Analytical calculation of two leading exponents of the dilute Potts model

Analytical calculation of two leading exponents of the dilute Potts model

B Nienhuis†

Laboratorium voor Technische Natuurkunde THD Postbus 5046, 2600 GA Delft, The Netherlands and Department of Electronics, The Weizmann Institut of Science, Rehovot, Israel

Abstract. A Potts model on a square lattice with two- and four-spin interaction and site and bond dilution is shown to be dual to itself. The model is mapped onto a vertex problem which in turn is equivalent to a solid-on-solid model. By means of these mappings the dilute Potts model can be written as a Gaussian-like model with staggered and direct periodic fields. The leading and next-to-leading exponents of the Potts model are calculated, subject to the validity of certain assumptions.

1. Introduction

The last decade has brought considerable progress in the understanding of the two-dimensional q-state Potts (1952) model. By virtue of an equivalence with a staggered vertex model, demonstrated by Temperley and Lieb (1971), Baxter (1973) showed that the model with $q > 4$ has a first-order transition and that with $q \leq 4$ a second-order transition. This fact was not readily verifiable by numerical methods (Kim and Joseph 1975, Dasgupta 1977, den Nijs 1978). den Nijs (1979) conjectured a specific dependence of the thermal exponents on q for $q \leq 4$, based mainly on numerical results. Nienhuis *et al* (1979, 1980), introducing the concept of the diluted Potts model or Potts lattice gas, provided a renormalisation group understanding of the change of the order of the transition at $q = 4$. This led Rebbi and Swendsen (1980) to explain why previous numerical work had not detected the first-order transition. Recently Black and Emery (1981) verified the original conjecture of den Nijs for the thermal critical exponents by a direct calculation.

It is the purpose of this paper to generalise some exact results for the Potts model to the dilute Potts model, in order to calculate the tricritical and next-to-leading exponents. It is shown that a suitably generalised Potts lattice gas is again self-dual and can be mapped onto a vertex model and ultimately onto a Gaussian-type model in fields. With this mapping it can be shown that the extension of the den Nijs conjecture gives the tricritical thermal exponent. The cross-over tricritical exponent and a correction to scaling exponent for the critical point turn out to be given by a formula previously conjectured by Burkhardt (1980).

† Address after September 1, 1981: The James Franck Institute, University of Chicago, 5640 Ellis Avenue, Chicago Illinois 60637, USA.

2. The model

Each site i of a Potts lattice gas is either vacant or occupied by a Potts variable s_i which can assume q values. The occupation is described by a variable t_i which is 0 (vacant) or 1 (occupied). The most general nearest-neighbour Hamiltonian for this model on a square lattice is

$$-\beta H_0 = \sum_{\text{edges}} [Jt_i t_j \delta_{s_i s_j} + K(1-t_i)(1-t_j)] + \sum_{\text{sites}} \Delta(1-t_i). \tag{1}$$

The summations are over the edges (i, j) and the sites i of the lattice. The couplings J and K are Potts and lattice gas interactions, and the chemical potential Δ controls the concentration of vacancies. The pure Potts model ($\Delta = -\infty$) on a square lattice is self-dual and is equivalent to a six-vertex model (Temperley and Lieb 1971). These nice properties are lost for the dilute model, and we shall attempt to restore them.

It has been recognised (Kondor 1980, Wu 1981) that vacancies map under duality onto multi-spin interaction. It turns out that the lattice gas coupling of vacancies can be described by bond dilution in the dual Hamiltonian. We shall show that the following Hamiltonian is form-invariant under duality:

$$H = H_0 + H_1, \tag{2}$$

$$-\beta H_1 = \sum_{\text{faces}} L t_i t_j t_k t_l \tau_{ij} \tau_{jk} \tau_{il} \tau_{lk} \delta_{s_i s_j} \delta_{s_j s_l} \delta_{s_j s_k} + \sum_{\text{edges}} B(1-\tau_{ij}), \tag{3}$$

in which the edge variables $\tau_{ij} = 0$ or 1. The bond dilution is governed by B, and L is the strength of four-spin interaction of the elementary squares or faces $(ijkl)$. In this Hamiltonian the bond dilution does not affect the two-spin coupling, but the free energy of this problem can be expressed readily in that of the system where the two-spin interaction term in (1) is multiplied by τ_{ij}.

The partition sum

$$Z = \sum_{\{t,s,\tau\}} \left(\prod_{\text{sites}} [t_i + (1-t_i)e^{\Delta}] \right) \left(\prod_{\text{edges}} \exp[K(1-t_i)(1-t_j) + B(1-\tau_{ij})][1 + (e^J - 1)t_i t_j \delta_{s_i s_j}] \right)$$

$$\times \left(\prod_{\text{faces}} [1 + (e^L - 1)t_i t_j t_k t_l \tau_{ij} \tau_{jk} \tau_{lk} \tau_{il} \delta_{s_i s_j} \delta_{s_j s_k} \delta_{s_j s_l}] \right). \tag{4}$$

The summand (4) contains besides multiplicative factors two terms for each site, generically t and $(1-t)e^{\Delta}$, two for each edge, 1 and $(e^J - 1)tt\delta$, and two for each face, 1 and $(e^L - 1)tttt\tau\tau\tau\tau\delta\delta\delta$. Expanding the product, one obtains a number of terms, each of which can be represented by a diagram on the lattice, by means of the following rules.

(i) When the term $(1-t)e^{\Delta}$ is taken, a dot is placed on the corresponding site.

(ii) The term $(e^L - 1)tttt\tau\tau\tau\tau\delta\delta\delta$ is represented by a dot in the middle of the face and bonds are placed on its surrounding edges.

(iii) On the four edges surrounding a dotted face, the terms $(e^J - 1)tt\delta$ and 1 are taken together, adding up to e^J.

(iv) On the remaining edges bonds are placed whenever the term $(e^J - 1)tt\delta$ is chosen.

A typical diagram is shown in figure 1.

