

Algèbres, Intégrabilité et Modèles Exactement Solubles

Jesper Lykke Jacobsen et Yacine Ikhlef

Mars 2013

1 Avant-propos

Ce cours est une introduction au monde merveilleux des modèles exactement solubles en deux dimensions. Il est destiné avant tout aux étudiants du niveau M2, mais peut sans doute avoir une certaine utilité un peu avant ou après.

Une solution exacte d'un problème non trivial en physique théorique a une importance considérable. En opposition avec une méthode approximative ou un calcul numérique, un résultat exact est pérenne et ne saura être mis en question ultérieurement. Son auteur entrera dans les annales et y restera pour toujours. Très souvent, une solution exacte est rendue possible par une avancée technique qui permettra d'ouvrir tout un nouveau domaine de recherche.

Les solutions présentées dans ce cours sont motivées par mes propres intérêts en physique statistique, mais très souvent des modèles similaires ou équivalents sont étudiés dans des contextes très différents : matière condensée, calcul quantique, physique de hautes énergies... ou sous un angle mathématique: topologie, probabilités, théorie des graphes... On se rappelle qu'un modèle 2D classique est toujours équivalent à un modèle 1D quantique (chaîne de spins) et nous allons souvent exploiter ce dernier point de vue.

Quant au contenu, on commence par des solutions de nature combinatoire (pavages de dimères et modèle d'Ising). Ces solutions permettront d'établir un étalon pour ce qu'on veut dire par un résultat exact. Par ailleurs, elles mettent progressivement en évidence les symétries et le contenu algébrique qui rendent une solution exacte possible. Nous allons ensuite exposer deux

techniques importantes pour la recherche et la classification de solutions exactes en deux dimensions : l'intégrabilité (aussi connue sous le nom de diffusion quantique inverse) et les théories conformes des champs. Ces techniques ont des points communs importants que nous essayerons de mettre en évidence. En cours de route, nous rencontrerons d'autres modèles qui servent à diriger la discussion et les applications vers des questions qui sont d'actualité dans la recherche contemporaine.

Les notes de cours sont rédigées en anglais afin de mieux permettre aux étudiants de s'habituer à la lecture d'articles de recherche.

Des calculs intermédiaires ou facultatifs (souvent propices pour des exercices de TD) apparaissent sur fond gris.

2 Dimer coverings

The dimer problem occurs experimentally when a diatomic gas is adsorbed onto a crystalline substrate. Given some lattice, we ask for the number of ways that its vertices can be completely covered by non-overlapping *dimers* that each occupy two neighbouring vertices. More generally, we can assign a fugacity to each type of dimers, for instance depending on its orientation on the lattice (or maybe even on its exact position), and the goal is to compute the corresponding partition function.

Dimer are also known in mathematics as dominos. When the lattice is bipartite,¹ dimer coverings are known in graph theory as perfect matchings.

The dimer problem was solved in 1961, almost simultaneously and independently by Kasteleyn [Ka61], Fisher [Fi61] and Temperley [TF61]. In all cases it was crucial to realise that the partition function can be conveniently expressed as a Pfaffian (whose square is an ordinary determinant). The crux of the problem was to get all configurations counted with the same sign.

The detailed solution for the square lattice was presented in [Ka61, Fi61]. The approach of [Ka61] to deal with the sign problem appears to generalise most easily to other lattices, and we shall follow this method with subsequent simplifications [Ka63]. Several fine points were further discussed in chapter 4 of the book by McCoy and Wu [MW73]. Certain correlation functions were obtained by Fisher and Stephenson [FS63] as Toeplitz determinants.

Results on other lattices, scattered throughout the literature, have been reviewed in [Wu06]. Recent generalisations of the problem include quantum dimer models [RK88] with applications to superconductivity, aligning interactions [Al05], and the inclusion of a single monomer on the boundary. The solution with an single monomer in an arbitrary position is a difficult open problem. Including several monomers appears to be intractable, although the essential physics is known in the continuum limit [Al06].

2.1 Determinant formulation

Let us first define the problem. We consider a graph $G = (V, E)$ where V is the set of vertices and E is the set of edges. A dimer configuration \mathcal{C} on G is a subset of edges which covers all the vertices, and where no overlap

¹By definition a bipartite lattice is one for which the set of vertices V can be written as a disjoint union $V = A \cup B$, so that any edge connects an A -vertex with a B -vertex.

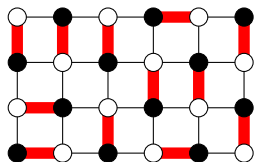


Figure 1: Dimer covering of a portion of square lattice.

occurs between two edges. An example is shown in Fig. 1. Each edge e of the graph carries a positive weight $\pi(e)$, and to each configuration \mathcal{C} , we assign a Boltzmann weight

$$\pi(\mathcal{C}) := \prod_{e \in \mathcal{C}} \pi(e).$$

The partition function is defined as

$$Z_G := \sum_{\mathcal{C}} \pi(\mathcal{C}) = \sum_{\mathcal{C}} \prod_{e \in \mathcal{C}} \pi(e), \quad (2.1)$$

where the sum runs over all possible dimer configurations on G . In particular, if we set $\pi(e) = 1$ for every edge, Z_G is the number of dimer configurations on G .

We now restrict the discussion to the case when G is bipartite. This means that the vertices of G can be coloured in black and white, in such a way that every white vertex is only adjacent to black vertices, and vice-versa. Examples of bipartite graphs are the square lattice and the honeycomb lattice. The triangular lattice is *not* bipartite. We denote by (w_1, \dots, w_N) the white vertices, and (b_1, \dots, b_N) the black ones. Also, we choose an orientation of the edges of G , and we introduce the $N \times N$ weighted adjacency matrix K , defined by:

$$K_{ij} = \begin{cases} \pi(w_i, b_j) & \text{if } w_i \rightarrow b_j \\ -\pi(w_i, b_j) & \text{if } w_i \leftarrow b_j \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

Consider the determinant of this matrix:

$$\det K = \sum_{\sigma \in \mathcal{S}_N} \text{sgn}(\sigma) K_{1, \sigma(1)} \dots K_{N, \sigma(N)},$$

where \mathcal{S}_N is the group of permutations of N elements. The permutations σ which contribute to $\det K$ are those which satisfy the condition

$$\forall i \in \{1, \dots, N\}, \quad w_i \text{ is adjacent to } b_{\sigma(i)}.$$

This is equivalent to saying that the set of edges

$$\mathcal{C}(\sigma) := \{(w_1, b_{\sigma(1)}), \dots, (w_N, b_{\sigma(N)})\}$$

is a dimer configuration! Also, the contribution of σ to $\det K$ equals the weight $\pi[\mathcal{C}(\sigma)]$, up to a sign. We will now show that, if the orientation of G is well chosen, every contribution to $\det K$ picks the same sign, and we have

$$Z_G = |\det K|. \tag{2.3}$$

To show this, we take two arbitrary dimer configurations \mathcal{C} and \mathcal{C}' , corresponding to the permutations σ and σ' , and draw their superposition, as in Fig. 2. The resulting graph is made of doubly-covered edges, together with closed cycles of even length, and is called the transition graph. For simplicity, let us assume first that there is only one closed cycle of length 2ℓ . If we number the vertices around this cycle $(w_1, b_1, w_2, b_2, \dots, w_\ell, b_\ell)$, it is easy to see that the permutations σ and σ' differ by a cyclic permutation:

$$\sigma' = \begin{pmatrix} 1 & 2 & 3 & \dots & \ell \\ \ell & 1 & 2 & \dots & \ell - 1 \end{pmatrix} \circ \sigma,$$

and hence $\text{sgn } \sigma' = (-1)^{\ell+1} \text{sgn } \sigma$. In order to compensate the sign $(-1)^{\ell+1}$ with those coming from (2.2), we impose the condition:

In any cycle of length 2ℓ , $\begin{cases} \# \text{ edges oriented } b \rightarrow w \\ (\ell + 1) \end{cases}$ have the same parity.

An orientation satisfying this condition is called a Kasteleyn orientation. Actually, it is sufficient to impose the above condition on the elementary cycles, *i.e.* cycles which enclose a single face of G . In the determinant of the corresponding matrix K , every contribution has the same sign, and (2.3) follows.

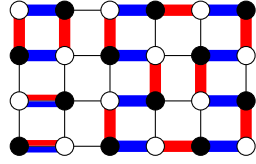


Figure 2: Superposition of two dimer configurations.

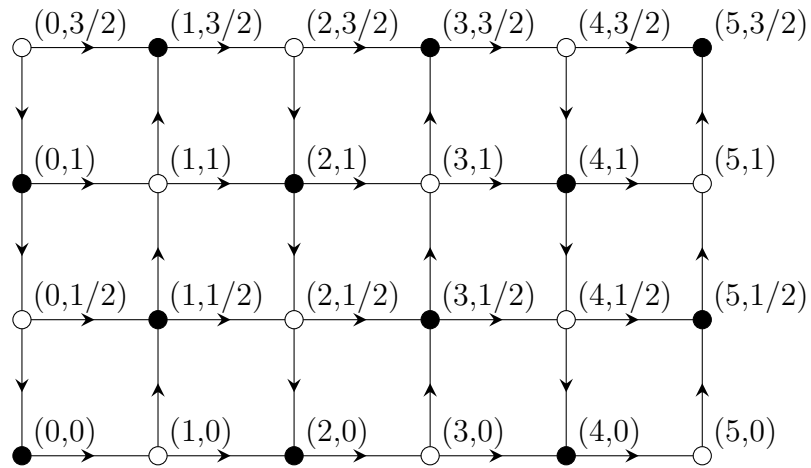


Figure 3: A Kasteleyn orientation of the square lattice. The coordinates of black and white vertices are indicated.

2.2 Explicit computation on the square lattice

We consider a portion of the square lattice G_{mn} of size $m \times n$. For simplicity, we restrict to even n and m . Once a Kasteleyn orientation is found, we can use (2.3) to compute the partition function of dimers on the square lattice. We shall use the orientation shown in Fig. 3, and give a weight $\pi(e) = z_1$ to every horizontal edge e and $\pi(e) = z_2$ to every vertical edge e . The resulting adjacency matrix is

$$K_{(x,y),(x',y')} = z_1(\delta_{x+1,x'} - \delta_{x-1,x'})\delta_{yy'} + z_2\delta_{xx'}(-1)^x(\delta_{y+1/2,y'} - \delta_{y-1/2,y'}),$$

where (x, y) are the coordinates of a white vertex, and (x', y') are the coordinates of a black vertex. Let us rewrite K as

$$K_{(x,y),(x',y')} = (-1)^x(z_1 U_{xx'} \delta_{yy'} + z_2 \delta_{xx'} V_{yy'}),$$

with

$$U_{xx'} := (-1)^x(\delta_{x+1,x'} - \delta_{x-1,x'}), \quad V_{yy'} := (\delta_{y+1/2,y'} - \delta_{y-1/2,y'}).$$

Note that the two terms in the bracket correspond to two commuting operators, and hence we can diagonalise them simultaneously. One can show that the eigenvalues λ_k (resp. μ_ℓ) of U (resp. V) are

$$\begin{aligned} \lambda_k &= 2 \cos \frac{\pi k}{m+1}, & k &\in \{1, 2, \dots, m\}, \\ \mu_\ell &= 2i \cos \frac{\pi \ell}{n+1}, & \ell &\in \{1, 2, \dots, n/2\}. \end{aligned}$$

Using the identity $\lambda_{m+1-k} = -\lambda_k$, we have

$$Z_{mn}(z_1, z_2) = |\det K| = 2^{mn/2} \prod_{k=1}^{m/2} \prod_{\ell=1}^{n/2} \left[z_1^2 \cos^2 \left(\frac{\pi k}{m+1} \right) + z_2^2 \cos^2 \left(\frac{\pi \ell}{n+1} \right) \right]. \quad (2.4)$$

In particular we can find the number of ways to tile a chessboard by 32 dominos [TF61]:

$$12\,988\,816 = 2^4 \times 17^2 \times 53^2. \quad (2.5)$$

2.3 Thermodynamical limit

The above combinatorial derivation has produced an expression (2.4) for the partition function which is explicit and exact in finite size. This is a rather *unusual* situation in statistical physics, where more often than not one can obtain exact results only in the thermodynamical limit $m, n \rightarrow \infty$.

In that limit one is typically interested in the free energy per site

$$f(z_1, z_2) = \lim_{m, n \rightarrow \infty} \frac{1}{mn} \log Z_{mn}(z_1, z_2) \quad (2.6)$$

for which one expects a finite limit. The result (2.4) needs some manipulation in order to extract an analytical expression for f .

Replacing first

$$\frac{1}{m} \sum_{k=1}^{m/2} \rightarrow \frac{1}{\pi} \int_0^{\pi/2} d\omega,$$

we obtain

$$f(z_1, z_2) = \frac{1}{\pi^2} \int_0^{\pi/2} d\omega \int_0^{\pi/2} d\omega' \log [4(z_1^2 \cos^2 \omega + z_2^2 \cos^2 \omega')] . \quad (2.7)$$

This expression can be simplified by using a few tricks of analysis.

To this end we first perform explicitly the integral over ω . Defining the ratio $\zeta = \frac{z_2}{z_1}$ we have

$$\begin{aligned} & \frac{1}{\pi} \int_0^{\pi/2} d\omega \log [4(z_1^2 \cos^2 \omega + z_2^2 \cos^2 \omega')] \\ &= \frac{1}{2} \log(2z_2 \cos \omega')^2 + \frac{1}{\pi} \int_0^{\pi/2} d\omega \log \left(1 + \frac{\cos^2 \omega}{\zeta^2 \cos^2 \omega'} \right) \\ &= \log z_1 + \log(2\zeta \cos \omega') + \log \left(\frac{1 + \sqrt{1 + \frac{1}{\zeta^2 \cos^2 \omega'}}}{2} \right) \\ &= \log z_1 + \log \left(\zeta \cos \omega' + \sqrt{1 + \zeta^2 \cos^2 \omega'} \right) . \end{aligned}$$

Renaming $\omega' \mapsto \omega$, (2.7) then becomes

$$f(z_1, z_2) = \frac{1}{2} \log z_1 + \frac{1}{\pi} \int_0^{\pi/2} d\omega g(\zeta \cos \omega) \quad (2.8)$$

where we have defined

$$g(x) = \log(x + \sqrt{1 + x^2}).$$

We can suppose that $|z_2| \leq |z_1|$, since if this were not the case we could simply exchange z_1 and z_2 . Setting $x = \zeta \cos \omega$ we have then $|x| < 1$ (except for $\omega = 0$, which does not matter under the integral). Therefore the integrand $g(x)$ can be replaced by its expansion as an entire series and we can integrate term by term. After a little work one finds

$$g(x) = \sum_{j=0}^{\infty} \binom{2j}{j} \frac{(-1)^j}{(2j+1)2^{2j}} x^{2j+1}.$$

Using now

$$\int_0^{\pi/2} d\omega (\zeta \cos \omega)^{2j+1} = \zeta^{2j+1} \frac{\sqrt{\pi} \Gamma(j+1)}{2\Gamma(j+\frac{3}{2})} = \zeta^{2j+1} \frac{j! 2^j}{(2j+1)!!}$$

we arrive at

$$\begin{aligned} \int_0^{\pi/2} d\omega g(\zeta \cos \omega) &= \sum_{j=0}^{\infty} \binom{2j}{j} \frac{(-1)^j}{(2j+1)2^{2j}} \frac{j! 2^j}{(2j+1)!!} \zeta^{2j+1} \\ &= \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)^2} \zeta^{2j+1}. \end{aligned}$$

This expression is reminiscent of $\arctan \zeta$. Indeed we have

$$\arctan x = \sum_{j=0}^{\infty} \frac{(-1)^j}{2j+1} x^{2j+1}$$

and so

$$\int_0^{\zeta} dx \frac{\arctan x}{x} = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)^2} \zeta^{2j+1}.$$

Putting the pieces together we obtain

$$f(z_1, z_2) = \frac{1}{2} \log z_1 + \frac{1}{\pi} \int_0^{\zeta} dx \frac{\arctan x}{x}. \quad (2.9)$$

To simplify further we can use that

$$\arctan x = \frac{1}{2i} \log \left(\frac{1+ix}{1-ix} \right),$$

as is easily seen by taking tan on both sides. The integral in (2.9) can therefore be expressed in terms of the Euler dilogarithm²

$$\text{Li}_2(u) = - \int_0^u dx \frac{\log(1-x)}{x} = \sum_{k=1}^{\infty} \frac{u^k}{k^2}$$

as follows

$$\int_0^\zeta dx \frac{\arctan x}{x} = \frac{1}{2i} (\text{Li}_2(i\zeta) - \text{Li}_2(-i\zeta)) \equiv \text{Ti}_2(\zeta), \quad (2.10)$$

where we have finally introduced the inverse tangent integral $\text{Ti}_2(\zeta)$.

The final result thus reads

$$f(z_1, z_2) = \frac{1}{2} \log z_1 + \frac{1}{\pi} \text{Ti}_2(\zeta), \quad \zeta = \frac{z_2}{z_1}. \quad (2.11)$$

For the combinatorial counting problem $\zeta = 1$. Recalling the series expansion (2.10) of $\text{Li}_2(u)$ we obtain

$$f(1, 1) = \frac{1}{\pi} \text{Ti}_2(1) = \frac{1}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} \equiv \frac{G}{\pi}, \quad (2.12)$$

where we have introduced the Catalan constant G . The number of arrangements per dimer (sometimes known in the theoretical chemistry literature as the *molecular freedom*) is then

$$e^{2f(1,1)} = \exp \left(\frac{2G}{\pi} \right) \simeq 1.791\,622\dots \quad (2.13)$$

This is of course smaller than 2, since the effective number of arrangements of a given dimer is constrained by the other dimers.

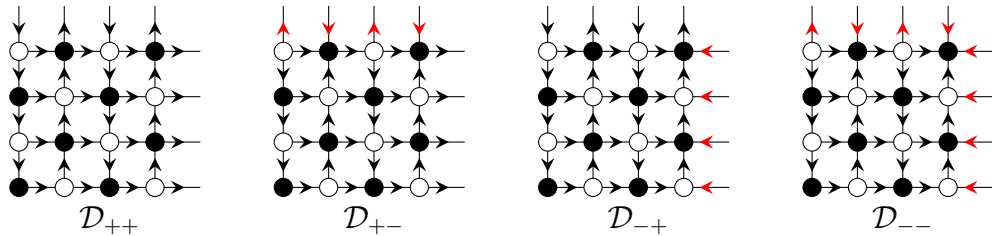


Figure 4: Four different orientations of the square lattice with toroidal boundary conditions.

2.4 Toroidal boundary conditions

Consider now a square lattice \tilde{G}_{mn} embedded in the torus. The results of Sec. 2.1 do not apply directly, since the lattice is no longer planar. In particular, one cannot find a Kasteleyn orientation for this lattice. It is however still possible to express the partition function (2.1) as a sum of *four* different determinants.

Consider the four different orientations of \tilde{G}_{mn} shown in Fig. 4. The first orientation \mathcal{D}_{++} is just the Kasteleyn orientation of G_{mn} endowed with periodic boundary conditions (The other three orientations are obtained by applying antiperiodic boundary conditions in one or both directions).

Pick an arbitrary dimer configuration \mathcal{C}_0 , say the blue dimers in Fig. 2, and construct the transition graph between \mathcal{C}_0 and the other dimer configurations. Orientation \mathcal{D}_{++} gives the correct parity to all transition cycles that are homotopic to a point. However, an incorrect sign is given to some of the dimer configurations for which the transition cycles have non-trivial homotopy. Note that since different transition cycles cannot intersect, all non-trivial transition cycles have in fact the same homotopy, *i.e.*, they all wrap the horizontal and vertical directions the same number of times.

Let us divide the possible dimer configurations on \tilde{G}_{mn} into four disjoint classes. Class (e,e) comprises dimer configuration for which the set of transition cycles wrap both the horizontal and vertical directions an even number of times. Similarly we define classes (o,e), (e,o) and (o,o), where e = even and o = odd, and the first (resp. second) symbol refers to the horizontal (resp. vertical) direction. We denote by $Z_{ee}, Z_{oe}, Z_{eo}, Z_{oo}$ the corresponding contributions to the partition function. By examining one example per case and arguing that local deformations of the transition cycles do not alter the

results, one establishes that the signs with which the four different classes of dimer configurations are counted in the four different determinants are as follows:

$$\begin{aligned}\det \mathcal{D}_{++} &= +Z_{ee} - Z_{oe} - Z_{eo} - Z_{oo}, \\ \det \mathcal{D}_{+-} &= +Z_{ee} - Z_{oe} + Z_{eo} + Z_{oo}, \\ \det \mathcal{D}_{-+} &= +Z_{ee} + Z_{oe} - Z_{eo} + Z_{oo}, \\ \det \mathcal{D}_{--} &= +Z_{ee} + Z_{oe} + Z_{eo} - Z_{oo}.\end{aligned}$$

It follows that

$$Z_{mn}(z_1, z_2) = \frac{1}{2} (-\det \mathcal{D}_{++} + \det \mathcal{D}_{+-} + \det \mathcal{D}_{-+} + \det \mathcal{D}_{--}). \quad (2.14)$$

The four determinants in (2.14) can be computed explicitly as before, except the set of eigenvalues of U and V are slightly different: they are typically of the form $\lambda_k = 2 \cos \frac{2\pi k}{m}$. Explicit results are given in [Ka61].

2.5 Height mapping

To any dimer covering we can associate a height mapping on the dual lattice which is defined as follows. First note that the square lattice is bipartite. When encircling a white vertex in the positive (counterclockwise) direction, the height h changes by $+1/4$ upon crossing an empty edge and by $-3/4$ upon crossing an edge that is covered by a dimer. The same rule holds when encircling a black vertex in the negative direction. By fixing the height at the origin, e.g., $h(0,0) = 0$, these rules define the entire height function $h(x,y)$ uniquely. An example is shown in Fig. 5.

The height mapping provides a link with the conformal field theory (CFT) description of dimer coverings. We briefly outline some of the essential elements. In the continuum limit the height function becomes a free bosonic field with effective action

$$S_{\text{eff}} = \pi g \int dx dy \left\{ \left(\frac{\partial h(x,y)}{\partial x} \right)^2 + \left(\frac{\partial h(x,y)}{\partial y} \right)^2 \right\}. \quad (2.15)$$

Here g is the coupling constant, which controls the stiffness of the interface model. It is a priori unknown.

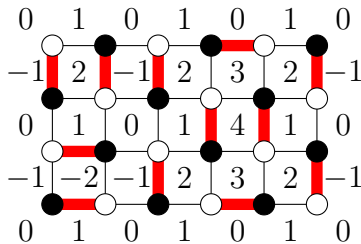


Figure 5: Dimer covering of the lattice $G_{6,4}$ and the corresponding height function $4h(x, y)$ defined on the dual lattice.

Note that the standard configuration of Fig. 2 corresponds to a flat height interface. The central argument behind the “derivation” of (2.15) is that the dimer configurations represent bounded fluctuations around this flat state.

The two-point correlation functions decay asymptotically—i.e., for distances r satisfying $1 \ll r \ll N$ on an $N \times N$ lattice, and with operator positions far from the boundaries—as $r^{-2X(e,m)}$, where the critical exponent

$$X(e, m) = \frac{e^2}{2g} + \frac{g}{2}m^2. \quad (2.16)$$

Here e and m are integer valued electric and magnetic charges. Electric charges e correspond to vertex operators $V_e(\mathbf{r}) =: \exp[2\pi i e h(\mathbf{r})]$: appearing in the Fourier expansion of any operator periodic in the coarse-grained height field. Dual magnetic charges correspond to a dislocation of m in the height field.

For example, two monomer defects on opposite sublattices correspond to $m = \pm 1$. It is known from exact results [FS63] that $X(0, 1) = \frac{1}{4}$, and this fixes $g = \frac{1}{2}$ from (2.16). The exponents for correlation function of all possible charges then follow from (2.16). In particular, the dimer-dimer correlation function—i.e., the probability that two widely separated dimers have the same orientation, after subtraction of the trivial $r \rightarrow \infty$ limit of $\frac{1}{2}$ —then decays with exponent $X(1, 0) = 1$, and this is confirmed by the exact solution [FS63].

Let us also note that in CFT there is a link between (2.14) and modular invariant partition functions for the free boson on the torus.

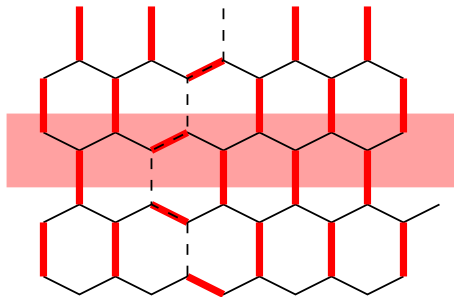


Figure 6: Dimers on the honeycomb lattice. The shaded region corresponds to the action of the row-to-row transfer matrix T . The dotted line represents the trajectory of a particle.

2.6 Transfer matrix formulation

In this section, we consider the dimer problem on the honeycomb lattice, with edge weights z_1, z_2, z_3 on the three different types of edges, as shown in Fig. 6. Moreover, we assume periodic boundary conditions in the horizontal direction, and we denote by L the number of vertical edges (type z_3) in any row.

The state of a row of vertical edges is described by the sequence $\alpha = (\alpha_1, \dots, \alpha_L)$, where $\alpha_j = 1$ (resp. $\alpha_j = 0$) if the edge j is occupied by a dimer (resp. empty). Between two rows of type α , the intermediary state is described by $\beta = (\beta_1, \dots, \beta_{2L})$, where β_j is defined similarly in terms of occupied/empty edges.

The row-to-row transfer matrix T acts on an α state, and adds a row to the system. More precisely, the matrix element $T_{\alpha, \alpha'}$ is defined as the Boltzmann weight of all possible intermediary states from α' to α :

$$T_{\alpha, \alpha'} = \sum_{\beta | \alpha, \alpha'} \prod_{j=1}^L z_1^{\beta_{2j-1}} z_2^{\beta_{2j}} z_3^{(\alpha_j + \alpha'_j)/2},$$

where the sum is over the possible intermediary states between configurations α' and α . In our conventions, T acts from bottom to top, so α is placed above α' .

2.6.1 Solution by free fermions

Let us now reformulate the problem in terms of a closed system of particles evolving in the vertical direction. Consider the reference configuration \mathcal{C}_0 where all vertical edges are occupied by a dimer, and define the particles associated to any dimer configuration \mathcal{C} as follows: if the dimer occupation of an edge e is the same in \mathcal{C} as in \mathcal{C}_0 , we say it carries no particle, whereas if the dimer occupations of e are different in \mathcal{C} and \mathcal{C}_0 , we say it carries a particle. It is easy to see that the matrix element $T_{\alpha,\alpha'}$ vanishes, unless α and α' have the same number of particles: this means that the number of particles is conserved by the action of T . Thus, we can look for the eigenvectors and eigenvalues of T in sectors of fixed number of particles n . A left eigenvector ψ of T is defined by the condition

$$\sum_{\alpha} \psi_{\alpha} T_{\alpha,\alpha'} = \Lambda \psi_{\alpha'}. \quad (2.17)$$

For $n = 0$, there is only one vacuum state, with eigenvalue z_3^L for T . We shall use this to normalise our transfer matrix, and set $\tau := z_3^{-L} T$.

For $n = 1$, the states are specified by the position x of a single particle. Notice that $x \in \{1, 2, \dots, L\}$ for even rows, and $x \in \{1/2, 3/2, \dots, L - 1/2\}$ for odd rows. We denote by $\tau^{(1)}$ the transfer matrix from even to odd rows, and $\tilde{\tau}^{(1)}$ the transfer matrix from odd to even rows. The matrix elements of τ read, for $(x, y) \in (\mathbb{N} + 1/2, \mathbb{N})$:

$$\tau_{xy}^{(1)} = (z_1/z_3) \delta_{x,y-1/2} + (z_2/z_3) \delta_{x,y+1/2},$$

and $\tilde{\tau}_{xy}$ has the same expression, but with $(x, y) \in (\mathbb{N}, \mathbb{N} + 1/2)$. We wish to find the left eigenvectors of τ and $\tilde{\tau}$. The eigenvalue equations read

$$\sum_{x=1}^L \psi(x - 1/2) \tau_{x-1/2,y}^{(1)} = \Lambda \psi(y), \quad (2.18)$$

$$\sum_{x=1}^L \psi(x) \tilde{\tau}_{x,y-1/2}^{(1)} = \Lambda \psi(y - 1/2), \quad (2.19)$$

where we have identified the points $(L+1/2)$ and $1/2$. for any $y \in \{1, \dots, L\}$. Let us focus on (2.18) first. Since $\tau^{(1)}$ is invariant by cyclic translations, ψ must have the form of a plane wave:

$$\psi_k(x) := \exp(ikx), \quad \text{with} \quad -\pi < k \leq \pi.$$

The corresponding eigenvalue is

$$\Lambda_k = (z_1/z_3)e^{-ik/2} + (z_2/z_3)e^{+ik/2}.$$

Moreover, periodic boundary conditions impose that $e^{ikL} = 1$. It is easy to see that ψ_k is also an eigenvector of $\tilde{\tau}^{(1)}$, with the same eigenvalue.

For $n = 2$, particle states are labelled by two positions (x_1, x_2) , with $x_1 < x_2$. The matrix element takes into account the avoiding constraint between particles:

$$\tau_{(x_1, x_2), (y_1, y_2)}^{(2)} = \tau_{x_1, y_1}^{(1)} \tau_{x_2, y_2}^{(1)} - (z_1 z_2 / z_3^2) \delta_{x_1 x_2} \delta_{x_1, y_1 + 1/2} \delta_{x_2, y_2 - 1/2}.$$

If we set $\psi_{12}(x_1, x_2) = \psi_{k_1}(x_1)\psi_{k_2}(x_2)$, the left-hand side of (2.17) reads

$$\begin{aligned} \sum_{x_1, x_2} \psi_{12}(x_1, x_2) \tau_{(x_1, x_2), (y_1, y_2)}^{(2)} &= \Lambda_{k_1} \Lambda_{k_2} \psi_{12}(y_1, y_2) \\ &\quad - (z_1 z_2 / z_3^2) \delta_{y_1 + 1, y_2} \psi_{k_1}(y_1 + 1/2) \psi_{k_2}(y_1 + 1/2). \end{aligned}$$

Similarly, for $\psi_{21}(x_1, x_2) = \psi_{k_2}(x_1)\psi_{k_1}(x_2)$, we get

$$\begin{aligned} \sum_{x_1, x_2} \psi_{21}(x_1, x_2) \tau_{(x_1, x_2), (y_1, y_2)}^{(2)} &= \Lambda_{k_2} \Lambda_{k_1} \psi_{21}(y_1, y_2) \\ &\quad - (z_1 z_2 / z_3^2) \delta_{y_1 + 1, y_2} \psi_{k_2}(y_1 + 1/2) \psi_{k_1}(y_1 + 1/2). \end{aligned}$$

We can thus simply combine ψ_{12} and ψ_{21} to eliminate the $\delta_{y_1 + 1, y_2}$ terms. We define

$$\psi(x_1, x_2) := \psi_{k_1}(x_1)\psi_{k_2}(x_2) - \psi_{k_2}(x_1)\psi_{k_1}(x_2), \quad (2.20)$$

and we have

$$\sum_{x_1, x_2} \psi(x_1, x_2) \tau_{(x_1, x_2), (y_1, y_2)}^{(2)} = \Lambda_{k_1} \Lambda_{k_2} \psi(y_1, y_2), \quad (2.21)$$

which is the eigenvalue equation in the two-particle sector. We recognise that (2.20) is a fermionic two-body wave function. When a particle goes around the system, the wavefunction picks a factor (-1) , and hence the momenta satisfy:

$$e^{iLk_1} = e^{iLk_2} = -1,$$

with the additional constraint that $k_1 \neq k_2$. This suggests that the particles behave like free fermions.

For a general value of n , states are specified by the sequence $x_1 < \dots < x_n$ of particle positions. One can extend the previous discussion, and (2.20) is replaced by

$$\psi(x_1, \dots, x_n) = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \exp[ik_{\sigma(1)}x_1 + \dots + ik_{\sigma(n)}x_n], \quad (2.22)$$

where \mathcal{S}_n is the set of permutation of n elements, and the momenta k_1, \dots, k_n take n distinct values, subject to the conditions:

$$\exp(iLk_j) = (-1)^{n+1}. \quad (2.23)$$

Again, (2.22) is a fermionic wavefunction. The corresponding eigenvalue of τ is

$$\Lambda = \Lambda_{k_1} \dots \Lambda_{k_n}, \quad \text{where} \quad \Lambda_k = (z_1/z_3)e^{-ik/2} + (z_2/z_3)e^{+ik/2}. \quad (2.24)$$

It is important to work we *left* eigenvectors, for the following reason. In the $n = 2$ sector, if we were dealing with the right eigenvector with $\sum_{\alpha'} \tau_{\alpha\alpha'} \psi(\alpha') = \Lambda \psi(\alpha)$, we would get unwanted terms of the form $\delta_{y_1+1, y_2} \psi_{k_1}(y_1) \psi_{k_2}(y_1 + 1)$. Hence, the right eigenvector is given by $\psi = e^{-ik_2} \psi_{12} - e^{-ik_1} \psi_{21}$, which is not antisymmetric in the positions like (2.20). Of course, the eigenvalues are the same, but the relation to a free Fermi system is less obvious.

2.6.2 Dominant eigenvalues

We first restrict to the isotropic case $z_1 = z_2 = z_3$. The one-particle eigenvalues then read $\Lambda_k = 2 \cos(k/2)$. The maximal eigenvalue (2.24) is obtained when we choose the values of k for which $|\Lambda_k| > 1$. This corresponds to a Fermi sea in the interval $|k| < 2\pi/3$.

Let us show how to compute the dominant eigenvalue Λ_{\max} of T as an expansion in $1/L$. For simplicity, we consider the case when L is a multiple of 3. Then Λ_{\max} is obtained by taking $n = 2L/3$ particles, with momenta $k_j = 2\pi q_j/L$, and

$$\{q_1, \dots, q_n\} = -\frac{n-1}{2}, -\frac{n-3}{2}, \dots, \frac{n-1}{2}. \quad (2.25)$$

To evaluate (2.24), we can use the Euler-McLaurin formula

$$h \times \left[\frac{f(a) + f(b)}{2} + \sum_{j=1}^{N-1} f(a + jh) \right] \simeq \int_a^b f(k) dk + \frac{h^2}{12} [f'(b) - f'(a)], \quad (2.26)$$

where $h := (b - a)/N$. This yields the asymptotic expansion:

$$\frac{1}{L(\sqrt{3}/2)} \log \Lambda_{\max} \simeq \frac{1}{\pi\sqrt{3}} \int_{-2\pi/3}^{+2\pi/3} \log[2 \cos(k/2)] dk + \frac{\pi}{6L^2}. \quad (2.27)$$

We have normalised the left-hand side by the aspect ratio of one row. On the right-hand side, the first term is the free energy density f for the honeycomb lattice. The second term is a universal correction, which is predicted by Conformal Field Theory (this will be explained in future lectures).

Proof of (2.27).

In (2.26), we set $f(k) := \log[2 \cos(k/2)]$, $b = -a := (n - 1)2\pi/L$ and $N = n - 1$. We have immediately $f(\pm 2\pi/3) = 0$. Then, we approximate $f(a) \simeq f'(2\pi/3)h/2$, $f'(a) \simeq f'(2\pi/3)$ and similarly for $f(b)$ and $f'(b)$. Moreover, we write

$$\int_a^b f(k) dk \simeq \int_{-2\pi/3}^{2\pi/3} f(k) dk + \frac{h^2}{8} [f'(2\pi/3) - f'(-2\pi/3)].$$

Inserting these approximations into (2.26), we get the above result.

We can repeat the calculation for the subdominant eigenvalue Λ_1 , corresponding to the momentum distribution (2.25), but with one particle removed: $n = 2L/3 - 1$. In terms of dimers, when superposing configurations with $n = 2L/3$ and $n = 2L/3 - 1$, we get one transition line propagating in the transfer direction, which is equivalent to a shift of the dimers by one site along this line. Therefore, we can identify Λ_1 to the insertion of one monomer at each end of the transition line. The asymptotic expansion for Λ_1 is

$$\frac{1}{L(\sqrt{3}/2)} \log \Lambda_1 \simeq \frac{1}{\pi\sqrt{3}} \int_{-2\pi/3}^{+2\pi/3} \log[2 \cos(k/2)] dk - \frac{\pi}{3L^2}.$$

The corresponding scaling dimension X_{mon} can be extracted from the CFT prediction

$$\frac{1}{L(\sqrt{3}/2)} \log \frac{\Lambda_1}{\Lambda_{\text{max}}} \simeq -\frac{2\pi X_{\text{mon}}}{L^2},$$

and so we find $X_{\text{mon}} = 1/4$.

Exercise: use the Euler-McLaurin formula (2.26) to compute the critical exponent corresponding to the sector $n = 2L/3 - m$, where m is a finite positive integer. Introduce a second integer e , and compute the exponent corresponding to the shift $q_j \rightarrow q_j + e$ in (2.25). Compare with (2.16).

Finally, suppose we take anisotropic Boltzmann weights, of the form

$$z_1 = z_2 = w^{1/2}, \quad z_3 = w^{-1/2}, \quad \text{where } w \geq \frac{1}{2}.$$

The eigenvalues for single-particle states are then $\Lambda_k = 2w \cos k/2$. The Fermi level k_F is now defined by the relation

$$2 \cos(k_F/2) = w^{-1}, \quad 0 < k_F < \pi,$$

and the free energy density is given by

$$f = \frac{1}{\pi\sqrt{3}} \int_{-k_F}^{+k_F} \log[2 \cos(k/2)] dk.$$

TD: derive CFT prediction $\log \Lambda/\Lambda_{\text{max}} = 2\pi X/L$. Compute CG spectrum from finite-size analysis.

3 Dimer appendix

3.1 Pfaffian formulation

We consider the dimer problem on an $m \times n$ square lattice Q_{mn} . Obviously a dimer covering exists only if mn is even, so we shall suppose m even. An example on Q_{64} is shown in Fig. 1.

The number of dimer coverings will of course depend on the boundary conditions. In this section we concentrate on free boundary conditions (i.e., free both along the horizontal and vertical directions). In this case the result

can be written as the Pfaffian of an appropriate matrix. An expression in terms of a single Pfaffian also exists for cylindrical boundary conditions (i.e., free along one lattice direction and periodic along the other). In Sec. 2.4 we shall show that toroidal boundary conditions (i.e., periodic in both directions) leads to a linear combination of *four* different Pfaffians. More generally, the number of Pfaffians needed will be 4^g for a lattice embedded into a surface of genus g .

Let the fugacity of horizontal and vertical dimers be respectively z_1 and z_2 . The weight of the configuration shown in Fig. 1 is then $(z_1)^4(z_2)^8$. Let $g(N_1, N_2)$ be the number of dimer coverings of Q_{mn} using N_1 horizontal and N_2 vertical dimers. We have necessarily $g(N_1, N_2) = 0$ unless $2(N_1 + N_2) = mn$. The goal is then to compute the partition function

$$Z_{mn}(z_1, z_2) = \sum_{N_1, N_2} g(N_1, N_2) z_1^{N_1} z_2^{N_2}. \quad (3.1)$$

The Pfaffian of an $2N \times 2N$ skew-symmetric matrix A with elements $a(k, k') = -a(k', k)$ is defined by

$$\begin{aligned} \text{Pf } A &= \sum_P' \varepsilon(P) a(k_1, k_2) a(k_3, k_4) \cdots a(k_{2N-1}, k_{2N}) \\ &= \frac{1}{N! 2^N} \sum_P \varepsilon(P) a(k_1, k_2) a(k_3, k_4) \cdots a(k_{2N-1}, k_{2N}). \end{aligned} \quad (3.2)$$

Here \sum_P runs over all permutations $P : (1, 2, \dots, 2N) \rightarrow (k_1, k_2, \dots, k_{2N})$, whereas \sum_P' is constrained to those permutations satisfying the constraint

$$\begin{aligned} k_1 < k_2, \quad k_3 < k_4, \quad \cdots, \quad k_{2N-1} < k_{2N} \\ k_1 < k_3 < k_5 < \cdots < k_{2N-1} \end{aligned} \quad (3.3)$$

and $\varepsilon(P) = \pm 1$ is the sign of P . It is easy to see that there are $\frac{(2N)!}{N! 2^N} = (2N-1)!!$ terms in the sum \sum_P' .

The constraint (3.3) is very natural from the point of view of dimers. Suppose we assign to the vertices of the lattice (i, j) , where $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$, some numbering, for instance

$$(i, j) \mapsto k = (j-1)m + i. \quad (3.4)$$

Let a configuration of dimers be denoted

$$C = [k_1, k_2] [k_3, k_4] \cdots [k_{2N-1}, k_{2N}], \quad (3.5)$$

where $[k, k']$ means that there is a dimer covering vertices k and k' . Then the constraint (3.3) expresses simply that C is considered modulo exchanges of dimers, and modulo exchanges of the two end points within each dimer.

This suggests an obvious strategy for computing $Z_{mn}(z_1, z_2)$ as a Pfaffian. Indeed we will have

$$|\text{Pf } D| = Z_{mn}(z_1, z_2) \quad (3.6)$$

provided that we can define a $2N \times 2N$ skew-symmetric matrix D , with $2N = mn$, that fulfills three requirements:

1. There should be a bijection between the non-vanishing contributions to $\text{Pf } D$ and the dimer configurations on Q_{mn} .
2. The weight of each non-vanishing contribution to $\text{Pf } D$ should be equal, *up to a sign*, to the corresponding statistical weight in $Z_{mn}(z_1, z_2)$.
3. All contributions to $\text{Pf } D$ should have *the same sign*.

Requirements 1–2 are easy to fulfill. To satisfy requirement 1, we simply set $d(k, k') = 0$ if the vertices k and k' are not adjacent in Q_{mn} . To satisfy requirement 2, we set $d(k, k') = -d(k', k) = \pm z_{kk'}$ if k and k' are adjacent, where $z_{kk'}$ is the desired fugacity of a dimer that covers k and k' . Note that the liberty in choosing $z_{kk'}$ makes it possible to tackle the most general situation of edge-dependent fugacities; in the sequel we shall however only need z_1 and z_2 as in (3.1).

3.2 Solving the sign problem

The tricky part is requirement 3: how to choose the correct sign of $d(k, k')$ when k and k' are adjacent? It is convenient to represent the signs of the matrix elements $d(k, k')$ by an orientation of the edges of Q_{mn} . If the edge (kk') is oriented from k to k' (resp. from k' to k), we take $d(k, k') = +z_{kk'}$ (resp. $d(k, k') = -z_{kk'}$).

The question is then whether a lattice orientation exists that will fulfill requirement 3. The answer is positive, not only for Q_{mn} but in fact for any planar graph. The corresponding orientation is known as a *Kasteleyn orientation*. The goal of this section is to characterise precisely this orientation.

Consider superposing two different dimer configurations C and C' of Q_{mn} . The resulting *transition graph* is made up of doubly occupied edges, where the dimers of C and C' coincide, and of *transition cycles*, which are cycles

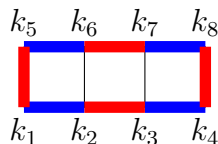


Figure 7: Example of a transition cycle on Q_{42} between the standard configuration C (in blue) and another configuration C' (in red). The vertices are labelled according to the canonical order (3.4).

of even length along which the dimers from C and C' alternate. This is shown in Fig. 2. Since the length of a transition cycle is even, the number of clockwise and anticlockwise oriented edges in the cycle is either both even, or both odd: we call this the *orientation parity*.

To turn configuration C into C' one needs to shift the dimers one unit along each transition cycle. The factor $\varepsilon(P)$ appearing in (3.2) can then be shown (see below) to produce a minus sign for each transition cycle. This sign must be cancelled by another one coming from the signs of the entries $d(k, k') = \pm z_{kk'}$. The terms representing C and C' will therefore have equal signs if *the orientation parity of all transition cycles is odd*.³ If we can find an orientation of Q_{mn} satisfying this requirement, the sign problem is solved.

For the sake of definiteness, let us consider the case where C is the standard configuration shown in blue in Fig. 2, and C' is another arbitrary dimer configuration. We focus on the contribution to $\varepsilon(P)$ of a single transition cycle in $C \cup C'$.

Orient the transition cycle in the counterclockwise direction. It will then pass through the edges of C in any fixed column exactly as many times in the right direction (i.e., in the direction of increasing labelling k) as in the left direction (i.e., in the direction of decreasing labelling k), since otherwise the cycle would not be a closed polygon. So *a fortiori* this is true for the passages through any edge of C . Let r be the number of right (and hence left) passages.

