5.3}$$

Unfortunately we have no independent calculations of $W$ for small positive values of $q$. If $\Lambda$ is still the numerically largest eigenvalue of the transfer matrix $V$, then (5.3) gives $W$ via (3.32) (presumably from $q=0$ up to the next Beraha number $q=1$, i.e. $\pi>\theta>2\pi/3$). However, it is possible that another contributing eigenvalue exceeds this one when $q$ increases through zero. To eliminate (or confirm) this possibility, we either need to prove that we have found the largest eigenvalue (which would be very difficult), or to make a numerical study of the eigenvalues of $V$ and/or the zeros of the chromatic polynomial $Z_{\text{Potts}}$.

In the complex $q$ plane, the zeros of $Z_{\text{Potts}}$ will presumably tend to be distributed along lines dividing the plane into domains. Once one is confident that one knows $W$ inside the various domains, their boundaries (and the large $N$ limiting distributions of zeros) can be obtained by comparing the form of $W$ on opposite sides.

**Acknowledgments**

The author is indebted to Professors A J Guttmann and A D Sokal for very useful discussions.

*Note added in proof.* The author thanks Dr V E Korepin for pointing out that the loop model of (1.1) and (1.3*a*) is a special case of the solvable Izergin-Korepin three-state vertex model on a square lattice of $N$ sites (Izergin A G and Korepin V E 1980 *LOMI preprint E380, Leningrad*; 1981 *Commun. Math. Phys.* **79** 303-16).

Very recent numerical results for triangular lattices of up to eight rows and columns, with free boundary conditions, suggest that in the large-lattice limit the complex zeros of the chromatic polynomial form continuous distributions on two curves A and B, where A is approximately the circle $|q-2|=2$; B is approximately an interior vertical chord of A, crossing the real axis at $q=q_l$, where $q_l \simeq 3.7$. There are also isolated real zeros at the Beraha numbers $q=2+2\cos(2\pi/n)$ to the left of B, i.e. for $0 \leq q \leq q_l$.

The curves A and B divide the complex $q$ plane into three domains: the exterior of A, the interior to the right of B and the interior to the left of B. The results (1.7), (1.9) and (5.3) should apply in these three domains, respectively; except that (5.3) fails at the relevant Beraha numbers (e.g. $q=3$), where $W$ is discontinuous.

For $q$ real, this implies that (5.3) holds for $0<q<q_l$, $q \neq 2+2\cos(2\pi/n)$; while (1.9) is true only for $q_l<q<4$. (It seems that it is an accident that (1.9) agrees with the finite-lattice $q=3$ result.) It is intended to report and discuss these numerical results separately.

## References


Baxter R J 1970 *J. Math. Phys.* **11** 784-9

—— 1982a *Exactly solved models in statistical mechanics* (New York: Academic)

—— 1982b *J. Stat. Phys.* **28** 1-41

Baxter R J, Kelland S B and Wu F Y 1976 *J. Phys. A: Math. Gen.* **9** 397-406

Baxter R J, Temperley H N V and Ashley S E 1978 *Proc. R. Soc.* A **358** 535-59

Beraha S and Kahane J 1979 *J. Combin. Theory* B **27** 1-12

Beraha S, Kahane J and Weiss N J 1978 *Studies in Foundations and Combinatorics* ed C G Rota (New York: Academic) pp 213-32

—— 1980 *J. Combin. Theory* B **28** 52-65

Domany E and Schaub B 1984 *Phys. Rev.* B **29** 4095-107

Guttmann A J 1984 *J. Phys. A: Math. Gen.* **17** 455-68

Kasteleyn P W and Fortuin C M 1969 *J. Phys. Soc. Japan Suppl.* **26** 11-4

Kim D and Enting I G 1979 *J. Combin. Theory* B **26** 327-36

Lieb E H 1967 *Phys. Rev.* **162** -72

Lieb E H and Wu F Y 1972 *Phase Transitions and Critical Phenomena* vol 1, ed C Domb and M S Green (New York: Academic) pp 32-490

Nienhuis B 1982 *Phys. Rev. Lett.* **49** 1062-5

—— 1984 *J. Stat. Phys.* **34** 731-61

Nienhuis B, Hilhorst H J and Blöte H W J 1984 *J. Phys. A: Math. Gen.* **17** 3559-81

Stephenson J 1964 *J. Math. Phys.* **5** 1009-24

Tutte W T 1973 *Can. J. Math.* **25** 426-47

—— 1982 *Can. J. Math.* **34** 741-58

Wannier G H 1950 *Phys. Rev.* **79** 357-64