For each diagram the summation over $\{t, s, \tau\}$ can be performed trivially to obtain the contribution to Z. The sum over t variables is inhibited by the distribution of dotted

sites. Besides, each dotted site contributes a factor e^Δ and each nearest-neighbour pair a factor e^K. The s variables on sites connected by bonds must be equal. Each dotted face contributes a factor $e^L - 1$ and constrains the τ on its surrounding edges. In addition since the s variables that surround a dotted face must be equal, the surrounding edges contribute e^J. From the remaining edges we obtain a factor $e^J - 1$ or 1 depending on the presence of a bond. For the total partition sum this results in

$$Z = \sum_g e^{\Delta v + K p_v}(e^J - 1)^{b-4w+p_w}(e^L - 1)^w \ e^{J(4w-p_w)}(e^B + 1)^{E-4w+p_w}q^{c-v}. \quad (5)$$

Here v is the number of vacancies or dotted sites, p_v the number of nearest-neighbour pairs of vacancies; w is the number of dotted faces and p_w the number of nearest-neighbour pairs of them; b is the number of bonds, c the number of connected components (regarding an isolated site as a component) and E the total number of edges. The summation in (5) is over all graphs consisting of bonds, dotted sites and dotted faces in which each dotted face is enclosed in four bonds and no bond is incident at a dotted site.

The dual of the original square lattice $\mathscr{S}$ is again a square lattice $\tilde{\mathscr{S}}$. As usual sites, edges and faces of $\mathscr{S}$ correspond to faces, edges and sites respectively of $\tilde{\mathscr{S}}$. Therefore a dotted site of $\mathscr{S}$ can be interpreted as a dotted face of the dual lattice, and vice versa. If a bond is placed on those edges of $\tilde{\mathscr{S}}$ to which the corresponding edge of $\mathscr{S}$ is empty, this completes a 1–1 map of graphs on $\mathscr{S}$ to graphs on $\tilde{\mathscr{S}}$. As an example figure 2 shows the dual graph of figure 1. Note that also on the dual lattice each dotted face is enclosed in four bonds and no bond touches on a dotted site.

Each component of the original graph is enclosed in a loop or circuit of the dual graph. The number of components c is therefore equal to the number of circuits $\tilde{l}$ of the dual graph. Let l be the number of circuits of the original graph, and let the symbols introduced in equation (5) provided with a ˜ denote the corresponding quantities of the dual graph; then the dual map of graphs is characterised by the equations

$$\tilde{v} = w, \qquad \tilde{p}_v = p_w, \qquad \tilde{w} = v, \qquad \tilde{p}_w = p_v, \qquad \tilde{l} = c, \qquad \tilde{c} = l, \qquad \tilde{b} = E - b. \quad (6)$$

These numbers are not all independent, but related by Euler's equation $N - b + l - c = 0$, in which N is the total number of sites. The sum over all allowed graphs of $\mathscr{S}$ can be written as a sum over the same class of graphs on $\tilde{\mathscr{S}}$. Thus the partition sum satisfies the duality relation

$$(e^{\tilde{B}} + 1)^{-E}(e^{\tilde{J}} - 1)^{-N}Z_{\tilde{\mathscr{S}}}(\tilde{J}, \tilde{K}, \tilde{\Delta}, \tilde{B}, \tilde{L}) = (e^B + 1)^{-E}(e^J - 1)^{-N} Z_{\mathscr{S}}(J, K, \Delta, B, L), \quad (7)$$

where the subscript on the partition sums Z refers to the lattice on which it is evaluated, and where

$$e^{\tilde{J}} - 1 = q(e^J - 1)^{-1}, \qquad\qquad\qquad\qquad\qquad (8a)$$

$$e^{\tilde{K}} = (1 - e^{-J})(e^B + 1), \qquad\qquad\qquad\qquad\qquad (8b)$$

$$e^{\tilde{\Delta}} = q(e^L - 1)(1 - e^{-J})^{-4}(e^B + 1)^{-4}, \qquad\qquad\qquad (8c)$$

$$e^{\tilde{B}} + 1 = (e^K/q)(e^J + q - 1), \qquad\qquad\qquad\qquad (8d)$$

$$e^{\tilde{L}} - 1 = (1/q) e^{\Delta+4K}. \qquad\qquad\qquad\qquad\qquad (8e)$$

In the space of five parameters J, K, Δ, B and L a two-dimensional subspace of self-dual Hamiltonians is found. The critical subspace is likely to be larger. In the limit $\Delta = -\infty$,
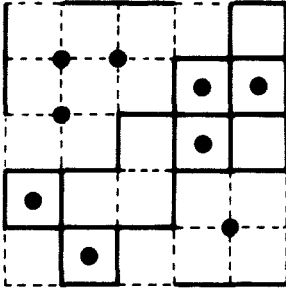
Figure 1. An example of a graph in the expansion of equation (4). The broken lines represent the edges of the lattice. Bonds placed on the edges are indicated by full lines. Heavy dots placed on the sites or faces represent vacancies or four-spin coupling terms respectively.
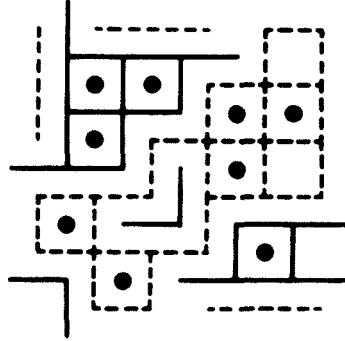

Figure 2. The dual graph of figure 1. The bonds of the graph shown in figure 1 are indicated by broken lines. The full lines represent bonds of the dual graph.
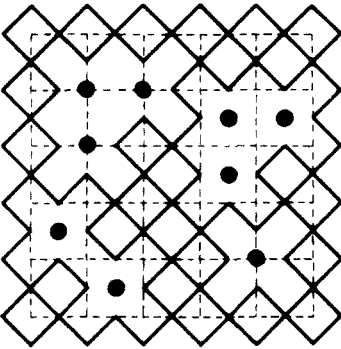

Figure 3. The surrounding lattice $\mathscr{L}'$ associated with the dot configuration of figure 1 is indicated by the full lines. The broken lines show the original lattice $\mathscr{L}$. $\mathscr{L}'$ consists of the surrounding lattice of $\mathscr{L}$ from which for each dot the four surrounding edges are deleted.

K is a redundant parameter, and likewise for $L = 0$, B is redundant. These redundancies probably continue to exist for finite L and Δ, thus giving rise to a four-dimensional critical manifold.