Let us now describe a 5-stage process that permutes the C -term into the C' -term. To follow the argument, it is useful to consider in parallel an

³Indeed, if the product of all the signs around the cycle is -1 , then the products of the subsets of signs corresponding to C and to C' must be opposite.

example with $r = 2$ on Q_{42} , i.e., with a single transition cycle of length $4r = 8$ (see Fig. 7). The initial configuration C is then

$$[k_1, k_2] [k_3, k_4] [k_5, k_6] [k_7, k_8]. \quad (3.7)$$

The stages are as follows:

1. Reverse the r pairs of points within each doublet that correspond to a left passage, so that the order within each pair now corresponds to the cyclical rather than the canonical order (3.4):

$$[k_1, k_2] [k_3, k_4] [k_6, k_5] [k_8, k_7]. \quad (3.8)$$

This produces a factor $(-1)^r$.

2. Permute the doublets as required to produce the perfect cyclic order:

$$[k_1, k_2] [k_3, k_4] [k_8, k_7] [k_6, k_5]. \quad (3.9)$$

Since only doublets are permuted, this results in a factor $+1$.

3. Permute all $4r$ points cyclically:

$$[k_2, k_3] [k_4, k_8] [k_7, k_6] [k_5, k_1]. \quad (3.10)$$

This produces a factor $(-1)^{4r-1}$. We now have the C' -term as desired, but the rules (3.3) are violated.

4. Permute the doublets so as to respect the second part of rule (3.3):

$$[k_5, k_1] [k_2, k_3] [k_4, k_8] [k_7, k_6]. \quad (3.11)$$

This is the “opposite of stage 2” and gives a factor $+1$.

5. Reverse r pairs of points within each doublet so as to respect the first part of rule (3.2):

$$[k_1, k_5] [k_2, k_3] [k_4, k_8] [k_6, k_7]. \quad (3.12)$$

This is the “opposite of stage 1” and gives a factor $(-1)^r$.

The total sign change is then

$$(-1)^r \times (-1)^{4r-1} \times (-1)^r = -1 \quad (3.13)$$

as claimed.

This correct choice of orientation parity can indeed be made for any planar graph. This relies on a number of properties that can rather easily be proved by induction in the size of the graph. Let us call a cycle that surrounds a single face of the lattice a *mesh cycle*. The relevant properties are then:

1. A planar graph can be oriented such that the orientation parity of all even mesh cycles is odd.
2. For a planar graph with such an orientation, the orientation parity of any even cycle whose interior contains an even (resp. odd) number of vertices is odd (resp. even).
3. In a planar graph the interior of any transition cycle contains an even number of vertices.

Leaving this generality and returning to the square lattice Q_{mn} , a possible Kasteleyn orientation is shown in Fig. 3.

Let us finally remark, that it is relatively easy to find a Kasteleyn orientation for any regular (Archimedean) lattice. However, despite of the above existence result, there does not seem to be a simple constructive approach for an arbitrary planar graph.

3.3 Evaluation of the Pfaffian

We have now established (3.6) when the matrix D is chosen according to requirements 1–2 and the Kasteleyn orientation of Fig. 3. This reads explicitly

$$d(i, j; i', j') = z_1 (\delta_{i+1, i'} - \delta_{i-1, i'}) \delta_{j, j'} + (-1)^i z_2 (\delta_{j+1, j'} - \delta_{j-1, j'}) \delta_{i, i'}, \quad (3.14)$$

where the subtractions guarantee the proper antisymmetrisation. All of this would be of little avail if the Pfaffian were difficult to compute. Fortunately its square is just a standard determinant. Thus

$$[Z_{mn}(z_1, z_2)]^2 = [\text{Pf } D]^2 = \det D. \quad (3.15)$$

Proving this relation is a little lengthy, and we only give a short outline (full details are provided in [MW73]). Introducing the cofactors D_{jk} , one first applies the Jacobi theorem $D_{jj}D_{kk} - D_{jk}D_{kj} = D_{jk, jk} \det D$ to the skew-symmetric matrix D . An induction argument then shows that $(\det D)^{1/2}$ is a rational function—and actually even a polynomial—of the matrix elements. Exploiting this finally leads to the desired relation with $\text{Pf } D$.

Let us illustrate the main points on a trivial example on Q_{22} . We first choose to orient the edges anticlockwise ($1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 1$). Note that this is *not* a Kasteleyn orientation. We choose the most general position-dependent edge weights:

$$\det \begin{pmatrix} 0 & z_{12} & -z_{13} & 0 \\ -z_{12} & 0 & 0 & z_{24} \\ z_{13} & 0 & 0 & -z_{34} \\ 0 & -z_{24} & z_{34} & 0 \end{pmatrix} = (z_{13}z_{24} - z_{12}z_{34})^2. \quad (3.16)$$

Changing the orientation of any one edge turns this into a Kasteleyn orientation and makes the two terms have the same sign.

The goal is therefore to compute $\det D$. If D were a cyclic matrix (i.e., with entries that depended on the indices i and j in a periodic fashion) this could be rather easily accomplished by bringing it into diagonal form via a Fourier transformation (see Sec. 2.4 for such a computation). In the present case there exists a slightly more complicated transformation that will turn D into a direct sum of 2×2 matrices.

Let us write D as a direct product of $m \times m$ and $n \times n$ matrices that describe the dependence of the weight on the horizontal and vertical coordinates respectively:

$$D = z_1(Q_m \otimes I_n) + z_2(F_m \otimes Q_n) \quad (3.17)$$

Here I_n is the $n \times n$ unit matrix, whereas

$$\begin{aligned}
Q_m(i, i') &= \delta_{i+1, i'} - \delta_{i-1, i'} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}, \\
F_m(i, i') &= (-1)^i \delta_{i, i'} = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{3.18}
\end{aligned}$$

Note that the matrix Q is only almost cyclic, since the elements in its upper-right and lower-left corners are zero.

The transformation that we need is

$$\begin{aligned}
\tilde{D} &= U^{-1} D U, \\
U &= U_m \otimes U_n, \tag{3.19}
\end{aligned}$$

where

$$U_n(l, l') = \sqrt{\frac{2}{n+1}} i^l \sin\left(\frac{l' \pi}{n+1}\right). \tag{3.20}$$

Let us use this transformation to find an explicit formula for $Z_{mn}(z_1, z_2)$. First we note the following orthogonality identity:

$$\frac{2}{n+1} \sum_{l''=1}^n \sin\left(\frac{l'' \pi}{n+1}\right) \sin\left(\frac{l'' l' \pi}{n+1}\right) = \delta_{l, l'}, \tag{3.21}$$

which can be easily proved by writing out the sines in terms of complex exponentials, multiplying out, and summing up the resulting geometrical series. This implies that the corresponding inverse matrix is

$$U_n^{-1}(l, l') = \sqrt{\frac{2}{n+1}} (-i)^l \sin\left(\frac{l' \pi}{n+1}\right). \tag{3.22}$$

Using this we first diagonalise the matrix Q . We have

$$(QU)(l, l') = \sum_{l''=1}^n Q(l, l'')U(l'', l') = U(l+1, l') - U(l-1, l')$$

and further

$$\begin{aligned} (U^{-1}QU)(l, l') &= \sum_{l''=1}^n U^{-1}(l, l'')(QU)(l'', l') \\ &= \sum_{l''=1}^n \{U^{-1}(l, l'')U(l''+1, l') - U^{-1}(l, l'')U(l''-1, l')\}. \end{aligned}$$

Inserting (3.20) and (3.22) this becomes

$$\begin{aligned} \dots &= \frac{2i}{n+1} \sum_{l''=1}^n \sin\left(\frac{ll''\pi}{n+1}\right) \left\{ \sin\left(\frac{l'(l''+1)\pi}{n+1}\right) + \sin\left(\frac{l'(l''-1)\pi}{n+1}\right) \right\} \\ &= \frac{4i}{n+1} \sum_{l''=1}^n \sin\left(\frac{ll''\pi}{n+1}\right) \cos\left(\frac{l'\pi}{n+1}\right) \sin\left(\frac{l'l''\pi}{n+1}\right) \\ &= 2i \cos\left(\frac{l'\pi}{n+1}\right) \delta_{l, l'}, \end{aligned} \quad (3.23)$$

where we have first used an addition formula and then the orthogonality relation (3.21).

Another identity that can be proved in the same way as (3.21) is the following:

$$\frac{2}{n+1} \sum_{l''=1}^n (-1)^{l''} \sin\left(\frac{ll''\pi}{n+1}\right) \sin\left(\frac{l''l'\pi}{n+1}\right) = \delta_{l+l', n+1}, \quad (3.24)$$

where the right-hand side is the “mirrored” identity matrix. This can be used to diagonalise the matrix F . We find:

$$(U^{-1}FU)(l, l') = \delta_{l+l', n+1}. \quad (3.25)$$

The “diagonalised” D -matrix now reads⁴, using (3.23) and (3.25),

$$\begin{aligned} \tilde{D}(k, l; k', l') &= 2iz_1 \delta_{k, k'} \delta_{l, l'} \cos\left(\frac{k\pi}{m+1}\right) \\ &+ 2iz_2 \delta_{k+k', m+1} \delta_{l, l'} \cos\left(\frac{l\pi}{n+1}\right). \end{aligned} \quad (3.26)$$

This is indeed diagonal in l -space, but not quite in k -space. Rather we have a matrix with the shape

$$\begin{pmatrix} w & & & w' \\ & w & & w' \\ & & \ddots & \\ & w' & & w \\ w' & & & w \end{pmatrix}.$$

By changing the labelling $1, 2, 3, 4, 5, 6, \dots, m$ of both rows and columns into $1, m, 2, m-1, 3, m-2, \dots, m/2, m/2+1$ (recall that m is even) this is turned into the block-diagonal matrix

$$\begin{pmatrix} w & w' & & & & \\ w' & \tilde{w} & & & & \\ & & \ddots & & & \\ & & & w & w' & \\ & & & w' & \tilde{w} & \end{pmatrix}.$$

Note that in this process some of the entries change sign ($\tilde{w} = -w$), since when $k \mapsto k' \equiv m+1-k$ we get

$$\cos\left(\frac{k'\pi}{m+1}\right) = -\cos\left(\frac{k\pi}{m+1}\right).$$

Therefore we obtain the result

$$\det D = \det \tilde{D} = \prod_{k=1}^{m/2} \prod_{l=1}^n \left| \begin{array}{cc} 2iz_1 \cos\left(\frac{k\pi}{m+1}\right) & 2iz_2 \cos\left(\frac{l\pi}{n+1}\right) \\ 2iz_2 \cos\left(\frac{l\pi}{n+1}\right) & -2iz_1 \cos\left(\frac{k\pi}{m+1}\right) \end{array} \right|. \quad (3.27)$$

Using finally (3.15) we arrive at

$$Z_{mn}(z_1, z_2) = 2^{\frac{mn}{2}} \prod_{k=1}^{m/2} \prod_{l=1}^n \sqrt{z_1^2 \cos^2\left(\frac{k\pi}{m+1}\right) + z_2^2 \cos^2\left(\frac{l\pi}{n+1}\right)}. \quad (3.28)$$

In particular we can find the number of ways to tile a chessboard by 32 dominos [TF61]:

$$12\,988\,816 = 2^4 \times 17^2 \times 53^2. \quad (3.29)$$

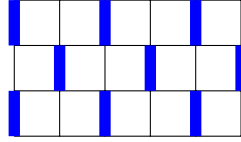


Figure 8: A particular dimer configuration C_0 in which all dimers sit on even vertical edges.

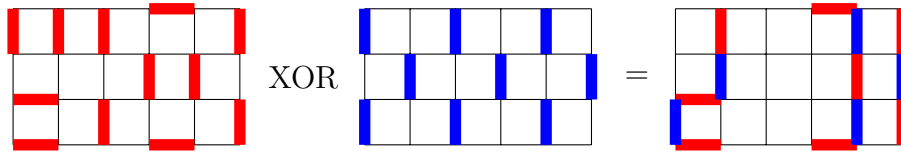


Figure 9: Conserved strings in a dimer configuration on $Q_{6,4}$.

3.4 Transfer matrix for the square lattice

On the square lattice one can assign a definite parity to the vertical edges by alternating even and odd edges throughout a given row, and alternating the convention between even and odd rows. Fig. 8 shows a particular dimer configuration C_0 in which all dimers sit on even vertical edges. Note that this corresponds to a maximal height gradient between the left and right rims of the lattice.

Consider now superposing a generic dimer configuration C with C_0 by means of an exclusive or (XOR) operation. For example, when C is the configuration of Fig. 1 the resulting superposition is shown in Fig. 9.

This superposition consists in a certain number s of strings (here $s = 3$) along which dimers from C and C_0 alternate. The dynamics under which these strings propagate in the vertical direction has interesting properties:

1. The number of strings is conserved, and
2. When moving from one horizontal layer to the next, a string can either go straight or move exactly one step to the left or to the right.

These properties follow directly from the definition of the XOR operation and from the definition of a valid dimer covering C .

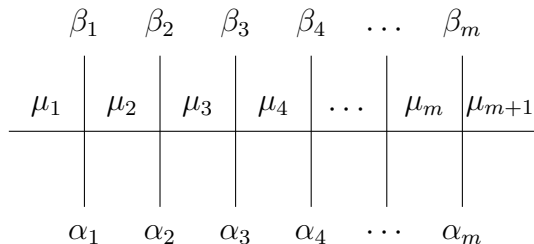


Figure 10: Labelling of edges used to define the row-to-row transfer matrix.

The properties of strings suggest to view dimer configurations as a discrete time evolution process, where the time increases along the vertical direction. In what follows it is convenient to refer to the horizontal (resp. vertical) direction as *space* (resp. *time*). The time evolution is then accomplished by a linear operator, called the row-to-row transfer matrix $T_{\beta\alpha}$, that we now define.

Let us label the edges of two consecutive rows of vertical edges, as well as the intermediate row of horizontal edges, as shown in Fig. 10. The state of a row is specified by the occupation numbers $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$, where $\alpha_i = 0$ (resp. $\alpha_i = 1$) means that the i th vertical edge is empty (resp. occupied by a dimer). Given the states α and β of two consecutive rows, the transfer matrix element $T_{\beta\alpha}$ is the part of the Boltzmann weight in (2.1) seen locally, summed over all possible intermediate states μ compatible with α and β . Thus

$$T_{\beta\alpha} = \sum_{\mu|(\alpha,\beta)} (z_1)^{\sum_i \mu_i} (z_2)^{\frac{1}{2} \sum_i (\alpha_i + \beta_i)}. \quad (3.30)$$

The compatibility criterion $\mu|(\alpha, \beta)$ can be expressed formally as

$$\forall i \in \{1, \dots, m\} : \mu_i + \alpha_i + \beta_i + \mu_{i+1} = 1, \quad (3.31)$$

meaning simply that the sum of occupation numbers around any one vertex is one.

Boundary conditions in the space direction can be specified through an additional constraint on the μ variables. Free boundary conditions mean $\mu_1 = \mu_{m+1} = 0$; periodic boundary conditions are obtained by identifying $\mu_{m+1} \equiv \mu_1$. This implies of course that the transfer matrix is different in the two cases.

Boundary conditions in the time direction are specified by constraints on the first and last row states. Let $|0\rangle$ denote the empty row state, i.e., such that $\forall i \in \{1, \dots, m\} : \alpha_i = 0$. The partition function (2.1) with free boundary conditions in the time direction is then

$$Z_{mn}(z_1, z_2) = \langle 0|T^n|0\rangle, \quad (3.32)$$

whereas periodic boundary conditions in the time direction lead to

$$Z_{mn}(z_1, z_2) = \text{Tr } T^n. \quad (3.33)$$

In all cases, the free energy per site can be related to the leading eigenvalue of T , and critical exponents can be inferred from various subleading eigenvalues. The eigenvalues can be found either by numerical diagonalisation, or analytically through the Bethe Ansatz technique. For both purposes it is useful to discuss more closely the structure of the transfer matrix.

3.5 Sparse matrix factorisation

For all but the smallest m it is inefficient (both in analytical and numerical calculations) to write down the whole transfer matrix in a single go. It is preferable to write $T_{\beta\alpha}$ as a product of matrices $R_{\beta_i, \mu_{i+1}; \mu_i, \alpha_i}$ that act locally by adding only the i th vertex in a given row. Organising the pairs of index values in binary order (00, 01, 10, 11) this reads explicitly

$$R = I \otimes \dots \otimes I \otimes \begin{bmatrix} 0 & \sqrt{z_2} & \sqrt{z_1} & 0 \\ \sqrt{z_1} & 0 & 0 & 0 \\ \sqrt{z_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \otimes I \dots \otimes I, \quad (3.34)$$

where the identity matrices mean that the action elsewhere in the tensor product of states is trivial.

Before building a row of the lattice, one needs to insert the leftmost horizontal space—sometimes called *auxiliary space*—corresponding to the variable μ_1 . Then follows the action of m factors of $R_{\beta_i, \mu_{i+1}; \mu_i, \alpha_i}$, each propagating an α_i to a β_i , starting by $i = 1$ and ending by $i = m$. And finally the rightmost horizontal space, corresponding to μ_{m+1} , must be removed. For free boundary conditions in the space direction the insertion and removal of the auxiliary space simply amounts to enforcing $\mu_1 = \mu_{m+1} = 0$. Periodic

boundary conditions are slightly more tricky, and require allowing for both possibilities $\mu_1 = 0, 1$ in the insertion, keeping a copy of μ_1 when acting with the factors of R , and finally enforcing $\mu_1 = \mu_{m+1}$ upon removal of the auxiliary space. To say it shortly, one “traces over the auxiliary space”.

The advantage of this procedure in numerical calculations is that each application of R generates at most 2 out-states for each in-state, and hence takes time proportional to the dimension of the state space. If the entire T were applied at once, each in-state would produce an exponentially large (in m) number of out-states.

3.6 Sector decomposition

Naively it appears that $\dim T = 2^m$. The effective dimension is however greatly reduced by exploiting the conservation of strings. Let the number of strings be $s = m/2 + Q$, with $Q = -m/2, \dots, m/2$, and denote the corresponding block in the transfer matrix by $T^{(Q)}$. We have then

$$T = \bigoplus_{Q=-\frac{m}{2}}^{\frac{m}{2}} T^{(Q)}. \quad (3.35)$$

Diagonalising T amounts to diagonalising separately each term $T^{(Q)}$ —sometimes called a *sector*—in the direct sum. The states contributing to $T^{(Q)}$ can be specified by giving the position of the strings, whence

$$\dim T^{(Q)} = \binom{m}{m/2 + Q}. \quad (3.36)$$

An explicit characterisation of the row states follows by noting that that

$$Q = \sum_{i \text{ odd}} \alpha_i - \sum_{i \text{ even}} \alpha_i. \quad (3.37)$$

In terms of the height mapping, the meaning of the conserved “charge” Q is the height difference Δh between the left and right rims of the lattice. Obviously for free time-like boundary conditions only the $Q = 0$ sector will contribute to $Z_{mn}(z_1, z_2)$, whereas for periodic time-like boundary conditions all sectors participate.

The sectors with $Q \neq 0$ can be used to define correlation functions. For instance, a monomer defect leads to $\Delta h = 1$. The exponential decay of the

monomer-monomer correlation function $\mathcal{C}(n)$ in a cylinder geometry (m fixed and $n \rightarrow \infty$) is given by the ratio of the largest eigenvalues

$$\mathcal{C}(n) \sim \left(\frac{\Lambda_{\max}^{(Q=1)}}{\Lambda_{\max}^{(Q=0)}} \right)^n \quad (3.38)$$

of the transfer matrix sectors $T^{(1)}$ and $T^{(0)}$. Using a standard CFT result (viz., conformally mapping the cylinder to the complex plane) this can be used to infer the corresponding critical exponent $X(0, 1)$ as in (2.16). This requires obviously finding the $m \rightarrow \infty$ limit of the participating eigenvalues, which can be achieved by using Bethe Ansatz techniques.

It is a useful exercise at this point to write down explicitly the row states contributing to the sector $Q = 0$ for a moderately small system, say $m = 4$.

The six possible row states $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ of $T^{(0)}$ read:

$$(0, 0, 0, 0) \quad (1, 1, 0, 0) \quad (0, 0, 1, 1) \quad (1, 0, 0, 1) \quad (0, 1, 1, 0) \quad (1, 1, 1, 1)$$

and the transfer matrix is

$$T^{(0)} = \begin{pmatrix} z_1^2 & z_1 z_2 & z_1 z_2 & z_1 z_2 & 0 & z_2^2 \\ z_1 z_2 & 0 & z_2^2 & 0 & 0 & 0 \\ z_1 z_2 & z_2^2 & 0 & 0 & 0 & 0 \\ z_1 z_2 & 0 & 0 & 0 & z_2^2 & 0 \\ 0 & 0 & 0 & z_2^2 & 0 & 0 \\ z_2^2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can use (3.32) to compute $Z_{44}(1, 1)$. The result is

$$Z_{4,4}(z_1, z_2) = z_1^8 + 9z_1^6 z_2^2 + 16z_1^4 z_2^4 + 9z_1^2 z_2^6 + z_2^8.$$

and this agrees with the exact result (2.4). In particular $Z_{44}(1, 1) = 36$.

4 Ising model

The so-called Ising model was suggested by Wilhelm Lenz in 1920 [Le20] as a simple model of ferromagnetism. The one-dimensional case was studied in detail by Lenz' Ph.D. student Ernst Ising in 1925 [Is25], who found that it exhibits no phase transition⁵ at $T > 0$.

The situation in two dimensions is however much richer. The exact transition temperature on the square lattice was found in 1941 through a duality argument by Hendrik Kramers and Gregory Wannier [KW41]. This was followed by the exact solution for the free energy in 1944 by Lars Onsager [On44]. The expression for the spontaneous magnetisation

$$M = \left(1 - (\sinh(2\beta J_1) \sinh(2\beta J_2))^{-2}\right)^{\frac{1}{8}} \quad (4.1)$$

and hence the exact value of the critical exponent $\beta = \frac{1}{8}$ was announced by Onsager in siblylline form in the discussion section at a conference in 1949 [On49], but a proof by Chen Ning Yang only appeared in written form in 1952 [Ya52].

Onsager's solution in terms of quaternion algebras is not easy reading, and it took researchers many years to extract from it the simplest and most convenient formulation of the algebraic facts that make an exact solution possible.⁶ Also, it was not easy to see if there was any hope of generalising the solution to the experimentally most relevant case of three dimensions, or to solve more general classes of models (such as the Potts model).

For these reasons, many alternative—and simpler—solutions appeared subsequently. Among the most influential and useful we can mention the derivation of correlation functions in terms of Pfaffians by Montroll, Potts and Ward [MPW63], and the formulation of the Ising model as a quantum spin chain involving fermion operators

$$aa^\dagger + a^\dagger a = 1 \quad (4.2)$$

by Schultz, Mattis and Lieb [SML64]. It is this latter work that most clearly characterises the field-theoretical content of the Ising model: it is a theory of free fermions. The exact way in which the fermion sign problem is solved

⁵Perhaps as a consequence, he then decided to quit physics!

⁶With hindsight, one can now see in Onsager's paper [On44] the germs of what was later to be known as the Yang-Baxter equations—the most important ingredient in the study of integrable systems.

in the quantum spin chain makes precise the current understanding that the Ising model is not solvable in three or higher dimensions.

A closely related—but somehow simpler—fermionic formulation in terms of Grassmann variables

$$aa^* + a^*a = 0 \tag{4.3}$$

was found by Berezin [Be69]. We shall present this approach (following [Pl88]) below. A rather different alternative that makes direct contact with Chapter 2 is an elegant reformulation by Kasteleyn [Ka63] of the Ising model as a dimer covering problem.

In some sense, the Ising model is to statistical physics what the hydrogen atom is to atomic physics. Although originally solved on a square lattice with periodic boundary conditions, the Ising model remains solvable when defined on other lattices, or when subjected to various kinds of modifications (such as the inclusion of certain interactions with the boundary or certain multi-spin interactions). For this reason it roles as a testing bed on which new theoretical ideas, approximation schemes or numerical calculations can be tried out.

Note also that despite of all the activity mentioned above (see the book [MW73] for a rather complete account as of 1973), seemingly simple questions about the Ising model remain unanswered to this day. For example, an exact solution in a rectangle with free boundary conditions does not seem to have been uncovered yet.

4.1 Duality transformation

The simplest two-dimensional version of the Ising model is defined on the square lattice through the Hamiltonian

$$\mathcal{H} = - \sum_{m,n} (J_1 \sigma_{m,n} \sigma_{m+1,n} + J_2 \sigma_{m,n} \sigma_{m,n+1}) , \tag{4.4}$$

where J_1 (resp. J_2) measures the aligning interaction between horizontal (resp. vertical) neighbouring spins. The goal is then to compute the partition function

$$Z = \sum_{\{\sigma\}} \exp(-\beta\mathcal{H}) , \tag{4.5}$$

the sum being over all $\sigma_{m,n} = \pm 1$, as well as various correlation functions.

Unless stated otherwise, we consider doubly periodic boundary conditions. In some of the arguments we shall—just for the sake of notational simplicity—consider the isotropic case $J_1 = J_2 \equiv J$ and write the set of nearest neighbours $\langle ij \rangle$.

4.1.1 High-temperature expansion

Since $\sigma_{i,j} = \pm 1$ takes only two different values, we have the identity

$$\exp(\beta J \sigma_i \sigma_j) = \cosh(\beta J) + \sinh(\beta J) \sigma_i \sigma_j = \cosh(\beta J) (1 + w \sigma_i \sigma_j), \quad (4.6)$$

where $w = \tanh(\beta J)$. The partition function then reads

$$Z = (\cosh \beta J)^{2N} \sum_{\{\sigma\}} \prod_{\langle ij \rangle} (1 + w \sigma_i \sigma_j), \quad (4.7)$$

where $2N$ is the number of edges of a square lattice with N vertices.

A graphical representation can be associated with the development of the product as follows. We color any given edge if the term $w \sigma_i \sigma_j$ is taken, and leave the edge empty if we take the term 1. The contribution of graphs in which any vertex is incident on an *odd* number of coloured edges then vanishes upon taking the sum $\sum_{\{\sigma\}}$. In other words, non-zero contributions correspond to graphs consisting of closed polygons (which have the possibility of touching at corners, i.e., to have coordination number 4).

This leads to

$$Z = 2^N (\cosh \beta J)^{2N} \sum_{p=0}^N w^{2p} g_N(p), \quad (4.8)$$

where $g_N(p)$ is the number of closed (not necessarily connected) polygons of total length $2p$ that can be drawn on a square lattice of N vertices. This is often referred to as a high-temperature expansion, since $w = \tanh(\beta J) \ll 1$ when $\beta J \ll 1$, but we stress that this is an exact rewriting of Z that holds for any β .

4.1.2 Low-temperature expansion

On the other hand, one can expand Z around one of the the totally ordered states, say $\sigma \equiv 1$. The excitations are then domain walls surrounding domains of the opposite spin value ($\sigma = -1$).

The domain walls live on the dual lattice, but since this is again a square lattice, the walls are exactly the same polygons as before. The totally ordered state has energy $-2J \times N$ and an excitation raises this by $2J$ times the total length of the polygons. Hence

$$Z = (e^{2\beta J})^N (\cosh \beta J)^{2N} \sum_{p=0}^N w^{2p} g_N(p) \quad (4.9)$$

with $w = \exp(-2\beta J)$.

To have a precise definition of $g_N(p)$ one needs to specify the boundary conditions. If we take free boundary conditions, the low-temperature expansion (4.9) around the ordered state $\sigma = 1$ is obtained by surrounding the finite system by a layer of fictitious spins $\sigma_0 = 1$ on the boundary. In particular, the other ordered state $\sigma = -1$ will define a polygon that surrounds the entire system.

Suppose on the other hand that one wishes to work on a finite lattice with doubly periodic boundary conditions. When the size of the polygons becomes comparable to the size of the lattice, the length $2p$ is ambiguous since one may “wrap” the polygon around a periodic direction. To keep things well defined in that case, the limit of an infinitely large lattice should therefore be taken *before* considering polygons of arbitrarily large p .

4.1.3 Duality relation

Up to unimportant multiplicative factors, the partition functions in the high-temperature expansion at inverse temperature β and in the low-temperature expansion at β^* therefore coincide, provided that

$$\exp(-2\beta^* J) = \tanh(\beta J). \quad (4.10)$$

This is a duality relation relating exactly the partition functions at high and low temperatures. Equivalently this can be rewritten

$$\sinh(2\beta J_1) \sinh(2\beta^* J_2) = 1, \quad (4.11)$$

where we have reintroduced the distinction between horizontal and vertical couplings.

The duality relation permits us to determine the critical temperature exactly. Namely, if there is a singularity in $\log Z(\beta)$ for some β , there should be another singularity with the same critical exponents at β^* . If we assume that the Ising model has a unique critical point β_c , we must conclude that it is given by the fixed point condition $\beta = \beta^*$. Thus $\sinh(2\beta_c J) = 1$, or

$$\beta_c J = \frac{1}{2} \log(\sqrt{2} + 1) = 0.440\,686\,\dots \quad (4.12)$$

There is a nice generalisation of the argument to the Q -state Potts model, defined by

$$\mathcal{H} = -J_{\text{Potts}} \sum_{\langle i,j \rangle} \delta(\sigma_i, \sigma_j) \quad (4.13)$$

and $\sigma_i = 1, 2, \dots, Q$. Note that for $Q = 2$ and $J_{\text{Potts}} = 2J$ this is equivalent to the Ising model. The self-duality criterion reads in this case

$$\beta_c J_{\text{Potts}} = \log(\sqrt{Q} + 1). \quad (4.14)$$

Regular lattices other than the square lattice are not self-dual. Therefore, the duality argument related the partition function $Z(\beta)$ of the original model to the partition function on the dual lattice $Z^*(\beta^*)$. A duality relation can only be extracted if there is some independent means of relating $Z^*(\beta^*)$ to a partition function $Z(\tilde{\beta})$ again defined on the original lattice. The critical point β_c is then determined by $\beta = \tilde{\beta}$.

In the case of the triangular lattice, this can be done by performing a partial summation (decimation) over the spins on the even sublattice of the dual, hexagonal lattice. This is known as a star-triangle transformation.

Such transformations were first used by Kennelly [Ke99] as early as 1899 in order to simplify networks of electrical resistors. They are now seen as the precursors of the celebrated Yang-Baxter equations in the theory of integrable systems.

The argument for the Ising model can again be generalised to the Potts model. In terms of the variable $v = \exp(\beta J_{\text{Potts}}) - 1$ the result is

$$v_c^3 + 3v_c^2 = Q. \quad (4.15)$$

Figure 11: Relation between polygon configurations on the square lattice (left) and dimer coverings on a decorated square lattice (right).

For the Ising case $Q = 2$, two of the solutions to this cubic read $\exp(2\beta_c J) = \pm\sqrt{3}$. The sign is of no importance, since the length of domain walls on the dual, hexagonal lattice is necessarily even. Hence

$$\beta_c J = \frac{1}{4} \log 3 = 0.274\,653\cdots . \quad (4.16)$$

That this is less than the result (4.12) on the square lattice makes perfect sense, since it is easier to align spins on a triangular lattice, due to the larger number of nearest neighbours.

The last solution to (4.15) for $Q = 2$ is $\exp(2\beta_c J) = 0$. This antiferromagnetic problem at zero temperature is actually quite interesting, since it means sending to zero the Boltzmann weight of neighbouring spins that align. On the triangular lattice it is not possible to avoid aligned neighbouring spins (think of the configurations around one lattice face), and the best one can do is to forbid any configuration in which the three spins around a face are aligned. The ground state is thus not unique, but a linear combination of all states with no aligned faces. This is an interesting combinatorial problem, which is critical and corresponds to a universality class *different* from that of the usual Ising model.

4.2 Relation to dimer coverings

The Ising model, viewed as polygon configurations on the square lattice, can be related to dimer configurations on a decorated square lattice, as shown in Fig. 11. In each line, the occupation of external edges by polygons or dimers is identical.

There appears to be two problems about this bijection. First, the correspondence is not bijective, since a vertex with no polygons corresponds to three (not one) dimer configurations on the internal decoration. Second, the decorated lattice is non-planar and so it is not guaranteed to possess a Kasteleyn orientation.

Fortunately it turns out that these two apparent complications compensate one another, resulting in an exact equivalence. To see this, consider

Figure 12: An oriented square lattice that permits us to solve the square-lattice Ising model as a dimer covering problem.

the orientation of the decorated lattice shown in Fig. 12. One can verify that with this orientation the orientation parity of all even cycles *without self-intersections* is odd. Thus, the orientation parity of the transition cycles connecting non-intersecting dimer configurations is odd as required. On the other hand, in the third line of Fig. 11, the orientation parity connecting either of the first two configurations with the third one is *even*, meaning that the third configuration is counted with a minus sign. This implies that the total count of the three configurations is $1 + 1 - 1 = 1$, and the bijection between polygon configurations and dimer coverings is established.

One can write down the matrix elements $d(k, k')$ of the matrix D by using the orientation of Fig. 12. When doing this, one can readily distinguish horizontal and vertical couplings, as in (4.4). It turns out that D is not easily diagonalised for an $M \times N$ lattice with free boundary conditions. However, on the torus this is easily done (as usual one needs then four Pfaffians). Going through the analysis one finds finally the same expression for the free energy in the $M, N \rightarrow \infty$ limit as obtained by other methods [On44, SML64, Be69].

The problem of computing the Ising partition function on a rectangle with free boundary conditions appears to be an open problem to this date. It can however be done in the conformal limit [KV92].

The configurations of the zero-temperature antiferromagnetic Ising model on the triangular lattice, discussed after (4.15), are bijectively related to dimer coverings of the hexagonal lattice.

4.3 Solution using Grassmann variables

We now present a detailed solution of the Ising model using the rather different approach of Grassmann integrations [Be69]. The main motivation for this approach is that it will enable us to make a precise connection between the Ising model and free fermions. More generally, it is always convenient to understand exact solvability as the consequence of some underlying algebraic structure, and we want to make this algebraic link clear.

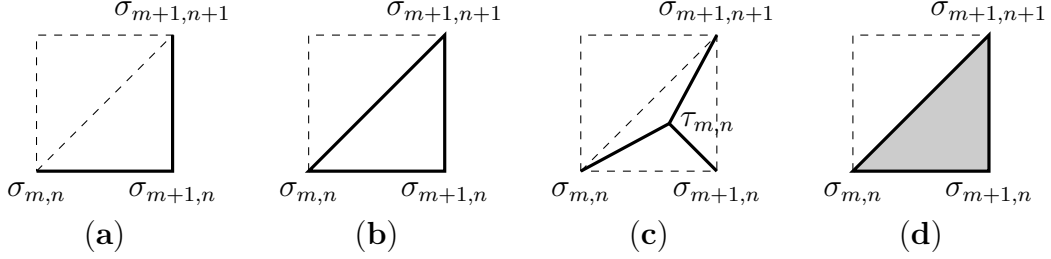


Figure 13: Four possible choices for the interactions within an elementary cell. This permits us to treat the a) square, b) triangular and c) hexagonal lattices in one single calculation. In d) the grey region stands for an arbitrary interaction between *pairs* of spins, possibly including one or more internal spins τ .

Rather than restricting to the square-lattice Hamiltonian (4.4) we might as well consider a more general situation [P188] in which a general class of interactions between *pairs* of spins, all situated within the shaded triangle in Fig. 4.3.d, take place within the elementary cells of an underlying square lattice.

Define the normalised trace over a spin $\sigma_{m,n} = \pm 1$ at lattice position (m, n) as

$$\text{Tr}_{\sigma_{m,n}} (\dots) = \frac{1}{2} \sum_{\sigma_{m,n}=\pm 1} (\dots) \quad (4.17)$$

and the normalised trace over all spins on the $M \times N$ lattice as

$$\text{Tr}_{\sigma} = \prod_{m=1}^M \prod_{n=1}^N \text{Tr}_{\sigma_{m,n}} . \quad (4.18)$$

4.3.1 Parameterisation of the chosen lattice

We first show that the partition function of any Ising model with interactions as in Fig. 4.3 can be written, up to an unimportant multiplicative factor, as

$$Z = \text{Tr}_{\sigma} \left\{ \prod_{mn} (\alpha_0 + \alpha_1 \sigma_1 \sigma_2 + \alpha_2 \sigma_2 \sigma_3 + \alpha_3 \sigma_1 \sigma_3)_{mn} \right\} , \quad (4.19)$$

where

$$(\sigma_1, \sigma_2, \sigma_3)_{mn} = (\sigma_{m,n}, \sigma_{m+1,n}, \sigma_{m+1,n+1}) \quad (4.20)$$

and α_i is a set of four coefficients which can easily be determined for any given lattice.

Consider as an example the triangular lattice (see Fig. 4.3.b) with Hamiltonian

$$\mathcal{H} = - \sum_{m,n} (J_1 \sigma_{m,n} \sigma_{m+1,n} + J_2 \sigma_{m+1,n} \sigma_{m+1,n+1} + J_3 \sigma_{m,n} \sigma_{m+1,n+1}) . \quad (4.21)$$

Using the identity (4.6) three times, we find that the Boltzmann weight of an elementary cell is

$$(e^{\beta(J_1 \sigma_1 \sigma_2 + J_2 \sigma_2 \sigma_3 + J_3 \sigma_1 \sigma_3)})_{mn} = R [(1 + t_1 \sigma_1 \sigma_2)(1 + t_2 \sigma_2 \sigma_3)(1 + t_3 \sigma_1 \sigma_3)]_{mn} \quad (4.22)$$

where we have set

$$\begin{aligned} R &= \cosh(\beta J_1) \cosh(\beta J_2) \cosh(\beta J_3) , \\ t_i &= \tanh(\beta J_i) , \text{ for } i = 1, 2, 3 . \end{aligned}$$

Expanding (4.22), and using $(\sigma_i)^2 = 1$, we find that indeed the partition function on the triangular lattice reads

$$Z_{\text{tri}} = (2R)^{MN} Z , \quad (4.23)$$

where Z is the general form (4.19) with coefficients

$$\begin{aligned} \alpha_0 &= 1 + t_1 t_2 t_3 , \\ \alpha_i &= t_i + t_{i+1} t_{i+2} , \text{ for } i = 1, 2, 3 \pmod{3} . \end{aligned} \quad (4.24)$$

In the most general setting of Fig. 4.3.d one must first obtain the form (4.19) by tracing over the spins in the interior of the shaded region. The simplest example of this is the hexagonal lattice, shown in Fig. 4.3.c. It is easily shown that

$$Z_{\text{hex}} = (4R)^{MN} Z , \quad (4.25)$$

with R as before and coefficients

$$\begin{aligned} \alpha_0 &= 1 , \\ \alpha_i &= t_i t_{i+1} , \text{ for } i = 1, 2, 3 \pmod{3} . \end{aligned} \quad (4.26)$$

4.3.2 Grassmann variables

We now introduce two Grassmann variables per site, $c_{m,n}$ and $c_{m,n}^*$. By definition, any two Grassmann variables anticommute

$$c_i c_j + c_j c_i = 0, \quad (4.27)$$

and in particular these variables are nilpotent

$$c_i^2 = c_j^2 = 0. \quad (4.28)$$

Any function of Grassmann variables is defined by its Taylor expansion, and because of the nilpotency such expansions are automatically truncated to the first order. The most general function of k Grassmann variables is defined by the coefficients of the 2^k possible monomials, e.g.,

$$f(c, c^*) = f_0 + f_1 c + f_2 c^* + f_3 c c^*. \quad (4.29)$$

Finally we introduce an integration measure, such that

$$\int dc \cdot 1 = 0, \quad \int dc \cdot c = 1 \quad (4.30)$$

and extended by linearity. The differentials dc are themselves Grassmann variables and thus anticommute. Note that Grassmann integration works like ordinary differentiation.

In particular one finds

$$\int dc^* dc e^{\lambda c c^*} f(c, c^*) = \lambda f_0 + f_3. \quad (4.31)$$

4.3.3 Density matrix

In parallel with the trace over spin variables (4.17)–(4.18) we introduce the trace over a pair of Grassmann variables

$$\text{Tr}_{c_{m,n}} (\dots) = \int dc_{m,n}^* dc_{m,n} e^{\lambda c_{m,n} c_{m,n}^*} (\dots) \quad (4.32)$$

as a Gaussian integral with some weight factor λ . We shall also need the trace over all $2MN$ Grassmann variables, and it turns out that the relevant weight factor is $\lambda = \alpha_0$. We therefore define

$$\text{Tr}_c (\dots) = \int \prod_{m=1}^M \prod_{n=1}^N dc_{m,n}^* dc_{m,n} e^{\alpha_0 c_{m,n} c_{m,n}^*} (\dots). \quad (4.33)$$

Because of the anticommuting nature of the differentials, it is quite important to understand what order of integrations is implied by our notation. In (4.32) the first integral is over $dc_{m,n}$, followed by a second integration over $dc_{m,n}^*$. More generally, the first integration is over the rightmost differential. On the other hand, any fixed *pair* of Grassmann variables commutes with the whole algebra, so the order in which the product in (4.33) is written out with respect to the indexing variable has no importance. We shall soon encounter situations where this order is of the highest importance, so let us define that $\prod_{m=1}^M$ means writing the first term (with $m = 1$) to the left. Sometimes we shall need the opposite order, and in that case we would write $\prod_{m=M}^1$.

Let us first concentrate on the Boltzmann weight of a single elementary cell in (4.19), viz.

$$(P_{123})_{mn} = (\alpha_0 + \alpha_1\sigma_1\sigma_2 + \alpha_2\sigma_2\sigma_3 + \alpha_3\sigma_1\sigma_3)_{mn}. \quad (4.34)$$

This can be rewritten in factorised form

$$(P_{123})_{mn} = \int dc_{mn}^* dc_{mn} e^{\alpha_0 c_{mn} c_{mn}^*} B_{m,n}^{(1)} B_{m+1,n}^{(2)} B_{m+1,n+1}^{(3)}, \quad (4.35)$$

where we have defined

$$\begin{aligned} B_{m,n}^{(1)} &= 1 + \frac{\alpha_1}{\sqrt{\eta}} c_{m,n} \sigma_{m,n}, \\ B_{m+1,n}^{(2)} &= 1 + \sqrt{\eta} (c_{m,n} + c_{m,n}^*) \sigma_{m+1,n}, \\ B_{m+1,n+1}^{(3)} &= 1 + \frac{\alpha_2}{\sqrt{\eta}} c_{m,n}^* \sigma_{m+1,n+1}. \end{aligned} \quad (4.36)$$

We have here defined $\eta = \frac{\alpha_1\alpha_2}{\alpha_3}$. Note that in these expressions for B , the subscript refers to the spin variable σ , whereas the subscript of the fermion variables c is always m, n .

To prove this, we first rewrite (4.34) as

$$P_{123} = (\alpha_0 - \eta) + \eta \left(\frac{\alpha_1}{\eta} \sigma_1 + \sigma_2 \right) \left(\sigma_2 + \frac{\alpha_2}{\eta} \sigma_3 \right),$$

where the subscript m, n has been omitted. By (4.31) we have

$$\begin{aligned} P_{123} &= \int dc^* dc e^{(\alpha_0 - \eta) cc^*} \left[1 + c \sqrt{\eta} \left(\frac{\alpha_1}{\eta} \sigma_1 + \sigma_2 \right) \right] \times \\ &\quad \left[1 + c^* \sqrt{\eta} \left(\sigma_2 + \frac{\alpha_2}{\eta} \sigma_3 \right) \right]. \end{aligned}$$

Using now the nilpotency, each $[\dots]$ can be factorised:

$$\begin{aligned} \left[1 + c\sqrt{\eta} \left(\frac{\alpha_1}{\eta} \sigma_1 + \sigma_2 \right) \right] &= \left(1 + c \frac{\alpha_1}{\sqrt{\eta}} \sigma_1 \right) (1 + c\sqrt{\eta} \sigma_2) \\ \left[1 + c^* \sqrt{\eta} \left(\sigma_2 + \frac{\alpha_2}{\eta} \sigma_3 \right) \right] &= (1 + c^* \sqrt{\eta} \sigma_2) \left(1 + c^* \frac{\alpha_2}{\sqrt{\eta}} \sigma_3 \right). \end{aligned}$$

Multiplying now the 2nd and 3rd factors

$$(1 + c\sqrt{\eta} \sigma_2) (1 + c^* \sqrt{\eta} \sigma_2) = 1 + \sqrt{\eta}(c + c^*) \sigma_2 + \eta c c^*$$

one gets a commuting term $\eta c c^*$ which can be absorbed into the integration measure:

$$e^{(\alpha_0 - \eta) c c^*} + \eta c c^* = e^{\alpha c c^*}.$$

Assembling the pieces, this implies

$$\begin{aligned} P_{123} &= \int d c^* d c e^{\alpha_0 c c^*} \left(1 + c \frac{\alpha_1}{\sqrt{\eta}} \sigma_1 \right) \times \\ &\quad (1 + \sqrt{\eta}(c + c^*) \sigma_2) \left(1 + c^* \frac{\alpha_2}{\sqrt{\eta}} \sigma_3 \right). \end{aligned}$$

This establishes the factorisation (4.35)–(4.36).