3. Equivalent vertex model

The dilute Potts model defined by (2) will be shown to be equivalent to a vertex model by means of the graphical method introduced by Baxter *et al* (1976). For the detailed arguments and definitions we refer to their article, but we shall try to remind the reader familiar with their concepts.

Consider the graphical expansion (5) of Z. Each graph is a configuration of bonds on edges and of dots on sites and faces satisfying certain restrictions. Let the dots be placed on the lattice first, and also bonds on all the edges that surround the dots on faces. While these are held fixed, a part of the total sum can be performed by summation over bond configurations on the remaining edges not touching on vacancies. In formula:

$$Z = \sum_{\text{dots}} e^{\Delta v + K p_v}(e^L - 1)^w \, e^{J(4w - p_w)}(e^B + 1)^{E - 4w + p_w} W(\text{dots}) \tag{9}$$

where

$$W(\text{dots}) = \sum_{\text{bonds}} (e^J - 1)^{b-4w+p_w} q^{c-v}.$$ (10)

The first sum (9) is over all configurations of dots and the second (10) over all configurations of bonds compatible with the dots.

With each such configuration of dots we associate a restricted lattice $\mathcal{L}$. The sites of $\mathcal{L}$ are all non-vacant sites of $\mathcal{S}$ or groups of sites coupled via one or more of those bonds that surround dotted faces. Thus each dotted face couples four sites of $\mathcal{S}$ to form one site of $\mathcal{L}$, and if neighbouring faces have dots, one site of $\mathcal{L}$ may correspond to even more sites of $\mathcal{S}$. The edges of $\mathcal{L}$ are all those edges of $\mathcal{S}$ that do not touch a vacancy or a dotted face. $W(\text{dots})$ is the partition sum of a pure Potts model on $\mathcal{L}$. Z can therefore be viewed as a grand-canonical partition sum of Potts models.

Baxter *et al* begin with a planar, not necessarily regular lattice $\mathcal{L}$. They introduce its surrounding lattice $\mathcal{L}'$ consisting of polygons surrounding the sites of $\mathcal{L}$. Sites as well as faces of $\mathcal{L}$ correspond to faces of $\mathcal{L}'$. The sites of $\mathcal{L}'$ have either two or four neighbours and are called external or internal sites respectively. Internal sites correspond to edges of $\mathcal{L}$. The edges of $\mathcal{L}'$ (but not necessarily of $\mathcal{L}$) must be straight. An ice-type vertex model is then defined by placing an arrow on each of the edges of $\mathcal{L}'$ so that an equal number of arrows point in and out at each site of $\mathcal{L}'$. A weight is associated with each site depending on the relative angle of the four (or two) edges, and on the configuration of arrows on these edges. It is then proven that the partition sum of the vertex model is equal to that of the Potts model.

Following Baxter *et al*, we associate with the square lattice $\mathcal{S}$ its surrounding lattice $\mathcal{S}'$, the sites of which are the midpoints of the edges of $\mathcal{S}$. Since sites as well as faces of $\mathcal{S}$ are associated with faces of $\mathcal{S}'$, a configuration of dots on both sites and faces of $\mathcal{S}$ corresponds to dots on faces only of $\mathcal{S}'$.

With each restricted lattice $\mathcal{L}$ we associate as surrounding lattice $\mathcal{L}'$ a restriction of $\mathcal{S}'$. The edges of $\mathcal{S}'$ not directly surrounding a dot are also edges of $\mathcal{L}'$. Figure 3 shows a dot configuration on $\mathcal{S}$ and the corresponding surrounding lattice $\mathcal{L}'$.

As $W(\text{dots})$ is the partition sum of a Potts model on $\mathcal{L}$ it is equal to the partition sum of a suitably defined ice-type model on $\mathcal{L}'$. Therefore Z can be viewed as a grand-canonical ensemble of ice models. It is convenient however, to consider it as a partition sum of a generalised ice-type model on $\mathcal{S}'$, which can be defined as follows.

(i) Place dots on some faces of $\mathcal{S}'$ so that no pair of neighbouring faces have dots simultaneously.

(ii) Place arrows on all those edges of $\mathcal{S}'$ that do not constitute a dotted face, so that at each site an equal number of arrows point in and out. At each site there are sixteen possible arrangements of dots and arrows which are shown as lines (b) and (b') of figure 4.

(iii) Associate a weight ω_k with each site, as given by the expressions (a) and (a') shown in figure 4 corresponding to the configuration of arrows and dots at the site.

(iv) The partition sum of the model is the sum over all such dot and arrow configurations of the product of the weights of all sites.

The weights ω_{1-6} are those given by Baxter *et al*, with

$$x = (e^{J-1})q^{-1/2}, \qquad z^{2\pi} + z^{-2\pi} = q^{1/2},$$ (11)

and ω_{7-16} are designed to take care of the weight of the distribution of dots in

(a) $\omega_1 = 1$ $\omega_2 = 1$ $\omega_3 = x$ $\omega_4 = x$ $\omega_5 = z^{\pi} + xz^{-\pi}$ $\omega_6 = z^{-\pi} + xz^{\pi}$ $\omega_7 = QO^2$ $\omega_8 = RP^2$

(b)

(c)

(d)
$$
\begin{array}{cccccccc}
+ & - & + & + & + & + & 0 & + \\
+\ +\ - & +\ +\ + & +\ +\ + & +\ -\ + & +\ +\ - & -\ -\ + & +\ +\ 0 & 0\ 0\ + \\
& & + & + & + & + & 0 & \\
\end{array}
$$

(a') $\omega_9 = Oz^{\pi/2}$ $\omega_{10} = Oz^{-\pi/2}$ $\omega_{11} = Oz^{\pi/2}$ $\omega_{12} = Oz^{-\pi/2}$ $\omega_{13} = Pz^{\pi/2}$ $\omega_{14} = Pz^{-\pi/2}$ $\omega_{15} = Pz^{\pi/2}$ $\omega_{16} = Pz^{\pi/2}$

(b')

(c')

(d')
$$
\begin{array}{cccccccc}
+ & - & 0 & 0 & + & + & + & + \\
+\ +\ 0 & +\ +\ 0 & +\ +\ + & +\ +\ - & -\ 0\ + & +\ 0\ + & 0\ -\ + & 0\ +\ + \\
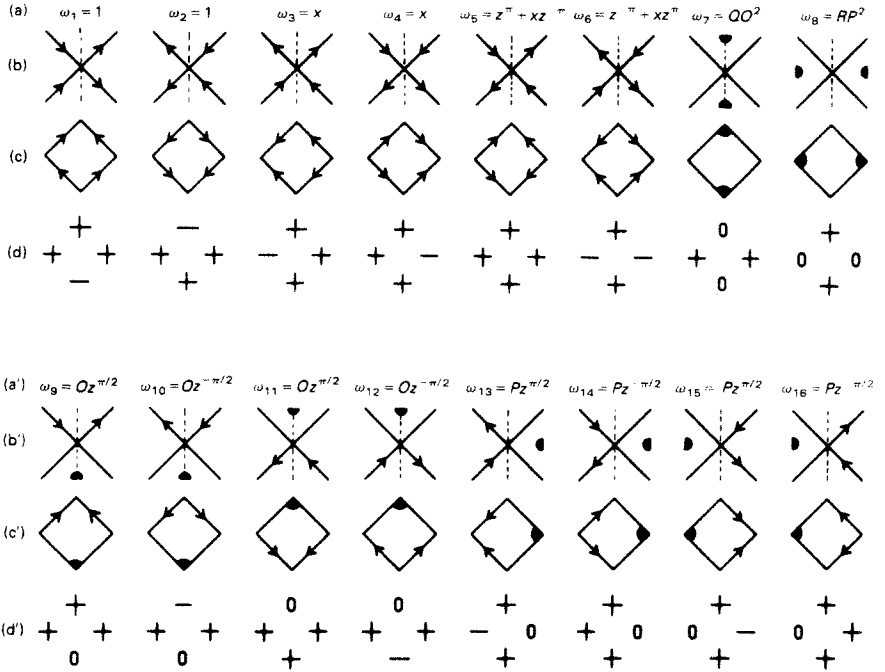\end{array}
$$

Figure 4. Weights and configurations. Lines (b) and (b') show the 16 possible configurations of the ice-type model equivalent with a dilute Potts model. Lines (a) and (a') represent the weights of each of the configurations. The weights depend on the orientation of the arrows relative to the edge of the original lattice $\mathscr{S}$ indicated by a broken line. Lines (c) and (c') $((d)$ and $(d'))$ show the corresponding configurations in the SRM (Ising model) as defined in § 4.

equation (9):

$$O^4 = e^{\Delta}q^{-1/2}, \qquad P^4 = (e^L - 1)(e^B + 1)^{-4}\,e^{4J}q^{-3/2}, \tag{12}$$

$$Q = e^K, \qquad R = (1 - e^{-J})q^{1/2}.$$

With these definitions the partition sum Z' of the generalised ice-type model is given by

$$Z' = (e^B + 1)^E\, q^{N/2} Z. \tag{13}$$

The parameter z is real for $q > 4$ and complex with $|z| = 1$ for $q < 4$.

4. Equivalent roughening and Ising models

The generalisation of the ice model defined above can be readily translated into a roughening model which is a natural generalisation of the BCSOS model (Van Beijeren 1977).

Consider the ice-type configurations shown as lines (b) and (b') of figure 4. Turn all the edges to the left over an angle $\pi/2$. This results in the configurations shown as lines (c) and (c') in the same figure. Note that the curl of the arrow arrangements thus obtained is zero. Therefore the figures $4(c)$ and (c') can be viewed as configurations of

four columns of different integer and half-integer heights, so that edges without arrows link columns of equal height, and edges with arrows connect columns that differ in height by $\frac{1}{2}$, the arrow pointing in the direction of the higher column. If the weights (a) and (a') are assigned to the column configurations (c) and (c'), this completes the definition of a roughening model equivalent to the dilute Potts model given in (2). We shall refer to this roughening model as SRM (staggered roughening model).

It is convenient to differentiate between two sublattices 1 and 2, consisting of the sites of $\tilde{\mathscr{S}}$ and $\mathscr{S}$ respectively. The height of columns on sublattice s will be denoted by h_s or h'_s.

A special case of the SRM is the BCSOS model defined by Van Beijeren (1977), in which only configurations 1–6 are allowed. In those configurations let h_1 be integer and h_2 half-integer. With that convention the dots in the other configurations denote columns of the 'wrong' height, i.e. integer h_2 and half-integer h_1. Note that a column of the wrong height is always equal to all its neighbours.

In passing we observe that lines (d) and (d') of figure 4 show configurations of a spin-1 Ising model which again is equivalent to the dilute Potts model (2). The spins σ_s on sublattice s are defined as

$$\sigma_1 = \cos(h_1\pi), \qquad \sigma_2 = \sin(h_2\pi). \tag{14}$$

For integer and half-integer values of h, σ is 0 or ± 1. By a shift of h to $h + 1$, which does not affect the energy, all σ variables change sign, so that the 16 configurations listed in figure 4 are all degenerate with their spin-inverted twins. The total number of allowed configurations of an elementary square of four spins is 32.

Vacancies in the Potts model correspond to instances of $\sigma_2 = 0$, while four-spin interaction corresponds to $\sigma_1 = 0$. In the context of this Ising model we shall use the terminology of vacancies and dilution for all $\sigma = 0$, irrespective of the sublattice.

5. Reformulation in periodic fields

The variables of the above-defined SRM can assume integer and half-integer values. A system with discrete variables can be described as a model with continuous variables in which a field, enhancing the original discrete values, is taken to infinity. The advantage of such description in the case of the SRM is that models with continuous variables of infinite range are relatively well understood. A rather large family of such models is in the universality class of the gaussian model, in which an infinite hierarchy of critical exponents can be calculated explicitly (Kandanoff and Brown 1979). Therefore in this section we shall rewrite the SRM as a model with continuous variables h in the presence of fields periodic in h. The relation with the gaussian model will be discussed in § 6.