In terms of the trace (4.33) we have therefore

$$Z = \text{Tr}_\sigma \hat{Q}, \quad (4.37)$$

$$\hat{Q} = \text{Tr}_c \left\{ \prod_{m,n} \left(B_{m,n}^{(1)} B_{m+1,n}^{(2)} B_{m+1,n+1}^{(3)} \right) \right\}, \quad (4.38)$$

where \hat{Q} will be referred to as the density matrix.

4.3.4 Mirror factorisation

The strategy will now be to perform Tr_σ while keeping Tr_c , so as to obtain a Grassmann representation of Z . This cannot be done directly with the

form (4.38), since factors referring to the same $\sigma_{m,n}$ do not occur in adjacent positions in the product (cf. the different subscripts on the B factors). We therefore first aim at rearranging the product (4.38) in order to obtain the required adjacency.

It is convenient in this subsection to omit writing the integration over Grassmann variables. We thus write instead of (4.35)

$$(P_{123})_{mn} = B_{m,n}^{(1)} B_{m+1,n}^{(2)} B_{m+1,n+1}^{(3)}, \quad (4.39)$$

keeping in mind that the result will eventually be integrated. Note that the Grassmann pair $(c_{m,n}, c_{m,n}^*)$ occurs *only* in this factor, and since terms in $c_{m,n}$ and $c_{m,n}^*$ will vanish under the integration—as in (4.31)—eventually only the commuting terms 1 and $c_{m,n}c_{m,n}^*$ will matter anyway. In this sense, the factor (4.39) can be considered to commute with the whole algebra.

Suppose that the products $\mathcal{O}_i \mathcal{O}_i^*$ are commuting terms, whereas individual factors are not. Then we can write

$$(\mathcal{O}_1 \mathcal{O}_1^*)(\mathcal{O}_2 \mathcal{O}_2^*)(\mathcal{O}_3 \mathcal{O}_3^*) = (\mathcal{O}_1 \mathcal{O}_1^*)(\mathcal{O}_2(\mathcal{O}_3 \mathcal{O}_3^*)\mathcal{O}_2^*) = (\mathcal{O}_1(\mathcal{O}_2(\mathcal{O}_3 \mathcal{O}_3^*)\mathcal{O}_2^*)\mathcal{O}_1^*)$$

and more generally

$$\prod_{i=1}^L \mathcal{O}_i \mathcal{O}_i^* = \prod_{i=1}^L \mathcal{O}_i \cdot \prod_{i=L}^1 \mathcal{O}_i^*. \quad (4.40)$$

Using this property repeatedly suffices to bring (4.38) into the desired form.

Let us see in details how this is done. It is convenient sometimes to include factors $B_{m,n}^{(i)}$ in the product for which the indices $m = 0$ or $m = M + 1$ (and $n = 0$ or $n = N + 1$). Imposing formally that spins “beyond the boundary” vanish ($\sigma_{m,n} = 0$) we have $B_{m,n}^{(i)} = 1$, and the inclusion of such factors does not alter the result.

Consider first the product of a row of $(P_{123})_{mn}$ for fixed n . Using (4.40) we have

$$\prod_{m=0}^M (P_{123})_{mn} = \prod_{m=0}^M B_{m,n}^{(1)} B_{m+1,n}^{(2)} \cdot \prod_{m=M}^0 B_{m+1,n+1}^{(3)}.$$

In the both factors on the right-hand side we can move the parenthesis and

eliminate boundary terms:

$$\begin{aligned} \prod_{m=0}^M B_{m,n}^{(1)} B_{m+1,n}^{(2)} &= B_{0,n}^{(1)} \cdot \prod_{m=1}^M B_{m,n}^{(2)} B_{m,n}^{(1)} \cdot B_{M+1,n}^{(2)} = \prod_{m=1}^M B_{m,n}^{(2)} B_{m,n}^{(1)}, \\ \prod_{m=M}^0 B_{m+1,n+1}^{(3)} &= B_{M+1,n+1}^{(3)} \cdot \prod_{m=M}^1 B_{m,n+1}^{(3)} = \prod_{m=M}^1 B_{m,n+1}^{(3)}. \end{aligned} \quad (4.41)$$

Thus we have for one row (neglecting boundary effects)

$$\prod_{m=1}^M (P_{123})_{mn} = \prod_{m=1}^M B_{m,n}^{(2)} B_{m,n}^{(1)} \cdot \prod_{m=M}^1 B_{m,n+1}^{(3)},$$

where now the m indices are nicely organised.

The n indices are still not the same, but this is settled by taking the product over n and rearranging the expression in the same way as we just saw. The result is

$$\prod_{n=1}^N \prod_{m=1}^M (P_{123})_{mn} = \prod_{n=1}^N \left[\prod_{m=M}^1 B_{m,n}^{(3)} \cdot \prod_{m=1}^M B_{m,n}^{(2)} B_{m,n}^{(1)} \right]. \quad (4.42)$$

Putting back the trace over Grassmann variables, the density matrix therefore has the mirror factorised form

$$\hat{Q} = \text{Tr}_c \left\{ \prod_{n=1}^N \left[\prod_{m=M}^1 B_{m,n}^{(3)} \cdot \prod_{m=1}^M B_{m,n}^{(2)} B_{m,n}^{(1)} \right] \right\}. \quad (4.43)$$

4.3.5 Fermionic representation of Z

At the junction of the two products in (4.43) we have three B with the same subscripts $(m, n) = (1, n)$, i.e., referring to the same spin $\sigma_{m,n}$. We can now trace over that spin:

$$\begin{aligned} \text{Tr}_{\sigma_{m,n}} \left\{ B_{m,n}^{(3)} B_{m,n}^{(2)} B_{m,n}^{(1)} \right\} &= \frac{1}{2} \sum_{\sigma_{m,n}=\pm 1} \left(1 + \frac{\alpha_2}{\sqrt{\eta}} c_{m-1,n-1}^* \sigma_{m,n} \right) \times \\ &\quad \left(1 + \sqrt{\eta} (c_{m-1,n} + c_{m-1,n}^*) \sigma_{m,n} \right) \left(1 + \frac{\alpha_1}{\sqrt{\eta}} c_{m,n} \sigma_{m,n} \right). \end{aligned}$$

Only terms in 1 and $\sigma_{m,n}^2$ survive the trace:

$$\begin{aligned} \dots &= 1 + (c_{m-1,n} + c_{m-1,n}^*)(\alpha_1 c_{m,n} - \alpha_2 c_{m-1,n-1}^*) + \alpha_3 c_{m-1,n-1}^* c_{m,n} \\ &= \exp \left[\alpha_3 c_{m-1,n-1}^* c_{m,n} + (c_{m-1,n} + c_{m-1,n}^*)(\alpha_1 c_{m,n} - \alpha_2 c_{m-1,n-1}^*) \right]. \end{aligned}$$

For that same reason, this expression is quadratic in the fermion operators, hence commutes with the algebra. It can therefore be taken out in front of the expression.

In the remainder of the product the three B factors with $(m, n) = (2, n)$ are now adjacent, and we can trace next over that $\sigma_{m,n}$. Repeating the operation until nothing remains, we arrive at a purely fermionic expression for the partition function

$$\begin{aligned} Z &= \int \prod_{m=1}^M \prod_{n=1}^N dc_{m,n}^* dc_{m,n} \exp \left\{ \sum_{m=1}^M \sum_{n=1}^N \left[\alpha_0 c_{m,n} c_{m,n}^* \right. \right. \\ &\quad \left. \left. + (c_{m-1,n} + c_{m-1,n}^*)(\alpha_1 c_{m,n} - \alpha_2 c_{m-1,n-1}^*) + \alpha_3 c_{m-1,n-1}^* c_{m,n} \right] \right\}. \end{aligned} \quad (4.44)$$

4.3.6 Diagonalisation and thermodynamical limit

One can now diagonalise by performing a discrete Fourier transformation of the Grassmann variables. Let us suppose that $M = N \equiv L$:

$$c_{mn} = \frac{1}{L} \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} \tilde{c}_{pq} \exp \left(\frac{2\pi i}{L} (mp + nq) \right), \quad (4.45)$$

$$c_{mn}^* = \frac{1}{L} \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} \tilde{c}_{pq}^* \exp \left(-\frac{2\pi i}{L} (mp + nq) \right). \quad (4.46)$$

This is a rather standard exercise. Neglecting a few delicate effects having to do with the boundary, the result is

$$\begin{aligned} Z &= \prod_{p=0}^{L-1} \prod_{q=0}^{L-1} \left[(\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2) - 2(\alpha_0 \alpha_1 - \alpha_2 \alpha_3) \cos \left(\frac{2\pi p}{L} \right) \right. \\ &\quad \left. - 2(\alpha_0 \alpha_2 - \alpha_1 \alpha_3) \cos \left(\frac{2\pi q}{L} \right) - 2(\alpha_0 \alpha_3 - \alpha_1 \alpha_2) \cos \left(\frac{2\pi(p+q)}{L} \right) \right]^{1/2}. \end{aligned} \quad (4.47)$$

In the thermodynamical limit we then have for the free energy

$$\frac{\log Z}{L^2} \stackrel{L \rightarrow \infty}{=} \frac{1}{2} \int_0^{2\pi} \frac{dp}{2\pi} \int_0^{2\pi} \frac{dq}{2\pi} \quad (4.48)$$

$$\log \left[(\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2) - 2(\alpha_0\alpha_1 - \alpha_2\alpha_3) \cos p \right. \\ \left. - 2(\alpha_0\alpha_2 - \alpha_1\alpha_3) \cos q - 2(\alpha_0\alpha_3 - \alpha_1\alpha_2) \cos(p+q) \right].$$

This expression can be specialised to yield the result for the square, triangular, hexagonal and other lattices by inserting the coefficients α_i .

One could think that this exact result could be converted into an enumeration of the lattice polygons of the low/high temperature expansions. This does not appear to be feasible in practice.

4.3.7 Critical point

There are several obvious symmetries in the expressions (4.47)–(4.48) allowing arbitrary permutations of the α_i and sign changes of any two of them, provided that one simultaneously shifts the integration angles p and q . A less obvious symmetry is to change α_i to the conjugated parameters α_i^* defined by

$$\begin{bmatrix} \alpha_0^* \\ \alpha_1^* \\ \alpha_2^* \\ \alpha_3^* \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}. \quad (4.49)$$

Specialising to the square and triangular lattices, this transformation turns out to coincide with the Kramers-Wannier duality transformation.

At the critical point, (4.48) must exhibit a singularity. A moment's reflection shows that this can happen if and only if the argument $Q(p, q)$ of the logarithm vanishes for certain exceptional modes $(p, q) = (0, 0), (0, \pi), (\pi, 0), (\pi, \pi)$.⁷ One can check the following rewriting:

$$Q(p, q) = \begin{bmatrix} \bar{\alpha}_0^2 & \bar{\alpha}_1^2 & \bar{\alpha}_2^2 & \bar{\alpha}_3^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \cos p \\ \cos q \\ \cos(p+q) \end{bmatrix}, \quad (4.50)$$

⁷This will become clear when we compute the singular part of the free energy f_{sing} below.

where we have defined

$$\bar{\alpha}_i = \alpha_0 - \alpha_i^*, \text{ for } i = 0, 1, 2, 3. \quad (4.51)$$

In the product of the matrix and the right vector, three out of four trigonometric polynomials (i.e., the terms with coefficient $\bar{\alpha}_i^2$ for $i = 0, 1, 2, 3$) vanish at any one of the exceptional (p, q) values. Therefore $Q(p, q) = 0$ if and only if the prefactor $(\alpha_0 - \alpha_i^*)^2$ of the last trigonometric polynomial vanishes. The criticality criterion can thus be written in compact form as

$$\bar{\alpha}_0 \bar{\alpha}_1 \bar{\alpha}_2 \bar{\alpha}_3 = 0 \quad (4.52)$$

or equivalently

$$\alpha_0 \alpha_1 \alpha_2 \alpha_3 = \alpha_0^* \alpha_1^* \alpha_2^* \alpha_3^*. \quad (4.53)$$

4.4 Critical exponents

The singularity in the free energy at the transition can be inferred by integrating (4.48) around one of the exceptional (p, q) values. Consider for instance (p, q) near $(0, 0)$ and small $(\alpha_0 - \alpha_0^*)^2$. Taylor expanding the cosines to second order⁸ we find for the free energy

$$\begin{aligned} -\beta f &= \frac{1}{8\pi^2} \iint_{0 \leq p^2 + q^2 \leq r^2} dp dq \\ &\log \left[4(\alpha_0 - \alpha_0^*)^2 + A_1 \frac{p^2}{2} + A_2 \frac{q^2}{2} + A_3 \frac{(p+q)^2}{2} \right] + \dots \end{aligned} \quad (4.54)$$

with $A_1 = 2(\alpha_0 \alpha_1 - \alpha_2 \alpha_3)$, $A_2 = 2(\alpha_0 \alpha_2 - \alpha_1 \alpha_3)$, and $A_3 = 2(\alpha_0 \alpha_3 - \alpha_1 \alpha_2)$. Integrating this yields

$$-\beta f = \frac{(2\bar{\alpha}_0)^2}{4\pi \sqrt{A_1 A_2 + A_2 A_3 + A_1 A_3}} \log \left(\frac{1}{(2\bar{\alpha}_0)^2} \right) + \dots, \quad (4.55)$$

where the argument of the square root can be evaluated at criticality.

More generally, for the critical point $\bar{\alpha}_i = 0$ the dominant singular behaviour of the free energy reads

$$-\beta f_{\text{sing}} = \frac{(2\bar{\alpha}_i)^2}{16\pi \sqrt{(\alpha_0 \alpha_1 \alpha_2 \alpha_3)_c}} \log \left(\frac{1}{(2\bar{\alpha}_i)^2} \right). \quad (4.56)$$

⁸It is indeed the absence of first order terms that leads to the singular behaviour.

We infer that the singularity of the specific heat $C = -\beta^2 \frac{\partial^2}{\partial \beta^2}(\beta f)$ is

$$C_{\text{sing}} \stackrel{T \rightarrow T_c}{\sim} A_c \log \left| \frac{T_c}{T - T_c} \right|, \quad (4.57)$$

where the specific-heat critical amplitude is

$$A_c = \frac{\beta_c^2}{\pi \sqrt{(\alpha_0 \alpha_1 \alpha_2 \alpha_3)_c}} \left(\frac{d\bar{\alpha}_j(\beta)}{d\beta} \right)_c^2. \quad (4.58)$$

Thus, the standard critical exponent $\alpha = 0$, but with a logarithmic divergence (i.e., weaker than the usual power-law one).

The Grassmann approach can also be used to compute correlation functions. For example, the one-point function gives the spontaneous magnetisation, $M = \langle \sigma_{m,n} \rangle$. One finds [P188] a surprisingly simple expression for the eight power:

$$M^8 = \begin{cases} 1 - \frac{\alpha_0^* \alpha_1^* \alpha_2^* \alpha_3^*}{\alpha_0 \alpha_1 \alpha_2 \alpha_3} & \text{for } T \leq T_c \\ 0 & \text{for } T \geq T_c \end{cases} \quad (4.59)$$

This implies that

$$M \stackrel{T \rightarrow T_c}{\sim} \left| \frac{T - T_c}{T_c} \right|^{1/8}, \quad (4.60)$$

so that the critical exponent $\beta = \frac{1}{8}$.

4.5 Fermionic action

To exhibit precisely the fermionic nature of the Ising model, we wish to find the continuum limit of the action S appearing under the exponential in (4.44).

We first rewrite the expression for S so that finite differences appear wherever possible:

$$S = \sum_{m,n} \left\{ \tilde{m} c_{m,n} c_{m,n}^* + \alpha_1 c_{m,n} (c_{m,n}^* - c_{m-1,n}^*) \right. \\ \left. + \alpha_2 c_{m,n} (c_{m,n}^* - c_{m,n-1}^*) + \alpha_3 c_{m,n} (c_{m,n}^* - c_{m-1,n-1}^*) \right. \quad (4.61)$$

$$\left. + \alpha_1 c_{m,n} (c_{m,n} - c_{m-1,n}) + \alpha_2 c_{m,n}^* (c_{m,n}^* - c_{m,n-1}^*) \right\}, \quad (4.62)$$

where $\tilde{m} = \alpha_0 - \alpha_1 - \alpha_2 - \alpha_3 = 2\bar{\alpha}_0$. We have here used nilpotency and anticommutativity, and some of the summation indices have been shifted by one unit. Introduce now finite lattice derivatives

$$\partial_m x_{m,n} = x_{m,n} - x_{m-1,n}, \quad \partial_n x_{m,n} = x_{m,n} - x_{m,n-1}, \quad (4.63)$$

so that $x_{m,n} - x_{m-1,n-1} = (\partial_m + \partial_n - \partial_m \partial_n) x_{m,n}$. This gives

$$\begin{aligned} S = & \sum_{m,n} \{ \tilde{m} c_{mn} c_{mn}^* + \lambda_1 c_{mn} \partial_m c_{mn}^* + \lambda_2 c_{mn} \partial_n c_{mn}^* \\ & - \alpha_3 c_{mn} \partial_m \partial_n c_{mn}^* + \alpha_1 c_{mn} \partial_m c_{mn} + \alpha_2 c_{mn}^* \partial_n c_{mn}^* \}, \end{aligned} \quad (4.64)$$

where $\lambda_1 = \alpha_1 + \alpha_3$ and $\lambda_2 = \alpha_2 + \alpha_3$.

We now take the continuum limit $(m, n) \rightarrow (x_1, x_2) \equiv x$, so that $\partial_m \rightarrow \frac{\partial}{\partial x_1} \equiv \partial_1$ and $\partial_n \rightarrow \frac{\partial}{\partial x_2} \equiv \partial_2$. The continuum limit of the Grassmann variables defines a two-component field: $c_{mn} \rightarrow \psi(x)$ and $c_{mn}^* \rightarrow \bar{\psi}(x)$. This leads to a slightly non-standard form of the fermionic action. However, if we rotate the derivatives

$$\partial = \frac{1}{2}(\partial_1 - i\partial_2), \quad \bar{\partial} = \frac{1}{2}(\partial_1 + i\partial_2) \quad (4.65)$$

and similarly rotate the field components, we arrive at

$$S = \int d^2x \{ i\mathcal{M}\psi(x)\bar{\psi}(x) + \psi(x)\partial\psi(x) + \bar{\psi}(x)\bar{\partial}\bar{\psi}(x) \}, \quad (4.66)$$

with the rescaled mass

$$\mathcal{M} = \frac{\alpha_0 - \alpha_1 - \alpha_2 - \alpha_3}{\left(2\sqrt{(\alpha_0\alpha_1\alpha_2\alpha_3)_c}\right)^{1/2}}. \quad (4.67)$$

Note that (4.66) has the standard form for the action of a Majorana fermion.

According to (4.52) the mass \mathcal{M} vanishes at the critical point. We can therefore conclude that the critical Ising model corresponds, in the continuum limit, to a massless Majorana fermion. This furnishes a direct link between the lattice model and (conformal) field theory.

We see moreover that all different lattices that we have treated on the same footing are in the same universality class, since the coefficients α_i only enter into the mass \mathcal{M} .

Needless to say, one needs to be a little more careful when all the coupling constants are not positive. For instance, it is an interesting exercise to track down where the preceding argument should be changed when dealing with antiferromagnetic cases, such as the $T = 0$ Ising model on the triangular lattice (which is known to belong to a different universality class).

5 Coordinate Bethe Ansatz

Given a 2D model of equilibrium statistical physics, one would like to compute the partition function and certain correlation functions. This is rarely possible on a finite lattice—with dimer coverings being a notable exception—but in many cases exact results can be obtained in the thermodynamical limit. The question arises under which conditions such computations are possible.

Via the transfer matrix formulation, a 2D model can be recast as the time evolution of a 1D quantum spin chain. A first step in the analysis of a given problem is to identify the particles being described by the spin chain as well as their exact dynamics. The free energy and the correlation functions are related in the usual way to the ground state and the excitations of the spin chain Hamiltonian. The goal is then to diagonalise exactly this Hamiltonian, i.e., to identify its eigenvectors and eigenvalues.

If there were only one particle in a spin chain with periodic boundary conditions, its eigenvectors would be plane waves e^{ikx} of a certain momentum k . When more than one particle is present, a product of plane waves would be an appropriate wave function only if the particles moved independently. This is of course not the case in any non-trivial model. But a natural idea, pioneered by Bethe [Be31] in his 1931 paper on the Heisenberg model (also known as the XXX spin chain), is to try an Ansatz of coupled plane waves.

In this chapter we shall illustrate this approach on a more general model, known as the six-vertex model in its 2D incarnation, or equivalently as the XXZ spin chain. This model has important relations to the Potts model and to the Temperley-Lieb lattice algebra, as will be discussed in later chapters.

The six-vertex model was solved by Lieb [Li67]. It is a special case of the eight-vertex model which was later solved by Baxter [Ba72].

The Bethe Ansatz technique comes in several variants. In this chapter we focus on the so-called coordinate Bethe Ansatz, following roughly chapter 8 of Baxter's book [Ba82a]. The procedure is here to construct the eigenvectors explicitly for n -particle states, by identifying the relations—known as the Bethe Ansatz equations (BAE)—under which the so-called unwanted terms cancel out. Studying in detail the cases $n = 1$ and $n = 2$ usually gives crucial insight into the general form of these relations, and one proceeds to the general case by reasonable guesswork (which for the simplest models can be justified by detailed arguments).

In later chapters we shall rederive those results in an algebraic framework, culminating in the so-called algebraic Bethe Ansatz. It will gradually emerge

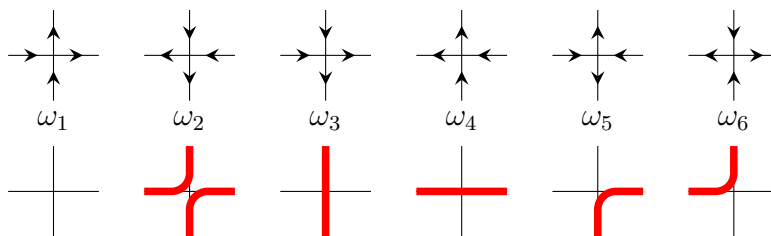


Figure 14: The allowed arrow arrangements (top) around a vertex that define the six-vertex model, with the corresponding particle trajectories (bottom).

that the exact solvability of a model hinges on the possibility to factorise any multi-particle scattering as a product of two-particle scatterings. This solvability condition is encoded in the celebrated Yang-Baxter equation.

5.1 Six-vertex model

The six-vertex model is defined by placing arrows on the edges of a square lattice, in such a way that every vertex is adjacent on two incoming and two outgoing arrows. The six possible configurations around a vertex are shown in the first line of Fig. 14 along with their respective Boltzmann weights $\omega_1, \dots, \omega_6$. The corresponding energies are denoted ϵ_i , and we have $\omega_i = \exp(-\frac{\epsilon_i}{k_B T})$. If there are n_i vertices of type i on the given lattice, the goal is to compute the asymptotic behaviour in the limit of a large lattice of the partition function

$$Z = \sum_{\text{arrows}} \prod_{i=1}^6 (\omega_i)^{n_i}. \quad (5.1)$$

In the transfer matrix formulation, we impose for the moment periodic boundary conditions along the horizontal lattice direction. The row-to-row transfer matrix then conserves the net arrow flux in the time direction. To fully exploit this conservation law we move to an equivalent particle picture. Recall that in the R -matrix factorisation of the transfer matrix, time flows in the North-Eastern direction. We therefore define that an edge is occupied by a particle if and only if it sustains a right-pointing or an up-pointing arrow. This is shown in the second line of Fig. 14.

The resulting world-lines of particles are conserved, and moreover they have a very simple dynamics. Following them from the bottom of the system

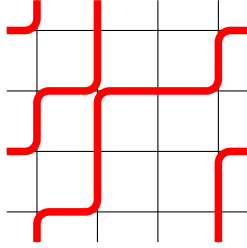


Figure 15: A possible configuration of world-lines on a 4×4 lattice.

to the top, they can only move up or to the right. This is illustrated in Fig. 15.

The six-vertex model is solvable when the weights are chosen to be invariant under a global reversal of arrows. The weights are traditionally denoted

$$\omega_1 = \omega_2 = a, \quad \omega_3 = \omega_4 = b, \quad \omega_5 = \omega_6 = c. \quad (5.2)$$

Note however that since the arrow arrangement 5 (resp. 6) acts as a sink (resp. source) of horizontal arrows, we must have the same number of each in every line. In particular, $n_5 = n_6$. For any $k \neq 0$ we can therefore take instead $\omega_5 = kc$ and $\omega_6 = k^{-1}c$ without changing Z . This “gauge symmetry” will turn out useful in later chapters.

5.2 Transfer matrix

The row-to-row transfer matrix T conserves the number n of world-lines when going from one row to the next. The positions x_i of the lines specifies a state $|x_1, x_2, \dots, x_n\rangle$, where we have assumed $x_1 < x_2 < \dots < x_n$. The transfer matrix element

$$\langle y_1, y_2, \dots, y_n | T | x_1, x_2, \dots, x_n \rangle$$

equals the product of Boltzmann weights along the row, provided that the state of the upper row $\langle y_1, y_2, \dots, y_n |$ is compatible with the state of the lower row $|x_1, x_2, \dots, x_n\rangle$. If the two states are not compatible, the matrix element is defined to be zero.

The two states are compatible if and only if the positions in the upper row y_i *interlace* those in the lower row x_i . The precise meaning of interlacing is the following:



Figure 16: The two ways in which $\langle y_1, y_2, \dots, y_n |$ can interlace $|x_1, x_2, \dots, x_n\rangle$, here shown for $n = 2$.

- If no line goes through the periodic lattice direction we have $x_i \leq y_i \leq x_{i+1}$ for $i = 1, 2, \dots, n - 1$, and $x_n \leq y_n \leq N$.
- If a line goes through the periodic direction, we have $1 \leq y_1 \leq x_1$, and $x_i \leq y_{i+1} \leq x_{i+1}$ for $i = 1, 2, \dots, n - 1$.

By self-avoidance, at most one line can go through the periodic direction. The two possible interlacings are shown for $n = 2$ in Fig. 16.

We now wish to construct n -particle states

$$|\Psi_n\rangle = \sum_{1 \leq x_1 < \dots < x_n \leq N} g(x_1, \dots, x_n) |x_1, \dots, x_n\rangle, \quad (5.3)$$

which are eigenvectors of T :

$$T|\Psi_n\rangle = \Lambda|\Psi_n\rangle. \quad (5.4)$$

To this end we try an Ansatz of the form

$$g(x_1, \dots, x_n) = \sum_{p \in \mathfrak{S}_n} A_p z_{p(1)}^{x_1} z_{p(2)}^{x_2} \cdots z_{p(n)}^{x_n}, \quad (5.5)$$

where the sum runs over all permutations $p \in \mathfrak{S}_n$ of the particle labels $\{1, 2, \dots, n\}$. The complex numbers z_j are related to the so-called quasi-momenta k_j through the relation $z_j = \exp(ik_j)$. For the moment this Ansatz can be considered loosely as “coupled plane waves”; we shall come back to its physical interpretation in due course.

5.2.1 Sector with $n = 0$ particles

When $n = 0$, the unique state is completely empty. The horizontal row of edges is either empty (i.e., completely filled with ω_1 vertices) or filled by



Figure 17: The four possible transitions between row states in the $n = 1$ particle sector.

a horizontal line (i.e., completely filled with ω_4 vertices). Thus T is one-dimensional and takes the value

$$\Lambda = a^N + b^N. \quad (5.6)$$

5.2.2 Sector with $n = 1$ particle

When $n = 1$, $\dim T = N$. The eigenvectors (5.3) read $|\Psi\rangle = \sum_x g(x)|x\rangle$, and the Ansatz (5.5) is $g(x) = z^x$. The particle at position x can undergo four different processes as shown in Fig. 17. The eigenvalue equation (5.4) projected on a basis state of the eigenvector (5.3) then becomes⁹

$$\begin{aligned} \Lambda z^x &= a^{N-1} b z^x + \sum_{y=x+1}^N a^{N-(y-x+1)} b^{y-x-1} c^2 z^y \\ &+ ab^{N-1} z^x + \sum_{y=1}^{x-1} a^{x-y-1} b^{N-(x-y+1)} c^2 z^y. \end{aligned} \quad (5.7)$$

Recalling now the geometric series

$$\sum_{n=N_1}^{N_2} \omega^n = \frac{\omega^{N_1} - \omega^{N_2+1}}{1 - \omega}$$

the first sum in (5.7) reads

$$a^N \left(\frac{a}{b}\right)^{x+1} \left(\frac{c}{a}\right)^2 \frac{\left(\frac{bz}{a}\right)^{x+1} - \left(\frac{bz}{a}\right)^{N+1}}{1 - \frac{bz}{a}}, \quad (5.8)$$

⁹To be precise, (5.7) contains the transition probabilities that x becomes y , hence it is the transcription of $\langle \Psi_n | T | x \rangle = \Lambda \langle \Psi_n | x \rangle$.

while the second sum is

$$b^N \left(\frac{a}{b}\right)^{x+1} \left(\frac{c}{a}\right)^2 \frac{\left(\frac{bz}{a}\right)^1 - \left(\frac{bz}{a}\right)^x}{1 - \frac{bz}{a}}. \quad (5.9)$$

Collecting everything, we first have the “wanted terms” multiplying $a^N z^x$. They come from the first term in (5.7) and the first term in (5.8), and their coefficient is

$$\frac{b}{a} + \frac{c^2 z}{a(a-bz)} = \frac{ab + (c^2 - b^2)z}{a(a-bz)} \equiv L(z). \quad (5.10)$$

Another class of wanted terms, multiplying $b^N z^x$, comes from the third term in (5.7) and the second term in (5.9). The coefficient of these terms is

$$\frac{a}{b} - \frac{c^2}{b(a-bz)} = \frac{a^2 - c^2 - abz}{b(a-bz)} \equiv M(z). \quad (5.11)$$

The remaining terms, namely the second term in (5.8) and the first term in (5.9), are “unwanted boundary terms” that read

$$\frac{a^{x-1} b^{N-x} c^2 z}{a-bz} (1 - z^N). \quad (5.12)$$

The unwanted terms cancel out provided we impose the following condition on the allowed quasi-momenta:

$$z^N = 1. \quad (5.13)$$

Note that this has precisely N solutions, and we have thus determined N eigenvectors for the N -dimensional matrix T . Its eigenvalue is given by the wanted terms and reads simply

$$\Lambda = a^N L(z) + b^N M(z). \quad (5.14)$$

5.2.3 Sector with $n = 2$ particles

When $n = 2$, $\dim T = \binom{N}{2} = \frac{N(N-1)}{2}$. The two possibilities that y_1, y_2 can interlace x_1, x_2 are shown in Fig. 16. Let us define a function $E(x, y)$ that contains the weight of a world-line entering at x and exiting at $y \geq x$

$$E(x, y) = \begin{cases} \frac{b}{c} & \text{if } y = x \\ cb^{y-x-1} & \text{if } y > x \end{cases} \quad (5.15)$$

and another function $D(y, x)$ giving the weight of the empty segment between two world-lines

$$D(y, x) = \begin{cases} \frac{a}{c} & \text{if } x = y \\ ca^{x-y-1} & \text{if } x > y \end{cases} . \quad (5.16)$$

The eigenvalue equation (5.4) projected on one basis state $|x_1, x_2\rangle$ of the eigenvector (5.3) then reads

$$\begin{aligned} \Lambda g(x_1, x_2) &= \sum_{y_1=x_1}^{x_2} \sum_{y_2=x_2}^N a^{x_1-1} E(x_1, y_1) D(y_1, x_2) E(x_2, y_2) ca^{N-y_2} g(y_1, y_2) \\ &+ \sum_{y_1=1}^{x_1} \sum_{y_2=x_1}^{x_2} b^{y_1-1} D(y_1, x_1) E(x_1, y_2) D(y_2, x_2) cb^{N-x_2} g(y_1, y_2) , \end{aligned}$$

where the two terms correspond to the situations shown in Fig. 16. Note that the special cases when one of the x coincides with one of the y are already provided for in the definitions of $E(x, y)$ and $D(y, x)$. However, the double sums must be constrained to exclude terms with $y_1 = y_2$. This is best done by first computing the sums without the constraint, then subtracting off the disallowed contribution $y_1 = x_2 = y_2$ to the first double sum, and $y_1 = x_1 = y_2$ to the second.

Despite of (5.5) we first insert the simpler Ansatz

$$g(x_1, x_2) = A_{12} z_1^{x_1} z_2^{x_2} .$$

It is convenient to introduce the short-hand notations

$$L_j \equiv L(z_j) , \quad M_j \equiv M(z_j) , \quad \rho_j \equiv \rho(z_j) = \frac{c^2 z_j}{a(a - bz_j)}$$

and to define the function

$$R_j(x_1, x_2) = L_j a^{x_2-x_1} z_j^{x_1} + M_j b^{x_2-x_1} z_j^{x_2} .$$

As before the terms coming from the constrained double summations are of several types:

Wanted terms

These wanted terms are those proportional to $z_1^{x_1} z_2^{x_2}$:

$$A_{12} (a^N L_1 L_2 + b^N M_1 M_2) z_1^{x_1} z_2^{x_2}$$

and they determine the eigenvalue

$$\Lambda = a^N L_1 L_2 + b^N M_1 M_2. \quad (5.17)$$

Unwanted internal terms

There are of the form $(z_1 z_2)^{x_2}$ or $(z_1 z_2)^{x_1}$. One can verify that both such terms are proportional to

$$M_1 L_2 - 1 = -\frac{c^2 s_{12}}{(a - bz_1)(a - bz_2)}, \quad (5.18)$$

where

$$s_{12} = 1 - 2\Delta z_2 + z_1 z_2, \quad (5.19)$$

$$\Delta = \frac{a^2 + b^2 - c^2}{2ab}. \quad (5.20)$$

The quantities s_{12} (scattering phase) and Δ (anisotropy parameter) play a very important role in the solution of the six-vertex model, and in the physical interpretation of the scattering theory described by the Bethe Ansatz. We shall come back to this later.

Unwanted boundary terms

These come from the $y_2 = N$ or the $y_1 = 1$ summation limits and their sum is

$$A_{12} a^{x_1} b^{N-x_2} (R_2(x_1, x_2) \rho_1 - R_1(x_1, x_2) \rho_2 z_2^N). \quad (5.21)$$

Elimination of the unwanted terms

The justification of the complete Ansatz (5.5) is precisely that it permits us to eliminate the unwanted terms. We therefore set

$$g(x_1, x_2) = A_{12} z_1^{x_1} z_2^{x_2} + A_{21} z_2^{x_1} z_1^{x_2}.$$

The unwanted internal terms cancel under the condition

$$s_{12}A_{12} + s_{21}A_{21} = 0. \quad (5.22)$$

The sum of the unwanted boundary terms is

$$a^{x_1}b^{N-x_2} \{ \rho_2 R_1(x_1, x_2)(A_{21} - z_2^N A_{12}) + \rho_1 R_2(x_1, x_2)(A_{12} - z_1^N A_{21}) \}$$

and this will vanish under the conditions

$$\begin{aligned} z_1^N &= \frac{A_{12}}{A_{21}} = -\frac{s_{21}}{s_{12}}, \\ z_2^N &= \frac{A_{21}}{A_{12}} = -\frac{s_{12}}{s_{21}}. \end{aligned} \quad (5.23)$$

5.3 Bethe Ansatz equations

The structure of the solution for the case of general n is very much visible in the above detailed treatment for $n = 2$. By generalising the argument (or proceeding by educated guesswork) it emerges that the eigenvalue is

$$\Lambda = a^N L_1 L_2 \cdots L_n + b^N M_1 M_2 \cdots M_n. \quad (5.24)$$

The condition for the vanishing of the unwanted internal terms becomes

$$s_{p_j, p_{j+1}} A_{p_1, \dots, p_j, p_{j+1}, \dots, p_n} + s_{p_{j+1}, p_j} A_{p_1, \dots, p_{j+1}, p_j, \dots, p_n} = 0 \quad (5.25)$$

for each $j = 1, 2, \dots, n - 1$ and all permutations $p \in \mathfrak{S}_n$. Finally, the condition for the vanishing of the unwanted boundary terms reads

$$z_{p_1}^N = \frac{A_{p_1, p_2, \dots, p_{n-1}, p_n}}{A_{p_2, p_3, \dots, p_n, p_1}} \quad (5.26)$$

for all $p \in \mathfrak{S}_n$.

There is a nice alternative way of deriving (5.26) using the consideration of translational invariance. Indeed the eigenstate must be unchanged upon taking any of the particles through the periodic boundary condition and back to its original position. In particular

$$g(x_1, x_2, \dots, x_{n-1}, x_n) = g(x_2, x_3, \dots, x_n, x_1 + N). \quad (5.27)$$

Note that this respects our conventions that the arguments of g must be written in increasing order. Using this, the form of the Ansatz (5.5) implies (5.26).

To see this, consider for simplicity the case of $n = 3$ particles. We have then

$$\begin{aligned} g(x_1, x_2, x_3) &= A_{123} z_1^{x_1} z_2^{x_2} z_3^{x_3} + A_{132} z_1^{x_1} z_3^{x_2} z_2^{x_3} + \dots, \\ g(x_2, x_3, x_1 + N) &= A_{231} z_2^{x_2} z_3^{x_3} z_1^{x_1 + N} + A_{321} z_3^{x_2} z_2^{x_3} z_1^{x_1 + N} + \dots \end{aligned}$$

Since this must be valid for all x_j we can identify terms

$$\begin{aligned} A_{231} z_1^N &= A_{123}, \\ A_{321} z_1^N &= A_{132}. \end{aligned} \tag{5.28}$$

This proves (5.26) for the case $n = 3$.

Obviously (5.25)–(5.26) provides many more equations than the n unknown quasi-momenta z_1, z_2, \dots, z_n . Generalising (5.19) we define

$$s_{ij}(z_i, z_j) = 1 - 2\Delta z_j + z_i z_j. \tag{5.29}$$

One then easily verifies that (5.25) is solved by

$$A_{p_1, p_2, \dots, p_n} = \epsilon_p \prod_{1 \leq i < j \leq n} s_{p_j, p_i}, \tag{5.30}$$

where ϵ_p is the signature of the permutation $p \in \mathfrak{S}_n$. Inserting this into (5.26) gives

$$z_{p_1}^N = (-1)^{n-1} \prod_{l=2}^n \frac{s_{p_l, p_1}}{s_{p_1, p_l}}$$

for all $p \in \mathfrak{S}_n$. But since the right-hand side is symmetric in p_2, p_3, \dots, p_n there are actually only n distinct equations:

$$z_j^N = (-1)^{n-1} \prod_{\substack{l=1 \\ l \neq j}}^n \frac{s_{l,j}}{s_{j,l}} \quad \text{for } j = 1, 2, \dots, n. \tag{5.31}$$

These are the Bethe Ansatz equations (BAE) for the six-vertex model.

The progress obtained by now is considerable. Rather than diagonalising a transfer matrix of dimension 2^N we have to solve only a set of n coupled (but non-linear) equations for each $n = 1, 2, \dots, N$.

5.3.1 Scattering phases and the Yang-Baxter equation

It is useful to define the modified scattering phases

$$\hat{S}_{ij}(z_i, z_j) = -\frac{s_{ij}}{s_{ji}} = -\frac{1 - 2\Delta z_j + z_i z_j}{1 - 2\Delta z_i + z_i z_j}. \quad (5.32)$$

The Bethe Ansatz equations can be written in the suggestive form

$$z_j^N = \prod_{\substack{l=1 \\ l \neq j}}^n \hat{S}_{lj}(z_l, z_j) \quad \text{for } j = 1, 2, \dots, n. \quad (5.33)$$

This can be interpreted physically as follows. When the particle j is taken around the periodic direction and back to its original position, it picks up a scattering phase \hat{S}_{lj} each time it crosses another particle l . These phases are also known as the S -matrix elements of the scattering theory.

We now derive some important physical properties of the S -matrix. Let us again focus on the case of $n = 3$ particles. Eliminating z_1^N from (5.28) we obtain $\frac{A_{123}}{A_{231}} = \frac{A_{132}}{A_{321}}$. Doing the same for z_2^N and z_3^N , and making some rearrangements, we arrive at

$$\frac{A_{213}}{A_{123}} = \frac{A_{321}}{A_{312}}, \quad \frac{A_{312}}{A_{132}} = \frac{A_{231}}{A_{213}}, \quad \frac{A_{321}}{A_{231}} = \frac{A_{132}}{A_{123}}. \quad (5.34)$$

This tells us that the interchange of two particles (e.g., 1 and 2 in the first relation) is independent of the position of the third particle (which on the left-hand side of the relations is to the right of the two particles being interchanges, and vice versa).

It thus emerges that the S -matrix possesses a locality property, according to which the scattering amplitude of n quasi-particles factorises into a product of $\binom{n}{2}$ two-particle S -matrices. To make this more precise, consider the following relations which follow from (5.30):

$$A_{321} = \begin{cases} \hat{S}_{12} A_{312} = \hat{S}_{12} \hat{S}_{13} A_{132} = \hat{S}_{12} \hat{S}_{13} \hat{S}_{23} A_{123} \\ \hat{S}_{23} A_{231} = \hat{S}_{23} \hat{S}_{13} A_{213} = \hat{S}_{23} \hat{S}_{13} \hat{S}_{12} A_{123} \end{cases}$$

Eliminating A_{123} yields the so-called Yang-Baxter relation

$$\hat{S}_{12} \hat{S}_{13} \hat{S}_{23} = \hat{S}_{23} \hat{S}_{13} \hat{S}_{12}, \quad (5.35)$$

which can be represented diagrammatically in terms of the world-lines of the particles:

$$\begin{array}{ccc}
 \begin{array}{c} \diagup \\ \diagdown \end{array} & \begin{array}{c} | \\ | \\ | \end{array} & \begin{array}{c} \diagdown \\ \diagup \end{array} \\
 \hat{S}_{13} & \hat{S}_{12} \\
 \hat{S}_{23} & \\
 1 & 2 & 3
 \end{array}
 =
 \begin{array}{ccc}
 \begin{array}{c} | \\ | \\ | \end{array} & \begin{array}{c} \diagdown \\ \diagup \end{array} & \begin{array}{c} \diagup \\ \diagdown \end{array} \\
 \hat{S}_{23} & \hat{S}_{13} \\
 \hat{S}_{12} & \\
 1 & 2 & 3
 \end{array}
 \tag{5.36}$$

The graphical reading of this diagram is that any world-line can be moved across the intersection of two other world-lines.

Obviously the above argument is at most suggestive, since after all the factors in (5.35) are just scalars, and as such the identity is trivial. We shall however see later that the same relation holds for matrix-valued quantities (operators), such as R -matrices and monodromy (\simeq transfer) matrices.

The Yang-Baxter equation (5.35) is at the heart of the algebraic approach to the Bethe Ansatz and we shall return to it extensively in later chapters.

5.4 Phase diagram

The six-vertex model has a non-trivial phase diagram, and it is hardly surprising that its thermodynamic limit depends on the parameter Δ . In fact the Bethe Ansatz equations (5.31) show that the limit depends *only* on Δ . Although the phase diagram can be derived in details, let us first discuss a few qualitative arguments.

If either a or b is large compared to the other weights, the system will freeze into a unique state in which all vertical arrows and all horizontal arrows point in the same direction. It turns out that this freezing occurs whenever $\Delta > 1$. The largest eigenvalue is that of the $n = 0$ particle sector, whence trivially $\Lambda_{\max} = a^N + b^N$. The free energy per vertex is therefore $f = \min(\epsilon_1, \epsilon_3)$.

If c is very large compared to the other weights, the predominant configuration is the one where all vertices on the even (resp. odd) sublattice are of the type ω_5 (resp. ω_6). The system is however not frozen, meaning that it exhibits fluctuations around the predominant configuration.¹⁰ One would

¹⁰We shall discuss the nature of these fluctuations more carefully when dealing with the Coulomb gas. Suffice it to say here that the least possible change of a configuration is to reverse a path of consistently oriented arrows. Any such path has infinite length in the

expect non-critical behaviour as long as c remains reasonably large. This is indeed the case: when $\Delta < -1$ the system is in a non-critical phase.

The most interesting phase occurs for $-1 < \Delta < 1$. In that range the six-vertex model is critical, and it turns out that the critical exponents depend continuously on Δ . This is an interesting counter-example to naive ideas of universality.

We shall concentrate most of the subsequent discussion on the critical case $-1 < \Delta < 1$, where the free energy can be expressed in terms of Fourier integrals. (The non-critical case $\Delta < -1$ can also be worked out in details and calls instead for the use of Fourier series.)

5.5 Thermodynamic limit for $\Delta < 1$

It is not known how to solve the Bethe Ansatz equations (5.31) for finite n and N . This situation is quite common in the study of integrable systems. By contrast, we have seen that the partition function of dimer coverings can be exactly computed on a finite lattice—a highly unusual situation.

Nevertheless, it turns out that the six-vertex model is exactly solvable in the thermodynamic limit. By this we mean precisely that the free energy $f = -\frac{1}{\beta N} \log \Lambda_{\max}$, or equivalently the ground state energy in the spin chain, can be determined analytically for $N \rightarrow \infty$. The same is true for the low-lying excitations, but for the moment we concentrate on the ground state.