Let the variables h_1 and h_2 reside on the sites of two interpenetrating square sublattices 1 and 2, as in the SRM except that h_1 and h_2 now may assume all real values. The Hamiltonian in the absence of fields is

$$-\beta H_G = -\beta \sum_{nn} V_1(h_1 - h_2) - \beta \sum_{nnn} V_2(h - h'). \tag{15}$$

The summations are over nearest-neighbour and next-nearest-neighbour pairs of variables. Therefore V_2 is an interaction within each sublattice, and V_1 couples the two sublattices. Both $V_1(h)$ and $V_2(h)$ are even functions increasing with $|h|$ and analytic at

$h = 0$. $V_1(h)$ is ∞ for $h \geqslant 1$. The precise form of V_1 and V_2 is not important but for convenience we will take $V_2(h) = h^2$.

Table 1 shows a number of interaction terms that will be added to the reduced Hamiltonian $-\beta H_G$ in order to turn it into the Hamiltonian of the SRM. These interactions will be discussed in the order in which they are listed.

Table 1. Fields listed in the first column are conjugate to periodic functions of the heights, shown in the second column. The last column displays the same functions in Ising language. The variables h_1, h_2, h_1' and h_2' reside on the corners of an elementary square.

F_1	$\cos(4\pi h)$	—
F_2	$\sin(\pi h_1)^2 \cos(\pi h_2)^2$	$(1 - \sigma_1^2)(1 - \sigma_2^2)$
Δ_-	$\cos(2\pi h_1) - \cos(2\pi h_2)$	$\sigma_1^2 + \sigma_2^2$
Δ_+	$\cos(2\pi h_1) + \cos(2\pi h_2)$	$\sigma_1^2 - \sigma_2^2$
AT	$\cos(\pi h_1)\cos(\pi h_1') - \sin(\pi h_2)\sin(\pi h_2')$	$\sigma_1 \sigma_1' - \sigma_2 \sigma_2'$
P_+	$\cos(\pi h_1)\sin(\pi h_2)$	$\sigma_1 \sigma_2$
P_-	$\cos(\pi h_1')\sin(\pi h_2')[\cos(2\pi h_1) + \cos(2\pi h_2)]$	$(\sigma_1^2 - \sigma_2^2)\sigma_1'\sigma_2'$
$K_\pm$	$\cos(2\pi h_1)\cos(2\pi h_1') \pm \cos(2\pi h_2)\cos(2\pi h_2')$	$\sigma_1^2 \sigma_1'^2 \mp \sigma_2^2 \sigma_2'^2$

The field F_1 should be taken to infinity so that all non-integer values of $2h$ are suppressed. This field acts on both sublattices alike. Given the remaining values of h, the corresponding Ising spins (14) $\sigma_{1,2}$ are 0 or ± 1. They provide a convenient language to refer to height configurations and interactions, which we will use below. The second and third columns of table 1 give the same function but expressed in h and σ respectively.

The field F_2 couples to a function of two neighbouring spins which is 1 if both are vacant and 0 otherwise. Taking this field to $-\infty$ the only surviving spin configurations are the ones listed in figures 4(d) and (d') and their opposites. The remaining fields in table 1 are necessary to reproduce the weights ω_k.

The field Δ_- controls the number of vacancies, and Δ_+ the difference in dilution between the two sublattices. In the Potts model $\Delta_+ + \Delta_-$ corresponds to the four-spin interaction L and $\Delta_+ - \Delta_-$ to the chemical potential Δ.

The natural ordering of the Ising version of the SRM is the prevalence of non-zero spins equal to their second neighbours. The ordering of both sublattices together is governed by the inverse temperature β in equation (15). The difference in ordering of the two sublattices is controlled by the field AT. The field P_+ controls the orientation of σ_1 relative to σ_2. AT and P_+ are not independent but both correspond to the parameter J in the dilute Potts model.

Another field that controls the relative orientations of σ_1 and σ_2 is P_-, but it only acts when a neighbouring spin σ' is zero. Depending on the sublattice of σ', it aligns or anti-aligns σ_1 and σ_2. This complicated interaction is somewhat more natural in the SRM: if a basic square of the lattice contains three columns of equal height and one different, P_- tends to make the single different one the higher.

Finally $K_+ + K_-$ and $K_+ - K_-$ are the lattice gas couplings between vacancies in sublattices 1 and 2 respectively.

Table 2 shows the relation between the fields in the continuous model, the weights of the SRM and the interaction parameters of the dilute Potts model. This table intends to give a physical intuition for the related quantities, not exact equations. The precise relations between the parameters of the dilute Potts model and the SRM are given in figures 4(a) and (a') together with the equations (11) and (12).

Table 2. Correspondence between fields in the gaussian model, weights of the staggered roughening model and interaction parameters of the dilute Potts model. ω's with multiple subscripts stand for geometric averages, e.g. $\omega_{5,6} = (\omega_5\omega_6)^{1/2}$ and $\omega_{1-4} = (\omega_1\omega_2\omega_3\omega_4)^{1/4}$.

Gaussian fields	SRM weights	Dilute Potts interactions
β	$\omega_{5,6}/\omega_{1-4}$	q
$\Delta_+ + \Delta_-$	ω_8/ω_{1-6}	L
$\Delta_+ - \Delta_-$	ω_7/ω_{1-6}	Δ
AT	$\omega_{3,4}/\omega_{1,2}$	J
P_+	ω_5/ω_6	J
P_-	$\omega_{9,11,13,15}/\omega_{10,12,14,16}$	$\cosh^{-1}(\tfrac{1}{2}q^{1/2})$
$K_+ + K_-$	$(\omega_8\omega_{1-6})/\omega_{9-12}^2$	B
$K_+ - K_-$	$(\omega_7\omega_{1-6})/\omega_{13-16}^2$	K

6. Critical behaviour

It is the purpose of this section to utilise these equivalences between models to calculate some critical exponents of the Potts model. This is done by showing under what conditions the SRM may flow towards the gaussian model under renormalisation. The extensive knowledge of the gaussian critical exponents then serves to analyse the critical behaviour of the SRM and therewith of the dilute Potts model.

Unlike the previous sections, the arguments here are not all rigorous. Some assumptions have to be made on the way, but given their validity two Potts exponents can be calculated as functions of q analytically.

Realising that it is the rough phase of a roughening model that flows towards the gaussian model, it is helpful to gain a physical intuition for the location of the rough phase, and its relation to the critical regime of the Potts model.

The phase transition in the Potts model can be driven first order by increasing q or by increasing the number of vacancies. Large q corresponds to large values of ω_5 and ω_6 and likewise large dilution translates into large ω_7 and ω_8. Observe that configurations 5–8 are exactly the ones without a net gradient. A predominance of those four configurations therefore drives the model into the smooth phase. This indicates that the first-order transition of the Potts model corresponds to the smooth phase of the SRM.

Also the SRM corresponding to a pure Potts model for small q can be found in a smooth phase. If for example x is small, height gradients are the result mainly of configurations 1 and 2. However these correspond to gradients on sublattice 2 only, and therefore cannot develop large fluctuations over large distances. Only when $x \approx 1$, that is in the critical region of the Potts model, and for sufficiently small q, one can expect the SRM to be rough. These considerations support the idea that the second-order phase transition on the Potts model corresponds to the rough phase of the SRM and can therefore be described by the gaussian model. In order to calculate exponents the identification between operators in the Potts and gaussian models needs to be made more precise.

The first step is to see that the Hamiltonian H_G without fields does renormalise towards a gaussian. By virtue of the analyticity of V_1, H_G deviates from a gaussian only by terms like $(h - h')^{2n}$ with $n > 1$. It is an easy check to verify that these terms are irrelevant in the vicinity of the gaussian fixed line. Numerical calculations (Jose *et al* 1977) indicate that also finite non-gaussian terms vanish under renormalisation. On the

basis of these indications we assume that renormalisation flows do lead H_G to a gaussian form.

The fields of table 1 are analysed in the same spirit. If a field F is irrelevant at the gaussian line, it is assumed that the presence of F in the initial Hamiltonian does not prevent its flow to a gaussian renormalised Hamiltonian.

Let y_F be the exponent of a field F defined by the behaviour of the singular part of the free energy as F^{2/y_F}. The scaling index of the operator O_F conjugate to F is related to y_F by $x_F + y_F = 2$. The correlation function of O_F at criticality is governed by x_F via

$$\langle O_F(0)\, O_F(r)\rangle \propto |r|^{-2x_F}, \tag{16}$$

where the vectors 0 and r are positions in the lattice. The exponents of the fields in table 1 can thus be caluclated by means of the associated correlation function. The correlation functions in the gaussian model needed to calculate the Potts exponents are of the form

$$\langle \exp(2\pi ikh(r) + 2\pi ilh(r') + 2\pi imh(0))\rangle = \delta_{k+l+m,0}\, |r|^{-2t(k^2+kl)}|r'|^{-2t(l^2+kl)}|r-r'|^{2tkl}. \tag{17}$$

Here t is proportional to the temperature of the gaussian model. Since the actual renormalisation flows are not known, t is an unknown function of the initial temperature β and of the fields. All critical exponents however can be parametrised in the one parameter t.

The operators in table 1 can be rewritten as linear combinations of sines and cosines of various heights, for example the operator conjugate to F_2,

$$\sin(\pi h_1)^2 \cos(\pi h_2)^2$$
$$= \tfrac{1}{4}\{1 + \cos(2\pi h_2) - \cos(2\pi h_1) - \tfrac{1}{2}\cos[2\pi(h_1+h_2)] - \tfrac{1}{2}\cos[2\pi(h_1-h_2)]\}. \tag{18}$$

The terms $\cos(2\pi h_2) - \cos(2\pi h_1)$ represent a field staggered in both directions of the lattice. The correlation function of this staggered operator between different elementary squares of the lattice can be found from equation (17) by differentiation with respect to both components of r and of r' and setting $k = -l = 1$ and $m = 0$. The operator therefore has a scaling index $2 + t$ and is consequently irrelevant for all t. Under renormalisation flow some potentially relevant operators will be generated, but in the SRM these are already present. The term $\cos[2\pi(h_1+h_2)]$ is equivalent to $\cos(4\pi h)$ and has scaling index $4t$. Finally the $\cos[2\pi(h_1-h_2)]$ term acts as the square of a gradient of h and thus modifies the effective temperature of the gaussian model.

The same analysis on all the operators of table 1 shows that they can all be written as direct and staggered spin-wave operators and gradients. This implies that all the exponents associated with the parameters in the Potts Hamiltonian are linear functions of t. The most relevant fields are Δ_+, AT and P_+ to which the dominant contributions are $\cos(2\pi h)$ or $\sin(2\pi h)$. The exponent of these twofold fields is $2 - t$. The dominant contribution to the other fields is like $\cos(4\pi h)$ with exponent $2-4t$. These will be called fourfold fields. Less relevant terms like $\cos(8\pi h)$ will be generated under renormalisation but the Potts Hamiltonian contains no adjustable parameters to control their amplitude. For $\tfrac{1}{2} < t < 2$ only the twofold fields are relevant and for $t < \tfrac{1}{2}$ both the twofold and the fourfold fields are relevant. It will become clear that the singly and doubly unstable segments of the gaussian line control critical and tricritical behaviour respectively.

In the pure limit and in the absence of the P_+ field a simple duality on one of the sublattices of the Ising version of the SRM reveals that AT corresponds to the temperature variable of the Ashkin–Teller (1943) model. The exponent of AT has been calculated (Kadanoff and Brown 1979, Knops 1980) as a function of z which parametrises the critical line $x = 1$. This results in an explicit relation between t and z and therefore between t and q:

$$2 - 1/t = 4 \arg(z), \qquad 2 \cos(\pi/2t) = -q^{1/2}, \tag{19}$$

for z on the unit circle, i.e. $q \leqslant 4$.

According to the universality hypothesis the critical exponents of a ferromagnetic Potts model depend on q only. This implies that vacancies and four-spin interaction do not affect the validity of equation (19). In addition we make the usual assumption that the renormalisation group which brings the SRM towards the gaussian model is analytic. Then the initial parameters of the SRM are analytic in the final temperature of the gaussian model. Therefore equation (19) is valid not only when $t > \frac{1}{2}$ but also on the doubly unstable segment $t < \frac{1}{2}$.