5.5.1 Location of the quasi-momenta

The BAE (5.31) possess many solutions for the quasi-momenta z_j . It is not a priori clear which one corresponds to the ground state. In what follows we shall admit the following fact:

- The solution of the BAE (5.31) that maximises the eigenvalue Λ is such that z_1, z_2, \dots, z_n are distinct, lie on the unit circle, are distributed symmetrically about unity, and are packed as closely as possible.

This can actually be proved quite rigorously, using some lengthy analysis [YY66]. An easier method, that usually works quite well for more general

frozen phase, whereas the $\Delta < -1$ has an exponential number of short paths (of length four).

integrable models, is to study numerically the solutions for low values of N —confronting the results with exact diagonalisations of the transfer matrix—until the pattern has become clear.

5.5.2 Transformation to a set of real equations

We introduce the momenta $k_j \in \mathbb{R}$ and the function $\Theta(p, q)$ so that

$$z_j = \exp(ik_j), \quad (5.37)$$

$$\frac{s_{i,j}}{s_{j,i}} = \exp(-i\Theta(k_j, k_i)). \quad (5.38)$$

By (5.32) we have then

$$e^{-i\Theta(p,q)} = \frac{1 - 2\Delta e^{ip} + e^{i(p+q)}}{1 - 2\Delta e^{iq} + e^{i(p+q)}}. \quad (5.39)$$

To see that $\Theta(p, q)$ is a real function, it suffices to notice that

$$\tan\left(\frac{1}{2}\Theta(p, q)\right) = -i \frac{1 - e^{-i\Theta(p,q)}}{1 + e^{-i\Theta(p,q)}} = \frac{\Delta \sin\left(\frac{p-q}{2}\right)}{\cos\left(\frac{p+q}{2}\right) - \Delta \cos\left(\frac{p-q}{2}\right)},$$

where the right-hand side is manifestly real.

Including the term $l = j$ obviously leaves the right-hand side of (5.31) unchanged, so we can rewrite it as

$$\exp(iNk_j) = (-1)^{n-1} \prod_{l=1}^n \exp(-i\Theta(k_j, k_l)),$$

where now both sides of the equation are unimodular. Taking logarithms we have

$$Nk_j = 2\pi I_j - \sum_{l=1}^n \Theta(k_j, k_l), \quad (5.40)$$

where I_j ranges between $\pm\left(\frac{n-1}{2}\right)$, hence is an integer (resp. half an odd integer) if n is odd (resp. even). Note that both sides of this equation are real.

The hypothesis that k_1, k_2, \dots, k_n be distinct, symmetrically distributed about the origin, and packed as closely as possible implies that the ground state is obtained by choosing

$$I_j = j - \frac{1}{2}(n+1), \quad \text{for } j = 1, 2, \dots, n. \quad (5.41)$$

5.5.3 Continuum limit

The thermodynamic limit is obtained by sending $n, N \rightarrow \infty$, while keeping the ratio n/N fixed and finite. This ratio describes the (fixed) ratio of up-pointing arrows in each row of the lattice. The distribution function $\rho(k)$ of Bethe roots is defined so that $N\rho(k) dk$ is the number of k_j lying between k and $k + dk$. By assumption $\rho(k)$ has support on a symmetric interval $[-Q, Q]$, where Q will be determined later. Thus

$$\int_{-Q}^Q \rho(k) dk = \frac{n}{N}. \quad (5.42)$$

For a given value k_j of k , the quantity $I_j + \frac{1}{2}(n + 1) = N \int_{-Q}^k \rho(k') dk'$ is the number of momenta k_l with $l < j$. Passing from sums to integrals in (5.40)—and denoting k_j simply as k —then produces

$$Nk = -\pi(n + 1) + 2\pi N \int_{-Q}^k \rho(k') dk' - N \int_{-Q}^Q \Theta(k, k') \rho(k') dk'.$$

Taking derivatives with respect to k , and dividing by N , then leads to a linear integral equation for $\rho(k)$

$$2\pi\rho(k) = 1 + \int_{-Q}^Q \frac{\partial\Theta(k, k')}{\partial k} \rho(k') dk'. \quad (5.43)$$

The free energy is then given by (5.24) as

$$f = -\frac{1}{\beta} \max \left\{ \log a + \frac{1}{N} \sum_{j=1}^n \log L(z_j), \log b + \frac{1}{N} \sum_{j=1}^n \log M(z_j) \right\}.$$

In the thermodynamic limit this becomes

$$f = -\frac{1}{\beta} \max \left\{ \log a + \int_{-Q}^Q [\log L(e^{ik})] \rho(k) dk, \log b + \int_{-Q}^Q [\log M(e^{ik})] \rho(k) dk \right\}. \quad (5.44)$$

5.6 Free energy for $-1 < \Delta < 1$

A natural strategy for solving the linear integral equation (5.43) would be to use Fourier transformation. This is however only possible if we can find a transformation to a difference kernel. Fortunately this is possible.

5.6.1 Difference kernel transformation

For $-1 < \Delta < 1$ we parameterise

$$\Delta = -\cos \mu, \quad \text{with } 0 < \mu < \pi \quad (5.45)$$

and we trade k for a new variable α defined by

$$e^{ik} = \frac{e^{i\mu} - e^\alpha}{e^{i\mu+\alpha} - 1}. \quad (5.46)$$

Note that $\alpha \in \mathbb{R}$. Indeed, supposing this, it is easily seen that $|e^{ik}|^2 = 1$ from the right-hand side of (5.46), as is consistent with the hypothesis that $k \in \mathbb{R}$.

Differentiating logarithmically—i.e., using $\frac{d}{d\alpha} e^{ik} = ie^{ik} \frac{dk}{d\alpha}$ to isolate $\frac{dk}{d\alpha}$ —we find

$$\frac{dk}{d\alpha} = \frac{\sin \mu}{\cosh \alpha - \cos \mu}. \quad (5.47)$$

This proves in particular that $k(\alpha) \in \mathbb{R}$ is a monotonically increasing function (since $0 < \mu < \pi$), and by (5.46) it maps the interval $(-\infty, \infty)$ onto $(\mu - \pi, \pi - \mu)$. It follows directly from (5.46) that $k(-\alpha) = -k(\alpha)$, i.e., the function is odd.

The change of variables (5.46) is not as miraculous as it may first appear. Indeed consider the map $z : \mathbb{C} \mapsto \mathbb{C}$ given by

$$z(w) = \frac{a_{11}w + a_{12}}{a_{21}w + a_{22}}. \quad (5.48)$$

This conformal transformation, variously known as a projective map or a Möbius transformation, will play a major role in chapter 10. It has the property that it preserves the set of circles and straight lines (the latter being considered circles of infinite radius). After a normalisation, usually taken as $\det a_{ij} = 1$, it depends on three complex parameters.

To make contact with (5.46) we set $z = e^{ik}$ and $w = e^\alpha$. We then have $|z| = 1$ (a circle), whereas $w \in \mathbb{R}_+$ (a straight line). We also abbreviate the free parameter in (5.46) as $\rho = e^{i\mu}$. The desired symmetry property $k(-\alpha) = -k(\alpha)$ can be rewritten $z(\frac{1}{w}) = \frac{1}{z(w)}$. Using (5.48), this is solved by $a_{11} = a_{22}$ and $a_{12} = a_{21}$. If we now fix the global scale by setting $z(0) = \frac{a_{12}}{a_{11}} = \frac{1}{z(\infty)} = -\rho$, we obtain precisely (5.46).

Let us define $p = k(\alpha)$ and $q = k(\beta)$, so that

$$e^{ip} = \frac{e^{i\mu} - e^\alpha}{e^{i\mu+\alpha} - 1}, \quad e^{iq} = \frac{e^{i\mu} - e^\beta}{e^{i\mu+\beta} - 1}.$$

Inserting this into (5.39) the scattering phase becomes

$$e^{-i\Theta(p,q)} = \frac{e^{\alpha-\beta} - e^{2i\mu}}{e^{\beta-\alpha} - e^{2i\mu}}. \quad (5.49)$$

Crucially, this depends only on the difference $\alpha - \beta$ (and on the constant μ).

We shall need the root density function $R(\alpha)$ transformed to the α variable (and renormalised by $\frac{1}{2\pi}$ for later convenience)

$$R(\alpha) d\alpha = 2\pi\rho(k) dk. \quad (5.50)$$

Plugging this into (5.43) leads to

$$R(\alpha) = \frac{dk}{d\alpha} - \frac{1}{2\pi} \int_{-Q_1}^{Q_1} \frac{\partial\Theta(\alpha, \beta)}{\partial\beta} R(\beta) d\beta,$$

where we note that there is a new integration range $(-Q_1, Q_1)$ corresponding to the α variable. Notice also the sign change on the second term, because the dependence is on $\alpha - \beta$ and we now derive with respect to β . Computing the derivatives from (5.47) and (5.49) finally leads to

$$R(\alpha) = \frac{\sin \mu}{\cosh \alpha - \cos \mu} - \frac{1}{2\pi} \int_{-Q_1}^{Q_1} \frac{\sin(2\mu)}{\cosh(\alpha - \beta) - \cos(2\mu)} R(\beta) d\beta, \quad (5.51)$$

and the normalisation condition (5.42) for the root density function now reads

$$\frac{1}{2\pi} \int_{-Q_1}^{Q_1} R(\alpha) d\alpha = \frac{n}{N}. \quad (5.52)$$

5.6.2 Parameterisation

We have already parameterised $\Delta = -\cos \mu$ in (5.45). The Bethe Ansatz equations (5.31)—and hence the universality class of the six-vertex model—depend only on the initial vertex weights a, b, c through $\Delta = \frac{a^2+b^2-c^2}{2ab}$. We must therefore choose a parameterisation of the two independent ratios $a :$

$b : c$ that respects this latter constraint (we “uniformise the spectral curve”). This can be done in this case using trigonometric functions:

$$a : b : c = \sin\left(\frac{\mu - w}{2}\right) : \sin\left(\frac{\mu + w}{2}\right) : \sin \mu, \quad -\mu < w < \mu \quad (5.53)$$

defining another parameter w .

The eigenvalues of the transfer matrix are determined by (5.24) through the functions $L(z)$ and $M(z)$ given by (5.10)–(5.11). Recalling that $z = e^{ik}$ is parameterised by (5.46), the parametric form of these functions now becomes

$$\begin{aligned} L(e^{ik}) &= \frac{e^{i(w+\mu)} - e^{\alpha-i\mu}}{e^\alpha - e^{iw}}, \\ M(e^{ik}) &= \frac{e^{i(w-\mu)} - e^{\alpha+i\mu}}{e^\alpha - e^{iw}}. \end{aligned} \quad (5.54)$$

5.6.3 Solution by Fourier integrals

The integral equation (5.51) that determines the root density function now has a difference kernel. One can therefore solve it by Fourier transformation, provided that $Q_1 = \infty$. Let us suppose that this is so, and justify the assumption below. The Fourier transformed root density function then reads

$$\tilde{R}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\alpha) e^{ix\alpha} d\alpha. \quad (5.55)$$

Let us define the function

$$\phi_\mu(\alpha) = \frac{\sin \mu}{\cosh \alpha - \cos \mu}, \quad (5.56)$$

which is often referred to as the *source term* of the Bethe Ansatz equations. The difference kernel equation (5.51) then decouples upon Fourier transformation, since the Fourier transform of a convolution is the product of Fourier transforms. Explicitly, multiplying both sides of (5.51) by $\frac{1}{2\pi} e^{ix\alpha}$ and integrating over α leads to

$$\tilde{R}(x) = \tilde{\phi}_\mu(x) - \tilde{\phi}_{2\mu}(x) \cdot \tilde{R}(x). \quad (5.57)$$

Exercise: Show that the Fourier transform of the source term is

$$\widetilde{\phi}_\mu(x) = \frac{\sinh((\pi - \mu)x)}{\sinh(\pi x)}. \quad (5.58)$$

Inserting this yields

$$\widetilde{R}(x) = \frac{\sinh((\pi - \mu)x)}{\sinh(\pi x)} - \frac{\sinh((\pi - 2\mu)x)}{\sinh(\pi x)} \widetilde{R}(x) \quad (5.59)$$

and we can then finally isolate

$$\widetilde{R}(x) = \frac{1}{2 \cosh(\mu x)}. \quad (5.60)$$

The normalisation condition (5.52) is such that $\widetilde{R}(0) = \frac{n}{N}$, and evaluating (5.60) we arrive at

$$\frac{n}{N} = \frac{1}{2}. \quad (5.61)$$

This simple result justifies the assumption $Q_1 = \infty$ a posteriori. Indeed, the largest sector of the transfer matrix precisely corresponds to the case where there are as many up-pointing as down-pointing arrows. By a simple entropic reasoning, this is also the ground state sector.¹¹

From (5.54) one obtains

$$\begin{aligned} |L(e^{ik})|^2 &= \frac{\cos(w + 2\mu) - \cosh \alpha}{\cos w - \cosh \alpha}, \\ |M(e^{ik})|^2 &= \frac{\cos(w - 2\mu) - \cosh \alpha}{\cos w - \cosh \alpha}. \end{aligned}$$

This implies that $|L| > |M|$ for $w < 0$, and $|L| < |M|$ for $w > 0$.

Suppose in the sequel that $w < 0$; a similar calculation for $w > 0$ can be shown to lead to exactly the same end result. The free energy is then given by the first term in (5.44):

$$f = -\frac{1}{\beta} \left(\log a + \frac{1}{2\pi} \int_{-\infty}^{\infty} [\log |L(e^{ik})|] R(\alpha) d\alpha \right) \quad (5.62)$$

¹¹A variant argument is obtained by examining (5.24).

Using (5.60) and the fact that parity is conserved by Fourier transformation, we see that $R(\alpha)$ is an even function. Under the integral we can therefore replace the other factor $\log |L(e^{ik})|$ by its even part, which is also its real part. From (5.54) we get

$$\operatorname{Re} L(e^{ik}) = -\cos \mu + \frac{\sin \mu \sin w}{\cos w - \cosh \alpha}, \quad (5.63)$$

and the Fourier transform of $\log |L(e^{ik})|$ becomes

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix\alpha} \log |L(e^{ik})| d\alpha = \frac{\sinh((\mu + w)x) \sinh((\pi - \mu)x)}{x \sinh(\pi x)}. \quad (5.64)$$

Exercise: Detail this computation!

To compute (5.62) we can use that the Fourier transform of a product is the convolution of Fourier transforms. The end result follows by combining (5.60) and (5.64):

$$f = -\frac{1}{\beta} \left(\log a + \int_{-\infty}^{\infty} \frac{\sinh((\mu + w)x) \sinh((\pi - \mu)x)}{2x \cosh(\mu x) \sinh(\pi x)} dx \right). \quad (5.65)$$

As already stated, exactly the same result is found for $w > 0$. We have therefore found, for any $w \in (-\mu, \mu)$, the free energy of the six-vertex model in the critical region $\Delta \in (-1, 1)$.

5.6.4 Ice model

The equal-weighted case $a = b = c = 1$ is of special interest, both combinatorially and historically (it was solved by Lieb [Li67] before the general case). Physically it can be interpreted as a two-dimensional model of ice: The arrows on the edges represent to which side the electron cloud is pushed by the hydrogen bonding, and the six-vertex constraint corresponds to local charge neutrality.

In this case $w = 0$ and $\mu = \frac{2\pi}{3}$ from (5.53). The integral (5.65) can then be performed explicitly by contour integration; this is true more generally whenever μ is a rational fraction of π .

We have

$$-\beta f = \int_{-\infty}^{\infty} \frac{\sinh\left(\frac{\pi x}{3}\right) \tanh\left(\frac{2\pi x}{3}\right)}{2x \sinh(\pi x)} dx. \quad (5.66)$$

The integrand is regular for $x \rightarrow 0$, even, and decays like $\frac{1}{2x}e^{-2\pi x/3}$ for $\text{Re } x \gg 1$. We can therefore close the contour in the upper-half plane and use the Cauchy integration theorem.

The residues are all on the imaginary axis. They read for $p \in \mathbb{N}$:

$$\begin{aligned} \frac{-3i}{4(3p + \frac{3}{4})\pi} & \text{ at } x = (3p + \frac{3}{4})i, \\ \frac{3i}{4(3p + 1)\pi} & \text{ at } x = (3p + 1)i, \\ \frac{3i}{4(3p + 2)\pi} & \text{ at } x = (3p + 2)i, \\ \frac{-3i}{4(3p + \frac{9}{4})\pi} & \text{ at } x = (3p + \frac{9}{4})i. \end{aligned}$$

Thus

$$\begin{aligned} -\beta f &= 2\pi i \sum_{\text{Im}(x)>0} \text{Res } f(x) \\ &= -\frac{3}{2} \sum_{p=0}^{\infty} \left(-\frac{1}{3p + \frac{3}{4}} + \frac{1}{3p + 1} + \frac{1}{3p + 2} - \frac{1}{3p + \frac{9}{4}} \right) \\ &= \frac{5}{2} \sum_{p=0}^{\infty} \frac{(1 + 2p)}{(1 + 3p)(2 + 3p)(1 + 4p)(3 + 4p)} \\ &= \frac{3}{2} \log \left(\frac{4}{3} \right). \end{aligned} \tag{5.67}$$

The effective number of configurations per vertex

$$Z^{1/MN} = \exp(-\beta f) = \left(\frac{4}{3} \right)^{\frac{3}{2}} \simeq 1.539\,600 \dots \tag{5.68}$$

is known as Lieb's constant [Li67].

6 Yang-Baxter equation

In the preceding chapter we have seen how to solve the six-vertex model by the coordinate Bethe Ansatz. A natural next step is to identify the structures that make such a solution possible, and then to exploit those structures to solve more general classes of models. One would expect the structures in question to be of algebraic nature.

Obviously this programme will require to first take a slightly more abstract point of view. We shall expose the necessary ingredients in the first half of this chapter, and then connect them concretely to the six-vertex model in the second half.

6.1 R -matrix

Let us start by highlighting some crucial results from our analysis of the six-vertex model:

1. The three Boltzmann weights of the six-vertex model define two independent ratios $a : b : c$. From these we have defined two parameters:
 - A parameter $\Delta = \frac{a^2+b^2-c^2}{2ab}$ which characterises the universality class of the model. Only this combination appears in the Bethe Ansatz equations, via the scattering amplitudes.
 - Another uniformising parameter—called w in (5.53)—on which the ratios $a : b : c$ depend, but which does not change the value of Δ .
2. The scattering amplitudes—encoded in the S -matrix—are such that that multi-particle scattering of several quasi-particles factorise into a product of two-particle scattering amplitudes.

We now attempt to formalise these properties for statistical models of a particular type, defined on a so-called *Baxter lattice*. By this we mean any lattice that can be drawn in the plane as a collection of lines (one can think of them as straight lines, but this is not necessary) that undergo only pairwise intersections. The lattice does not at all need to be regular. The degrees of freedom live on the edges of the lattice, and interactions take place at the vertices.

Moreover, we suppose that to each line is associated a so-called *spectral parameter*, analogous to the uniformising parameter discussed above. The Boltzmann weights for a vertex where two lines with spectral parameters u and v intersect is supposed to have the *difference property*: it must depend only on the difference $u - v$. The weights of course also depends on Δ , and on the states of the degrees of freedom defined on the edges adjacent on the vertex.

The Boltzmann weights can be encoded as the matrix element of a linear operator, called the R -matrix:

$$\begin{array}{c}
 \beta_i \\
 | \\
 \mu_i \xrightarrow{u} \text{---} \mu_{i+1} \\
 | \\
 v \uparrow \\
 \alpha_i
 \end{array}
 = R_{\mu_i \alpha_i}^{\mu_{i+1} \beta_i}(u - v) = {}_a \langle \mu_{i+1} | \otimes {}_i \langle \beta_i | R_{ai}(u - v) | \mu_i \rangle_a \otimes | \alpha_i \rangle_i$$

(6.1)

At the risk of appearing pedantic, let us explain very carefully this notation:

- The indices μ_i, α_i and μ_{i+1}, β_i label the statistical degrees of freedom defined on the lattice edges. The notation corresponds exactly to what we have seen in the definition of the transfer matrix for the dimer problem. The first pair of indices is the in-state, and the second pair is the out-state. The time evolution thus goes in the North-East direction.
- A spectral parameter u is attached to the horizontal line, and v to the vertical line. The R -matrix depends on the difference $u - v$. To avoid any confusion about the sign, the lines carry an orientation.¹² When looking along the direction of the time evolution, the argument of R is the spectral parameter seen on one's left (namely u) *minus* the spectral parameter seen on one's right (namely v), both counted with a sign that reflects the orientation of the lines along the direction of sight.
- We shall often refer to the spaces α and β as *quantum* and to the space μ as *auxiliary*.
- The R -matrix acting between the auxiliary space and the i th quantum is denoted by R_{ai} . Its components are denoted $R_{\mu_i \alpha_i}^{\mu_{i+1} \beta_i}$. Note that the

¹²These arrows should of course not be confused with the (six-vertex) arrow degrees of freedom that live on lattice edges!

order of the out-indices has been permuted: this convention defines the R -matrix. One sometimes encounter the opposite, unpermuted convention: this defines what is called the \check{R} -matrix.

More formally, the R -matrix is a linear operator

$$R_{ai} : V_a \otimes V_i \mapsto V_a \otimes V_i, \quad (6.2)$$

where the vector spaces V_a (auxiliary) and V_i (quantum) carry the edge degrees of freedom. For instance, in the six-vertex model they are both equal to the spin- $\frac{1}{2}$ representation space \mathbb{C}^2 , since each arrow can be in two possible states: the R -matrix is then a 4×4 matrix.

The transfer matrix t is an endomorphism on the tensor product of all quantum spaces

$$t : V_1 \otimes V_2 \otimes \cdots \otimes V_L \mapsto V_1 \otimes V_2 \otimes \cdots \otimes V_L. \quad (6.3)$$

It can be written as

$$t = \text{Tr}_a (R_{aL} R_{aL-1} \cdots R_{a2} R_{a1}), \quad (6.4)$$

where Tr_a denotes the trace over the auxiliary space V_a . For simplicity we have not written the dependence on the spectral parameters. Indeed, one has the possibility of taking *different* spectral parameters for each quantum space V_i , and also for V_a , which will correspond to a completely inhomogeneous lattice model. The matrix elements of t can be written very explicitly as

$$\langle \beta | t | \alpha \rangle = \sum_{\mu_1, \dots, \mu_L} R_{\mu_L \alpha_L}^{\mu_1 \beta_L} R_{\mu_{L-1} \alpha_{L-1}}^{\mu_L \beta_{L-1}} \cdots R_{\mu_2 \alpha_2}^{\mu_3 \beta_2} R_{\mu_1 \alpha_1}^{\mu_2 \beta_1}. \quad (6.5)$$

Note that μ_1 appears both in the rightmost and the leftmost factor, so we indeed perform the operator Tr_a .

Before going on, it is legitimate to ask oneself whether this formalism is general enough to accommodate “all” statistical models of interest:

- A first question concerns the generality of the Baxter lattice. This of course encompasses all regular lattices for which all vertices are of degree four. What then about other lattices, such as the hexagonal and triangular lattices, which are widely used in statistical physics? One can usually find one’s way out by making suitable transformations.

For instance, on the hexagonal lattice one can shrink all edges along one of the three principal direction to zero, thus regrouping pairs of 3-valent vertices to form 4-valent vertices:¹³ This turns the hexagonal lattice into a square lattice. The triangular lattice can be turned into a hexagonal lattice by duality, or into a Kagomé lattice (which is a Baxter lattice) by going to the medial lattice. Other possible tricks include decimation procedures.

- A second question is how to deal with situations where the statistical degrees of freedom are not defined on the edges, but rather on the vertices. Going to suitably defined dual variables will usually enable us to transform such a model into one involving edge variables.
- Finally, how could one deal with long-range interactions, mediated by spatially extended objects, such as clusters or loops? One possibility is to define the R -matrix on appropriate representation spaces that take into account the non-locality of the interaction. Another option is to transform the model into one with local interactions. We shall see both possibilities at play in our subsequent treatment of the Potts model, which can be represented either as a Temperley-Lieb loop model with non-local interactions, or transformed into a six-vertex model with complex Boltzmann weights.

6.2 Commuting transfer matrices

A statistical model defined on a Baxter lattice is said to be integrable provided its R -matrix satisfies the Yang-Baxter equation and the inversion relation. The Yang-Baxter relation reads pictorially:

$$(6.6)$$

¹³This is a lattice version of the Hubbard-Stratonovich transformation used in field theory.

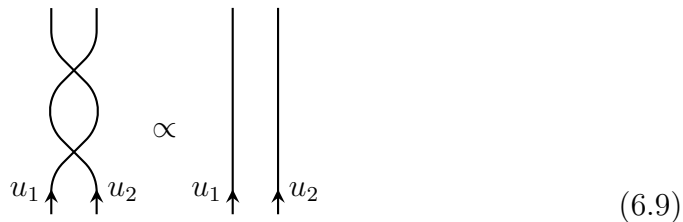
The algebraic transcription is

$$R_{12}(u)R_{13}(u+v)R_{23}(v) = R_{23}(v)R_{13}(u+v)R_{12}(u), \quad (6.7)$$

where we have set $u = u_1 - u_2$ and $v = u_2 - u_3$, so that $u_1 - u_3 = u + v$. We stress again that the spectral parameters u_i always follow the lines of the Baxter lattice. The same is true for the labels of the representation spaces, that appear as subscripts for the R -matrix. It is sometimes convenient to have these labels stay well-ordered in space (i.e., with 1 on the left, 2 in the middle, and 3 on the right) at all times (in the diagram time flows upwards). In that case one uses instead the \check{R} -matrix, for which the Yang-Baxter equation reads

$$\check{R}_{23}(u)\check{R}_{12}(u+v)\check{R}_{23}(v) = \check{R}_{12}(v)\check{R}_{23}(u+v)\check{R}_{12}(u). \quad (6.8)$$

The inversion relation can be represented pictorially as



$$\begin{array}{c} \text{Figure-eight diagram} \\ \propto \\ \text{Two parallel lines diagram} \end{array} \quad (6.9)$$

and reads algebraically

$$R_{12}(u)R_{12}(-u) \propto I. \quad (6.10)$$

The constant of proportionality could of course be set to unity by a suitable rescaling of R . Note also how the sign convention for spectral parameters comes into use when writing (6.10).

We shall show below that the R -matrix of the six-vertex model indeed satisfies (6.7) and (6.10).

The relations (6.7) and (6.10) imply the commutation of two transfer matrices corresponding to different choices of spectral parameters on the

auxiliary lines. This is best demonstrated graphically:

$$(6.11)$$

The first picture represents the product $t(u_2)t(u_1)$, since the two crossings to the left amount to the identity by (6.10).¹⁴ In the second picture we have used (6.7) to push the v_1 line to the left. This is repeated in the third picture for the next v_2 line. Repeating this operation L times, we finally arrive at the last picture, which represents the product $t(u_1)t(u_2)$, apart from the crossings on the left and right. But the right crossing can be taken around the periodic boundary condition (more formally: we are using the cyclicity of the trace), and using once more (6.10) the two crossings annihilate. Summarising, we have shown that

$$t(u_2)t(u_1) = t(u_1)t(u_2). \quad (6.12)$$

The existence of an infinite family of commuting transfer matrices has important consequences. Indeed the Bethe Ansatz technique permits us to diagonalise all these transfer matrices simultaneously.

Moreover, we can take derivatives of (6.12) with respect to u_2 . All these derivatives commute with $t(u_1)$, hence are conserved by the time evolution process. In other words, an integrable system has an infinite number of conserved quantities. The first few derivatives can be identified with the Hamiltonian, the momentum operator, and so on. We shall present explicit examples below.

Note also that the various vector spaces in which the R -matrices act need not be isomorphic. In particular, one can have different representations on the quantum and auxiliary spaces. From a basic integrable model—such as

¹⁴The transfer matrices depend also on the spectral parameters v_1, v_2, \dots, v_L of the quantum spaces, but we omit this dependence for notational convenience.

the six-vertex model in the spin- $\frac{1}{2}$ representation—one can construct higher-spin solutions by appropriate fusions of representation spaces. One speaks in that case of descendent models.

6.3 Six-vertex model

Let us parametrise the weights of the six-vertex model (cf. Fig. 14) as follows:

$$\begin{aligned}\omega_1 = \omega_2 &= \sin(\gamma - u), \\ \omega_3 = \omega_4 &= \sin u, \\ \omega_5 &= e^{-i(u-\eta)} \sin \gamma, \\ \omega_6 &= e^{i(u-\eta)} \sin \gamma.\end{aligned}\tag{6.13}$$

We have then $\Delta = -\cos \gamma$. The gauge parameter η can be chosen at will, since vertices of type 5 and 6 appear in pairs, and only the value of $\sqrt{\omega_5 \omega_6}$ enters the computation of the partition function.

Denote by $\mathbb{C}^2 = 0, 1$ the occupation number of each edge in the particle picture defined by the lower half of Fig. 14. The \check{R} -matrix can then be written in the basis $(\mathbb{C}^2)^2 = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as

$$\check{R} = \begin{bmatrix} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_5 & \omega_4 & 0 \\ 0 & \omega_3 & \omega_6 & 0 \\ 0 & 0 & 0 & \omega_2 \end{bmatrix}.\tag{6.14}$$

In the gauge $\eta = 0$ this becomes

$$\check{R}(u) = \begin{bmatrix} \sin(\gamma - u) & 0 & 0 & 0 \\ 0 & e^{-iu} \sin \gamma & \sin u & 0 \\ 0 & \sin u & e^{iu} \sin \gamma & 0 \\ 0 & 0 & 0 & \sin(\gamma - u) \end{bmatrix}.\tag{6.15}$$

We now claim that this \check{R} -matrix satisfies the Yang-Baxter equation (6.8), where the uniformising parameter u has been identified with the spectral parameter.

It is an instructive exercise to verify this. In tensor notation (6.8) reads in the space $(\mathbb{C}^2)^3$

$$(I \otimes \check{R}(u))(\check{R}(u+v) \otimes I)(I \otimes \check{R}(v)) = (\check{R}(v) \otimes I)(I \otimes \check{R}(u+v))(\check{R}(u) \otimes I).\tag{6.16}$$

This identity between 8×8 matrices is greatly simplified by the symmetries of the problem. First, the number of particles is conserved. Second, the weights are invariant under a global negation ($0 \leftrightarrow 1$) of the occupation numbers.

The only equations to be verified thus concern a 1×1 matrix (in the 0-particle space $|000\rangle$) and a 3×3 matrix (in the 1-particle space $|100\rangle, |010\rangle, |001\rangle$). Only the latter gives rise to non-trivial equations.

The inversion relation (6.10) reads

$$\check{R}(u)\check{R}(-u) = \sin(\gamma - u)\sin(\gamma + u)I. \quad (6.17)$$

6.3.1 Temperley-Lieb algebra

The \check{R} -matrix (6.15) can be decomposed as

$$\check{R}(u) = \sin(\gamma - u)I + \sin(u)E, \quad (6.18)$$

where I is the identity operator in $V^2 = (\mathbb{C}^2)^2$ and

$$E = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & e^{-i\gamma} & 1 & 0 \\ 0 & 1 & e^{i\gamma} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (6.19)$$

The operator E satisfies the basis-independent relations

$$\begin{aligned} E^2 &= 2 \cos \gamma E, \\ (E \otimes I)(I \otimes E)(E \otimes I) &= E \otimes I, \\ (I \otimes E)(E \otimes I)(I \otimes E) &= I \otimes E, \end{aligned} \quad (6.20)$$

where now I is the identity in V . For a system of width L one defines on $V_L = V^{\otimes L}$ a family of $L - 1$ such operators:

$$E_m = I^{\otimes m-1} \otimes E \otimes I^{\otimes L-m-1}, \quad \text{for } m = 1, 2, \dots, L - 1. \quad (6.21)$$

They verify the relations

$$\begin{aligned} (E_m)^2 &= 2 \cos \gamma E_m, \\ E_m E_{m\pm 1} E_m &= E_m, \\ E_m E_{m'} &= E_{m'} E_m \text{ for } |m - m'| > 1 \end{aligned} \quad (6.22)$$

defining the so-called Temperley-Lieb (TL) algebra.

The TL algebra plays a major role in lattice models of statistical mechanics, the Potts model in particular. In addition to the above spin- $\frac{1}{2}$ arrow representation, the TL algebra can be represented in terms of Fortuin-Kasteleyn clusters, their surrounding loops, domain walls of Potts spins, two-row Young tableaux, and much more. We shall come back to some of those issues in a later chapter.

6.3.2 Pauli matrices

It is convenient to make manifest the spin- $\frac{1}{2}$ nature of the six-vertex model by reexpressing things in terms of the Pauli matrices

$$\sigma^x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (6.23)$$

The arrow conservation then means that the transfer matrix $t(u)$ commutes with the total magnetisation

$$S^z = \frac{1}{2} \sum_{m=1}^L \sigma_m^z. \quad (6.24)$$

The Temperley-Lieb generator can be written as

$$E_m = \frac{1}{2} \left[\sigma_m^x \sigma_{m+1}^x + \sigma_m^y \sigma_{m+1}^y - \cos \gamma (\sigma_m^z \sigma_{m+1}^z - I) - i \sin \gamma (\sigma_m^z - \sigma_{m+1}^z) \right]. \quad (6.25)$$

6.3.3 Spectral parameter and anisotropy

The physical meaning of the spectral parameter u is that it controls the spatial anisotropy of the system. To see this qualitatively, note that in the $u \rightarrow 0$ limit, the \tilde{R} -matrix is proportional to the identity by (6.18). The transfer matrix $t(u)$ thus acts on a state just by shifting all spins one unit to the right (with periodic boundary conditions); note that this follows from the fact that time propagates in the North-East direction.

In a 1+1 dimensional quantum mechanical analogy, the $u \rightarrow 0$ limit thus means that interactions between spins happen very slowly. Equivalently, the time direction has been stretched with respect to the spatial direction. A homogeneous system can be retrieved by rescaling time by a certain anisotropy factor $\zeta(u)$.

It does not appear feasible to determine $\zeta(u)$ without invoking certain results of conformal field theory. Suffice it here to say that one finds

$$\zeta(u) = \sin\left(\frac{\pi u}{\gamma}\right). \quad (6.26)$$

This predicts that the isotropic point $\zeta(u) = 1$ occurs for $u = \frac{\gamma}{2}$. One then has by (6.18) that $\tilde{R} \propto I + E$, a fact that can be accounted for geometrically within the loop representation of the Temperley-Lieb algebra.

6.3.4 Spin chain hamiltonian

Using (6.18) we thus see that in the completely anisotropic limit $u \rightarrow 0$ the transfer matrix becomes

$$t(0) = \sin^L(\gamma) e^{-iP}, \quad (6.27)$$

where e^{-iP} is the shift operator that translate the lattice sites one unit to the right. Equivalently, P can be interpreted as the momentum operator.

We know from the path-integral formalism that the transfer matrix (the time evolution operator) is the exponential of the quantum hamiltonian. To make things completely precise, note that to first order in u , one may omit one of the factors $\sin(\gamma - u)I$ in (6.18) and take $\sin(u)E$ instead. The correct development in the limit $u \rightarrow 0$ therefore reads

$$t(u) \simeq t(0) \exp\left[-\frac{u}{\sin \gamma} H\right], \quad (6.28)$$

where H is the Hamiltonian of the spin chain. Equivalently

$$H = -\sin \gamma \left. \frac{\partial}{\partial u} \log t(u) \right|_{u \rightarrow 0} = -\sin \gamma t(0)^{-1} t'(0). \quad (6.29)$$

Here the inverse $t(0)^{-1} = (\sin \gamma)^{-L} e^{iP}$ is just the shift in the opposite (left) direction. The derivative $t'(0)$ gives L terms, one for each of the factors in the product (6.5). Using (6.18) we have $\tilde{R}'(0) = -\cos \gamma I + E$. Therefore

$$H = L \cos \gamma I - \sum_{m=1}^L E_m. \quad (6.30)$$

Inserting the expression (6.25) for the TL generators in terms of Pauli matrices, the piece in $i \sin \gamma (\sigma_m^z - \sigma_{m+1}^z)$ simplifies by telescopy. With open boundary condition it would become a surface magnetic field acting on the first and last spins. We consider instead periodic boundary conditions, so this term vanishes altogether. One is left with

$$H = -\frac{1}{2} \sum_{m=1}^L [\sigma_m^x \sigma_{m+1}^x + \sigma_m^y \sigma_{m+1}^y + \Delta(\sigma_m^z \sigma_{m+1}^z + I)] , \quad (6.31)$$

where we recall that $\Delta = -\cos \gamma$.

We thus arrive at the Hamiltonian of a Heisenberg-type spin chain, where however the interaction is anisotropic along the z -direction. For that reason, this is called the *XXZ spin chain* with anisotropy parameter Δ .¹⁵

Let us emphasize that due to the commutativity of transfer matrices, the eigenvectors of the six-vertex model transfer matrix and of the XXZ spin chain Hamiltonian are *identical*. It is thus equivalent to diagonalise one or the other, and in that sense the two models are equivalent.

¹⁵We are here referring to an anisotropy between the different components of the interaction in the space direction. This should not be confused with the space-time anisotropy linked with the spectral parameter u .

7 Algebraic Bethe Ansatz

It is possible to continue the formal developments of the previous chapter and obtain a completely algebraic derivation of the Bethe Ansatz equations (5.31) and the corresponding expression (5.24) for the transfer matrix eigenvalues. There are at least two reasons for following this route.

The first reason is that the elimination of unwanted terms is somewhat tricky, and it might be hard to see if one has taken into account all possible processes when generalising from $n = 1, 2, \dots$ computations to the general case. The algebraic approach will add clarity to this step.

The second reason is that the algebraic Bethe Ansatz approach makes contact with a rich mathematical structure known as affine Hopf algebras. Within this structure, results about Lie algebras (and even Lie superalgebras) make possible the generalisation from the spin- $\frac{1}{2}$ six-vertex model to infinite classes of higher-spin integrable models. Exposing this in some detail requires to make contact also with the underlying quantum group, which for the six-vertex model is $U_q(sl(2))$. Since this is a more advanced subject, we refrain from treating it in this introduction.

The word “algebraic” should of course not conceal the fact that once the Bethe Ansatz equations for a given model have been found, there is also—as we have seen in section 5.6—a fair amount of *analytical* work to be done in order to derive the free energy in the thermodynamic limit and the critical exponents.

7.1 Monodromy matrix

We define the monodromy matrix $T(u)$ as the same product over R -matrices as used in defining the transfer matrix (6.4), but without the trace over the auxiliary space:

$$T(u) = R_{aL}R_{aL-1} \cdots R_{a2}R_{a1}. \quad (7.1)$$

Thus $T(u)$ is an endomorphism on the auxiliary space V_a and we have

$$t(u) = \text{Tr}_a T(u). \quad (7.2)$$

When several auxiliary spaces are involved we shall sometimes use the notation $T_a(u)$ to make clear what space is involved. We shall also denote matrix elements of $T(u)$ using the same convention as for the R -matrix, and

sometimes represent them graphically as

$$T_i^j(u) = i \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} j \quad (7.3)$$

These matrix elements are operators acting in the quantum spaces, here shown symbolically as a double line.

We can repeat the reasoning of (6.11)

$$\begin{array}{c} u_2 \rightarrow \\ u_1 \rightarrow \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} v_1 \\ v_2 \\ \cdots \\ v_L \end{array} = \begin{array}{c} u_2 \rightarrow \\ u_1 \rightarrow \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} v_1 \\ v_2 \\ \cdots \\ v_L \end{array} = \\ \begin{array}{c} u_2 \rightarrow \\ u_1 \rightarrow \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} v_1 \\ v_2 \\ \cdots \\ v_L \end{array} = \begin{array}{c} u_2 \rightarrow \\ u_1 \rightarrow \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} v_1 \\ v_2 \\ \cdots \\ v_L \end{array} \quad (7.4)$$

to establish that

$$R_{ab}(u-v)(T_a(u)T_b(v)) = (T_b(v)T_a(u))R_{ab}(u-v). \quad (7.5)$$

This identity is known popularly as the RTT relation. Using the double line convention of (7.3) it can also be written pictorially

$$\begin{array}{c} \begin{array}{c} \diagdown \\ \diagup \end{array} \\ \begin{array}{c} u_1 \\ u_2 \\ u_3 \end{array} \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} T_{12} \\ T_{23} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} = \begin{array}{c} \begin{array}{c} \diagdown \\ \diagup \end{array} \\ \begin{array}{c} u_1 \\ u_2 \\ u_3 \end{array} \end{array} \begin{array}{c} \parallel \\ \parallel \\ \text{---} \\ \parallel \\ \parallel \end{array} \begin{array}{c} T_{23} \\ T_{12} \end{array} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \quad (7.6)$$

7.2 Co-product and Yang-Baxter algebra

A Yang-Baxter algebra \mathcal{A} is a couple (R, T) satisfying the RTT relation (7.5). Its generators are the matrix elements $T_i^j(u)$. It is equipped with a product,

obtained graphically by stacking two monodromy matrices along a common quantum space (represented as a double line).

In addition to this product, \mathcal{A} is also equipped with a co-product Δ ,¹⁶ obtained graphically by glueing together two monodromy matrices along a common auxiliary space (represented as a single line). We have

$$\Delta(i \text{---} \parallel \text{---} j) = \sum_k i \text{---} \parallel \text{---} k \text{---} \parallel \text{---} j \quad (7.7)$$

The co-product thus serves to map the algebra \mathcal{A} into the tensor product $\mathcal{A} \otimes \mathcal{A}$:

$$\begin{aligned} \Delta & : \mathcal{A} \rightarrow \mathcal{A} \otimes \mathcal{A} \\ T_i^j(u) & \mapsto \sum_k T_i^k(u) \otimes T_k^j(u) \end{aligned} \quad (7.8)$$

while preserving the algebraic relations of \mathcal{A} .

In particular, the co-product ΔT_i^j must again satisfy the RTT relation (7.5). It is a nice exercise to understand what this means and to prove it.

An algebra equipped with a product and a co-product is called a *bi-algebra*. To be precise, we need a little more structure (co-associativity, existence of a co-unit, ...). If in addition we have an antipode (and if various diagrams commute) one arrives at a Hopf algebra.

7.3 Six-vertex model

When the auxiliary space is \mathbb{C}^2 , the matrix elements of the monodromy matrix are usually denoted as follows:

$$T_0^0(u) = A(u), \quad T_1^0(u) = B(u), \quad T_0^1(u) = C(u), \quad T_1^1(u) = D(u). \quad (7.9)$$

Recall that the structure constants of a Lie algebra provide a representation, known as the adjoint. In the same way, the R -matrix provides a representation of dimension 2 of the Yang-Baxter algebra. Indeed, in the

¹⁶This Δ should not be confused with the anisotropy parameter of the six-vertex model (XXZ spin chain).

special case where the double line is just a single line, the monodromy matrix reduces to the R -matrix:

$$(T_i^j(u))_l^k = R_{il}^{jk}(u). \quad (7.10)$$

The RTT relation is then nothing but the Yang-Baxter relation for the R -matrix.

The notation (7.9) just amounts to reading the R -matrix as a 2×2 matrix of blocks of size 2×2 . As in (6.14) we have

$$R = \left[\begin{array}{cc|cc} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_3 & \omega_6 & 0 \\ \hline 0 & \omega_5 & \omega_4 & 0 \\ 0 & 0 & 0 & \omega_2 \end{array} \right] = \begin{bmatrix} A(u) & B(u) \\ C(u) & D(u) \end{bmatrix}. \quad (7.11)$$

We recall the usual weights a, b, c (which now depend on the spectral parameter u), and we take the gauge $\eta = u$ in (6.15):¹⁷

$$a(u) = \sin(\gamma - u), \quad b(u) = \sin u, \quad c(u) = \sin \gamma. \quad (7.12)$$

We have then in explicit notation, and in terms of Pauli matrices:

$$\begin{aligned} A(u) &= \begin{bmatrix} a(u) & 0 \\ 0 & b(u) \end{bmatrix} = \frac{a(u) + b(u)}{2} I + \frac{a(u) - b(u)}{2} \sigma^z, \\ B(u) &= \begin{bmatrix} 0 & 0 \\ c(u) & 0 \end{bmatrix} = \frac{c(u)}{2} (\sigma^x - i\sigma^y) = c(u) \sigma^-, \\ C(u) &= \begin{bmatrix} 0 & c(u) \\ 0 & 0 \end{bmatrix} = \frac{c(u)}{2} (\sigma^x + i\sigma^y) = c(u) \sigma^+, \\ D(u) &= \begin{bmatrix} b(u) & 0 \\ 0 & a(u) \end{bmatrix} = \frac{a(u) + b(u)}{2} I - \frac{a(u) - b(u)}{2} \sigma^z. \end{aligned} \quad (7.13)$$

Note that $B(u)$ (resp. $C(u)$) acts as a creation (resp. annihilation) operator on the quantum space, with respect to the pseudo-vacuum in which all spins are up. We shall see below that this interpretation remains valid when taking co-products: $B(u)$ transforms n particle states into $n+1$ particle states (and vice versa for $C(u)$).

¹⁷To go from this convention to that of Gómez *et al*, take $u \rightarrow -iu$, $\gamma \rightarrow \pi - \gamma$, and divide all weights by $-i$.

7.3.1 Co-product

Establishing how the co-product acts on the operators $A(u)$, $B(u)$, $C(u)$ and $D(u)$ will turn out to be an important ingredient in the sequel. In more formal terms, we wish to obtain a representation of the six-vertex Yang-Baxter algebra \mathcal{A} on the space $V^{\otimes L}$.

Let us begin by examining the case $L = 2$ in details. Consider for instance the construction of $\Delta B(u)$. We have:

$$\Delta B(u)|00\rangle = \underbrace{1 \begin{array}{c} 1 \quad 0 \\ | \quad | \\ \hline 0 \quad 0 \end{array}}_{c(u)a(u)|10\rangle} + \underbrace{1 \begin{array}{c} 0 \quad 1 \\ | \quad | \\ \hline 0 \quad 0 \end{array}}_{b(u)c(u)|01\rangle} \quad (7.14)$$

Here the left and right indices define $B(u) = T_1^0(u)$, and the co-multiplication implies a sum over the middle index. The bottom (resp. top) indices define the in-state (resp. out-state) of the quantum spaces, here denoted as kets.

Proceeding in the same way for the three other in-states, we find that $\Delta B(u)$ can be written in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as

$$\begin{aligned} \Delta B(u) &= \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ b(u)c(u) & 0 & 0 & 0 \\ \hline c(u)a(u) & 0 & 0 & 0 \\ 0 & c(u)b(u) & a(u)c(u) & 0 \end{array} \right] \\ &= \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline c(u)a(u) & 0 & 0 & 0 \\ 0 & c(u)b(u) & 0 & 0 \end{array} \right] + \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ b(u)c(u) & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & a(u)c(u) & 0 \end{array} \right] \\ &= B(u) \otimes A(u) + D(u) \otimes B(u). \end{aligned} \quad (7.15)$$

It is actually simpler to avoid specifying the states of the quantum spaces altogether. Applying (7.7) directly one then obtains

$$\Delta^{L-1}B(u) = \underbrace{1 \begin{array}{c} \parallel \quad \parallel \\ | \quad | \\ \hline 0 \quad 0 \end{array}}_{\Delta^{L_1-1}B(u) \otimes \Delta^{L_2-1}A(u)} + \underbrace{1 \begin{array}{c} \parallel \quad \parallel \\ | \quad | \\ \hline 1 \quad 1 \end{array}}_{\Delta^{L_1-1}D(u) \otimes \Delta^{L_2-1}B(u)} \quad (7.16)$$

Note that this derivation applies for any bipartition $L_1 + L_2 = L$, and not only for $L_1 = L_2 = 1$.

Repeating the working for the three other operators, the complete co-multiplication table reads:

$$\begin{aligned}
\Delta A(u) &= A(u) \otimes A(u) + C(u) \otimes B(u), \\
\Delta B(u) &= B(u) \otimes A(u) + D(u) \otimes B(u), \\
\Delta C(u) &= C(u) \otimes D(u) + A(u) \otimes C(u), \\
\Delta D(u) &= D(u) \otimes D(u) + B(u) \otimes C(u).
\end{aligned} \tag{7.17}$$

To generalise this construction from $L = 2$ to arbitrary L it suffices to use the associativity of the co-multiplication. Indeed for $L \geq 2$ we have

$$\begin{aligned}
\Delta^{L-1} &: \mathcal{A} \rightarrow \mathcal{A}^{\otimes L} \\
\Delta^{L-1} &\mapsto (I^{\otimes L-2} \otimes \Delta)\Delta^{L-2}.
\end{aligned} \tag{7.18}$$

When making this definition, we have chosen to insert new tensorands from the right. Inserting them from the left would make no difference to the result, since in any case it can also be computed directly along the lines of (7.14). In the latter case, one has to sum over all $L - 1$ intermediate indices, of the type k in (7.7). It is a useful exercise to compute $\Delta^2 B(u)$ for $L = 3$ in all three ways and check that one obtains identical results.

In the following we shall simplify the notation and write, for example, $B(u)$ instead of $\Delta^{L-1}B(u)$. Thus $B(u)$ is an operator that acts on all L spaces in the tensor product $V^{\otimes L}$. Using (7.17)–(7.18) repeatedly it can be expanded in fully tensorised form, as an expression with 2^{L-1} terms. This expanded form is (7.17) for $L = 2$, and the expression for $L = 3$ is contained in the above exercise. The factors entering each term in the expanded form act on a single space V .

7.3.2 Commutation relations

The operators $A(u)$, $B(u)$, $C(u)$ and $D(u)$ satisfy a set of commutation relations which follow as a direct consequence of the RTT relation (7.5).

To see in details how this works, we first write out the RTT relation in component form:

$$\sum_{j_1, j_2} R_{j_1 j_2}^{k_1 k_2}(u-v) T(u)_{i_1}^{j_1} T(v)_{i_2}^{j_2} = \sum_{j_1, j_2} T(v)_{j_2}^{k_2} T(u)_{j_1}^{k_1} R_{i_1 i_2}^{j_1 j_2}(u-v). \quad (7.19)$$

This gives a relation for each choice of (k_1, k_2, i_1, i_2) . Consider for instance the choice $(0, 0, 1, 0)$:

$$R_{00}^{00}(u-v) T(u)_1^0 T(v)_0^0 = T(v)_1^0 T(u)_0^0 R_{10}^{01}(u-v) + T(v)_0^0 T(u)_1^0 R_{10}^{10}(u-v). \quad (7.20)$$

Insert now the R -matrix elements from (7.11)–(7.12) and the monodromy matrix elements from (7.9), recalling that the former are just scalars, whereas the latter are (non-commuting) operators. This gives

$$a(u-v)B(u)A(v) = c(u-v)B(v)A(u) + b(u-v)A(v)B(u). \quad (7.21)$$

Among all the possible commutation relations we shall actually only need a few. First, for two operators of the same type we have simply

$$\begin{aligned} A(u)A(v) &= A(v)A(u), & B(u)B(v) &= B(v)B(u), \\ C(u)C(v) &= C(v)C(u), & D(u)D(v) &= D(v)D(u). \end{aligned} \quad (7.22)$$

Second, to push an A or a D past a B we have

$$\begin{aligned} A(u)B(v) &= \frac{a(v-u)}{b(v-u)} B(v)A(u) - \frac{c(v-u)}{b(v-u)} B(u)A(v), \\ D(u)B(v) &= \frac{a(u-v)}{b(u-v)} B(v)D(u) - \frac{c(u-v)}{b(u-v)} B(u)D(v). \end{aligned} \quad (7.23)$$

The first of these relations follows from (7.21) after a relabelling $u \leftrightarrow v$ and some rearrangement. The second relation is obtained from a similar computation.

7.3.3 Algebraic Bethe Ansatz

We now have all necessary ingredients to treat the six-vertex model using the algebraic Bethe Ansatz.

As in the coordinate Bethe Ansatz, one starts from the pseudo-vacuum, or reference state, in which all spins point up and no particle world-lines are present. We denote this state as

$$|\uparrow\rangle = |\uparrow\uparrow\cdots\uparrow\rangle = |00\cdots 0\rangle. \quad (7.24)$$

Recall that $B(u)$ creates a particle (or equivalently, flips down one spin), whereas $C(u)$ annihilates a particle. Thus, an n -particle state (i.e., with n down spins) can be constructed as follows:

$$|\Psi_n\rangle = \prod_{i=1}^n B(u_i) |\uparrow\rangle. \quad (7.25)$$

The states (7.25) are called algebraic Bethe Ansatz states.

Our goal is to diagonalise the transfer matrix

$$t(u) = \text{Tr}_a T(u) = A(u) + D(u). \quad (7.26)$$

This means solving the eigenvalue equation

$$t(u)|\Psi_n\rangle = [A(u) + D(u)] \prod_{i=1}^n B(u_i) |\uparrow\rangle = \Lambda_n(u; \{u_i\}) \prod_{i=1}^n B(u_i) |\uparrow\rangle. \quad (7.27)$$

This can obviously only be done if the parameters $\{u_i\}$ satisfy certain conditions. These are precisely the Bethe Ansatz equations, and we shall rederive them now using the algebraic method.

This obviously implies that $\{u_i\}$ must somehow be related to the pseudo-momenta. The correct relation will be obtained from the algebraic method below, but we can obtain it already now by comparing our setup to that of the coordinate Bethe Ansatz.

We know from section 5.2.2 that a one-particle state reads $|\Psi_1\rangle = \sum_x g(x)|x\rangle$ with $g(x) = z^x = e^{ikx}$. In the algebraic framework—and setting $L = 2$ for simplicity—this same state follows from (7.25) and (7.14):

$$|\Psi_1\rangle = \Delta B(u) |\uparrow\rangle = c(u)a(u)|10\rangle + b(u)c(u)|01\rangle.$$

Comparing this with $|\Psi_1\rangle = \sum_x g(x)|x\rangle$ we identify¹⁸

$$z = e^{ik} = \frac{a(u)}{b(u)}. \quad (7.28)$$

(To be quite honest, for $L = 2$ one cannot by comparing $|01\rangle$ and $|10\rangle$ say whether the particle has moved to the right or to the left, i.e., distinguish z and z^{-1} . But the above result will be rederived below by other means.)

To compute $[A(u) + D(u)] \prod_{i=1}^n B(u_i) | \uparrow \rangle$ we use the commutation relations (7.23) to push $A(u)$ and $D(u)$ to the right, past the string of B 's. When they have been pushed completely to the right, one applies the relations

$$A(u) | \uparrow \rangle = a(u)^L | \uparrow \rangle, \quad D(u) | \uparrow \rangle = b(u)^L | \uparrow \rangle, \quad (7.29)$$

Note that (7.29) follows from the first and last lines of (7.17), generalised for $L = 2$ to arbitrary L . Consider for instance $\Delta A(u)$. It is easy to see that the right-hand side will contain a single term $A(u)^{\otimes L}$, and all remaining terms will contain at least one factor $C(u)$ in the tensor product. But this $C(u)$ will annihilate $| \uparrow \rangle$, so the only contribution is $a(u)^L | \uparrow \rangle$ indeed.

Each time we push $A(u)$ one position towards the right, we obtain two contributions from the right-hand side of (7.23). The unique term obtained by always choosing the first contribution is a *wanted A-term*. In this term, the arguments of the $B(u_i)$ remain unchanged, and $A(u)$ simply “goes through”. The remaining $2^n - 1$ terms are *unwanted A-terms*. In those terms, at least one of the arguments u_i of the B 's has been changed into u , and so the state is not of the form (7.25). Similarly, there is one wanted and $2^n - 1$ unwanted D -terms.

The wanted A -term and the wanted D -term produces the expression for the eigenvalue of the transfer matrix

$$\Lambda_n(u; \{u_i\}) = a(u)^L \prod_{i=1}^n \frac{a(u_i - u)}{b(u_i - u)} + b(u)^L \prod_{i=1}^n \frac{a(u - u_i)}{b(u - u_i)}. \quad (7.30)$$

This should of course coincide with (5.24). This means that we should identify

$$L_i = L(z_i) = \frac{a(u_i - u)}{b(u_i - u)}. \quad (7.31)$$

On the other hand we have the definition (5.10) according to which

$$L_i = \frac{a(u)b(u) + (c(u)^2 - b(u)^2)z_i}{a(u)(a(u) - b(u)z_i)}, \quad (7.32)$$

where $z_i = \frac{a(u_i)}{b(u_i)}$. Inserting the parameterisations (7.12) and simplifying we find indeed

$$L_i = \frac{\sin(\gamma - (u_i - u))}{\sin(u_i - u)} = \frac{a(u_i - u)}{b(u_i - u)}. \quad (7.33)$$

The result $z_i = \frac{a(u_i)}{b(u_i)}$ was obtained above; alternatively one can find it by taking the $u \rightarrow 0$ limit of (7.32).

In the same way we can verify that

$$M_i = M(z_i) = \frac{a(u - u_i)}{b(u - u_i)} \quad (7.34)$$

is in agreement with the definition (5.11).

The condition that the unwanted A -terms cancel the unwanted D -terms leads to the Bethe Ansatz equations (BAE)

$$\left(\frac{a(u_i)}{b(u_i)}\right)^L = \prod_{\substack{j=1 \\ j \neq i}}^n \frac{a(u_i - u_j)b(u_j - u_i)}{a(u_j - u_i)b(u_i - u_j)}. \quad (7.35)$$

Proof of (7.35). Let us abbreviate $A_0 \equiv A(u)$ and $A_i \equiv A(u_i)$ for $i = 1, 2, \dots, n$, and similarly for the other types of operators. We also set

$$\alpha_{ij} \equiv \frac{a(u_j - u_i)}{b(u_j - u_i)}, \quad \beta_{ij} \equiv -\frac{c(u_j - u_i)}{b(u_j - u_i)}, \quad (7.36)$$

so that the commutation relations (7.23) can be rewritten

$$\begin{aligned} A_i B_j &= \alpha_{ij} B_j A_i + \beta_{ij} B_i A_j, \\ D_i B_j &= \alpha_{ji} B_j D_i + \beta_{ji} B_i D_j. \end{aligned} \quad (7.37)$$

The unwanted A -terms (resp. D -terms) are those where A_0 (resp. D_0) exchanges its spectral parameter with one or more of the B 's and hence becomes some A_i (resp. D_i) with $i \geq 1$ as it is pushed to the right of $\prod_{i=1}^n B_i$. The sum of these unwanted A -terms is

$$\sum_{i=1}^n \bar{a}_i \left(\prod_{\substack{j=1 \\ j \neq i}}^n B_j \right) A_i | \uparrow \rangle = \sum_{i=1}^n \bar{a}_i a(u_i)^L \left(\prod_{\substack{j=1 \\ j \neq i}}^n B_j \right) | \uparrow \rangle. \quad (7.38)$$

At first sight, it may appear complicated to compute the coefficients \bar{a}_i , since the A -operator might exchange its rapidity with the B 's several times, as it is moved through the product. However, we can simplify the computation of \bar{a}_i dramatically by using (7.22) to rewrite (7.25) as

$$|\Psi_n\rangle = B_i \prod_{\substack{j=1 \\ j \neq i}}^n B_j |\uparrow\rangle. \quad (7.39)$$

The action of A_0 on this can then only produce an A_i on the right if the exchange of spectral parameter happens when A_0 is commuted through the very first factor B_i in (7.39). Therefore:

$$\bar{a}_i = \beta_{0i} \prod_{\substack{k=1 \\ k \neq i}} \alpha_{ik}. \quad (7.40)$$

By this simple trick, the total number of unwanted A -terms has been reduced from $2^n - 1$ to just n .

Similarly, the unwanted D -terms read

$$\sum_{i=1}^n \bar{d}_i \left(\prod_{\substack{j=1 \\ j \neq i}}^n B_j \right) D_i |\uparrow\rangle = \sum_{i=1}^n \bar{d}_i b(u_i)^L \left(\prod_{\substack{j=1 \\ j \neq i}}^n B_j \right) |\uparrow\rangle \quad (7.41)$$

with

$$\bar{d}_i = \beta_{i0} \prod_{\substack{k=1 \\ k \neq i}} \alpha_{ki}. \quad (7.42)$$

Due to (7.12) we have $\beta_{0i} = -\beta_{i0}$. Therefore the sum of (7.38) and (7.41) vanishes provided that

$$\left(\frac{a(u_i)}{b(u_i)} \right)^L = \prod_{\substack{k=1 \\ k \neq i}}^n \frac{\alpha_{ki}}{\alpha_{ik}}. \quad (7.43)$$

Plugging back (7.36) we arrive at (7.35). □

Alternatively (7.35) follows also from the form (7.30) of the eigenvalue, as we now argue. We set $z = e^{2iu}$ and $q = e^{i\gamma}$, and we define the shifted eigenvalue $\tilde{\Lambda} = (2ie^{iu})^L \Lambda_n(u, \{u_i\})$. Elementary computations then bring (7.35) into the form

$$\tilde{\Lambda} = (q - q^{-1}z)^L \prod_{i=1}^n \frac{q^{-1}z_i - qz}{z - z_i} + (z - 1)^L \prod_{i=1}^n \frac{qz_i - q^{-1}z}{z - z_i}. \quad (7.44)$$

Defining the polynomials

$$Q(z) = \prod_{i=1}^n (z - z_i), \quad \phi_N(z) = (z - 1)^L, \quad (7.45)$$

we obtain

$$\tilde{\Lambda}Q(z) = (-q)^{L-n} \phi_N(q^{-2}z)Q(q^2z) + (-q)^N \phi_N(z)Q(q^{-2}z). \quad (7.46)$$

Whenever the spectral parameter equals one of the Bethe roots, we have $Q(z_j) = 0$ on the left-hand side of (7.46). Therefore the right-hand side must also vanish. Working backwards through the change of variables then produces the BAE (7.35).

The form (7.35) is best compared with (5.33). We have therefore the identification of the scattering phases

$$\hat{S}_{ji} = \frac{a(u_j - u_i)b(u_i - u_j)}{a(u_i - u_j)b(u_j - u_i)} = -\frac{1 - 2\Delta z_i + z_i z_j}{1 - 2\Delta z_j + z_i z_j}. \quad (7.47)$$

Exercise: Verify this using the parameterisation (7.12).

7.4 Energy and momentum

We can compute the energy E of the Bethe Ansatz state (7.25). To this end we just need to recall the link (6.29) between the transfer matrix $t(u)$ and the Hamiltonian H . Taking expectation values with respect to the state (7.25) the operator H gets replaced by its expectation value E , and $t(u)$ gets replaced by the eigenvalue $\Lambda(u; \{u_i\})$. Therefore

$$E_n(\{u_i\}) = -\sin \gamma \left. \frac{\partial}{\partial u} \log \Lambda_n(u; \{u_i\}) \right|_{u \rightarrow 0}. \quad (7.48)$$

In (7.30) only the first term contributes in the $u \rightarrow 0$ limit:

$$\Lambda_n \simeq \sin^L(\gamma) \prod_{i=1}^n \frac{a(u_i - u)}{b(u_i - u)}. \quad (7.49)$$

Taking the derivative we arrive at

$$E_n(\{u_i\}) = L \cos \gamma + \sum_{i=1}^n \epsilon(u_i), \quad (7.50)$$

where the energy of a single particle with quasi-momentum

$$z_j = e^{ik_j} = \frac{a(u_j)}{b(u_j)}. \quad (7.51)$$

is simply

$$\epsilon(u_i) = -\frac{\sin^2(\gamma)}{\sin(u_i) \sin(\gamma - u_i)}. \quad (7.52)$$

In the same way we can express the momentum of the Bethe Ansatz state:

$$-iP = \log \left(\frac{t(0)}{\sin^L(\gamma)} \right) = \sum_{i=1}^n \log \left(\frac{a(u_i)}{b(u_i)} \right), \quad (7.53)$$

meaning that the quasi-momenta (7.51) just add up.

Note that the constant (i.e., independent of u_i) term in (7.50) is consistent with that of (6.30). If we normalise the Hamiltonian as $H = -\sum_{m=1}^L E_m$, the n -particle energy reads simply $E_n(\{u_i\}) = \sum_{i=1}^n \epsilon(u_i)$.

7.5 Further developments

- We have not exposed how to manipulate the Bethe roots to construct low-lying excitations over the ground state. This is best discussed after introducing some notions of conformal field theory. In particular it is possible to make an exhaustive comparison between excitations in the XXZ spin chain and in the Coulomb gas CFT for the compactified boson.

- In the case of an open, non-periodic spin chain one will in general need specific boundary interactions to guarantee integrability. These are encoded in a so-called K -matrix. The Yang-Baxter equations must then be supplemented with the Sklyanin equation that ensures the compatibility between R and K matrices.
- We have already mentioned the possibility of introducing Hopf algebras and quantum groups as a means of constructing algebraic Bethe Ansätze for more general systems. Indeed, the algebraic Bethe Ansatz solution of the 6-vertex model can essentially be generalised in two different ways, namely by increasing the rank, or by increasing the spin.
- Baxter's TQ-relation, the Y-system,
- Computation of correlation functions via algebraic Bethe Ansatz, following the works by J.-M. Maillet *et al.*

8 Potts model

When writing the integrable R -matrix of the six-vertex model, in (6.18), we have briefly come across a new type of algebraic structure: the Temperley-Lieb algebra. This is an example of a lattice algebra [Ma91], or more generally a partition algebra [HR05]. Other examples include the dilute Temperley-Lieb algebra, the Brauer algebra, and various types of multi-colour braid-monoid algebras [GP93].

The R -matrix based on each of these algebras generates the transfer matrix of a corresponding statistical mechanics model. Obviously one can gather important information about the statistical mechanics model by studying the underlying algebra and its representation theory.

In what follows we shall focus on the (open, or non-periodic) Temperley-Lieb (TL) algebra. This algebra has many different representations, each of which is related to a particular stat-mech model: Potts, Ising, six-vertex, restricted solid-on solid (RSOS) model, Historically, each model was introduced independently, but with hindsight the unifying algebraic framework can be used to understand better the relations among them.

Most of the corresponding representations of the TL algebra are not faithful, i.e., they obey additional relations than those defining the TL algebra. The Potts model—to be precise: in its formulation as a loop model—furnishes a faithful representation. Since it is also an extremely interesting and well-studied model in statistical mechanics, it is natural to study it in some detail. The selfdual Potts model on a square lattice will turn out to be closely related to the six-vertex model, so that the results of preceding chapters imply the exact solution of the Potts model. Also on the triangular lattice can the selfdual model be exactly solved.

For the mathematically inclined, let us briefly mention an important connection to representation theory. A central result, known as Schur-Weyl duality, states that:

1. The general linear group $GL_n(\mathbb{C})$ and the symmetric group \mathfrak{S}_k both act on the tensor product $V^{\otimes k}$ with $\dim V = n$. (We interpret $V^{\otimes k}$ as the quantum space.)
2. These two actions commute and each action generates the full centraliser of the other.
3. As a $(GL_n(\mathbb{C}), \mathfrak{S}_k)$ -bimodule, the tensor space has a *multiplicity free*

decomposition

$$V^{\otimes k} \simeq \bigoplus_{\lambda} L_{GL_n}(\lambda) \otimes L_{\mathfrak{S}_k}(\lambda), \quad (8.1)$$

where $L_{GL_n}(\lambda)$ are irreducible $GL_n(\mathbb{C})$ -modules and $L_{\mathfrak{S}_k}(\lambda)$ are irreducible \mathfrak{S}_k -modules.

Similar results hold when taking subgroups of $GL_n(\mathbb{C})$, in which case the centraliser algebras become bigger. The TL algebra occurs in this hierarchy of dualities, and its centraliser is the quantum algebra $U_q(\mathfrak{sl}_2)$.

In this chapter we define the Potts model in various representations and exhibit its equivalence to the six-vertex model. Even though we are mainly interested in the model defined on a square lattice, it turns out that many of the transformations that we need hold when the model on more general graphs. Since it hardly more complicated—and a lot more instructive—to work in the “correct” generality, we shall choose to do so and specialise only when needed.

8.1 Spin representation

Let $G = (V, E)$ be an arbitrary connected graph with vertex set V and edge set E . The Q -state Potts model is initially defined by assigning a spin variable σ_i to each vertex $i \in V$. Each spin can take Q different values, by convention chosen as $\sigma_i = 1, 2, \dots, Q$. We denote by σ the collection of all spin variables on the graph. Two spins i and j are called nearest neighbours if they are incident on a common edge $e = (ij) \in E$. In any given configuration σ , a pair of nearest neighbour spins is assigned an energy $-J$ if they take identical values, $\sigma_i = \sigma_j$. The Hamiltonian (dimensionless energy functional) of the Potts model is thus

$$\mathcal{H} = -K \sum_{(i,j) \in E} \delta(\sigma_i, \sigma_j). \quad (8.2)$$

where the Kronecker delta function is defined as

$$\delta(\sigma_i, \sigma_j) = \begin{cases} 1 & \text{if } \sigma_i = \sigma_j \\ 0 & \text{otherwise} \end{cases} \quad (8.3)$$

and $K = J/k_B T$ is a dimensionless coupling constant (interaction energy).

The case $Q = 2$ corresponds to the Ising model. Indeed, if $S_i = \pm 1$ we have

$$2\delta(S_i, S_j) = S_i S_j + 1. \quad (8.4)$$

The second term amounts to an unimportant shift of the interaction energy, and so the models are equivalent if we set $K_{\text{Potts}} = 2K_{\text{Ising}}$.

The thermodynamic information about the Potts model is encoded in the partition function

$$Z = \sum_{\sigma} e^{-\mathcal{H}} = \sum_{\sigma} \prod_{(ij) \in E} e^{K\delta(\sigma_i, \sigma_j)} \quad (8.5)$$

and in various correlation functions. By a correlation function we understand the probability that a given set of vertices are assigned fixed values of the spins.

In the ferromagnetic case $K > 0$ the spins tend to align at low temperatures ($K \gg 1$), defining a phase of ferromagnetic order. Conversely, at high temperatures ($K \ll 1$) the spins are almost independent, leading to a paramagnetic phase where entropic effects prevail. On physical grounds, one expects the two phases to be separated by a critical point K_c where the effective interactions between spins becomes long ranged.

For certain regular planar lattices K_c can be determined exactly by duality considerations. Moreover, K_c will turn out to be the locus of a second order phase transition if $0 \leq Q \leq 4$. In that case the Potts model enjoys conformal invariance in the limit of an infinite lattice, allowing its critical properties to be determined exactly by a variety of techniques. These properties turn out to be *universal*, i.e., independent of the lattice used for defining the model microscopically.

8.2 Fortuin-Kasteleyn cluster representation

The initial definition (8.2) of the Potts model requires the number of spins Q to be a positive integer. It is possible to rewrite the partition function and correlation functions so that Q appears only as a parameter. This makes it possible to assign to Q arbitrary real (or even complex) values.

Notice first that by (8.3) we have the identity

$$e^{K\delta(\sigma_i, \sigma_j)} = 1 + v\delta(\sigma_i, \sigma_j), \quad (8.6)$$

where we have defined $v = e^K - 1$. Now, it is obvious that for any edge-dependent factors h_e one has

$$\prod_{e \in E} (1 + h_e) = \sum_{E' \subseteq E} \prod_{e \in E'} h_e. \quad (8.7)$$

where the subset E' is defined as the set of edges for which we have taken the term h_e in the development of the product $\prod_{e \in E}$. In particular, taking $h_e = v\delta(\sigma_i, \sigma_j)$ we obtain for the partition function (8.5)

$$Z = \sum_{E' \subseteq E} v^{|E'|} \sum_{\sigma} \prod_{(ij) \in E'} \delta(\sigma_i, \sigma_j) = \sum_{E' \subseteq E} v^{|E'|} Q^{k(E')}, \quad (8.8)$$

where $k(E')$ is the number of connected components in the graph $G' = (V, E')$, i.e., the graph obtained from G by removing the edges in $E \setminus E'$. Those connected components are called *clusters*, and (8.8) is the Fortuin-Kasteleyn *cluster representation* of the Potts model partition function. The sum over spins σ in (8.5) has now been replaced by a sum over edge subsets, and Q appears as a parameter in (8.8) and no longer as a summation limit.

8.3 Duality of the partition function

Consider now the case where $G = (V, E)$ is a connected *planar* graph. Any planar graph possesses a *dual graph* $G^* = (V^*, E^*)$ which is constructed by placing a dual vertex $i^* \in V^*$ in each face of G , and connecting a pair of dual vertices by a dual edge $e^* \in E^*$ if and only if the corresponding faces are adjacent in G . In other words, there is a bijection between edges and dual edges, since each edge $e \in E$ intersects precisely one dual edge $e^* \in E^*$. Note that by the Euler relation

$$|V| + |V^*| = |E| + 2. \quad (8.9)$$

By construction, the dual graph is also connected and planar. Note also that duality is an involution, i.e., $(G^*)^* = G$.

The Euler relation can easily be proved by induction. If $E = \emptyset$, since G was supposed connected we must have $|V| = |V^*| = 1$, so (8.9) indeed holds. Each time a further edge is added to E , there are two possibilities. Either it connects an existing vertex to a new vertex, in which case $|V|$ increases

by one and $|V^*|$ is unchanged. Or it connects two existing vertices, meaning that a cycle is closed in G . In this case $|V|$ is unchanged and V^* increases by one. In both cases (8.9) remains valid.

Recalling the cluster representation (8.8)

$$\begin{aligned} Z_G(Q, v) &= \sum_{E_1 \subseteq E} v^{|E_1|} Q^{k(E_1)} \\ Z_{G^*}(Q, v^*) &= \sum_{E_2 \subseteq E^*} (v^*)^{|E_2|} Q^{k(E_2)} \end{aligned} \quad (8.10)$$

we now claim that it is possible to chose v^* so that

$$Z_G(Q, v) = k Z_{G^*}(Q, v^*) \quad (8.11)$$

where k is an unimportant multiplicative constant.

To prove this claim, we show that the proportionality (8.11) holds term by term in the summations (8.10). To this end, we first define a bijection between the terms by $E_2 = (E \setminus E_1)^*$, i.e., an edge is present in E_1 if its dual edge is absent from E_2 , and vice versa. This implies

$$|E_1| + |E_2| = |E|. \quad (8.12)$$

We have moreover the topological identity for the induced (not necessarily connected) graphs $G_1 = (V, E_1)$ and $G_2 = (V^*, E_2)$

$$k(E_1) = |V| - |E_1| + c(E_1) = |V| - |E_1| + k(E_2) - 1, \quad (8.13)$$

where we $k(E_1)$ and $c(E_1)$ are respectively the number of connected components and the number of independent cycles¹⁹ in the graph G_1 .

The proof of (8.13) is again by induction. If $E_1 = \emptyset$, we have $k(E_1) = |V|$, $c(E_1) = 0$ and $k(E_2) = 1$. Each time an edge is added to E_1 there are two possibilities. Either $c(E_1)$ stays constant, in which case $k(E_1)$ is reduced by one and $k(E_2)$ is unchanged. Or $c(E_1)$ increases by one, in which case $k(E_1)$ is unchanged and $k(E_2)$ increases by one. In both cases (8.13) remains valid.

¹⁹The number of independent cycles—also known as the circuit rank, or the cyclomatic number—is the smallest number of edges to be removed from a graph in order that no graph cycle remains.

Combining (8.12)–(8.13) gives

$$v^{|E_1|} Q^{k(E_1)} = k (v^*)^{|E_2|} Q^{k(E_2)}, \quad (8.14)$$

where we have defined

$$k = Q^{1-|V^*|} v^{|E|} = Q^{|V|-|E|-1} v^{|E|} \quad (8.15)$$

and $v^* = Q/v$. Comparing (8.14) with (8.10) completes the demonstration of (8.11) and furnishes the desired duality relation

$$vv^* = Q. \quad (8.16)$$

The duality relation (8.16) is particularly useful when the graph is self-dual, $G^* = G$. This is the case of the regular square lattice. Assuming the uniqueness of the phase transition, the critical point is given by the selfdual coupling:

$$v_c = \pm \sqrt{Q} \quad (\text{square lattice}) \quad (8.17)$$

In the Ising case $Q = 2$, the solution $v_c = +\sqrt{Q}$ gives $K_c = \log(\sqrt{2} + 1)$ in agreement with (4.12).

8.4 Special cases

One of the strengths of the Q -state Potts model is that it contains a large number of interesting special cases. Many of those make manifest the geometrical content of the partition function (8.8). The equivalence between $Q = 2$ and the Ising model has already been discussed. We shall concentrate here on a couple of more subtle equivalences, that explicitly exploit the fact that Q can now be used as a continuous variable.

8.4.1 Bond percolation

For $Q = 1$ the Potts model is seemingly trivial, with partition function $Z = (1 + v)^{|E|}$. Instead of setting $Q = 1$ brutally, one can however consider taking the *limit* $Q \rightarrow 1$. This leads to the important special case of *bond percolation*.

Let $p \in [0, 1]$ and set $v = p/(1 - p)$. We then consider the rescaled partition function

$$\tilde{Z}(Q) \equiv (1 - p)^{|E|} Z = \sum_{E' \subseteq E} p^{|E'|} (1 - p)^{|E|-|E'|} Q^{k(E')}. \quad (8.18)$$

We have of course $\tilde{Z}(1) = 1$. But formally, what is written here is that each edge is present in E' (i.e., percolating) with probability p and absent (i.e., non percolating) with probability $1 - p$. Appropriate correlation functions and derivatives of $\tilde{Z}(Q)$ in the limit $Q \rightarrow 1$ furnish valuable information about the geometry of the percolation clusters. For instance

$$\lim_{Q \rightarrow 1} Q \frac{d\tilde{Z}(Q)}{dQ} = \langle k(E') \rangle \quad (8.19)$$

gives the average number of clusters.

8.4.2 Trees and forests

Using (8.13), and defining $w = \frac{Q}{v}$, one can rewrite (8.8) as

$$\begin{aligned} Z &= \sum_{E_1 \subseteq E} \left(\frac{Q}{w} \right)^{|V| + c(E_1) - k(E_1)} Q^{|V| - |E_1| + c(E_1)} \\ &= v^{|V|} \sum_{E_1 \subseteq E} w^{k(E_1) - c(E_1)} Q^{c(E_1)}. \end{aligned} \quad (8.20)$$

Take now the limit $Q \rightarrow 0$ and $v \rightarrow 0$ in such a way that the ratio $w = Q/v$ is fixed and finite, and consider the rescaled partition function $\tilde{Z} = Zv^{-|V|}$. The limit $Q \rightarrow 0$ will suppress any term with $c(E_1) > 0$, and we are left with

$$\tilde{Z} = \sum'_{E_1 \subseteq E} w^{k(E_1)}, \quad (8.21)$$

where the prime indicates that the summation is over edge sets such that the graphs $G_1 = (V, E_1)$ have no cycles, $c(E_1) = 0$. Such graphs are known as forests, or more precisely (since the vertex set V is that of G), *spanning forests* of G . Each connected component carries a weight w .

For $w \rightarrow 0$, the surviving terms are *spanning trees*, i.e., forests with a single connected component. Note that the critical curve on the square lattice (8.17) goes through the point $(Q, v) = (0, 0)$ with a vertical tangent (i.e., $w \rightarrow 0$) and thus describes spanning trees.

8.5 Loop representation

We now transform the Potts model defined on a planar graph G into a model of self-avoiding loops on a related graph $\mathcal{M}(G)$, known in graph theory as

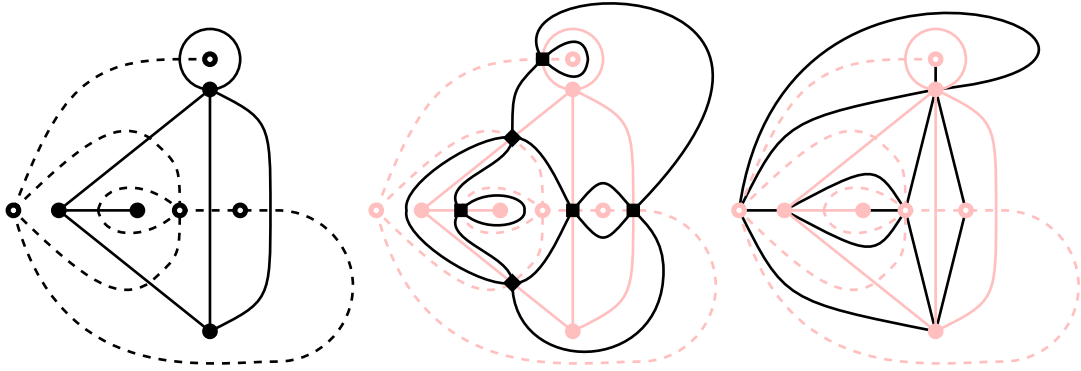


Figure 18: (a) A planar graph G (black circles and solid lines) and its dual graph G^* (white circles and dashed lines). (b) The medial graph $\mathcal{M}(G) = \mathcal{M}(G^*)$. (c) The plane quadrangulation $\hat{G} = \mathcal{M}(G)^*$.

the *medial graph*. Each term E' in the cluster representation (8.8) is in bijection with a term in the loop representation. The correspondence is, roughly speaking, that the loops turn around the connected components in $G' = (V, E')$ as well as their elementary internal cycles. More precisely, the loops separate the clusters from their duals.

To make this transformation precise, we first need to define the medial graph $\mathcal{M}(G) = (\tilde{V}, \tilde{E})$ carefully. Let $G = (V, E)$ be a connected planar graph with dual $G^* = (V^*, E^*)$. The pair (G, G^*) can be drawn in the plane such that each edge $e \in E$ intersects its corresponding dual edge $e^* \in E^*$ exactly once, see Figure 18a. To each of these intersections corresponds a vertex $\tilde{i} \in \tilde{V}$ of $\mathcal{M}(G)$.

Consider now the union $G \cup G^*$. This is in fact a quadrangulation of the plane. Each quadrangle consists of a pair of half-edges and one vertex from G , and a pair of half-edges and one vertex from G^* . These two pairs of half-edges meet in a pair of vertices from \tilde{V} . An edge of $\mathcal{M}(G)$ is drawn diagonally inside each quadrangle, joining the pair of vertices from \tilde{V} . This defines the edge set \tilde{E} and completes the definition of the medial graph. An example is shown in Figure 18b.

It is manifest in these definitions that G and G^* are used in a completely

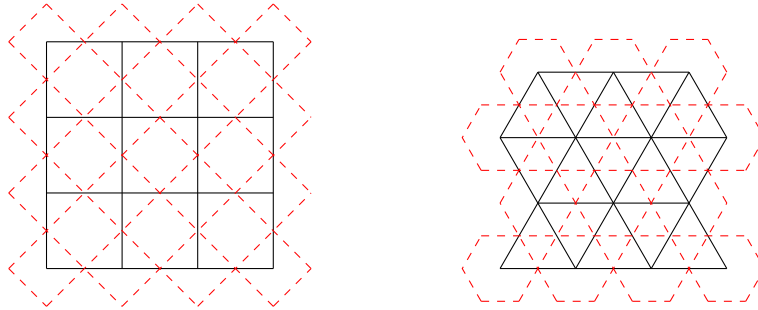


Figure 19: Square and triangular lattices (solid lines) with their corresponding medial lattices (dashed lines).

symmetric way. Thus, a graph and its dual has the same medial graph, $\mathcal{M}(G) = \mathcal{M}(G^*)$. Moreover, it is easy to see that every vertex of $\mathcal{M}(G)$ has degree four.²⁰

The medial of the square lattice is another (tilted) square lattice. The medial of the triangular lattice (or of its dual hexagonal lattice) is known as the kagome lattice.²¹ These two medial lattices, shown in Figure 19, are particularly important for subsequent applications.

To each term E' in appearing in the sum (8.8) we now define a system of self-avoiding loops that completely cover the edges of $\mathcal{M}(G)$. Let $\tilde{i} \in \tilde{V}$ be a vertex of $\mathcal{M}(G)$ and write its adjacent (half)edges from E , E^* and \tilde{E} in cyclic order as $\tilde{e}_1 e \tilde{e}_2 e^* \tilde{e}_3 e \tilde{e}_4 e^*$. Now if $e \in E'$, link up the half-edges of \tilde{E} in two pairs as $(\tilde{e}_4 \tilde{e}_1)(\tilde{e}_2 \tilde{e}_3)$. Conversely, if $e \in E \setminus E'$, we link $(\tilde{e}_1 \tilde{e}_2)(\tilde{e}_3 \tilde{e}_4)$. Note that we do not allow the non-planar (crossing) linking $(\tilde{e}_1 \tilde{e}_3)(\tilde{e}_2 \tilde{e}_4)$. The set of linkings at all vertices \tilde{V} defines the desired system of loops.

In concrete terms, this definition means that the loops bounce off all edges E' and cut through the corresponding dual edges. The complete correspondence is illustrated in Figure 20.

To complete the transformation, note that the number of loops $l(E')$ is the sum of the number of connected components $k(E')$ and the number of

²⁰This implies that the dual of $\mathcal{M}(G)$ is a quadrangulation \widehat{G} , which is however different from the quadrangulation $G \cup G^*$. See Figure 18c. The Potts model admits yet another representation, namely as a height model—or RSOS model—on \widehat{G} .

²¹Literally “eye basket”. This refers to a type of traditional Japanese wicker basket weave.

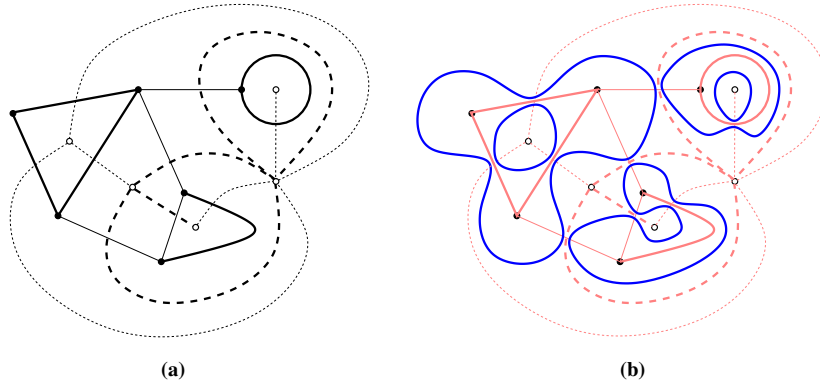


Figure 20: (a) A subset $E' \subseteq E$ (thick black solid lines) and its complementary subset $(E')^* \subseteq E^*$ (thick black dashed lines). (b) The corresponding system of self-avoiding loops on the medial graph (blue curves).

independent cycles $c(E')$:

$$l(E') = k(E') + c(E'). \quad (8.22)$$

Inserting this and the topological identity (8.13), which reads in the present notation

$$k(E') = |V| - |E'| + c(E'), \quad (8.23)$$

into (8.8) we arrive at

$$Z = Q^{|V|/2} \sum_{E' \subseteq E} x^{|E'|} Q^{l(E')/2}, \quad (8.24)$$

where we have defined $x = vQ^{-1/2}$.

This is the *loop representation* of the Potts partition function. Its importance stems from the fact that the loops, their local connectivities (called linkings in the above argument), and the non-local quantity $l(E')$ all admit an algebraic interpretation within the Temperley-Lieb algebra.

The duality transformation in the loop representation consists in shifting the linking at each vertex cyclically by one step. In terms of the x variables the duality relation (8.16) reads simply

$$xx^* = 1. \quad (8.25)$$

In the case of the square lattice, the self-dual points are $x_c = \pm 1$, and the usual critical point is $x_{c+} = 1$. The loop model (8.24) then becomes extremely simple: there is just a weight $n = \sqrt{Q}$ for each loop.

8.6 Vertex model representation

In the definition of the Q -state Potts model, Q was originally a positive integers. However, in the corresponding loop model (8.24) it appears as formal parameters and may thus take arbitrary complex values. The price to pay for this generalisation is the appearance of a non-locally defined quantity, the number of loops l . The locality of the model may be recovered by transforming it to a vertex model with complex Boltzmann weights as we now show.

The following argument supposes that $G = (V, E)$ is a (connected) planar graph. Most applications however suppose a regular lattice, a situation to which we shall return shortly.

Consider any model of self-avoiding loops defined on G (or some related graph, such as the medial graph $\mathcal{M}(G)$ for the Potts model). The Boltzmann weights are supposed to consist of a local piece—depending on if and how the loops pass through a given vertex—and a non-local piece of the form n^l , where n is the loop weight and l is the number of loops. In the case of the Potts model we have $n = \sqrt{Q}$.

In a first step, each loop is independently decorated by a global orientation $s = \pm 1$, which by planarity and self-avoidance can be described as either counterclockwise ($s = 1$) or clockwise ($s = -1$). If each oriented loop is given a weight $w(s)$, we have the requirement

$$n = w(1) + w(-1). \quad (8.26)$$

An obvious possibility, sometimes referred to as the *real loop ensemble*, is $w(1) = w(-1) = n/2$. This can be interpreted as an $O(n/2)$ model of complex spins.

We are however more interested in the *complex loop ensemble* with $w(s) = e^{is\gamma}$. Note that in the expected critical regime,

$$n = 2 \cos \gamma \in [-2, 2], \quad (8.27)$$

the parameter $\gamma \in [0, \pi]$ is real. Locality is retrieved by remarking that the weights $w(\pm 1)$ are equivalent to assigning a local weight $w(\alpha/2\pi)$ each time the loop turns an angle α (counted positive for left turns).

Note that a planar graph cannot necessarily be drawn in the plane in such a way that all edges are straight line segments. Therefore, the local weights $w(\alpha/2\pi)$ must in general be assigned both to vertices *and* to edges. However, it is certainly possible to redraw the graph so that each edge is a succession of several straight line segments. Introducing auxiliary vertices of degree two at the places where two segments join up, the weight for turning can be assigned to those auxiliary vertices. In that sense, any planar graph admits a local redistribution of the loop weight, with local weights $w(\alpha/2\pi)$ assigned only to vertices.