Since the parameters in the dilute Potts model all translate into spin-wave operators in the SRM, the exponents must be linear functions of t. Therefore knowing t as a function of q, the value of a Potts exponent at two different values of q suffices to determine its entire q dependence. For the two leading thermal exponents y_1 and y_2 at least two such values are known: the tricritical point of the $q = 1$ ($t = \frac{3}{8}$) model is like the Ising critical point with field (Nienhuis *et al* 1979) with $y_1 = \frac{15}{8}$ and $y_2 = 1$. At $q = 4$ ($t = \frac{1}{2}$) the second exponent must change sign $y_2 = 0$. The critical point of the Ising model ($q = 2$, $t = \frac{2}{3}$) has $y_1 = 1$. Hence the two leading thermal exponents are

$$y_1 = 3 - 3t \tag{20}$$

and

$$y_2 = 4 - 8t \tag{21}$$

where t is given by (19) with $t > \frac{1}{2}$ for the critical and $t < \frac{1}{2}$ for the tricritical point. These formulae have been conjectured partly on the basis of numerical evidence by den Nijs (1979) and Burkhardt (1980). Equation (20) has been calculated analytically for the pure Potts model by Black and Emery (1981) and Knops (1981). Equation (21) contains new crossover and correction-to-scaling exponents. The values of y_1 and y_2 for integer q are listed in table 3 and are discussed in § 7.

Although the above approach results in actual numbers for the critical exponents, it is unsatisfactory in that it does not provide a clue how those exponents come about. Also the derivation of equation (20) and (21) depends on the correct identification of

Table 3. The leading and next-to-leading thermal exponents y_1 and y_2 of the dilute Potts model. The first four columns represent critical exponents and the last four tricritical. The values for $q = 4$ are both critical and tricritical. The exponent y_2 governs corrections to scaling and crossover at the critical and tricritical points respectively.

	Critical					Tricritical			
q	0	1	2	3	4	3	2	1	0
y_1	0	$\frac{3}{4}$	1	$\frac{6}{5}$	$\frac{3}{2}$	$\frac{12}{7}$	$\frac{9}{5}$	$\frac{15}{8}$	2
y_2	-4	-2	$-\frac{4}{3}$	$-\frac{4}{5}$	0	$\frac{4}{7}$	$\frac{4}{5}$	1	$\frac{4}{3}$

exponents at different values of q. The following explanation of the values of y_1 and y_2, even in retrospect, may increase the confidence in their validity.

Both AT and P_+ correspond to the temperature variable of the Potts model (table 2). The most relevant operator conjugate to AT is $\cos(2\pi h)$ and to P_+ $\sin(2\pi h)$. Both have the exponent $2 - t$ which therefore would be the most obvious guess for the Potts exponent y_1. This gives the Ising model an exponent $\frac{4}{3}$ rather than 1, and therefore cannot be correct. The actual behaviour depends on the precise combination of AT and P_+. For $q < 4$ P_+ is imaginary while AT is real. There must be a ratio of P_+ and AT which at large length scales looks just like a pure power $\exp(2\pi ih)$. Since AT and P_+ will renormalise differently, especially initially, this ratio is not known. Black and Emery, however, verified by finite size scaling that in the pure Potts model P_+ and AT come just in that proportion. Therefore a Potts model almost at its critical temperature looks after many renormalisation steps as a gaussian model with a field $\exp(2\pi ih)$ and besides that only four- and more-fold fields. Then the first non-vanishing contribution to the energy–energy correlation function of the Potts model for large separation r is

$$\int \langle \exp(2\pi ih(0) + 2\pi ih(r) - 4\pi ih(r'))\rangle dr', \tag{22}$$

evaluated in the gaussian model. The asymptotic behaviour of this correlation function can be calculated by means of equation (17) and gives for the Potts exponent y_1 equation (20).

In this formulation of the Potts model the conspiracy between AT and P_+ appears very accidental. It is tempting to think that the model can be written such that the ratio between P_+ and AT maintains a special symmetry which is preserved under renormalisation.

Equation (21) for y_2 can be understood by the following reasoning. At the tricritical point of the Potts model the fourfold field is relevant. Therefore the tricritical point can be reached only by adjusting the dilution or the four-spin interaction so that the effective four-fold field h_4 vanishes. Crossover phenomena are governed by the flow of h_4. Consider the flow diagram of h_4 in the neighbourhood of the gaussian line, figure $5(a)$ and compare it with the phenomenological flow diagram for the Potts model as found by the approximate renormalisation group, figure $5(c)$ (Nienhuis *et al* 1979). The topological difference between the two patterns is the existence in figure $5(c)$ of flow lines commencing at the unstable part of the fixed line and terminating at the stable segment. However, when the flow pattern $5(a)$ is plotted in a h_4^2 versus t^{-1} diagram
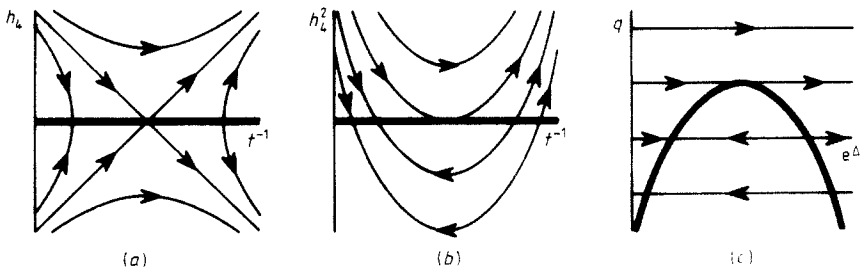

Figure 5. Renormalisation flow diagrams. The standard flow pattern of a fourfold field h_4 is shown as (a), while (b) represents the same figure plotted against h_4^2. The phenomenological flow diagram of a dilute Potts model is shown as (c).

(figure 5(b)) then these flow lines exist in the region of negative h_4^2. Figure 5(b) can be turned into 5(c) by using as the vertical axis a constant of the flow. This picture indicates that the second relevant parameter at the tricritical point of the dilute Potts model is proportional to h_4^2. The exponent associated to that variable is twice the exponent of h_4, which results in equation (21).