The loop model is now transformed into a *local vertex model* by assigning to each edge traversed by a loop the orientation of that loop. An edge not traversed by any loop is assigned no orientation. The total vertex weight is determined from the configuration of its incident oriented edges: it equals the above local loop weights summed over the possible linkings of oriented loops which are compatible with the given edge orientations. In addition, one must multiply this by any loop-independent local weights, such as x in (8.24).

8.7 Six-vertex model

To see how this is done, we finally specialise to the Potts model defined on the square lattice G . The loop model is defined on the corresponding medial lattice $\mathcal{M}(G)$ which is another (tilted) square lattice. Each edge of the lattice is visited by a loop, and two loop segments (possibly parts of the same loop) meet at each vertex. In the oriented loop representation, each vertex is therefore incident on two outgoing and two ingoing edges.

It is convenient for the subsequent discussion to make the couplings of the Potts model anisotropic. In its original spin formulation (8.5) we therefore let K_1 (resp. K_2) denote respectively the dimensionless coupling in the horizontal (resp. vertical) direction of the square lattice, and we let

$$x_1 = \frac{e^{K_1} - 1}{\sqrt{Q}}, \quad x_2 = \frac{e^{K_2} - 1}{\sqrt{Q}} \quad (8.28)$$

be the corresponding parameters appearing in the loop representation (8.24). Note that in all the results obtained this far it is straightforward to generalise to completely inhomogeneous edge dependent couplings, and the only reason that we have chosen not to present the results in this generality is that it tends to make notations slightly cumbersome.

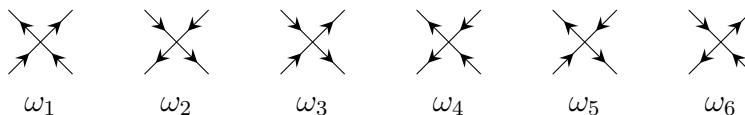


Figure 21: The allowed arrow arrangements (top) around a vertex that define the six-vertex model, with the corresponding particle trajectories (bottom).

The six possible configurations of arrows around a vertex of the medial lattice $\mathcal{M}(G)$ are shown in Fig. 21 (obtained simply by rotating Fig. 14 through $\frac{\pi}{4}$). The corresponding vertex weights are denoted ω_p (resp. ω'_p) on the even (resp. odd) sublattice of $\mathcal{M}(G)$. By definition, a vertex of the even (resp. odd) sublattice of $\mathcal{M}(G)$ is the mid point of an edge with coupling K_1 (resp. K_2) of the original spin lattice G . With respect to Figure 21 we define the even sublattice to be such that an edge $e \in E$ occupies the upper-left and lower-right quadrants, and the corresponding dual edge $e^* \in E^*$ occupies the lower-left and upper-right quadrants. For the odd sublattice, exchange e and e^* .

Using (8.24) we then have

$$Z = Q^{|V|/2} \sum_{\text{arrows}} \prod_{p=1}^6 (\omega_p)^{N_p} (\omega'_p)^{N'_p}, \quad (8.29)$$

where the sum is over arrow configurations satisfying the constraint “two in, two out” at each vertex, and N_p (resp. N'_p) is the number of vertices on the even (resp. odd) sublattice with arrow configuration p . Thus, the square-lattice Potts model has been represented as a *staggered six-vertex model*.²² The weights read explicitly

$$\omega_1, \dots, \omega_6 = 1, 1, x_1, x_1, e^{i\gamma/2} + x_1 e^{-i\gamma/2}, e^{-i\gamma/2} + x_1 e^{i\gamma/2} \quad (8.30)$$

$$\omega'_1, \dots, \omega'_6 = x_2, x_2, 1, 1, e^{-i\gamma/2} + x_2 e^{i\gamma/2}, e^{i\gamma/2} + x_2 e^{-i\gamma/2} \quad (8.31)$$

To see this, note that configurations $i = 1, 2, 3, 4$ are compatible with just one linking of the oriented loops:

$$\omega_1 = \text{loop} \quad (8.32)$$

²²The term *staggered* means that the weights alternate between sublattices.

whereas $i = 5, 6$ are compatible with two different linkings (and the weight is obtained by summing over these two):

$$\begin{array}{c} \begin{array}{c} \nearrow \\ \nwarrow \\ \nwarrow \\ \nearrow \end{array} \\ \omega_5 \end{array} = \begin{array}{c} \nearrow \\ \nwarrow \end{array} + \begin{array}{c} \nwarrow \\ \nearrow \end{array} \\ e^{i\gamma/2} \quad x_1 e^{-i\gamma/2}
 \end{array} \tag{8.33}$$

Note that the even and odd sublattices are related by a $\pi/2$ rotation of the vertices in Figure 21. This rotation interchanges configurations $(\omega_1, \omega_2) \leftrightarrow (\omega'_3, \omega'_4)$ and $\omega_5 \leftrightarrow \omega'_6$. On the level of the weights it corresponds to $x_1 \leftrightarrow x_2$.

The staggered six-vertex model is not exactly solvable in general. However, if we impose

$$x_2 = (x_1)^{-1} \tag{8.34}$$

we have $\omega'_i = (x_1)^{-1} \omega_i$ for any $i = 1, 2, \dots, 6$. The factors $(x_1)^{-1}$ from each ω'_i can be taken outside the summation in (8.29) and we have effectively $\omega'_i = \omega_i$. The six-vertex model then becomes homogeneous, hence solvable. Note that the solvability condition is nothing else than (8.25): the selfdual square-lattice Potts model is equivalent to a homogeneous six-vertex model. The results for the latter therefore apply. The anisotropy parameter is

$$\Delta = \frac{a^2 + b^2 - c^2}{2ab} = -\cos \gamma, \tag{8.35}$$

where we have replaced c^2 by $\omega_5 \omega_6$ by invoking the usual gauge symmetry.²³ This matches perfectly the parameterisation (6.13). Note that the spectral parameter u is precisely what allows us to take different horizontal and vertical couplings.

We stress once more that the square-lattice Potts model is solvable at its selfdual point, but not at arbitrary temperatures. This is in contrast with the Ising model, which is solvable at any temperature. In that sense the Ising model is a rather untypical integrable model.

However the integrable R -matrix of the six-vertex model satisfies the Yang-Baxter relations for any choice of the spectral parameters. There is one other choice that also corresponds to a Potts model. If one lets the horizontal spectral parameters alternate like $u, u + \frac{\pi}{2}, u, u + \frac{\pi}{2}, \dots$ and the

²³In the special case $x_1 = x_2 = 1$ of (8.34) we even have $\omega_5 = \omega_6$.

vertical like $v, v + \frac{\pi}{2}, v, v + \frac{\pi}{2}, \dots$ one obtains the *antiferromagnetic* transition curve of the Potts model:

$$x_1 = \frac{\sin(u)}{\sin(\gamma - u)}, \quad x_2 = -\frac{\cos(\gamma - u)}{\cos(u)}. \quad (8.36)$$

This has been analysed in [Ba82b, JS06, IJS08].

8.8 Twisted vertex model

Sometimes it is convenient to consider particular correlation functions in which the weight of some of the loops are changed. As an elementary example, consider the Potts loop model defined on a connected planar graph $G = (V, E)$ and let $i_1, i_2 \in V$ be a pair of root vertices. The partition function $Z(n)$ is given by (8.24) with loop weight $n = \sqrt{Q}$ and additional local weights at the vertices.

Define now a modified partition function $Z_1(n, n_1)$ as follows: loops on $\mathcal{M}(G)$ surrounding neither or both of the roots have an unchanged weight n , whereas those surrounding only one of the roots have a modified weight n_1 . This defines the two-point correlation function $Z_1(n, n_1)/Z(n)$. An interesting special case is provided by $n_1 = 0$, which expresses the probability that the two roots belong to the same cluster.

It is possible to produce $Z_1(n, n_1)$ in the vertex model representation, leading to a so-called *twisted vertex model*. To this end, let \mathcal{P}_{12} be an oriented self-avoiding path on G , going from i_1 to i_2 . Let us parametrise

$$n_1 = 2 \cos \gamma_1 \in [-2, 2] \quad (8.37)$$

with real $\gamma_1 \in [0, \pi]$. In the arrow formulation, we then associate a special weight \tilde{w} to any edge \tilde{e} of $\mathcal{M}(G)$ that crosses the path \mathcal{P}_{12} . The weight \tilde{w} depends on the orientation of the arrow on \tilde{e} : it equals $e^{i\gamma_1}$ (resp. $e^{-i\gamma_1}$) if the arrow points from left to right (resp. from right to left) upon viewing \tilde{e} along the direction given by \mathcal{P}_{12} .

The path \mathcal{P}_{12} is often called a *seam*, and the edges traversing it are referred to as *seam edges*.

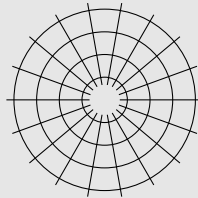
In the oriented loop representation, it is easy to see that a loop surrounding neither or both of the roots will traverse \mathcal{P}_{12} an even number of times, and the phase factors \tilde{w} will cancel out globally. However, a loop surrounding just one of the roots will have one excess factor $e^{\pm i\gamma_1}$ depending on its

global orientation (clockwise or counterclockwise), leading to (8.37) once the orientations have been summed over.

Note that the above construction of $Z_1(n, n_1)$ depends on the seam \mathcal{P}_{12} only through its end points i_1 and i_2 . In that sense, the exact shape of the seam is irrelevant and can be deformed at will.

Finally, the weights \tilde{w} can be absorbed in the vertex weights, by incorporating them in the weight of the vertex at the right (with respect to the orientation defined by \mathcal{P}_{12}) end point of \tilde{e} .

These considerations are important when discussing boundary effects. Suppose we wish to define the square-lattice Potts model on a cylinder with periodic boundary conditions. This is still a planar graph, but if we decide to draw it as such, i.e., in cobweb shape



(8.38)

the edges will be curved and additional complex phase factors will be picked up by the loops. These extra phases will cancel out for oriented loops that do not encircle the origin, and the usual weight $n = 2 \cos \gamma$ will result from summing over orientations. However, loops that do encircle the origin will finally a wrong weight $\bar{n} = 2$.

Alternatively, one may draw the cylinder as a standard square lattice with periodic boundary conditions across. In this version, oriented loops that are not homotopic to a point will not turn a total angle $\alpha = \pm 2\pi$, but rather $\bar{\alpha} = 0$. They thus get the weight $\bar{n} = 2$ as above: the two points of view are equivalent.

The introduction of a seam running from the origin to the point at infinity will change the weighting: $\bar{n} = n_1$. In particular, setting $n_1 = n$ we obtain the true Potts model. Such subtleties are important when discussing critical exponents, since these will in fact depend on n_1 .

Note finally that the case of doubly periodic (toroidal) boundary conditions is quite delicate, since most of the equivalences presented in this chapter depend crucially on the planarity of the graph.

9 Temperley-Lieb algebra

The Temperley-Lieb algebra $TL_N(n)$ is a unital associative algebra over \mathbb{C} . Its $N - 1$ generators are denoted E_m for $m = 1, 2, \dots, N - 1$. They satisfy the relations

$$\begin{aligned} (E_m)^2 &= n E_m, \\ E_m E_{m\pm 1} E_m &= E_m, \\ E_m E_{m'} &= E_{m'} E_m \text{ for } |m - m'| > 1. \end{aligned} \quad (9.1)$$

It will turn out useful to define the q -deformed numbers

$$[k]_q = \frac{q^k - q^{-k}}{q - q^{-1}} \quad (9.2)$$

and parameterise

$$n = q + q^{-1} = [2]_q. \quad (9.3)$$

Notice that in the previous chapter we had $n = 2 \cos \gamma$, so that $q = e^{i\gamma}$. For $k \in \mathbb{N}$ the q -number is actually a polynomial in n :

$$[k + 1]_q = U_k(n/2), \quad (9.4)$$

where $U_k(x)$ is the k th order Chebyshev polynomial of the second kind

$$U_k(\cos \theta) = \frac{\sin((k + 1)\theta)}{\sin(\theta)}. \quad (9.5)$$

The algebra $TL_N(n)$ can be represented in many ways. We shall be particularly interested in its *loop-model representation*, since this permits us to make contact with the previous chapter. In this representation, $TL_N(n)$ is viewed as an algebra of diagrams acting on N numbered vertical strands (for convenience depicted inside a dashed box) as

$$E_m = \begin{array}{c} \boxed{\begin{array}{ccccccc} | & | & \dots & \cup & \dots & | & | \\ & & & \cap & & & \\ 1 & 2 & & m & m+1 & & N-1 & N \end{array}} \end{array}$$

Multiplication in $TL_N(n)$ is defined by stacking diagrams vertically. More precisely, the product of two generators $g_2 g_1$ is defined by placing the diagram

for g_2 above the diagram for g_1 , identifying the bottom points of g_2 with the top points of g_1 . The resulting diagram is considered up to smooth isotopies that keep fixed the surrounding box, and any closed loop is replaced by the factor n .

In this way we have for instance (omitting strands on which the action is trivial)

$$(E_m)^2 = \boxed{\begin{array}{c} \text{---} \\ \cup \\ \text{---} \\ \circ \\ \text{---} \\ \cup \\ \text{---} \end{array}} = n \boxed{\begin{array}{c} \text{---} \\ \cup \\ \text{---} \\ \cup \\ \text{---} \end{array}} = n E_m$$

and

$$E_m E_{m+1} E_m = \boxed{\begin{array}{c} \text{---} \\ \cup \\ \text{---} \\ \cup \\ \text{---} \\ \cup \\ \text{---} \\ \cup \\ \text{---} \end{array}} = \boxed{\begin{array}{c} \text{---} \\ \cup \\ \text{---} \\ \cup \\ \text{---} \end{array}} = E_m.$$

It is thus readily seen that all the defining relations (9.1) are satisfied. Moreover, for generic values of n no further relations hold: the loop-model representation is faithful.

9.1 Integrable \check{R} -matrix

Starting from first principles, we now construct an integrable model based on the TL algebra. Let us suppose that the \check{R} -matrix has the form

$$\check{R}_{m,m+1}(u) = f(u)I + g(u)E_m, \quad (9.6)$$

where $f(u)$ and $g(u)$ are some functions of the spectral parameter u to be determined. Inserting this into the Yang-Baxter equation (6.8) yields

$$\begin{aligned} & (f(u)I + g(u)E_2) (f(u+v)I + g(u+v)E_1) (f(v)I + g(v)E_2) = \\ & (f(v)I + g(v)E_1) (f(u+v)I + g(u+v)E_2) (f(u)I + g(u)E_1). \end{aligned} \quad (9.7)$$

Using the algebraic relations (9.1) we can expand both sides of (9.7). The left-hand side produces

$$\begin{aligned} & f(u)f(u+v)f(v)I + f(u)g(u+v)f(v)E_1 + \\ & g(u)g(u+v)f(v)E_2E_1 + f(u)g(u+v)g(v)E_1E_2 + \\ & [g(u)g(v)(g(u+v) + nf(u+v)) + f(u+v)(f(u)g(v) + f(v)g(u))] E_2, \end{aligned}$$

and the right-hand side becomes

$$\begin{aligned} & f(v)f(u+v)f(u)I + f(v)g(u+v)f(u)E_2 + \\ & f(v)g(u+v)g(u)E_2E_1 + g(v)g(u+v)f(u)E_1E_2 + \\ & [g(u)g(v)(g(u+v) + nf(u+v)) + f(u+v)(f(u)g(v) + f(v)g(u))] E_1. \end{aligned}$$

These expressions must be identical in $TL_3(n)$, and so we can identify the coefficients for each of the five possible words in the algebra. The relations resulting from the words I , E_1E_2 and E_2E_1 are trivial. The relations coming from E_1 and E_2 are identical—related via an exchange of the left- and right-hand sides—and read

$$\begin{aligned} & g(u)g(v)(g(u+v) + nf(u+v)) + f(u+v)(f(u)g(v) + f(v)g(u)) = \\ & f(u)f(v)g(u+v). \end{aligned} \tag{9.8}$$

The functional relation (9.8) is a typical outcome of this way of solving the Yang-Baxter equations. It is in general not easy to solve this type of relation, and even if one finds solutions it is often difficult to make sure that one has found *all* the solutions. Worse, in more complicated cases than the one considered here the Ansatz for the \check{R} -matrix will involve more terms and the functions $f(u)$, $g(u)$, . . . must satisfy several coupled functional equations.

It is useful to rewrite (9.8) in terms of the parameters $z = e^{iu}$, $w = e^{iv}$ and $q = e^{i\gamma}$. That is, instead of the *additive* spectral parameters u, v we have now *multiplicative* spectral parameters z, w . Thus

$$\begin{aligned} & g(z)g(w)(g(zw) + (q + q^{-1})f(zw)) + f(zw)(f(z)g(w) + f(w)g(z)) = \\ & f(z)f(w)g(zw). \end{aligned} \tag{9.9}$$

It is tempting to set $f(z) = 1$, since the overall normalisation of the \check{R} -matrix is unimportant, but in general this is *not* a good idea. A time proven strategy is to suppose that $f(z)$ and $g(z)$ are polynomials of some small degree in the variables z , z^{-1} , q and q^{-1} . (In some cases one needs to try fractional powers of q as well.) In this case we are lucky: there is a solution of degree one

$$f(z) = \frac{q}{z} - \frac{z}{q}, \tag{9.10}$$

$$g(z) = z - z^{-1}. \tag{9.11}$$

Verifying that this *is* a solution is of course easy. Finding it from scratch already calls for the use of symbolic algebra software such as MATHEMATICA.

Going back to additive spectral parameters, we thus have a trigonometric solution of (9.7):

$$f(u) = \sin(\gamma - u), \quad g(u) = \sin(u). \quad (9.12)$$

Note that this agrees with (6.18).

In general, solutions to the Yang-Baxter equation turn out to be polynomial, trigonometric or elliptic (in order of increasing difficulty).

9.2 Transfer matrix decomposition

In the remainder of this chapter we shall be interested in the Q -state Potts model defined on an $L \times M$ annulus of width L spins and of circumference M spins. The boundary conditions are free in the space (L , horizontal) direction and periodic in the time (M , vertical) direction.

We work in the loop representation in order to make contact with the Temperley-Lieb algebra $TL_N(n)$ defined on $N = 2L$ strands and with loop weight $n = \sqrt{Q}$. The transfer matrix can be read off from (8.24):

$$T = Q^{L/2} \left(\prod_{m=1}^{L-1} (I + x_1 E_{2m}) \right) \left(\prod_{m=1}^L (x_2 I + E_{2m-1}) \right), \quad (9.13)$$

where x_1 (resp. x_2) defines the horizontal (resp. vertical) coupling constant through (8.28).

We have seen in (8.34) that the Potts model is solvable if $x_2 = (x_1)^{-1}$. In that case we have

$$T = \left(\frac{\sqrt{Q}}{x_1} \right)^L \left(\prod_{m=1}^{L-1} (I + x_1 E_{2m}) \right) \left(\prod_{m=1}^L (I + x_1 E_{2m-1}) \right). \quad (9.14)$$

We recognise here the integrable \check{R} -matrix (9.6) and identify

$$x_1 = \frac{g(u)}{f(u)} = \frac{\sin(u)}{\sin(\gamma - u)}. \quad (9.15)$$

According to (8.30) we have also $x_1 = \frac{\omega_3}{\omega_1}$, and we note that this agrees precisely with the parameterisation (6.13) used when studying the six-vertex model.

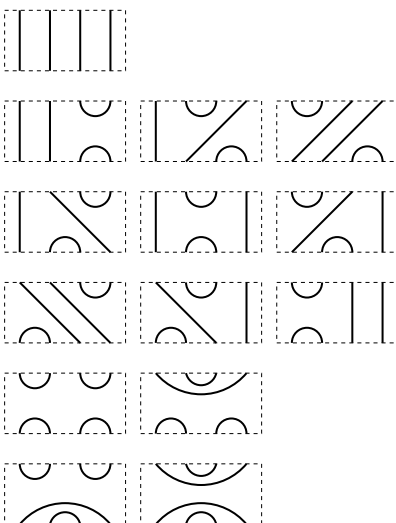


Figure 22: List of all TL states on $N = 4$ strands. Each row corresponds to a definite sector of the transfer matrix.

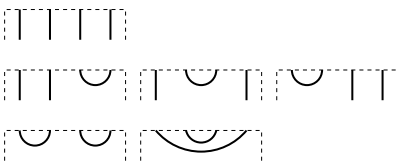


Figure 23: List of all TL reduced states on $N = 4$ strands. Each row corresponds to a definite sector of the transfer matrix.

The transfer matrix T acts on states which can be depicted diagrammatically as non-crossing link patterns within a box bordered by two horizontal rows, each of N points. The complete list of states for $N = 4$ is shown in Fig. 22. The bottom (resp. top) side of the box corresponds to time $t = 0$ (resp. $t = t_0$); the transfer matrix propagates the states from t_0 to $t_0 + 1$ and thus acts on the top of the box only.

A link joining the top and the bottom of the box is called a *string*, and any other link is called an *arc*. We denote by s the number of strings in a given state. Any state can be turned into a pair of *reduced states* by cutting all its strings and pulling apart the upper and lower parts. For convenience,

a cut string will still be called a string with respect to the reduced state. The complete list of reduced states for $N = 4$ is shown in Fig. 23.

Conversely, a state can be obtained by adjoining two reduced states, gluing together their strings in a unique fashion. Thus, if we define d_{2j} as the number of reduced states with $s = 2j$ strings, the number of states with $s = 2j$ strings is simply $(d_{2j})^2$.

The partition function $Z_{N,M}$ on an annulus of width N strands and height M units of time cannot be immediately expressed in terms of reduced states only, since these do not contain the information about how many loops (contractible or non-contractible) are formed when the periodic boundary condition is imposed. We can however write it in terms of states as

$$Z_{N,M}(n, \ell) = \langle u | T^M | v \rangle. \quad (9.16)$$

It is useful to slightly generalise the problem by giving the weight n to contractible loops and a different weight ℓ to non-contractible loops. Recalling (9.3) we shall parameterise the latter as²⁴

$$\ell = t + t^{-1} = [2]_t. \quad (9.17)$$

At time $t_0 = 0$ the top and the bottom of the box must be identified. Therefore, the right vector $|v\rangle$ is just the unit vector corresponding to the unique state that contains no arcs and N strings (i.e., each link connects a point on the bottom to the point immediately above it on the top).

At time $t_0 = M$ the top and the bottom of the box must be reglued. Therefore, the left vector $\langle u|$ is obtained by identifying the top and bottom sides for each state. Counting the number of loops of each type gives the corresponding weight as a monomial in the loop weights n and ℓ .

The reduced states can be ordered according to a decreasing number of strings. The states can be ordered first according to a decreasing number of strings, and next, for a fixed number of strings, according to its bottom half reduced state. These orderings are brought out by the rows in Figs. 22–23.

With this ordering, T has a blockwise lower triangular structure in the basis of reduced states, since the generator e_i can annihilate two strings (if their position on the top of the box are i and $i + 1$) but cannot create any strings.

In the basis of states, T is blockwise lower triangular with respect to the number of strings, for the same reason. Each block on the diagonal in

²⁴This t has of course nothing to do with the “time” discussed above.

this decomposition corresponds to a definite number of strings. The block corresponding to $s = 2j$ strings is denoted \widetilde{T}_{2j} . But since T acts only on the top of the box, each $\widetilde{T}_{2j} = T_{2j} \oplus \dots \oplus T_{2j}$ is in turn a direct sum of d_{2j} identical blocks T_{2j} which correspond simply to the action of T on the reduced states with $2j$ strings.

In particular, the eigenvalues of T are the union of the eigenvalues of T_{2j} , where the T_{2j} now act in the much smaller basis of reduced states. This observation is particularly useful in numerical studies.

9.3 The dimensions d_k and D_k

In spite of the periodic boundary conditions, $Z_{N,M}(n, \ell)$ is obviously not a usual matrix trace. It can however be decomposed on standard traces by constructing the transfer matrix blocks T_k algebraically within $TL_N(n)$. We shall come back to this issue in the following sections.

For each block T_k we define the corresponding character as

$$K_k = \text{tr} (T_k)^M, \quad (9.18)$$

where we stress that the trace is over *reduced* states. Obviously we have

$$K_k = \sum_{i=1}^{d_k} \left(\lambda_i^{(k)} \right)^M, \quad (9.19)$$

where $\lambda_i^{(k)}$ are the eigenvalues of T_k . We recall that $d_k = \dim T_k$.

The expression of the partition function in terms of transfer matrix eigenvalues is more involved, due essentially to the non-local nature of the loops, and reads

$$Z_{N,M} = \sum_{j=0}^L D_{2j} K_{2j}, \quad (9.20)$$

where D_k are some eigenvalue amplitudes to be determined. We shall provide the answer in the next sections, using algebraic means.

In view of the Schur-Weyl duality (mentioned briefly in the introduction to chapter 8) the D_k can also be interpreted as the (quantum) dimensions of the commutant of $TL_N(n)$, which is the quantum algebra $U_q(\mathfrak{sl}_2)$. In the corresponding bimodule, the partition function (9.20) therefore has a *multiplicity free* decomposition.

Determining the d_k is an exercise of elementary combinatorics that we deal with now. Let $E(j, k)$ denote the number of reduced states on $2j$ strands, and using $2k$ strings, so that $d_{2j} = E(L, j)$. The corresponding generating function reads

$$E^{(k)}(z) = \sum_{j=0}^{\infty} E(j, k) z^j, \quad (9.21)$$

where z is a formal parameter representing the weight of an arc, or of a pair of strings. When $k = 0$, a reduced state with no strings is either empty, or has a leftmost arc which divides the space into two parts (inside the arc and to its right) each of which can accommodate an independent arc state. The generating function $f(z) \equiv E^{(0)}(z)$ therefore satisfies $f(z) = 1 + zf(z)^2$ with regular solution

$$f(z) = \frac{1 - \sqrt{1 - 4z}}{2z} = \sum_{j=0}^{\infty} \frac{(2j)!}{j!(j+1)!} z^j. \quad (9.22)$$

When $k \neq 0$, the strings simply divide the space into $2k + 1$ parts each of which contains an independent arc state. Therefore,

$$E^{(k)}(z) = z^k f(z)^{2k+1} = \sum_{j=k}^{\infty} \left[\binom{2j}{j+k} - \binom{2j}{j+1+k} \right] z^j \quad (9.23)$$

and in particular we have

$$d_{2j} = E(L, j) = \binom{2L}{L+j} - \binom{2L}{1+L+j}. \quad (9.24)$$

Note that d_{2j} depends on the number of strands $N = 2L$, but we usually will not mention this explicitly.

The total number of reduced states is

$$\sum_{j=0}^L d_{2j} = \binom{2L}{L}, \quad (9.25)$$

while the total number of (non-reduced) states is

$$\sum_{j=0}^L (d_{2j})^2 = E(2L, 0) = \frac{1}{2L+1} \binom{4L}{2L}. \quad (9.26)$$

In particular for $N = 2L = 4$ we have

$$d_4 = 1, \quad d_2 = 3, \quad d_0 = 2,$$

in agreement with the number of reduced states shown in each row of Fig. 23. The total number of states is $1^2 + 3^2 + 2^2 = 14$ in agreement with Fig. 22.

9.4 Jones-Wenzl projectors

We have decomposed the full transfer matrix T to elementary blocks T_k that have the property that the number of strings is precisely k and cannot be lowered by the action of $TL_N(n)$. In more algebraic terms, T_k is the restriction of T to a representation with precisely k strings. This representation is known as the standard module \mathcal{V}_k . It can be shown that \mathcal{V}_k is irreducible when q is not a root of unity.

Within \mathcal{V}_k , the generator E_m annihilates any (reduced) state for which the strands at positions m and $m + 1$ are both strings, and acts in the usual way on any other state. There exists an algebraic object that imposes this restriction: the Jones-Wenzl (JW) projector.

The JW projector $P_k \in TL_k(n)$ is defined by the recursion relation

$$P_{k+1} = P_k - \frac{[k]_q}{[k+1]_q} P_k E_k P_k, \quad \text{for } k \geq 1, \quad (9.27)$$

and the initial condition $P_1 = I$. Both sides of this equation act in $TL_{k+1}(n)$. To keep the notation simple it is implicitly understood that the projector P_k acts only on the k leftmost strands.

The first few JW projectors read explicitly

$$\begin{aligned} P_1 &= I, \\ P_2 &= I - \frac{1}{n} E_1, \\ P_3 &= I - \frac{n}{n^2 - 1} (E_1 + E_2) + \frac{1}{n^2 - 1} (E_1 E_2 + E_2 E_1). \end{aligned} \quad (9.28)$$

In the sequel it will turn out useful to have a diagrammatic representation of P_k . We shall represent it as a bar across the strands being projected (here and in the following all pictures are for $k = 4$):

$$P_k = \begin{array}{c} | \quad | \quad | \quad | \\ \hline | \quad | \quad | \quad | \end{array} \quad (9.29)$$

The JW has two crucial properties. First, it is idempotent and bigger projectors swallow smaller ones:

$$P_m P_k = P_k P_m = P_k, \quad \text{for } 1 \leq m \leq k. \quad (9.30)$$

Second, no contractions are allowed among the strands having been projected:

$$E_m P_k = P_k E_m = 0, \quad \text{for } 1 \leq m \leq k-1. \quad (9.31)$$

The proof of the properties (9.30)–(9.31) is by induction in k . The case $k = 1$ is obvious: since $P_1 = I$, the first property (9.30) is trivial, and for the second property (9.31) there is nothing to be shown. Suppose therefore that both properties hold for k and let us show them for $k + 1$. For convenience we write $\alpha_k = [k]_q/[k+1]_q$.

We consider first (9.31). For $m < k$ we have

$$E_m P_{k+1} = \begin{array}{c} \text{diagram 1} \\ \text{diagram 2} \end{array} = \begin{array}{c} \text{diagram 3} \\ \text{diagram 4} \end{array} - \alpha_k \begin{array}{c} \text{diagram 5} \\ \text{diagram 6} \end{array} \quad (9.32)$$

where we have used (9.27). Both diagrams on the right-hand side are zero by the induction hypothesis (9.31), whence $E_m P_{k+1} = 0$ as required.

The argument for $m = k$ is slightly more involved. We first use (9.27) to write

$$E_k P_{k+1} = \begin{array}{c} \text{diagram 1} \\ \text{diagram 2} \end{array} - \alpha_k \begin{array}{c} \text{diagram 3} \\ \text{diagram 4} \end{array} \quad (9.33)$$

In the second term on the right-hand side, the small loop cannot yet be replaced by n since it is “trapped” by the projector. We therefore use (9.27) once more:

$$\begin{array}{c} \text{diagram 1} \\ \text{diagram 2} \end{array} = \begin{array}{c} \text{diagram 3} \\ \text{diagram 4} \end{array} - \alpha_{k-1} \begin{array}{c} \text{diagram 5} \\ \text{diagram 6} \end{array} = (n - \alpha_{k-1}) \begin{array}{c} \text{diagram 7} \\ \text{diagram 8} \end{array} \quad (9.34)$$

where in the last step use was made of the induction hypothesis (9.30). Inserting (9.34) into (9.33) we obtain

$$E_k P_{k+1} = \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} \cup - \alpha_k(n - \alpha_{k-1}) \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} \cup \quad (9.35)$$

but by (9.30) the two diagrams on the right-hand side are identical. To have $E_k P_{k+1} = 0$ as required, we therefore need the coefficient to vanish

$$1 - \alpha_k(n - \alpha_{k-1}) = 0,$$

which is easily shown to be equivalent to

$$[2]_q [k]_q = [k-1]_q + [k+1]_q. \quad (9.36)$$

But (9.36) is precisely the recursion relation satisfied by $[k]_q$, so (9.31) is proved.

Let us note that (9.34) is actually a quite useful identity. In algebraic terms, and using (9.36), it can be written

$$E_k P_k E_k = \frac{[k+1]_q}{[k]_q} E_k P_{k-1}. \quad (9.37)$$

We still need to prove (9.30) for $k+1$ and $m \leq k+1$. This is done by induction in m . Since $P_1 = I$ the statement is trivial for $m = 1$. Suppose now $m \leq k$ and that the statement has been proved for $m-1$. We have

$$P_m P_{k+1} = \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} = \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} - \alpha_{m-1} \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} \cup \quad (9.38)$$

After using the induction hypothesis on both diagrams on the right-hand side we obtain

$$\begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} = \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} - \alpha_{m-1} \begin{array}{c} | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \\ \text{---} \\ | \quad | \quad | \quad | \end{array} \cup \quad (9.39)$$

But the second diagram on the right-hand side vanishes by (9.31), so we have $P_m P_{k+1} = P_{k+1}$ as required.

Exactly the same argument applies when $m = k + 1$, so the proof is complete.

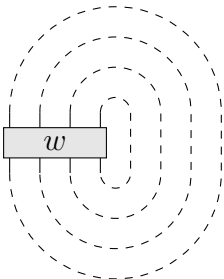
9.5 Markov trace

One of our main objectives is to compute the annulus partition function $Z_{N,M}(n, \ell)$ given by (9.16). This calls for algebraic way of imposing the periodic boundary conditions in the time direction. This motivates the following definition of the Markov trace.

Let $w \in TL_N(n)$ be a word in the TL algebra. We can represent w as a diagram in a box, cf. Fig. 22. The Markov trace of w is defined as

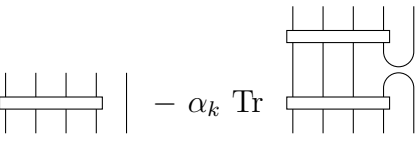
$$\text{Tr } w = n^{N_1} \ell^{N_2}, \quad (9.40)$$

where N_1 (resp. N_2) is the number of contractible (resp. non-contractible) loops formed when identifying the top and the bottom sides of the box. This definition is extended by linearity to the whole algebra $TL_N(n)$. Pictorially we can write

$$\text{Tr } w = \text{Diagram of } w \text{ in a box with } N_1 \text{ contractible and } N_2 \text{ non-contractible loops} \quad (9.41)$$


Contractible (resp. non-contractible) loops are those that cover an even (resp. odd) number of dashed lines. The corresponding weights have been defined as $n = [2]_q$ and $\ell = [2]_t$.

In particular wish to know the Markov trace of the JW projectors P_k . This can be found by using (9.27):

$$\text{Tr } P_{k+1} = \text{Tr} \left[\text{Diagram 1} \right] - \alpha_k \text{Tr} \left[\text{Diagram 2} \right] \quad (9.42)$$


In the second diagram on the right-hand side we can slide the uppermost projector across the periodic boundary condition to the bottom, where it gets swallowed by the other projector. We thus have

$$\mathrm{Tr} P_{k+1} = [2]_t \mathrm{Tr} P_k - \alpha_k \mathrm{Tr} \left(\begin{array}{c} | \\ | \\ | \\ | \\ \hline | \\ | \\ | \\ | \\ \circ \end{array} \right) \quad (9.43)$$

Note that the last diagram does *not* equal $\mathrm{Tr} P_{k-1}$. Indeed, the small loop on the right is contractible. For the moment it is “trapped” by the projector, but we can liberate it by repeating the argument of (9.34). We arrive at

$$\mathrm{Tr} P_{k+1} = [2]_t \mathrm{Tr} P_k - \alpha_k ([2]_q - \alpha_{k-1}) \mathrm{Tr} P_{k-1}. \quad (9.44)$$

Thanks to (9.36) we have $\alpha_k ([2]_q - \alpha_{k-1}) = 1$, and so we have the recursion relation

$$[2]_t \mathrm{Tr} P_k = \mathrm{Tr} P_{k-1} + \mathrm{Tr} P_{k+1} \quad (9.45)$$

with initial conditions $\mathrm{Tr} P_0 = [1]_t = 1$ and $\mathrm{Tr} P_1 = [2]_t = \ell$. Invoking again (9.36) the solution is

$$\mathrm{Tr} P_k = [k + 1]_t = U_k(\ell/2). \quad (9.46)$$

It is rather remarkable that this depends only on ℓ , and not on n .

9.6 Decomposition of the Markov trace

Assume that q is not a root of unity, so that the standard modules \mathcal{V}_j are irreducible. We now define a scalar product in \mathcal{V}_j . Given two reduced states $|v_1\rangle, |v_2\rangle \in \mathcal{V}_j$, cf. Fig. 23, each containing j strings. The scalar product $\langle v_1 | v_2 \rangle$ is obtained by reflecting $|v_1\rangle$ in a horizontal mirror, then gluing together the two states. We define $\langle v_1 | v_2 \rangle = 0$ unless each string in v_2 connects onto a string in v_1 . Otherwise we attribute a weight $[2]_q = n$ to each closed loop and 1 to each string in the compound diagram.

Let us give some examples in $TL_4(n)$. The Gram matrices of scalar products in \mathcal{V}_4 , \mathcal{V}_2 and \mathcal{V}_0 , cf. Fig. 23, read

$$M_4 = [1], \quad M_2 = \begin{bmatrix} n & 1 & 0 \\ 1 & n & 1 \\ 0 & 1 & n \end{bmatrix}, \quad M_0 = \begin{bmatrix} n^2 & n \\ n & n^2 \end{bmatrix}. \quad (9.47)$$

As a side remark we point out that the determinants of the Gram matrices may vanish if and only if q is a root of unity. We are supposing throughout that this is not the case, so that the representation theory is generic. In more technical terms, the algebra $TL_N(n)$ is supposed to be semi-simple.

A related observation is that the JW projectors are ill-defined when q is a root of unity.

Since \mathcal{V}_j are generic there exists a basis \mathcal{B}_j which is orthonormal with respect to the scalar product:

$$\forall b_k, b_l \in \mathcal{B}_j : \langle b_k | b_l \rangle = \delta_{k,l}. \quad (9.48)$$

We also extend our definition of the scalar product so that states belonging to different standard modules \mathcal{V}_j and $\mathcal{V}_{j'}$, with $j \neq j'$, are orthogonal.

The next step is to construct an element $P_{N,j} \in TL_N(n)$ that projects on \mathcal{V}_j . In other words, $P_{N,j}$ must act as the identity on \mathcal{V}_j and annihilate all states of $\mathcal{V}_{j'}$ with $j' \neq j$. Obviously $P_{N,N} = P_N$ is just the familiar JW projector. For arbitrary j we define (the diagrammatic representation shows the case $N = 4$ and $j = 2$)

$$P_{N,j} = \sum_{b \in \mathcal{B}_j} |b\rangle \circ P_j \circ \langle b| = \sum_{b \in \mathcal{B}_j} \begin{array}{c} \text{---} \\ | \\ \text{---} \\ b \\ \text{---} \\ | \\ \text{---} \\ b \\ \text{---} \\ | \\ \text{---} \end{array} \quad (9.49)$$

This has the required properties and satisfies the completeness relation

$$\sum_{j=0}^N P_{N,j} = I \in TL_N(n). \quad (9.50)$$

We are now in a position to attain the goal of decomposing the Markov trace of any element $w \in TL_N(n)$ over standard traces—i.e., traces with respect to the basis states $b \in \mathcal{B}_j$. We first decompose w using (9.50):

$$w = \sum_j P_{N,j} w = \sum_j \begin{array}{c} \text{---} \\ | \\ \text{---} \\ P_{N,j} \\ \text{---} \\ | \\ \text{---} \\ w \\ \text{---} \\ | \\ \text{---} \end{array} \quad (9.51)$$

Inserting the definition (9.49) and taking the Markov trace yields

$$\text{Tr} \begin{array}{c} P_{N,j} \\ w \end{array} = \sum_{b \in \mathcal{B}_j} \begin{array}{c} b \\ b \\ w \end{array} = \sum_{b \in \mathcal{B}_j} \begin{array}{c} b \\ w \\ b \end{array} \quad (9.52)$$

Returning to algebraic terms, this means that we have shown

$$\text{Tr} w = \sum_j \text{Tr} P_j \sum_{b \in \mathcal{B}_j} \langle b|w|b \rangle = \sum_j [j+1]_t \text{tr}_{\mathcal{V}_j} w. \quad (9.53)$$

The main result (9.53) applies in particular when $w = T^M$, the M th power of the Potts model partition function. We have thus finished the demonstration that $Z_{N,M}$ indeed has the form (9.20) and identified the eigenvalue amplitudes as $D_j = [j+1]_t$. The characters

$$K_j = \text{tr}_{\mathcal{V}_j} T^M \quad (9.54)$$

defined in () are usually written $K_{1,1+j}$ for reasons that will become clear later on. We can thus summarise our final result as

$$Z_{N,M}(n, \ell) = \sum_{j=0}^L [1+2j]_t K_{1,1+2j}(n). \quad (9.55)$$

The point is that the $K_{1,1+2j}$ can be computed exactly in the continuum limit, using CFT techniques. The expression (9.55) will then give access to—among many other things—exact crossing formulae in percolation.

Finally one should also note the sum rule

$$\sum_{j=0}^L d_{2j} D_{2j} = \ell^N, \quad (9.56)$$

which expresses the fact that there are ℓ degrees of freedom living on each site.

10 Basic aspects of CFT

An important break-through occurred in 1984 when Belavin, Polyakov and Zamolodchikov [BPZ84] applied ideas of conformal invariance to classify the possible types of critical behaviour in two dimensions. These ideas had emerged earlier in string theory and mathematics, and in fact go back to earlier (1970) work of Polyakov [Po70] in which global conformal invariance is used to constrain the form of correlation functions in d -dimensional theories. It is however only by imposing *local* conformal invariance in $d = 2$ that this approach becomes really powerful. In particular, it immediately permitted a full classification of an infinite family of conformally invariant theories (the so-called “minimal models”) having a finite number of fundamental (“primary”) fields, and the exact computation of the corresponding critical exponents. In the aftermath of these developments, conformal field theory (CFT) became for some years one of the most hectic research fields of theoretical physics, and indeed has remained a very active area up to this date.

This chapter focusses on the basic aspects of CFT, with a special emphasis on the ingredients which will allow us to tackle the geometrically defined loop models via the so-called Coulomb Gas (CG) approach. The CG technique will be exposed in the following chapter. The aim is to make the presentation self-contained while remaining rather brief; the reader interested in more details should turn to the comprehensive textbook [DMS87] or the Les Houches volume [LH89].

10.1 Global conformal invariance

A conformal transformation in d dimensions is an invertible mapping $\mathbf{x} \rightarrow \mathbf{x}'$ which multiplies the metric tensor $g_{\mu\nu}(\mathbf{x})$ by a space-dependent scale factor:

$$g'_{\mu\nu}(\mathbf{x}') = \Lambda(\mathbf{x})g_{\mu\nu}(\mathbf{x}). \quad (10.1)$$

Note that such a mapping preserves angles. Therefore, just as Wilson [Wi69] suggested using *global* scale invariance as the starting point for investigating a system at its critical point, Polyakov [Po70] proposed imposing the *local* scale invariance (10.1) as the fundamental requirement for studying a critical system in which the microscopic interactions are short ranged.

A priori, a geometrical model of self-avoiding objects such as loops does not seem to be governed by short-range interactions. However, we have

already seen in (8.30)–(8.31) how to transform it into a vertex model with local interactions (albeit the complex Boltzmann weights still point to its non-local origin). We shall see later that the critical exponents of the Potts model can indeed be derived by CFT and CG techniques.

10.1.1 The conformal group

We first investigate the consequences of (10.1) for an infinitesimal transformation of the form²⁵

$$x^\mu \rightarrow x'^\mu = x^\mu + \epsilon^\mu(\mathbf{x}). \quad (10.2)$$

To first order in ϵ the change in metric is given by

$$\begin{aligned} g'_{\mu\nu} &= \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta} \\ &= (\partial_\mu^\alpha - \partial_\mu \epsilon^\alpha)(\partial_\nu^\beta - \partial_\nu \epsilon^\beta) g_{\alpha\beta} \\ &= g_{\mu\nu} - (\partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu). \end{aligned} \quad (10.3)$$

The requirement (10.1) means that

$$\partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu = f(\mathbf{x}) g_{\mu\nu}, \quad (10.4)$$

where the factor $f(\mathbf{x})$ can be determined by taking traces on both sides of (10.4):

$$f(\mathbf{x}) = \frac{2}{d} \partial_\rho \epsilon^\rho. \quad (10.5)$$

We can assume that the conformal transformation amounts to an infinitesimal deformation of the standard Cartesian metric $g_{\mu\nu} = \eta_{\mu\nu}$, where $\eta_{\mu\nu}$ is the d -dimensional identity matrix. By differentiating (10.4), permuting indices and forming a linear combination one establishes

$$2\partial_\mu \partial_\nu \epsilon_\rho = \eta_{\mu\rho} \partial_\nu f + \eta_{\nu\rho} \partial_\mu f - \eta_{\mu\nu} \partial_\rho f \quad (10.6)$$

and contracting this with $\eta^{\mu\nu}$ we arrive at

$$2\partial^2 \epsilon_\mu = (2 - d) \partial_\mu f. \quad (10.7)$$

On the other hand, applying ∂_ν to (10.7) and ∂^2 to (10.4) gives

$$(2 - d) \partial_\mu \partial_\nu f = \eta_{\mu\nu} \partial^2 f, \quad (10.8)$$

²⁵Below we use the summation convention on repeated indices.

and contracting this with $\eta^{\mu\nu}$ leads to

$$(d-1)\partial^2 f = 0. \quad (10.9)$$

We are now ready to draw some important conclusions from (10.8)–(10.9) and valid in arbitrary dimension d . The case $d = 1$ is somewhat particular, since no constraints on f are implied: any smooth transformation is conformal. On the other hand, we are not likely to need CFT to solve simple short-ranged one-dimensional models! The case $d = 2$ is where CFT has the most to offer, and we shall discuss it in detail later.

For the moment we thus concentrate on the case $d \geq 3$. Eqs. (10.8)–(10.9) imply that $\partial_\mu \partial_\nu f = 0$, whence f is at most linear in the coordinates. Using (10.6) this means that $\partial_\mu \partial_\nu \epsilon_\rho$ is constant, whence

$$\epsilon_\mu = a_\mu + b_{\mu\nu} x^\nu + c_{\mu\nu\rho} x^\nu x^\rho \quad \text{with } c_{\mu\nu\rho} = c_{\mu\rho\nu}. \quad (10.10)$$

Since the above discussion holds for all \mathbf{x} , we may treat each power of the coordinates separately. The constant term

$$\epsilon_\mu = a_\mu \quad (10.11)$$

corresponds obviously to translations. For the linear term it is useful to distinguish between the diagonal and off-diagonal parts. The former

$$\epsilon_\mu = \lambda x^\mu \quad (10.12)$$

corresponds to dilatations, while the latter

$$\epsilon_\mu = \omega_{\mu\nu} x^\nu, \quad (10.13)$$

with $\omega_{\mu\nu} = -\omega_{\nu\mu}$ an antisymmetric tensor, corresponds to rotations.

The important new ingredient comes from the quadratic term which corresponds to the *special conformal transformation* (SCT). It can be written as (after some work)

$$x'^\mu = \frac{x^\mu - b^\mu \mathbf{x}^2}{1 - 2\mathbf{b} \cdot \mathbf{x} + b^2 \mathbf{x}^2}, \quad (10.14)$$

or equivalently as a translation, preceded and followed by an inversion $x^\mu \rightarrow x'^\mu = x^\mu / \mathbf{x}^2$, viz.

$$\frac{x'^\mu}{\mathbf{x}'^2} = \frac{x^\mu}{\mathbf{x}^2} - b^\mu. \quad (10.15)$$

The infinitesimal form of the SCT is found by developing (10.14) to linear order in b^μ :

$$x'^\mu = x^\mu + 2(\mathbf{x} \cdot \mathbf{b})x^\mu - b^\mu \mathbf{x}^2. \quad (10.16)$$

The corresponding scale factor is determined by

$$\left| \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \right| = \frac{1}{(1 - 2\mathbf{b} \cdot \mathbf{x} + b^2 \mathbf{x}^2)^d}. \quad (10.17)$$

In particular the distance separating two points \mathbf{x}_i and \mathbf{x}_j scales like

$$|\mathbf{x}'_i - \mathbf{x}'_j| = \frac{|\mathbf{x}_i - \mathbf{x}_j|}{(1 - 2\mathbf{b} \cdot \mathbf{x}_i + b^2 \mathbf{x}_i^2)^{1/2} (1 - 2\mathbf{b} \cdot \mathbf{x}_j + b^2 \mathbf{x}_j^2)^{1/2}}. \quad (10.18)$$

One can now write down the generators of infinitesimal conformal transformations and study their commutation relations. In this way one establishes that the conformal group is isomorphic to the pseudo-orthogonal group $\text{SO}(d+1, 1)$ with $\frac{1}{2}(d+1)(d+2)$ real parameters.

10.1.2 Correlation function of quasi-primary fields

The connection between a statistical mechanics model and quantum field theory is made as usual by writing the partition function and correlation functions of the former as functional integrals in the latter:

$$Z = \int \mathcal{D}\Phi e^{-S[\Phi]} \quad (10.19)$$

$$\langle \phi_1(\mathbf{x}_1) \dots \phi_k(\mathbf{x}_k) \rangle = Z^{-1} \int \mathcal{D}\Phi \phi_1(\mathbf{x}_1) \dots \phi_k(\mathbf{x}_k) e^{-S[\Phi]} \quad (10.20)$$

Here $S[\Phi]$ is the euclidean action, Φ the collection of fields, and $\phi_i \in \Phi$. In other words, $Z^{-1}e^{-S[\Phi]}\mathcal{D}\Phi$ is the Gibbs measure in the continuum limit. Paradoxically, in many cases the hypothesis of conformal invariance may permit one to classify and precisely characterise the possible continuum theories without ever having to write down explicitly the action $S[\Phi]$.

A field $\phi(\mathbf{x})$, here supposed spinless for simplicity, is called *quasi-primary* provided it transforms covariantly under the conformal transformation (10.1):

$$\phi(\mathbf{x}) \rightarrow \phi'(\mathbf{x}') = \left| \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \right|^{-\Delta/d} \phi(\mathbf{x}). \quad (10.21)$$

The number $\Delta = \Delta_\phi$ is a property of the field and is called its *scaling dimension*. Using this, conformal invariance completely fixes [Po70] the form of the two- and three-point correlation functions, as we shall now see.

The assumption of quasi-primarity implies the following covariance condition for a general two-point function

$$\langle \phi_1(\mathbf{x}_1)\phi_2(\mathbf{x}_2) \rangle = \left| \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_1}^{\Delta_1/d} \left| \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_2}^{\Delta_2/d} \langle \phi_1(\mathbf{x}'_1)\phi_2(\mathbf{x}'_2) \rangle. \quad (10.22)$$

Rotation and translation invariance imply that

$$\langle \phi_1(\mathbf{x}_1)\phi_2(\mathbf{x}_2) \rangle = f(|\mathbf{x}_1 - \mathbf{x}_2|), \quad (10.23)$$

and covariance under a scale transformation $\mathbf{x} \rightarrow \lambda \mathbf{x}$ fixes $f(\mathbf{x}) = \lambda^{\Delta_1 + \Delta_2} f(\lambda \mathbf{x})$. Therefore

$$\langle \phi_1(\mathbf{x}_1)\phi_2(\mathbf{x}_2) \rangle = \frac{C_{12}}{|\mathbf{x}_1 - \mathbf{x}_2|^{\Delta_1 + \Delta_2}} \quad (10.24)$$

for some constant C_{12} . Inserting now this into (10.22) and using the SCT with scale factor (10.17) we obtain

$$\frac{C_{12}}{|\mathbf{x}_1 - \mathbf{x}_2|^{\Delta_1 + \Delta_2}} = \frac{C_{12}}{\gamma_1^{\Delta_1} \gamma_2^{\Delta_2}} \frac{(\gamma_1 \gamma_2)^{(\Delta_1 + \Delta_2)/2}}{|\mathbf{x}_1 - \mathbf{x}_2|^{\Delta_1 + \Delta_2}}, \quad (10.25)$$

with

$$\gamma_i = 1 - 2\mathbf{b} \cdot \mathbf{x}_i + b^2 \mathbf{x}_i^2. \quad (10.26)$$

Equating powers of γ_i in (10.25) gives $2\Delta_1 = 2\Delta_2 = \Delta_1 + \Delta_2$ with the unique solution $\Delta_1 = \Delta_2$. This means that the two-point function vanishes unless the two fields have the same scaling dimension. Moreover it is conventional to normalise the fields so that $C_{12} = 1$. In conclusion

$$\langle \phi_1(\mathbf{x}_1)\phi_2(\mathbf{x}_2) \rangle = \frac{\delta_{\Delta_1, \Delta_2}}{x_{12}^{2\Delta_1}}, \quad (10.27)$$

where we have set $x_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$.

We next discuss the case of a three-point function. Covariance under rotations, translations and dilations imply that it must be of the form²⁶

$$\langle \phi_1(\mathbf{x}_1)\phi_2(\mathbf{x}_2)\phi_3(\mathbf{x}_3) \rangle = \frac{C_{123}}{x_{12}^a x_{23}^b x_{13}^c} \quad (10.28)$$

²⁶A priori the right-hand side of (10.28) may be replaced with a sum over several terms satisfying (10.29), but see below.

with

$$a + b + c = \Delta_1 + \Delta_2 + \Delta_3. \quad (10.29)$$

Covariance under SCT implies that

$$\frac{C_{123}}{x_{12}^a x_{23}^b x_{13}^c} = \frac{C_{123}}{\gamma_1^{\Delta_1} \gamma_2^{\Delta_2} \gamma_3^{\Delta_3}} \frac{(\gamma_1 \gamma_2)^{a/2} (\gamma_2 \gamma_3)^{b/2} (\gamma_1 \gamma_3)^{c/2}}{x_{12}^a x_{23}^b x_{13}^c}, \quad (10.30)$$

so that

$$a + c = 2\Delta_1, \quad a + b = 2\Delta_2, \quad b + c = 2\Delta_3.$$

This system has the unique solution

$$\langle \phi_1(\mathbf{x}_1) \phi_2(\mathbf{x}_2) \phi_3(\mathbf{x}_3) \rangle = \frac{C_{123}}{x_{12}^{\Delta_1 + \Delta_2 - \Delta_3} x_{23}^{\Delta_2 + \Delta_3 - \Delta_1} x_{31}^{\Delta_3 + \Delta_1 - \Delta_2}}. \quad (10.31)$$

The constants C_{123} are non-trivial parameters, which will reappear below as structure constants in the operator product expansion.

The complete determination (up to C_{123}) of two- and three-point functions is a consequence of the fact that (10.18) does not allow us to construct conformal invariants of two or three points. For $N \geq 4$ points one can however construct $N(N-3)/2$ independent invariants, known as *anharmonic ratios* or *cross-ratios*. For instance, the four-point function takes the form

$$\langle \phi_1(\mathbf{x}_1) \phi_2(\mathbf{x}_2) \phi_3(\mathbf{x}_3) \phi_4(\mathbf{x}_4) \rangle = f \left(\frac{x_{12} x_{34}}{x_{13} x_{24}}, \frac{x_{12} x_{34}}{x_{23} x_{14}} \right) \prod_{i < j}^4 x_{ij}^{\Delta/3 - \Delta_i - \Delta_j} \quad (10.32)$$

with $\Delta = \sum_{i=1}^4 \Delta_i$. We stress that the function f is not fixed solely by *global* conformal invariance.

10.1.3 Stress tensor and global Ward identity

The stress tensor $T^{\mu\nu}$ is the conserved Noether current associated with the conformal symmetry. It can be defined²⁷ as the response of the partition function to a local change in the metric:

$$T^{\mu\nu}(\mathbf{x}) = -\frac{1}{(2\pi)^{d-1}} \frac{\delta \log Z}{\delta g_{\mu\nu}(\mathbf{x})} \quad (10.33)$$

²⁷Note the analogy with the theory of integrable systems, where the conserved charges are obtained as derivatives of the transfer matrix with respect to the anisotropy (spectral parameter).

The power of 2π is conventional and will lead to convenient simplifications later.

Because of (10.19) this can be written equivalently as the variation of the local action $S[\phi]$ under the transformation (10.2):

$$\delta S = \frac{1}{(2\pi)^{d-1}} \int d^d x T^{\mu\nu}(x) \partial_\mu \epsilon_\nu(x). \quad (10.34)$$

This point of view will be useful later when we consider the response of the action to transformations that are only conformal in some parts of space. But for truly (global) conformal transformations we have obviously $\delta S = 0$. This immediately entails some important symmetry properties of $T^{\mu\nu}$.

For the translational invariance (10.11), $\epsilon_\nu(x) = a_\nu$, one has of course $\partial_\mu \epsilon_\nu(x) = 0$, whence $\delta S = 0$ as expected. But performing instead an integration by parts in (10.34), and using that a_ν is arbitrary, we obtain the conservation law

$$\partial_\mu T^{\mu\nu}(x) = 0. \quad (10.35)$$

Thus $T^{\mu\nu}(x)$ is indeed equivalent to the usual Noether current.

Regarding the rotational invariance (10.13), for the integral (10.34) to vanish, the stress tensor must be symmetric:

$$T^{\mu\nu}(x) = T^{\nu\mu}(x). \quad (10.36)$$

And finally the dilatation invariance (10.12) has $\partial_\mu \epsilon_\nu(x) = \delta_\mu^\nu$, so the stress tensor is traceless:

$$T^\mu_\mu(x) = 0. \quad (10.37)$$

The stress tensor also satisfies a very important constraint known as the Ward identity. This identity is most powerful in the case of local conformal invariance in $d = 2$ (see below), but the starting point is a global identity valid in any dimension that we derive now.

Consider the correlation function of a product of local fields $\phi_i(\mathbf{x}_i)$ that we denote for simplicity as

$$X = \phi_1(\mathbf{x}_1) \phi_2(\mathbf{x}_2) \cdots \phi_n(\mathbf{x}_n). \quad (10.38)$$

The correlation function $\langle X \rangle$ is a physical observable and does not change under an infinitesimal coordinate transformation (10.2). We have thus $\delta \langle X \rangle =$

0. A non-trivial identity however results from decomposing the various changes that add up to zero. There is an explicit variation of the fields,

$$\phi_i(\mathbf{x}_i) \rightarrow \phi'_i(\mathbf{x}_i) = \phi_i(\mathbf{x}_i) + \delta\phi_i(\mathbf{x}_i) \quad (10.39)$$

and the action pick up a variation δS given by (10.34). But since the correlation function is defined by the functional integral (10.20) there will be three types of changes to $\langle X \rangle$ induced by 1) the explicit field variation $\delta\phi_i(\mathbf{x}_i)$, 2) the variation δS in the correlation functional integral, and 3) the variation δS in the normalisation Z^{-1} . Summing these up leads to

$$\begin{aligned} 0 &= \sum_{i=1}^n \langle \phi_1(\mathbf{x}_1) \cdots \delta\phi_i(\mathbf{x}_i) \cdots \phi_n(\mathbf{x}_n) \rangle \\ &- \frac{1}{(2\pi)^{d-1}} \int d^d x \langle T^{\mu\nu}(x) X \rangle \partial_\mu \epsilon_\nu(x) \\ &- \left[\frac{1}{(2\pi)^{d-1}} \int d^d x \langle T^{\mu\nu}(x) \rangle \partial_\mu \epsilon_\nu(x) \right] \langle X \rangle. \end{aligned} \quad (10.40)$$

For a theory at its critical point only the identity operator has a non-zero one-point function. In particular $\langle T^{\mu\nu}(x) \rangle = 0$. The global Ward identity therefore takes the form

$$\sum_{i=1}^n \langle \phi_1(\mathbf{x}_1) \cdots \delta\phi_i(\mathbf{x}_i) \cdots \phi_n(\mathbf{x}_n) \rangle = \frac{1}{(2\pi)^{d-1}} \int d^d x \langle T^{\mu\nu}(x) X \rangle \partial_\mu \epsilon_\nu(x) \quad (10.41)$$

10.2 Two dimensions and local conformal invariance

Conformal invariance is especially powerful in two dimensions for reasons that we shall expose presently. For the moment, we work in the geometry of the Riemann sphere, i.e., the plane with a point at infinity, and we shall write the coordinates as $\mathbf{x} = (x^1, x^2)$. Under a general coordinate transformation $x^\mu \rightarrow x'^\mu = w^\mu(x^1, x^2)$ application of (10.1) implies the Cauchy-Riemann equations

$$\begin{aligned} \frac{\partial w^2}{\partial x^1} &= \pm \frac{\partial w^1}{\partial x^2}, \\ \frac{\partial w^1}{\partial x^1} &= \mp \frac{\partial w^2}{\partial x^2}, \end{aligned} \quad (10.42)$$

i.e., $\mathbf{w}(\mathbf{x})$ is either a holomorphic or an antiholomorphic function. Important simplifications will therefore result upon introducing the complex coordinates

$$z \equiv x^1 + ix^2, \quad \bar{z} \equiv x^1 - ix^2. \quad (10.43)$$

A conformal mapping then reads simply $z \rightarrow z' = w(z)$.

It is convenient to consider $(x^1, x^2) \in \mathbb{C}^2$, so that z and \bar{z} can be considered independent complex variables, not linked by complex conjugation. For that reason we can often concentrate on the transformations of z alone, bearing in mind that \bar{z} satisfies the same properties. Ultimately the relationship between the two—i.e., \bar{z} indeed *is* the complex conjugate of z —will be enforced through the constraint of modular invariance (see section 10.7).

The identification of two-dimensional conformal transformations with analytic maps $w(z)$ could have been anticipated from the well-known fact that the latter are angle-preserving. It should be noted that an analytic map is defined (via its Laurent series) by an *infinite* number of parameters. This does not contradict the result of section 10.1 that the set of global conformal transformations is defined by only $\frac{1}{2}(d+1)(d+2) = 6$ real parameters, since analytic maps are not necessarily invertible and defined in the whole complex plane.

Global conformal transformations in $d = 2$ take the form of the projective transformations

$$w(z) = \frac{a_{11}z + a_{12}}{a_{21}z + a_{22}} \quad (10.44)$$

with $a_{ij} \in \mathbb{C}$ and a normalisation constraint that we can take as $\det a_{ij} = 1$.

It is straightforward to verify that the composition of two projective transformations is again projective, with parameters $\{a_{ij}\}$ that correspond to multiplying those of the individual transformations as 2×2 matrices. In other words, $d = 2$ global conformal transformations form the group $\text{SL}(2, \mathbb{C}) \simeq \text{SO}(3, 1)$.

We can sketch an argument why the projective transformations (10.44) are the only globally defined invertible holomorphic mappings $f(z)$. First, for f to be single-valued it cannot have branch points. Second, for f to be invertible it cannot have essential singularities. Therefore $f(z) = P(z)/Q(z)$ must be a ratio of polynomials without common zeros. For the inverse image of zero to exist, $P(z)$ can only have a single zero. This cannot be a multiple

zero, since otherwise f would not be invertible. Therefore $P(z) = a_{11}z + a_{12}$. The same argument with zero replaced by infinity implies that $Q(z) = a_{21}z + a_{22}$.

In complex coordinates, the transformation law (10.21) becomes

$$\phi'(w, \bar{w}) = \left(\frac{dw}{dz}\right)^{-h} \left(\frac{d\bar{w}}{d\bar{z}}\right)^{-\bar{h}} \phi(z, \bar{z}) \quad (10.45)$$

where the *real* parameters (h, \bar{h}) are called the *conformal weights*. The combinations $\Delta = h + \bar{h}$ and $s = h - \bar{h}$ are called respectively the scaling dimension and the spin of ϕ . A field ϕ satisfying (10.45) for any projective transformation (resp. any analytic map) $w(z)$ is called *quasi-primary* (resp. *primary*). An example of a quasi-primary field which is not primary is furnished by the stress tensor (see below).

The expressions (10.27)–(10.31) for the two- and three-point correlation functions still hold true with the obvious modification that the dependence in $z_{ij} \equiv z_i - z_j$ (resp. in \bar{z}_{ij}) goes with the conformal weights h (resp. \bar{h}).

10.3 Stress tensor and local Ward identity

The change to complex coordinates implies that the conservation laws of $T^{\mu\nu}$ need some rewriting. Directly from (10.43) the corresponding derivatives read

$$\begin{aligned} \partial_z &= \frac{1}{2}(\partial_1 - i\partial_2), \\ \partial_{\bar{z}} &= \frac{1}{2}(\partial_1 + i\partial_2) \end{aligned} \quad (10.46)$$

with inverses

$$\begin{aligned} \partial_1 &= \partial_z + \partial_{\bar{z}}, \\ \partial_2 &= i(\partial_z - \partial_{\bar{z}}). \end{aligned} \quad (10.47)$$

The elements of the complex metric can be read off from the obvious rewriting of the line element in Euclidean $d = 2$ space:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = (dx^1)^2 + (dx^2)^2 = dz d\bar{z}. \quad (10.48)$$

This leads to $g_{zz} = g_{\bar{z}\bar{z}} = 0$ and $g_{z\bar{z}} = g_{\bar{z}z} = \frac{1}{2}$. In particular the components of the stress tensor read now in complex coordinates:

$$\begin{aligned} T_{zz} &\equiv T(z, \bar{z}) = \frac{1}{4}(T_{11} - T_{22} + 2iT_{12}), \\ T_{\bar{z}\bar{z}} &\equiv \bar{T}(z, \bar{z}) = \frac{1}{4}(T_{11} - T_{22} - 2iT_{12}) \\ T_{z\bar{z}} &= T_{\bar{z}z} = \frac{1}{4}(T_{11} + T_{22}) = \frac{1}{4}T_{\mu}^{\mu}. \end{aligned} \tag{10.49}$$

We can likewise rewrite the conservation law (10.35) in complex coordinates:

$$\begin{aligned} \partial_{\bar{z}}T(z, \bar{z}) + \frac{1}{4}\partial_z T_{\mu}^{\mu} &= 0, \\ \partial_z \bar{T}(z, \bar{z}) + \frac{1}{4}\partial_{\bar{z}} T_{\mu}^{\mu} &= 0. \end{aligned} \tag{10.50}$$

Recall that scale invariance further implies the tracelessness $T_{\mu}^{\mu} = 0$ from (10.37); in general the trace would be proportional to the beta function, which vanishes at a renormalisation group fixed point. At the fixed point we thus have

$$\partial_{\bar{z}}T(z, \bar{z}) = \partial_z \bar{T}(z, \bar{z}) = 0. \tag{10.51}$$

This means that T depends only on z , hence is an holomorphic function, and that \bar{T} depends only on \bar{z} , hence is an anti-holomorphic function. This is a very important element in the solvability of two-dimensional CFT.

To emphasize this crucial result we henceforth denote the two non-vanishing components of the stress tensor $T(z)$ and $\bar{T}(\bar{z})$, viz.

$$T(z) \equiv T_{zz}, \quad \bar{T}(\bar{z}) \equiv T_{\bar{z}\bar{z}}. \tag{10.52}$$

Following Fateev and Zamolodchikov [FZ87] it is even possible to go (much) further: CFT's in which the conformal symmetry is enhanced with other, so-called extended, symmetries (superconformal, parafermionic, W algebra, ...) can be constructed by requiring more analytic currents and making them coexist with $T(z)$ by imposing certain associativity requirements.

We now come back to the Ward identity (10.41) for a product of local operators $\phi_i(z_i, \bar{z}_i)$. In $d = 2$, if we suppose that these operators are *primary* and that the infinitesimal transformation is only *locally* conformal, we will get a much stronger local form of the Ward identity.

Let C be a circle centered at the origin of radius sufficiently large so as to surround all the points (z_i, \bar{z}_i) with $i = 1, 2, \dots, n$. Denote by n^μ its outgoing normal vector. We shall suppose that the infinitesimal transformation, $z' = z + \epsilon(z)$ and $\bar{z}' = \bar{z} + \bar{\epsilon}(\bar{z})$, is conformal only inside C , whereas on the outside it is merely a differentiable function that tends to zero sufficiently fast at infinity.

Consider first the right-hand side of (10.41). We can perform an integration by parts and invoke the conservation law (10.35) to get rid of the bulk part of this integral. Only remains the boundary terms. The boundary term at infinity vanishes due to the hypothesis that $\epsilon(z)$ and $\bar{\epsilon}(\bar{z})$ tend to zero sufficiently fast at infinity. The boundary term at C can be written as

$$\frac{1}{2\pi} \int_C d\Sigma n_\mu \epsilon_\mu \langle T^{\mu\nu} X \rangle = \frac{1}{2\pi i} \oint_C dz \epsilon(z) \langle T(z) X \rangle - \frac{1}{2\pi i} \oint_C d\bar{z} \bar{\epsilon}(\bar{z}) \langle \bar{T}(\bar{z}) X \rangle, \quad (10.53)$$

where Σ denotes the surface (actually line) element of the circle C and we recall that $\epsilon(z) = \epsilon^1 + i\epsilon^2$ and $\bar{\epsilon}(\bar{z}) = \epsilon^1 - i\epsilon^2$.

Consider next the left-hand side of (10.41). The transformation law (10.45) for the primary field $\phi_i(z_i, \bar{z}_i)$ can be written as

$$\phi'_i(z'_i, \bar{z}'_i) (dz')^{h_i} (d\bar{z}')^{\bar{h}_i} = \phi_i(z, \bar{z}) (dz)^{h_i} (d\bar{z})^{\bar{h}_i}, \quad (10.54)$$

where (h_i, \bar{h}_i) are the corresponding conformal weights. Developing the infinitesimal transformation to first order this reads

$$\delta\phi_i \equiv \phi(z_i, \bar{z}_i) - \phi'(z'_i, \bar{z}'_i) = [(h_i \partial_i \epsilon + \epsilon \partial_i) + (\bar{h}_i \bar{\partial}_i \bar{\epsilon} + \bar{\epsilon} \bar{\partial}_i)] \phi_i(z_i, \bar{z}_i). \quad (10.55)$$

Assembling these ingredients, and using the independence of the analytic and antianalytic parts of the expressions, we arrive at

$$\frac{1}{2\pi i} \oint_C dz \epsilon(z) \left[\sum_{i=1}^n \left(\frac{h_i}{(z - z_i)^2} + \frac{\partial_i}{z - z_i} \right) \langle X \rangle - \langle T(z) X \rangle \right] = 0 \quad (10.56)$$

We have here used the Cauchy theorem. There is a corresponding expression with bars. Since $\epsilon(z)$ is arbitrary the integrand must in fact vanish:

$$\langle T(z) X \rangle = \sum_{i=1}^n \left(\frac{h_i}{(z - z_i)^2} + \frac{\partial_i}{z - z_i} \right) \langle X \rangle. \quad (10.57)$$

This is the desired *conformal Ward identity*. On the right-hand side we see manifestly the singularities in each of the coordinates z_i of the primary fields

$\phi_i(z_i, \bar{z}_i)$ entering the product X . These are the expected short-distance singularities whenever $T(z)$ approaches any of the primary fields:

$$T(z)\phi_j(z_j, \bar{z}_j) = \frac{h_j}{(z - z_j)^2}\phi_j(z_j, \bar{z}_j) + \frac{1}{z - z_j}\partial_{z_j}\phi(z_j, \bar{z}_j) + \mathcal{O}(1). \quad (10.58)$$

The conformal Ward identity written in this local form is our first example of an *operator product expansion* (OPE), i.e., a formal power series in the coordinate difference that expresses the effect of bringing close together two operators.

Several remarks are in order:

1. It is tacitly understood that OPE's only have a sense when placed between the brackets $\langle \dots \rangle$ of a correlation function.
2. We generically expect singularities to arise when approaching two local operators in a quantum field theory; in particular the average of a field over some small volume will have a variance that diverges when that volume is taken to zero.
3. An OPE should be considered an exact identity (valid in a finite domain of the field coordinates) rather than an approximation, provided the formal expansion is written out to arbitrarily high order. In our example, (10.57) only determines the first two terms in the OPE (10.58).
4. Contracting any field ϕ with $T(z)$ and comparing with (10.58) is actually a useful practical means of determining its primarity and its conformal dimension h_ϕ .

It is not difficult to see from (10.33) that on dimensional grounds T itself is a quasi-primary field of conformal dimension $h = 2$, since the partition function Z is dimensionless. However, the average $\langle T(z_1)T(z_2) \rangle \sim (z_1 - z_2)^{-4}$ has no reason to vanish, and so the OPE of T with itself takes the form

$$T(z_1)T(z_2) = \frac{c/2}{(z_1 - z_2)^4} + \frac{2T(z_2)}{(z_1 - z_2)^2} + \frac{\partial T(z_2)}{z_1 - z_2} + \mathcal{O}(1). \quad (10.59)$$

In particular, T is *not* primary. The constant c appearing in (10.59) is called the *central charge*. Considering two non-interacting CFT's as a whole, one has from (10.33) that their stress tensors, and hence their central charges, add up, and so c can be considered as a measure of the number of quantum degrees of liberty in the CFT.

It is straightforward (but somewhat lengthy) exercise to establish that $c = \frac{1}{2}$ for a free fermion and $c = 1$ for a free boson. Details can be found in section 5.3 of [DMS87].

As T is not primary, it cannot transform like (10.45) under a finite conformal transformation $z \rightarrow w(z)$. We can always write the modified transformation law as

$$T'(w) = \left(\frac{dw}{dz}\right)^{-2} \left[T(z) - \frac{c}{12}\{w; z\}\right]. \quad (10.60)$$

To determine what $\{w; z\}$ represents, we use the constraint due to two successive applications of (10.60) and the fact that $\{w; z\} = 0$ for projective conformal transformations, since T is quasi-primary. The result is that $\{w; z\}$ is the Schwarzian derivative

$$\{w; z\} = \frac{d^3w/dz^3}{dw/dz} - \frac{3}{2} \left(\frac{d^2w/dz^2}{dw/dz}\right)^2. \quad (10.61)$$

10.4 Finite-size scaling on a cylinder

The central charge c is ubiquitous in situations where the CFT is placed in a finite geometry, i.e., interacts with some boundary condition. An important example is furnished by conformally mapping the plane to a cylinder of circumference L by means of the transformation

$$w(z) = \frac{L}{2\pi} \log z. \quad (10.62)$$

This transformation can be visualised by viewing the cylinder in perspective, with one rim contracting to the origin and the other expanding to form the point at infinity. Taking the expectation value of (10.60), and using the fact that $\langle T(z) \rangle = 0$ in the plane on symmetry grounds, one finds that $\langle T(w) \rangle = -\pi^2 c/6L^2$ on the cylinder. Applying (10.33) then implies that the free energy per unit area $f_0(L)$ satisfies [BCN86, Af86]

$$f_0(L) = f_0(\infty) - \frac{\pi c}{6L^2} + o(L^{-2}). \quad (10.63)$$

This is a very useful result for obtaining c for a concrete statistical model, since $f(L)$ can usually be determined from the corresponding transfer matrix,

either numerically for small L by using exact diagonalisation techniques, or analytically in the Bethe Ansatz context by using the Euler-Maclaurin formula.

One may note the clear analogy between (10.63) and the Casimir effect between two uncharged metallic plates. According to quantum electrodynamics, the vanishing of the wave function on the plates induces a force between them. This force may be attracting or repelling depending on the specific arrangement of the plates.

It is also of interest to study such finite-size effects on the level of the two-point correlation function of a primary field ϕ . Again using the mapping (10.62), the covariance property (10.45) and the form (10.27) of the correlator in the plane can be used to deduce its form on the cylinder. Assuming for simplicity $h = \bar{h} = \Delta/2$, and writing the coordinates on the cylinder as $w = t + ix$, with $t \in \mathbb{R}$ and $x \in [0, L)$, one arrives at

$$\langle \phi(t_1, x_1) \phi(t_2, x_2) \rangle = \left(\frac{2\pi}{L} \right)^{2\Delta} \left[2 \cosh \left(\frac{2\pi t_{12}}{L} \right) - 2 \cos \left(\frac{2\pi x_{12}}{L} \right) \right]^{-\Delta}, \quad (10.64)$$

where $t_{12} = t_1 - t_2$ and $x_{12} = x_1 - x_2$. In the limit of a large separation of the fields, $t_{12} \rightarrow \infty$, this decays like $e^{-t_{12}/\xi}$ with correlation length $\xi = L/2\pi\Delta$. But this decay can also be written $(\Lambda_\phi/\Lambda_0)^{-t_{12}}$, where Λ_0 is the largest eigenvalue of the transfer matrix, and Λ_ϕ is the largest eigenvalue compatible with the constraint that an operator ϕ has been inserted at each extremity $t = \pm\infty$ of the cylinder. Denoting the corresponding free energies per unit area $f(L) = -L^{-1} \log \Lambda$, we conclude that [Ca84a]

$$f_\phi(L) - f_0(L) = \frac{2\pi\Delta}{L^2} + o(L^{-2}). \quad (10.65)$$

This is as useful as (10.63) in (numerical or analytical) transfer matrix studies, since the constraint imposed by ϕ can usually be related explicitly to properties of the transfer matrix spectrum.

10.5 Virasoro algebra and its representation theory

Up to this point, we have worked in a setup where the fields were seen as functionals of the complex coordinates z, \bar{z} . To obtain an operator formalism, one must impose a quantisation scheme, i.e., single out a time and a

space direction. In this formalism a crucial role will be played by the mode operators of the stress tensor, defined by

$$L_n = \frac{1}{2\pi i} \oint_C z^{n+1} T(z) dz, \quad \bar{L}_n = \frac{1}{2\pi i} \oint_C \bar{z}^{n+1} \bar{T}(\bar{z}) d\bar{z}. \quad (10.66)$$

The transfer matrix then propagates the system from one time slice to the following and is written as the exponential of the Hamiltonian \mathcal{H} , i.e., the energy operator on a fixed-time surface. In the continuum limit, one may freely choose the time direction. In CFT this is most conveniently done by giving full honours to the scale invariance of the theory, viz., by using for \mathcal{H} the dilation operator

$$\mathcal{D} = \frac{1}{2\pi i} \oint_C z T(z) dz + \frac{1}{2\pi i} \oint_C \bar{z} \bar{T}(\bar{z}) d\bar{z} = L_0 + \bar{L}_0, \quad (10.67)$$

where C is a counterclockwise contour enclosing the origin. The following choice of additive and multiplicative normalisations defines \mathcal{H} precisely:

$$\mathcal{H} = (2\pi/L)(L_0 + \bar{L}_0 - c/12). \quad (10.68)$$

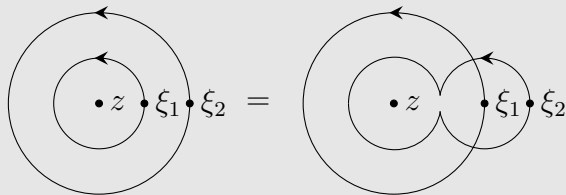
This is called the *radial quantisation* scheme: the constant-time surfaces are concentric circles around the origin. Under the map (10.62) the time becomes simply the coordinate along the cylinder axis. The usual time ordering of operators then becomes a prescription of radial ordering.

Using the radial ordering, the OPE (10.59) can be turned into a commutation relation $[L_n, L_m]$.

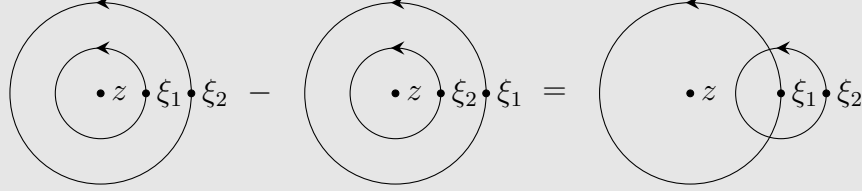
Consider first the action of $L_n L_m$ on an operator $\Phi(z)$. By (10.66) we have

$$L_n L_m \Phi(z) = \frac{1}{(2\pi i)^2} \oint_{C_{z, \xi_1}} d\xi_2 \oint_{C_z} d\xi_1 (\xi_2 - z)^{n+1} (\xi_1 - z)^{m+1} T(\xi_2) T(\xi_1) \Phi(z),$$

where the first integration contour C_z encircles only z , whereas the second C_{z, ξ_1} encircles both z and ξ_1 . We can deform the latter contour as follows:



so that



This implies that

$$[L_n, L_m]\Phi(z) = \frac{1}{(2\pi i)^2} \oint_{C_z} d\xi_1 (\xi_1 - z)^{m+1} \oint_{C_{\xi_1}} d\xi_2 (\xi_2 - z)^{n+1} T(\xi_2) T(\xi_1) \Phi(z).$$

But in the innermost integral—the one over C_{ξ_1} —the points ξ_1 and ξ_2 can be taken arbitrarily close, so that it is appropriate to replace $T(\xi_2)T(\xi_1)$ by the OPE (10.59). The innermost integral therefore reads

$$\begin{aligned} & \frac{1}{2\pi i} \oint_{C_{\xi_1}} d\xi_2 (\xi_2 - z)^{n+1} \left[\frac{c/2}{(\xi_2 - \xi_1)^4} + \frac{2T(\xi_1)}{(\xi_2 - \xi_1)^2} + \frac{\partial T(\xi_1)}{\xi_2 - \xi_1} + \mathcal{O}(1) \right] \\ &= \frac{c}{2} \frac{n+1}{3} \frac{n}{2} \frac{n-1}{1} (\xi_1 - z)^{n-2} + 2(n+1)(\xi_1 - z)^n T(\xi_1) + (\xi_1 - z)^{n+1} \partial T(\xi_1), \end{aligned}$$

where we have used the Cauchy theorem on each of the three singular terms. Performing now the outermost integral—the one over C_z —gives us back the mode operators (10.66):

$$[L_n, L_m] = \frac{c}{12} n(n^2 - 1) \delta_{n+m,0} + 2(n+1) L_{m+n} + \frac{1}{2\pi i} \oint_{C_z} d\xi_1 (\xi_1 - z)^{n+m+2} \partial T(\xi_1).$$

In this expression, the remaining integral can be found by partial integration:

$$-\frac{1}{2\pi i} \oint_{C_z} d\xi_1 (n+m+2) (\xi_1 - z)^{n+m+1} T(\xi_1) = -(n+m+2) L_{m+n}.$$

Inserting this gives us the final form of the commutation relations:

$$[L_n, L_m] = (n-m) L_{n+m} + \frac{c}{12} n(n^2 - 1) \delta_{n+m,0}. \quad (10.69)$$

A similar expression holds for $[\bar{L}_n, \bar{L}_m]$, whereas $[L_n, \bar{L}_m] = 0$. The algebra defined by (10.69) is called the *Virasoro algebra*. Importantly, the decoupling

into two isomorphic Virasoro algebras, one for L_n and another for \bar{L}_n , means that in the geometry chosen we can focus exclusively on L_n . It should be stressed that in the geometry of a torus, the two algebras couple non-trivially, in a way that is revealed by imposing modular invariance (see section 10.7 below).

We now describe the structure of the Hilbert space in radial quantisation. The vacuum state $|0\rangle$ must be invariant under projective transformations, whence $L_{\pm 1}|0\rangle = 0$, and we fix the ground state energy by $L_0|0\rangle = 0$. Non-trivial eigenstates of \mathcal{H} are created by action with a primary field, $|h, \bar{h}\rangle = \phi(0, 0)|0\rangle$. Translating (10.58) into operator language implies then in particular $L_0|h, \bar{h}\rangle = h|h, \bar{h}\rangle$. We must also impose the highest-weight condition $L_n|h, \bar{h}\rangle = \bar{L}_n|h, \bar{h}\rangle = 0$ for $n > 0$. Excited states with respect to the primary ϕ then read

$$\phi^{\{n, \bar{n}\}} \equiv L_{-n_1} L_{-n_2} \cdots L_{-n_k} \bar{L}_{-\bar{n}_1} \bar{L}_{-\bar{n}_2} \cdots \bar{L}_{-\bar{n}_k} |h, \bar{h}\rangle \quad (10.70)$$

with $1 \leq n_1 \leq n_2 \leq \cdots \leq n_k$ and similarly for $\{\bar{n}\}$. These states are called the *descendants* of ϕ at level $\{N, \bar{N}\}$, where $N = \sum_{i=1}^k n_i$. A primary state and its descendants form a highest weight representation (or Verma module) of the Virasoro algebra.

Correlation functions of descendent fields can be obtained by acting with appropriate differential operators on the correlation functions of the corresponding primary fields. To see this, consider first for $n \geq 1$ the descendent $(L_{-n}\phi)(w)$ of the primary field $\phi(w)$, and let $X = \prod_j \phi_j(w_j)$ be an arbitrary product of other primaries as in the conformal Ward identity (10.57). Using (10.66) and (10.58) we have then

$$\begin{aligned} \langle (L_{-n}\phi)(w)X \rangle &= \frac{1}{2\pi i} \oint_{C_z} dz (z-w)^{1-n} \langle T(z)\phi(w)X \rangle \\ &= -\frac{1}{2\pi i} \oint_{C_{\{w_j\}}} dz (z-w)^{1-n} \times \\ &\quad \sum_j \left\{ \frac{\partial_{w_j}}{z-w_j} + \frac{h_j}{(z-w_j)^2} \right\} \langle \phi(w)X \rangle, \end{aligned} \quad (10.71)$$

where the minus sign comes from turning the integration contour inside out, so that it surrounds all the points $\{w_j\}$. In other words, a descendent in a correlation function may be replaced by the corresponding primary

$$\langle (L_{-n}\phi)(w)X \rangle = \mathcal{L}_{-n} \langle \phi(w)X \rangle \quad (10.72)$$

provided that we act instead on the correlator with the linear differential operator

$$\mathcal{L}_{-n} \equiv \sum_j \left\{ \frac{(n-1)h_j}{(w_j-w)^n} - \frac{\partial_{w_j}}{(w_j-w)^{n-1}} \right\} \quad (10.73)$$

It is readily seen that a general descendent (10.70) is similarly dealt with by replacing each factor L_{-n_i} by the corresponding factor of \mathcal{L}_{-n_i} in (10.72).

We can now write the general form of the OPE of two primary fields ϕ_1 and ϕ_2 . It reads

$$\phi_1(z, \bar{z})\phi_2(0, 0) = \sum_p C_{12p} \sum_{\{n, \bar{n}\} \cup \{\emptyset, \emptyset\}} C_{12p}^{\{n, \bar{n}\}} z^{h_p-h_1-h_2+N} \bar{z}^{\bar{h}_p-\bar{h}_1-\bar{h}_2+\bar{N}} \phi_p^{\{n, \bar{n}\}}(0, 0), \quad (10.74)$$

where the summation is over a certain set of primaries $\phi_p \equiv \phi_p^{\{\emptyset, \emptyset\}}$ as well as their descendents. The coefficients $C_{12p}^{\{n, \bar{n}\}}$ (we have set $C_{12p}^{\{\emptyset, \emptyset\}} = 1$) can be determined by acting with all combinations of positive-index mode operators on both sides of (10.74) and solving the resulting set of linear equations.

In view of (10.69) it actually suffices to act with L_1 and L_2 . Determining the $C_{12p}^{\{n, \bar{n}\}}$ is then a nice exercise of contour integration. (The answer can be found in Appendix B of [BPZ84].)

In contradistinction, the coefficients C_{12p} are fundamental quantities. Contracting both sides of (10.74) with ϕ_p and using the orthogonality of two-point functions (10.27) we see that the coefficients C_{12p} coincide with those appearing in the three-point functions (10.31).

The C_{12p} can be computed by the so-called conformal bootstrap method, i.e., by assuming crossing symmetry of the four-point functions. In concrete terms, this amounts to writing a well-chosen four-point function, mapping three of its points to 0, 1, ∞ by means of a projective transformation (10.44), and comparing all possible limits of the remaining point z . When computing those limits, one successively uses (10.74).

10.6 Minimal models

Denote by $\mathcal{V}(c, h)$ the highest weight representation (Verma module) generated by the mode operators $\{L_n\}$ acting on a highest weight state $|h\rangle$ in a

CFT of central charge c . The Hilbert space of the CFT can then be written

$$\bigoplus_{h, \bar{h}} n_{h, \bar{h}} \mathcal{V}(c, h) \otimes \mathcal{V}(c, \bar{h}), \quad (10.75)$$

where the multiplicities $n_{h, \bar{h}}$ indicate the number of distinct primaries of conformal weights (h, \bar{h}) that are present in the theory. A *minimal model* is a CFT for which the sum in (10.75) is finite.

The Hermitian conjugate of a mode operator is defined by $L_n^\dagger = L_{-n}$; this induces an inner product on the Verma module. The *character* $\chi_{(c, h)}$ of the module $\mathcal{V}(c, h)$ can then be defined as

$$\chi_{(c, h)}(\tau) = \text{Tr} q^{L_0 - c/24}, \quad (10.76)$$

where $\tau \in \mathbb{C}$ is the so-called *modular parameter* (see section 10.7 below) and $q = e^{2\pi i \tau}$. Since the number of descendants of $|h\rangle$ at level N is just the number $p(N)$ of integer partitions of N , cf. (10.70), we have simply

$$\chi_{(c, h)}(\tau) = \frac{q^{h-c/24}}{P(q)}, \quad (10.77)$$

where

$$\frac{1}{P(q)} \equiv \prod_{n=1}^{\infty} \frac{1}{1 - q^n} = \sum_{n=0}^{\infty} p(n) q^n \quad (10.78)$$

is the generating function of partition numbers; this is also often expressed in terms of the Dedekind function

$$\eta(\tau) = q^{1/24} P(q). \quad (10.79)$$

However, the generic Verma module is not necessarily irreducible, so further work is needed.

For certain values of h , it may happen that a specific linear combination $|\chi\rangle$ of the descendants of $|h\rangle$ at level N is itself primary, i.e., $L_n |\chi\rangle = 0$ for $n > 0$. In other words, $|\chi\rangle$ is primary and descendent at the same time, and it generates its own Verma module $\mathcal{V}_\chi(c, h) \subset \mathcal{V}(c, h)$.