Though the exact equations connecting figures 5(a) and (c) cannot be calculated from the mapping between the Potts model and the SRM, some global features can be verified. Variation of h_4^2 at fixed $t = \frac{1}{2}$ in figure 5(b) corresponds roughly to variation of q at fixed dilution in 5(c). Values of $q > 4$ correspond to real h_4 and $q < 4$ to imaginary h_4. The actual value of h_4 is the result of several effectively fourfold fields, only one of which, P_-, may be imaginary. The value of P_- is $\cosh^{-1}(\frac{1}{2}q^{1/2})$ which is real for $q > 4$ and imaginary for $q < 4$.

Another check is the variation of h_4 with dilution for fixed q. The value of P_- is a function of q only, but the amount by which it contributes to the effective fourfold field h_4 varies with dilution. P_- only affects the weights ω_9–ω_{16}, which are most important at intermediate dilution. In the pure limit ω_1–ω_6 dominate and in the extreme dilute limit ω_7 and ω_8. Thus one might expect that increasing dilution corresponds to an increase followed by a decrease of imaginary h_4, which is indeed what comparison of figures 5(b) and (c) indicates.

7. Summary and discussion

We have demonstrated that a Potts model with two- and four-spin interaction and site and bond dilution is equivalent to a roughening model with various symmetry breaking interactions. For convenience the discussion is restricted to a square lattice, but without difficulty it can be generalised to any planar lattice. The sites of the roughening model are those of the original model and its dual lattice combined.

Subject to the assumption that the rough phase of the roughening model flows under renormalisation towards the gaussian model, it is shown that all the exponents of the original Potts model are linear functions of the renormalised temperature in the gaussian model. This renormalised temperature can be related to the intitial Potts Hamiltonian via a known exponent. A few exact solutions at special values of q furnish the information needed to determine the leading and next-to-leading thermal exponents. A quantitative picture is given to explain how the resulting exponents arise from the simple spin-wave operators. It should be stressed however that the arguments leading to the actual value of the exponents do not depend on the validity of this explanation, but only on the existence of an analytic renormalisation group from the roughening to the gaussian model.

Table 3 shows the exponents y_1 and y_2 for integer q. As y_1 was already known at least in the form of a conjecture, y_2 may be of most interest. The value $y_2 = -\frac{4}{3}$ for the Ising correction to scaling exponent is somewhat surprising, as the analyticity structure on the Onsager free energy suggests integer corrections to scaling exponents or none at all. It is of course always possible that the amplitude of the confluent singularity vanishes at the spin-$\frac{1}{2}$ nearest-neighbour Ising model. A variational renormalisation calculation (Nienhuis *et al* 1980, Burkhardt 1980) suggests another solution. At every integer q the exponents can be divided into a physical set, associated with configurations in which q or fewer spin states are represented, and an unphysical set associated with more than two spin states. The free energy is independent of the unphysical scaling

fields. The approximation for $q = 2$ gives physical exponents roughly equal to negative integers and an unphysical exponent of -1.2. This indicates that the exponent $-\frac{4}{3}$ does not show up as a correction to scaling to the free energy of any Ising model, but only to its derivative with respect to q.

The correction to scaling for the three-state Potts critical point may or may not affect the free energy of the hard-hexagon model (Baxter 1980). It does not show up in the order parameter.

The approach in this paper is similar to that of Black and Emery (1981), who calculated the critical exponent, y_1, of the Potts model. The main contribution of this presentation is the calculation of the critical exponent y_2 and of both the tricritical exponents y_1 and y_2. In addition the phenomenological picture of merging critical and tricritical points on which dilution is the marginal operator is confirmed.

The first term of an expansion of y_2 in half-integer powers of $4 - q$ has been correctly predicted by Nauenberg and Scalapino (1980) on the basis of a phenomenological renormalisation group. They also found a logarithmic correction for the four-state Potts model which agrees with the present analysis. However their variables ϕ and ψ contain non-integral powers of the spin-wave fields. This indicates that the renormalisation equations (1) and (2) in the paper of Nauenberg and Scalapino are the first terms in an expansion which may contain fractional powers. Alternatively, one may identify ϕ with the twofold fields and set $b = \frac{1}{4}$. Spin-wave charge conservation then requires that the free energy depends on ϕ only via $\phi^4(\mu^2 + \psi^2)$, which gives the correct exponents by means of a dangerous irrelevant variable.

Acknowledgments

References

Ashkin J and Teller E 1943 *Phys. Rev.* **64** 178–84
Baxter R J 1973 *J. Phys. C: Solid State Phys.* **6** 1445–8
—— 1980 *J. Phys. A: Math. Gen.* **13** L61–70
Baxter R J, Kelland S B and Wu F Y 1976 *J. Phys. A: Math. Gen.* **9** 397–406
Burkhardt T W 1980 *Z. Phys.* B **39** 159–84
Black J L and Emery V J 1981 *Phys. Rev.* B **23** 429–32
Dasgupta C 1977 *Phys. Rev.* B **15** 3460–4
José J V, Kadanoff L P, Kirkpatrick S and Nelson R D 1977 *Phys. Rev.* B **16** 1217–41
Kadanoff L P and Brown A C 1979 *Ann. Phys.* **121** 318–42
Kim D and Joseph R I 1975 *J. Phys. A: Math. Gen.* **8** 891–902
Knops H J F 1980 *Ann. Phys.* **128** 448–62
—— 1981 *Preprint Nijmegen*
Kondor I 1980 *J. Phys. C: Solid State Phys.* **13** L531–4
Nauenberg M and Scalapino D J 1980 *Phys. Rev. Lett.* **44** 837–40

Nienhuis B, Berker A N, Riedel E K and Schick M 1979 *Phys. Rev. Lett.* **43** 737–40
Nienhuis B, Riedel E K and Schick M 1980 *J. Phys. A: Math. Gen.* **13** L31–4, L189–92
den Nijs M 1978 *Physica* **95A** 449–72
—— 1979 *J. Phys. A: Math. Gen.* **12** 1857–68
Potts R B 1952 *Proc. Camb. Phil. Soc.* **48** 106–9
Rebbi C and Swendsen R H 1980 *Phys. Rev.* B **21** 4094–107
Temperley H N V and Lieb E H 1971 *Proc. R. Soc.* A **322** 251–80
Van Beijeren H 1977 *Phys. Rev. Lett.* **38** 993–5
Wu F Y 1981 *J. Phys. A: Math. Gen.* **14** L39–44