The states in $\mathcal{V}_\chi(c, h)$ are orthogonal to those in $\mathcal{V}(c, h)$,

$$\langle \chi | L_{-n_1} L_{-n_2} \cdots L_{-n_k} | h \rangle = \langle h | L_{n_k} \cdots L_{n_2} L_{n_1} | \chi \rangle^* = 0, \quad (10.80)$$

and so in particular they have zero norm. A Verma module $\mathcal{V}(c, h)$ containing one or more such *null fields* $|\chi\rangle$ is called reducible, and can be turned into

an irreducible Verma module $\mathcal{M}(c, h)$ by quotienting out the null fields, i.e., by setting $|\chi\rangle = 0$. The Hilbert space is then given by (10.75) with \mathcal{V} replaced by \mathcal{M} ; since it contains fewer states the corresponding characters (10.76) are *not* given by the simple result (10.77).

The concept of null states is instrumental in constructing *unitary* representations of the Virasoro algebra (10.69), i.e., representations in which no state of *negative* norm occurs. An important first step is the calculation of determinant of the Gram matrix of inner products between descendants at level N . This is known as the Kac determinant $\det M^{(N)}$. Its roots can be expressed through the following parameterisation:

$$\begin{aligned} c(m) &= 1 - \frac{6}{m(m+1)} \\ h(m) &= h_{r,s}(m) \equiv \frac{[(m+1)r - ms]^2 - 1}{4m(m+1)} \end{aligned} \quad (10.81)$$

where $r, s \geq 1$ are integers with $rs \leq N$. The condition for unitarity of models with $c < 1$, first found by Friedan, Qiu and Shenker [FQS84] reads: $m, r, s \in \mathbb{Z}$ with $m \geq 2$, and (r, s) must satisfy $1 \leq r < m$ and $1 \leq s \leq m$.

To get an idea of the origin of (10.81) it is instructive to compute the Kac determinant at the first few levels. For instance, at level $N = 1$ the only state is $L_{-1}|h\rangle$, while at level $N = 2$ there are two states: $L_{-1}^2|h\rangle$ and $L_{-2}|h\rangle$. The Kac determinants read

$$\begin{aligned} \det M^{(1)} &= 2h, \\ \det M^{(2)} &= 32(h - h_{1,1})(h - h_{1,2})(h - h_{2,1}). \end{aligned} \quad (10.82)$$

The general result is

$$\det M^{(N)} = \alpha_N \prod_{\substack{rs \leq N \\ r, s \geq 1}} [h - h_{r,s}(c)]^{p(N-rs)}, \quad (10.83)$$

where $p(n)$ was defined in (10.78) and $\alpha_N > 0$ is independent of h and c .

According to (10.72) the presence of a descendent field in a correlation function can be replaced by the action of a differential operator (10.73). Now

let

$$\chi(w) = \sum_{Y, |Y|=N} \alpha_Y L_{-Y} \phi(w) \quad (10.84)$$

be an arbitrary null state. Here, α_Y are some coefficients, and we have introduced the abbreviations

$$\begin{aligned} Y &= \{r_1, r_2, \dots, r_k\}, \\ |Y| &= r_1 + r_2 + \dots + r_k, \\ L_{-Y} &= L_{-r_1} L_{-r_2} \cdots L_{-r_k} \end{aligned} \quad (10.85)$$

with $1 \leq r_1 \leq r_2 \leq \dots \leq r_k$. A correlation function involving χ must vanish (since we have in fact set $\chi = 0$), and so

$$\langle \chi(w) X \rangle = \sum_{Y, |Y|=N} \alpha_Y \mathcal{L}_{-Y}(w) \langle \phi(w) X \rangle = 0. \quad (10.86)$$

Solving this N th order linear differential equation is a very useful practical means of computing the four-point correlation functions of a given CFT, provided that the level of degeneracy N is not too large. Indeed, since the coordinate dependence is through a single anharmonic ratio η , one has simply an ordinary linear differential equation.

Moreover, requiring consistency with (10.74) places restrictions on the primaries that can occur on the right-hand side of the OPE. One can then study the conditions under which this so-called *fusion algebra* closes over a finite number of primaries. The end result is that the minimal models are given by

$$\begin{aligned} c &= 1 - \frac{6(m - m')^2}{mm'} \\ h_{r,s} &= \frac{(mr - m's)^2 - (m - m')^2}{4mm'} \end{aligned} \quad (10.87)$$

with $m, m', r, s \in \mathbb{Z}$, and the allowed values of (r, s) are restricted by $1 \leq r < m'$ and $1 \leq s < m$. The corresponding $h_{r,s}$ are referred to as the *Kac table* of conformal weights. The corresponding fusion algebra reads (for clarity we omit scaling factors, structure constants, and descendents):

$$\phi_{(r_1, s_1)} \phi_{(r_2, s_2)} = \sum_{r, s} \phi_{(r, s)}, \quad (10.88)$$

where r runs from $1 + |r_1 - r_2|$ to $\min(r_1 + r_2 - 1, 2m' - 1 - r_1 - r_2)$ in steps of 2, and s runs from $1 + |s_1 - s_2|$ to $\min(s_1 + s_2 - 1, 2m - 1 - s_1 - s_2)$ in steps of 2.

The Kac table (10.87) is the starting point for elucidating the structure of the reducible Verma modules $\mathcal{V}_{r,s}$ for minimal models, and for constructing the proper irreducible modules $\mathcal{M}_{r,s}$. The fundamental observation is that

$$h_{r,s} + rs = h_{r,-s}. \quad (10.89)$$

This equation holds true for any value of c . It means that for $r, s \in \mathbb{Z}$ the Verma module $\mathcal{V}_{r,s}$ contains a singular vector at level rs that generates the submodule $\mathcal{V}_{r,-s}$. Quotienting out this submodule, we get an irreducible representation with character

$$K_{r,s}(\tau) = \frac{q^{-c/24}}{P(q)} (q^{h_{r,s}} - q^{h_{r,-s}}). \quad (10.90)$$

For $r, s \in \mathbb{Z}$ this replaces the generic character $\chi_{c,h}(\tau)$ defined in (10.77).

The case of minimal models is however different. Using the symmetry property $h_{r,s} = h_{m'-r,m-s}$ and the periodicity property $h_{r,s} = h_{r+m',s+m}$ it is seen that $h_{r,s} + rs = h_{m'+r,m-s}$ and that $h_{r,s} + (m' - r)(m - s) = h_{r,2m-s}$. This means that $\mathcal{V}_{r,s}$ contains *two* submodules, $\mathcal{V}_{m'+r,m-s}$ and $\mathcal{V}_{r,2m-s}$, at levels rs and $(m' - r)(m - s)$ respectively, and these must correspond to null vectors. To construct the irreducible module $\mathcal{M}_{r,s}$ one might at first think that it suffices to quotient out these two submodules. However, iterating the above observations, the two submodules are seen to share two sub-submodules, and so on. So $\mathcal{M}_{r,s}$ is constructed from $\mathcal{V}_{r,s}$ by an infinite series of inclusions-exclusions of pairs of submodules. This allows us in particular to compute the irreducible characters of minimal models as

$$\chi_{(r,s)}(\tau) = K_{r,s}^{(m,m')}(q) - K_{r,-s}^{(m,m')}(q), \quad (10.91)$$

where the infinite addition-subtraction scheme has been tucked away in the functions

$$K_{r,s}^{(m,m')}(q) = \frac{q^{-1/24}}{P(q)} \sum_{n \in \mathbb{Z}} q^{(2mm'n + mr - m's)^2 / 4mm'}. \quad (10.92)$$

This should be compared with the generic character (10.77) and with (10.90). Note also the similarity between (10.89) and (10.91) on the level of the indices.

It is truly remarkable that the above classification of minimal models has been achieved without ever writing down the action S appearing in (10.19). In fact, an effective Landau-Ginzburg Lagrangian description for the unitary minimal models ($m' = m+1$) has been suggested *a posteriori* by Zamolodchikov [Za86]. It suggests that the minimal models can be interpreted physically as an infinite series of multicritical versions of the Ising model. Indeed, the Ising model can be identified with the first non-trivial member in the series, $m = 3$, and the following, $m = 4$, with the tricritical Ising model.

To finish this section, we comment on the relation with self-avoiding walks and polygons. In section 11 we shall see that these (to be precise, the dilute $O(n \rightarrow 0)$ model) can be identified with the minimal model $m = 2$, $m' = 3$. Note that this is *not* a unitary theory. The central charge is $c = 0$, and the only field in the Kac table—modulo the symmetry property given after (10.89)—is the identity operator with conformal weight $h_{1,1} = 0$. Seemingly we have learnt nothing more than the trivial statement $Z = 1$. However, the operators of interest are of a *non-local* nature, and it is a pleasant surprise to find that their dimensions fit perfectly well into the Kac formula, although they are situated *outside* the “allowed” range of (r, s) values, and sometimes require the indices r, s to be half-integer. So the Kac formula, and the surrounding theoretical framework, is still a most useful tool for investigating these types of models.

10.7 Modular invariance

In section 10.3 we have seen that conformal symmetry makes the stress tensor decouple into its holomorphic and antiholomorphic components, $T(z)$ and $\bar{T}(\bar{z})$, implying in particular that the corresponding mode operators, L_n and \bar{L}_n , form two non-interacting Virasoro algebras (10.69). As a consequence, the key results of section 10.6 could be derived by considering only the holomorphic sector of the CFT. There are however constraints on the ways in which the two sectors may ultimately couple, the diagonal coupling (10.75) being just the simplest example in the context of minimal models. As first pointed out by Cardy [Ca86], a powerful tool for examining which couplings are allowed—and for placing constraints on the operator content and the conformal weights—is obtained by defining the CFT on a torus and imposing the constraint of *modular invariance*.

In this section we expose the principles of modular invariance and apply them to a CFT known as the *compactified boson*, which is going to play a central role in the Coulomb gas approach of section 11. Many other applications, including a detailed study of the minimal models, can be found in Ref. [DMS87].

Let $\omega_1, \omega_2 \in \mathbb{C} \setminus \{0\}$ such that $\tau \equiv \omega_2/\omega_1 \notin \mathbb{R}$. A torus is then defined as $\mathbb{C}/(\omega_1\mathbb{Z} + \omega_2\mathbb{Z})$, i.e., by identifying points in the complex plane that differ by an element in the lattice spanned by ω_1, ω_2 . The numbers ω_1, ω_2 are called the *periods* of the lattice, and τ the *modular parameter*. Without loss of generality we can assume $\omega_1 \in \mathbb{R}$ and $\Im\tau > 0$.

Instead of using the radial quantisation scheme of section 10.5 we now define the time (resp. space) direction to be the imaginary (resp. real) axis in \mathbb{C} . The partition function on the torus may then be written $Z(\tau) = \text{Tr} \exp[-(\Im\omega_2)\mathcal{H} - (\Re\omega_2)\mathcal{P}]$, where $\mathcal{H} = (2\pi/\omega_1)(L_0 + \bar{L}_0 - c/12)$ is the Hamiltonian and $\mathcal{P} = (2\pi/i\omega_1)(L_0 - \bar{L}_0 - c/12)$ the momentum operator. This gives

$$Z(\tau) = \text{Tr} \left(q^{L_0 - c/24} \bar{q}^{\bar{L}_0 - c/24} \right), \quad (10.93)$$

where we have defined $q = \exp(2\pi i\tau)$. Comparing with (10.75)–(10.76) we have also

$$Z(\tau) = \sum_{h, \bar{h}} n_{h, \bar{h}} \chi_{(c, h)}(\tau) \bar{\chi}_{(c, \bar{h})}(\tau). \quad (10.94)$$

An explicit computation of $Z(\tau)$ will therefore give information on the coupling $n_{h, \bar{h}}$ between the holomorphic and antiholomorphic sectors. In many cases, but not all, the coupling turns out to be simply diagonal, $n_{h, \bar{h}} = \delta_{h, \bar{h}}$.

The fundamental remark is now that $Z(\tau)$ is invariant upon making a different choice ω'_1, ω'_2 of the periods, inasmuch as they span the same lattice as ω_1, ω_2 . Any two set of equivalent periods must therefore be related by $\omega'_i = \sum_j a_{ij} \omega_j$, where $\{a_{ij}\} \in \text{Mat}(2, \mathbb{Z})$ with $\det a_{ij} = 1$. Moreover, an overall sign change, $a_{ij} \rightarrow -a_{ij}$ is immaterial, so the relevant symmetry group is the so-called *modular group* $\text{SL}(2, \mathbb{Z})/\mathbb{Z}_2 \simeq \text{PSL}(2, \mathbb{Z})$.

The remainder of this section is concerned with the construction of modular invariant partition functions for certain bosonic systems on the torus. As a warmup we consider the free boson, defined by the action

$$S[\phi] = \frac{g}{2} \int d^2\mathbf{x} (\nabla\phi)^2 \quad (10.95)$$

and $\phi(\mathbf{x}) \in \mathbb{R}$. Comparing (10.93) with (10.76)–(10.78), and bearing in mind that $c = 1$, we would expect the corresponding partition function to be of the form $Z_0(\tau) \propto 1/|\eta(\tau)|^2$. Fixing the proportionality constant is somewhat tricky [IZ86]. In a first step, ϕ is decomposed on the normalised eigenfunctions of the Laplacian, and $Z_0(\tau)$ is expressed as a product over the eigenvalues. This product however diverges, due to the presence of a zero-mode, and must be regularised. A sensible result is obtained by a shrewd analytic continuation, the so-called ζ -function regularisation technique [IZ86]:

$$Z_0(\tau) = \frac{\sqrt{4\pi g}}{\sqrt{\Im\tau} |\eta(\tau)|^2}. \quad (10.96)$$

The CFT which is of main interest for the CG technique is the so-called *compactified boson* in which $\phi(\mathbf{x}) \in \mathbb{R}/(2\pi aR\mathbb{Z})$. In other words, the field lives on a circle of radius aR (the reason for the appearance of *two* parameters, a and R , will become clear shortly). In this context, suitable periodic boundary conditions are specified by a pair of numbers, $m, m' \in a\mathbb{Z}$, so that for any $k, k' \in \mathbb{Z}$

$$\phi(z + k\omega_1 + k'\omega_2) = \phi(z) + 2\pi R(km + k'm'). \quad (10.97)$$

It is convenient to decompose $\phi = \phi_{m,m'} + \phi_0$, where

$$\phi_{m,m'} = \frac{2\pi R}{\bar{\tau} - \tau} \left[\frac{z}{\omega_1} (m\bar{\tau} - m') - \frac{\bar{z}}{\bar{\omega}_1} (m\tau - m') \right] \quad (10.98)$$

is the classical solution satisfying the topological constraint, and ϕ_0 represents the quantum fluctuations, i.e., is a standard free boson satisfying standard periodic boundary conditions.

Integrating over ϕ_0 as before, and keeping m, m' fixed, gives the partition function

$$Z_{m,m'}(\tau) = Z_0(\tau) \exp\left(-2\pi^2 g R^2 \frac{|m\tau - m'|^2}{\Im\tau}\right). \quad (10.99)$$

It is easy to see that this is not modular invariant. A modular invariant is however obtained by summing over all possible values of m, m' :

$$Z(\tau) \equiv \frac{R}{\sqrt{2}} Z_0(\tau) \sum_{m,m' \in a\mathbb{Z}} \exp\left(-2\pi^2 g R^2 \frac{|m\tau - m'|^2}{\Im\tau}\right) \quad (10.100)$$

The prefactor $R/\sqrt{2}$ is again a subtle effect of the zero-mode integration. It is actually most easily justified *a posteriori* by requiring the correct normalisation of the identity operator in (10.101) below.

A more useful, and more physically revealing, form of (10.100) is obtained by using the Poisson resummation formula to replace the sum over $m' \in a\mathbb{Z}$ by a sum over the dual variable $e \in \mathbb{Z}/a$. The result is

$$Z(\tau) = \frac{1}{|\eta(\tau)|^2} \sum_{e \in \mathbb{Z}/a, m \in a\mathbb{Z}} q^{h_{e,m}} \bar{q}^{\bar{h}_{e,m}}, \quad (10.101)$$

with

$$h_{e,m} = \frac{1}{2} \left(\frac{e}{R\sqrt{4\pi g}} + \frac{mR}{2} \sqrt{4\pi g} \right)^2, \quad \bar{h}_{e,m} = \frac{1}{2} \left(\frac{e}{R\sqrt{4\pi g}} - \frac{mR}{2} \sqrt{4\pi g} \right)^2. \quad (10.102)$$

Comparing now with (10.93) and (10.76)–(10.78) we see that (10.102) is nothing else than the conformal weights of the CFT at hand.

The requirement of modular invariance has therefore completely specified the operator content of the compactified boson system. An operator is characterised by two numbers, $e \in \mathbb{Z}/a$ and $m \in a\mathbb{Z}$, living on mutually dual lattices. A physical interpretation will be furnished by the CG formalism of section 11: e is the “electric” charge of a vertex operator (spin wave), and m is the “magnetic” charge of a topological defect (screw dislocation in the field ϕ). Let us write for later reference the corresponding scaling dimension and spin:

$$\Delta_{e,m} = \frac{e^2}{4\pi g R^2} + m^2 \pi g R^2, \quad s_{e,m} = em \quad (10.103)$$

Observe in particular that the spin is integer, as expected for a bosonic system.

The reader will notice that the three constants R , a and g are related by the fact that they always appear in the dimensionless combination $R^2 a^2 g$. Field-theoretic literature often makes the choice $a = 1$ and $g = 1/4\pi$ in order to simplify formulae such as (10.102). In the CG approach—the subject of section 11—one starts from a geometrical construction (mapping to a height model) in which a convention for a must be chosen. The compactification radius aR then follows from a “geometrical” computation (identification of the ideal state lattice), and the correct coupling constant g is only fixed in the

end by a field-theoretic argument (marginality requirement of the Liouville potential). Needless to say, the results, such as (10.103) for the dimensions of physical operators, need (and will) be independent of the initial choice made for a .

To conclude, note that the roles of e and m in (10.102) are interchanged under the transformation $Ra\sqrt{2\pi g} \rightarrow (Ra\sqrt{2\pi g})^{-1}$, which leaves (10.101) invariant. This is another manifestation of the electro-magnetic duality. Ultimately, the distinction between e and m comes down to the choice of transfer direction. In the geometry of the torus this choice is immaterial, of course. In sections 11.3–12.3 we shall compare the geometries of the cylinder and the annulus; these are related by interchanging the space and time directions, and accordingly the electric and magnetic charges switch role when going from one to the other.

11 Coulomb gas construction

It has been known since the 1970's [LP75, Ka78, KB79, Kn81] that the critical point of many two-dimensional models of statistical physics can be identified with a Gaussian free-field theory. A general framework for the computation of critical exponents was first given in 1977 by José et al. in the so-called spin wave picture [Jo77]. This was further elaborated in the early 1980's by den Nijs [Ni83, Ni84] and Nienhuis [Nh84] into what has become known as the Coulomb gas (CG) construction. These developments have been reviewed by Nienhuis [Nh87].

The CG approach is particularly suited to deal with the continuum limit of lattice models of closed self-avoiding loops, in which each loop carries a Boltzmann weight n . There are two prototype models which can be represented in terms of such loops. The Potts model has already been discussed at length in chapter 8. Another useful example is the $O(n)$ model, which can be reformulated elegantly as a loop model on the hexagonal lattice [Nh82].

The marriage between the CG and conformal field theory (CFT) happened in 1986–87, when Di Francesco, Saleur and Zuber [DSZ87a, DSZ87b] made the loop model \leftrightarrow CG correspondence more precise and showed how the ideas of modular invariance [Ca86, IZ86] can be put to good use in the study of loop models. At the same time, Duplantier and Saleur developed a range of applications to self-avoiding walks and polygons (see in particular [DS87]).

Any model of oriented self-avoiding loops is equivalent to a height model on the dual lattice. It is the continuum limit of this height which acts as the conformally invariant free field. The underlying lattice model implies that this height field is compactified, thus making contact with the modular invariance results of section 10.7. The naive free field action however needs to be modified with extra terms, traditionally known as background and screening electric charges [Nh87]. The resulting CFT, known as a Liouville field theory, is written down in section 11.2.

The requirement that the Liouville potential be RG marginal determines the coupling constant of the free field as a function of n , as first pointed out by Kondev [Ko97]. This is an important ingredient, since otherwise one would have to rely on an independent exact solution to fix the coupling. We discuss these developments in section 11.3.

11.1 From loops to a compactified boson

In chapter 8 we have seen how to transform the Potts model into a model of oriented loops on the medial lattice. The weight

$$n = \sqrt{Q} = 2 \cos \gamma \quad (11.1)$$

per loop was transformed into local complex weights at the vertices.

The oriented loop model can easily be turned into a *height model*. For this, assign a scalar variable $h(\mathbf{x})$ to each lattice face \mathbf{x} (i.e., to each vertex of the lattice *dual* to the one on which the loop model has been defined), so that h increases (resp. decreases) by a each time one traverses a left-going (resp. right-going) edge. This definition of the height h is consistent, since each vertex is incident on as many ingoing as outgoing edges. Since this defines only height differences, one may imagine fixing h completely by arbitrarily fixing $h(\mathbf{0}) = 0$.

In the continuum limit, we expect the local height field h to converge to a free bosonic field $\phi(\mathbf{x})$, whose entropic fluctuations are described by an action of the form (10.95), with coupling $g = g(n)$ which is a monotonically increasing function of n . In particular, for $n \rightarrow \infty$ the lattice model is dominated by the configuration where loops of the minimal possible length cover the lattice densely; the height field is then flat, $\phi(\mathbf{x}) = \text{constant}$, and the correlation length ξ is of the order of the lattice spacing. For finite but large n , ϕ will start fluctuating, loop lengths will be exponentially distributed, and ξ will be of the order of the linear size of the largest loop. When $n \rightarrow n_c^+$, for some critical n_c (we shall see that $n_c = 2$), this size will diverge, and for $n \leq n_c$ the loop model will be conformally invariant with critical exponents that depend on $g(n)$. The interface described by $\phi(\mathbf{x})$ is then in a *rough* phase. The remainder of this section is devoted to making this intuitive picture more precise, and to refine the free bosonic description of the critical phase.

As a first step towards greater precision, we now argue that $\phi(\mathbf{x})$ is in fact a *compactified boson*, cf. section 10.7. To see this, it is convenient to consider the *oriented* loop configurations that give rise to a maximally flat microscopic height h ; following Henley and Kondev [KH95] we shall refer to them as *ideal states*. For the Potts model (11.4), an ideal state is a dense packing of length-four loops, all having the same orientation. There are four such states, corresponding to two choices of orientation and two choices of the sublattice of lattice faces surrounded by the loops.

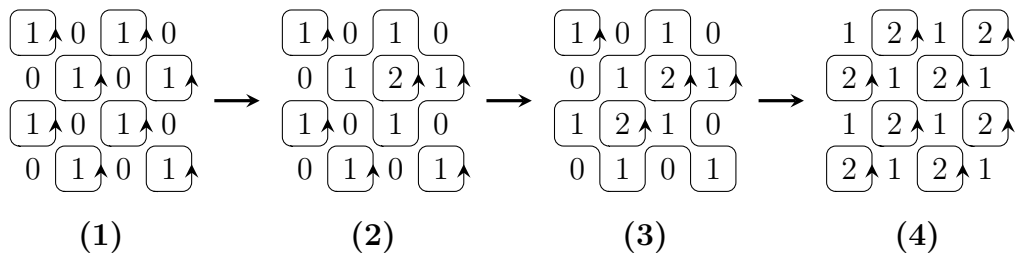


Figure 24: An ideal state (panel 1) in the oriented loop model is gradually changed into another (panel 4). The large loop created in panels 2 and 3 is annihilated via the periodic boundary conditions to obtain panel 4, which is a different ideal state. In the process the mean height changes from ϕ to $\phi + a$ (with $a = 1$ in the figure).

An ideal state can be gradually changed into another by means of $\sim N$ local changes of the transition system and/or the edge orientations. This is shown in Fig. 24. As a result, the mean height will change, $\phi \rightarrow \phi \pm a$. Repeating this, one sees that one may return to the initial ideal state whilst having $\phi \rightarrow \phi \pm 2a$. For consistency, we must therefore require $\phi(\mathbf{x}) \in \mathbb{R}/(2a\mathbb{Z})$, i.e., the field is compactified with radius $R = 1/\pi$, cf. (10.97).

In section 10.7 we have seen in detail that the normalisation constant a drops out from the final physical results. We shall therefore follow standard conventions and set $a = \pi$ in what follows.

11.2 Liouville field theory

The essence of the above discussion is that the critical properties of the loop model under consideration can be described by a continuum-limit partition function that takes the form of a functional integral

$$Z = \int \mathcal{D}\phi(\mathbf{x}) \exp(-S[\phi(\mathbf{x})]) . \quad (11.2)$$

Here $S[\phi(\mathbf{x})]$ is the Euclidean action of the compactified scalar field $\phi(\mathbf{x}) \in \mathbb{R}/(2\pi\mathbb{Z})$. The hypothesis that the critical phase is described by bounded elastic fluctuations around the ideal states means that S must contain a term

$$S_E = \frac{g}{4\pi} \int d^2\mathbf{x} (\nabla\phi)^2 \quad (11.3)$$

with coupling constant $g > 0$. Higher derivative terms that one may think of adding to (11.3) can be ruled out by the $\phi \rightarrow -\phi$ symmetry, or by arguing *a posteriori* that they are RG irrelevant in the full field theory that we are about to construct.

Note that the partition function (11.2) does not purport to coincide with that of the critical Potts model

$$Z = Q^{|V|/2} \sum_{E' \subseteq E} n^{l(E')} \quad (11.4)$$

on the scale of the lattice constant. (A similar remark holds true for the correlation functions that one may similarly write down.) We do however claim that their long-distance properties are the same. In that sense, the CG approach is an exact, albeit by no means rigorous, method for computing critical exponents and related quantities. A more precise equivalence between discrete and continuum-limit partition functions can however be achieved on a torus [DSZ87b] or on an annulus (see chapter 13).

The action (11.3) coincides with (10.95) for the compactified boson. To obtain the full physics of the loop model one however needs to add two more terms to the action, as we now shall see.

We consider the underlying lattice model as being defined on a cylinder, $\mathbf{x} = (x, t)$. This has the advantage of making direct contact with the radial quantisation formalism of section 10.5. The boundary conditions are thus periodic in the space direction, $x = x + L$, and free in the time (t) direction. Ultimately, the results obtained on the cylinder can always be transformed into other geometries by means of a conformal mapping.

We have seen in section 8.8 that with this geometry, in order to obtain the correct weighting of non-contractible loops, the corresponding six-vertex model must be twisted across a seam that runs along the cylinder. Consider now adding a term

$$S_B = \frac{ie_0}{4\pi} \int d^2\mathbf{x} \phi(\mathbf{x}) \mathcal{R}(\mathbf{x}) \quad (11.5)$$

to the effective action S , where \mathcal{R} is the scalar curvature²⁸ of the space \mathbf{x} . The parameter e_0 is known in CG language as the *background electric charge*. On the cylinder, one has simply $S_B = ie_0 (\phi(x, \infty) - \phi(x, -\infty))$, meaning that in

²⁸We consider the scalar curvature in a generalised sense, so that delta function contributions may be located at the boundaries. Implicitly, we are just applying the Gauss-Bonnet theorem.

the partition function (11.2) an oriented loop with winding number $q = 0, \pm 1$ (all other winding numbers are forbidden by the self-avoidance of the loops) can equivalently be assigned an extra weight of $\exp(i\pi q e_0)$. This leaves the weight n of non-winding loops unchanged, while winding loops get a modified weight

$$n_1 = 2 \cos \gamma_1, \quad \text{with } \gamma_1 = \pi e_0, \quad (11.6)$$

as in (8.37). The choice $\gamma_1 = \gamma$ will thus lead to $n_1 = n$. Note however that the possibility of having $n_1 \neq n$ is useful in some applications of the CG technique.

Verify the above argument for more than one non-contractible loop, bearing in mind that the orientation of each loop has to be summed over independently!

The object $e^{ie\phi}$ (or more precisely, its normal ordered product $: e^{ie\phi} :$) is known in CFT as a *vertex operator* of (electric) charge e . The boundary term (11.5) thus corresponds to the insertion of two oppositely charged vertex operators at either end of the cylinder (and more generally at the root vertices of section 8.8).

At this stage two problems remain: the field theory does not yet take account of the weight n of contractible loops, and the coupling constant g has not yet been determined. These two problems are closely linked, and allow [Ko97] us to fix exactly $g = g(n)$. The idea is to add a further *Liouville term*

$$S_L = \int d^2\mathbf{x} w[\phi(\mathbf{x})] \quad (11.7)$$

to the action, which then reads in full

$$S[\phi(\mathbf{x})] = S_E + S_B + S_L. \quad (11.8)$$

In (11.7), $e^{-w[\phi(\mathbf{x})]}$ is the scaling limit of the microscopic vertex weights w_i that we now identify.

Due to the compactification, $S_L[\phi]$ is a periodic functional of the field, and as such it can be developed as a Fourier sum over vertex operators

$$w[\phi] = \sum_{e \in \mathcal{L}_w} \tilde{w}_e e^{ie\phi}, \quad (11.9)$$

where \mathcal{L}_w is some sublattice of $\mathcal{L}_0 \equiv \mathbb{Z}$. Note that \mathcal{L}_w may be a proper sublattice of \mathcal{L}_0 if $w[\phi]$ has a higher periodicity than that trivially conferred by the compactification of ϕ . By inspecting Fig. 24 we see that this is indeed the case here: the (geometric) averages of the microscopic weights coincide on panels 1 and 4, indicating that the correct choice is $\mathcal{L}_w = 2\mathcal{L}_0$. This intuitive derivation of \mathcal{L}_w demonstrates the utility of the ideal state construction.

Some important properties of the compactified boson with action S_E have already been derived in section 10.7. In particular, its central charge is $c = 1$ and the dimension $\Delta_{e,m}$ of an operator with electromagnetic charge (e, m) is given by (10.103). Having now identified the electric charge e with that of the vertex operator $e^{ie\phi}$, one could alternatively rederive (10.103) by computing the two-point function $\langle e^{ie\phi(\mathbf{x})} e^{-ie\phi(\mathbf{y})} \rangle$ by standard Gaussian integration.

The physical interpretation of the magnetic charge m is already obvious from (10.97): it corresponds to dislocations in the height field ϕ due to the presence of defect lines. In section 11.4 we shall see how such defects are used in the computation of critical exponents.

It remains to assess how the properties of the compactified boson are modified by the inclusion of the term S_B . Physical reasoning consists in arguing that the vertex operators $e^{\pm ie_0\phi}$ will create a “floating” electric charge of magnitude $2e_0$ that “screens” that of the other fields in any given correlation function. We infer that (10.103) must be changed into

$$\Delta_{e,m} = \frac{1}{2} \left[\frac{e(e - 2e_0)}{g} + gm^2 \right]. \quad (11.10)$$

Note that to obtain (11.10) we have changed our normalisation so that both e and m are integers. This is consistent with the normalisation (11.3) of the coupling constant, rather than (10.95), which is the standard choice in the CG literature.

11.3 Marginality requirement

Following Kondev [Ko97] we now claim that the Liouville potential S_L must be exactly marginal. This follows from the fact that all loops carry the same weight n , independently of their size, and so the term S_L in the action that enforces the loop weight must not renormalise under a scale transformation. The most relevant vertex operator appearing in (11.9) has charge $e_w = 2\pi/a = 2$, and so $\Delta_{e_w,0} = 2$. Using (11.10), this fixes the coupling

constant as $g = 1 - e_0$. In other words, the loop weight has been related to the CG coupling as

$$n = \sqrt{Q} = -2 \cos(\pi g), \quad (11.11)$$

with $0 < g \leq 1$ for the critical Potts model.

The term S_B shifts the ground state energy with respect to the $c = 1$ theory described by S_E alone. The corrected central charge is then $c = 1 + 12\Delta_{e_0,0}$, where the factor of 12 comes from comparing (10.63) and (10.65). This gives

$$c = 1 - \frac{6(1-g)^2}{g}. \quad (11.12)$$

It should be noted that the choice $e_w = 2$ is not the only one possible. Namely, the coefficient \tilde{w}_{e_w} of the corresponding vertex operator in (11.9) may be made to vanish, for instance by driving the Potts model to tricriticality via the introduction of a carefully tuned coupling to non-magnetic vacancies. The next-most relevant choice is then $\tilde{e}_w = -2$, and going through the same steps as above we see that one can simply maintain (11.11), but take the coupling in the interval $1 \leq g \leq 2$ for the tricritical Potts model.

The electric charge e_w whose vertex operator is required to be exactly marginal is known as the *screening charge* in standard CG terminology.

The central charge (11.12) can now be formally identified with that of the Kac table (10.87), with $m' = m + 1$. The result is a formal relation between the minimal model index m and the CG coupling g , valid for integer m . We have

$$m = \begin{cases} \frac{g}{1-g} & \text{for the critical Potts model} \\ \frac{1}{g-1} & \text{for the tricritical Potts model} \end{cases} \quad (11.13)$$

11.4 Critical exponents

We shall now see how to use the Coulomb gas technology to compute a variety of critical exponents in loop models.

The watermelon exponents were derived by Nienhuis [Nh87] and by Duplantier and Saleur (see [DS87] and references therein). The issues of their relation to the standard exponents of polymer physics [Ge79], and to the Kac table (10.81), were discussed in [DS87].

Although the watermelon exponents are essentially magnetic-type exponents in the CG, they do not produce the standard magnetic exponent of the Potts model. The latter was derived by den Nijs [Ni83], but we present here a somewhat different argument.

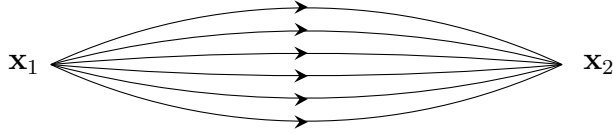


Figure 25: Watermelon configuration with $\ell = 6$ legs.

11.4.1 Watermelon exponents

An important object in loop models is the operator $\mathcal{O}_\ell(\mathbf{x}_1)$ that inserts ℓ oriented lines at a given point \mathbf{x}_1 . Microscopically, this can be achieved by violating the arrow conservation constraint at \mathbf{x}_1 . In the Potts model, or rather the equivalent six-vertex model, a vertex with four outgoing and zero ingoing arrows furnishes a microscopic realisation of the case $\ell = 2$. Higher ℓ can be obtained by inserting several defects in a small region around \mathbf{x}_1 .²⁹

If one had strict arrow conservation at all other vertices, the insertion of $\mathcal{O}_\ell(\mathbf{x}_1)$ would not lead to a consistent configuration. However, also inserting $\mathcal{O}_{-\ell}(\mathbf{x}_2)$, the operator that absorbs ℓ oriented lines in a small region around \mathbf{y} , will lead to consistent configurations (see Fig. 25) in which ℓ defect lines propagate from \mathbf{x}_1 to \mathbf{x}_2 . Let $Z_\ell(\mathbf{x}_1, \mathbf{x}_2)$ be the corresponding constrained partition function. One then expects

$$\langle \mathcal{O}_\ell(\mathbf{x}_1) \mathcal{O}_{-\ell}(\mathbf{x}_2) \rangle \equiv \frac{Z_\ell(\mathbf{x}_1, \mathbf{x}_2)}{Z} \sim \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|^{2\Delta_\ell}} \text{ for } |\mathbf{x}_1 - \mathbf{x}_2| \gg 1. \quad (11.14)$$

The corresponding critical exponents Δ_ℓ are known as watermelon (or *fuseau*, or ℓ -leg) exponents. To compute them we first notice that the sum of the height differences around a closed contour encircling \mathbf{x}_1 but not \mathbf{x}_2 will be $a\ell$. Equivalently, one could place the two defects at the extremities of a cylinder [i.e., taking $\mathbf{x}_1 = (x, -\infty)$ and $\mathbf{x}_2 = (x, \infty)$], and the height difference would be picked up by any non-contractible loop separating \mathbf{x}_1 and \mathbf{x}_2 . This latter formulation makes contact with the defect lines (10.97) introduced when studying the compactified boson, the equivalent magnetic charge being $m_\ell = \frac{\ell a}{2\pi} = \frac{\ell}{2}$.

A little care is needed to interpret the configurations of $Z_\ell(\mathbf{x}_1, \mathbf{x}_2)$ in the model of un-oriented loops. The fact that all ℓ lines are oriented away

²⁹The Potts model only allows for defects with even ℓ . In the closely related $O(n)$ model, any parity of ℓ is permitted.

from \mathbf{x}_1 prevents them from annihilating at any other vertex than \mathbf{x}_2 . One should therefore like to think about them as ℓ marked lines linking \mathbf{x}_1 and \mathbf{x}_2 , where each line carries the Boltzmann weight 1. This is consistent with not summing over the orientations of the defect lines in the oriented loop model. However, each oriented line can also pick up spurious phase factors $w(\alpha/2\pi)$, due to the local redistribution of loop weights, whenever it turns around the end points \mathbf{x}_1 and \mathbf{x}_2 . These factors are however exactly cancelled if we insert in addition a vertex operator $e^{ie_0\phi}$ (resp. $e^{-ie_0\phi}$) at \mathbf{x}_1 (resp. \mathbf{x}_2) [Nh84]. Note that these vertex operators do not modify the weighting of closed loops: these must encircle either none of both of $\mathbf{x}_1, \mathbf{x}_2$, since otherwise they would intersect the ℓ watermelon legs. We conclude that $\Delta_\ell = \Delta_{e_0, m_\ell}$, and using (11.10) this gives

$$\Delta_\ell = \frac{1}{8}g\ell^2 - \frac{(1-g)^2}{2g}. \quad (11.15)$$

Interestingly, these exponents can be attributed to the Kac table under the identification (11.13). One has

$$\Delta_\ell = \begin{cases} 2h_{0, \ell/2} & \text{for the dense } O(n) \text{ model} \\ 2h_{\ell/2, 0} & \text{for the dilute } O(n) \text{ model} \end{cases} \quad (11.16)$$

The Kac indices (r, s) appearing in $h_{r,s}$ are integer valued, since $\ell \in 2\mathbb{N}$. When the loop model coincides with a minimal model—i.e., when g is such that $m \in \mathbb{N}$ in (11.13)—some of these exponents are located outside the fundamental domain of the Kac table. This reflects the fact that the corresponding watermelon operators are of intrinsic non-local nature.

11.4.2 Application to percolation clusters

The watermelon exponents can be used to elucidate the fractal properties of the Fortuin-Kasteleyn (FK) clusters defined in section 8.2. Here we limit the discussion to the special case of percolation clusters.

We have seen in section 8.4.1 that bond percolation is the $Q \rightarrow 1$ limit of FK clusters. We have therefore $g = \frac{2}{3}$ from (11.11). The watermelon exponents (11.15) are therefore

$$\Delta_\ell = \frac{\ell^2 - 1}{12}. \quad (11.17)$$

Marking a point \mathbf{x} on the hull of a percolation cluster corresponds to the insertion of the operator $\mathcal{O}_2(\mathbf{x})$. The fractal dimension of the hull is therefore

$$d_h = 2 - \Delta_2 = \frac{7}{4}. \quad (11.18)$$

A *pivotal edge* is defined as an edge belonging to a percolation cluster which is such that the removal of the edge makes the cluster break into two connected components. In the literature on percolation pivotal edges are also known as *red bonds*. Cutting the loop strands on either side of any edge belonging to the cluster looks like an $\ell = 4$ leg insertion. Note however that only if the edge is pivotal will the four legs propagate to “infinity” without contracting among themselves. Therefore the fractal dimension of red bonds is

$$d_{rb} = 2 - \Delta_4 = \frac{3}{4}. \quad (11.19)$$

11.4.3 Magnetic exponent

The watermelon exponents can be said to be of the “magnetic” type, since they induce a magnetic type defect charge m_ℓ in the CG. The standard magnetic exponent, describing the decay of the spin-spin correlation function in the Potts model, is however not of the watermelon type. It can nevertheless be inferred from (11.10) as follows:

The probability that two spins situated at \mathbf{x}_1 and \mathbf{x}_2 are in the same Potts state is proportional, in the random cluster picture, to the probability that they belong to the same cluster. In the cylinder geometry this means that no winding loop separates \mathbf{x}_1 from \mathbf{x}_2 . This can be attained in the CG by giving a weight $n_1 = 0$ to such loops. We have seen that inserting a pair of vertex operators with charge $\pm e$ at \mathbf{x}_1 and \mathbf{x}_2 leads exactly to this situation with $n_1 = 2 \cos(\pi e)$, and so we need $e = \frac{1}{2}$. The scaling dimension of this excitation, with respect to the ground state which has $e = e_0$, is then

$$\Delta_m = \Delta_{\frac{1}{2},0} - \Delta_{e_0,0} = \frac{1 - 4(1 - g)^2}{8g}. \quad (11.20)$$

In particular for the Ising model, with $g = \frac{3}{4}$, this yields the magnetic exponent $\Delta_m = \frac{1}{8}$, or in standard notation

$$\beta = \frac{1}{8}. \quad (\text{Ising model}) \quad (11.21)$$

For bond percolation, with $g = \frac{2}{3}$, we find $\Delta_m = \frac{5}{48}$. The fractal dimension of a percolation cluster is thus

$$d_c = 2 - \Delta_m = \frac{91}{48}. \quad (11.22)$$

The location in the Kac table (10.81) of the magnetic exponent (11.20) can be found using (11.13):

$$\Delta_m = 2h_{1/2,0}. \quad (11.23)$$

Note that this differs from the lowest possible watermelon excitation $\Delta_2 = 2h_{0,1}$. Indeed, the two-leg excitation corresponds to a cluster that propagates along the length direction of the cylinder *without* wrapping around the transverse periodic direction. The dominant configurations participating in the magnetic correlation function have *no* propagating legs, since the cluster containing \mathbf{x}_1 and \mathbf{x}_2 will typically wrap around the cylinder.

12 Basic aspects of boundary CFT

The aspects of CFT exposed to this point pertain to unbounded geometries, either that of the infinite plane (Riemann sphere) or, in section 10.7, that of the torus (which is really a finite geometry made unbounded through the periodic boundary conditions). In contrast, boundary conformal field theory (BCFT) describes surface critical behaviour, i.e., a critical system confined to a bounded geometry. The simplest such geometry, is that of the upper half plane $\{z \mid \Im z \geq 0\}$, where the real axis \mathbb{R} acts as the boundary (one-dimensional “surface”).

The foundations of BCFT were set by Cardy [Ca84b] who also initiated many of the subsequent developments and applications (see [DMS87, Ca05] for reviews). A useful review of the status of boundary critical phenomena before the advent of CFT was given by Binder [Bi83].

12.1 Qualitative discussion

To convey an idea of which phase transitions may result from the interplay between bulk and boundary degrees of freedom, and what may be the corresponding boundary conditions, we begin by a qualitative discussion of a simple magnetic spin system. We denote the local order parameter (magnetisation) by ϕ . When the boundary spins enjoy free boundary conditions, they interact more weakly than the bulk spins, since microscopically they are coupled to fewer neighbouring spins. Upon lowering the temperature, the bulk will therefore order before the surface: this is the so-called *ordinary transition*. Now consider placing the system slightly below the bulk critical temperature. Then ϕ is non-zero deep inside the bulk, and will decrease upon approaching the boundary. One can argue that in the continuum limit ϕ will vanish exactly on the boundary. Thus, the Dirichlet boundary condition $\phi|_{\mathbb{R}} = 0$ is the appropriate choice for describing the ordinary transition.

Let us now introduce a coupling J_s between nearest-neighbour spins on the boundary which may be different from the usual bulk coupling constant J . Taking $J_s > J$ one may “help” the boundary to order more easily.³⁰ When J_s takes a certain critical value we are at the *special transition*, at which the bulk and the boundary order simultaneously. Finally, when $J_s \rightarrow \infty$ the boundary

³⁰A similar effect could be obtained by adding a surface magnetic field, but here we do not wish to break the symmetry of the model [typically $O(n)$ in applications to loop models].

spins are always completely ordered³¹, a fact which changes the nature of the ordering transition of the bulk, now referred to as the *extraordinary transition*. This corresponds to the Dirichlet boundary condition $\phi|_{\mathbb{R}} = \infty$ in the continuum limit. Note that in the application of boundary CFT to loop models (see section 12.5) the meaning of J_s is to give a specific fugacity to monomers on the boundary.

The control parameter J_s can be thought of in a renormalisation group sense, and is readily seen to be irrelevant at the ordinary and extraordinary transitions. Accordingly we expect a boundary RG flow to go from the special to either of the two other transitions. (In the case of the Ising model, the special and extraordinary transitions actually coincide.)

In our subsequent application to loop models (see section 12.5) we rather think of ϕ as a height field which is dual to the system of oriented loops (this construction is at the heart of the Coulomb gas approach, see section 11). In other words, the loops are level lines of ϕ . Dirichlet boundary conditions then describe a situation in which loops are reflected off the boundary, and adjoining two different Dirichlet conditions forces one or more “loop ends” to emanate from the boundary. One may also impose Neumann boundary conditions, $\partial\phi/\partial y|_{\mathbb{R}} = 0$, meaning that the “loops” coming close to the boundary must in fact terminate perpendicular to it. Clearly the non-local aspects of these situations call for a more detailed discussion, which will be postponed to section 12.5.

12.2 Comparison of bulk and boundary CFT

The formalism of boundary CFT is very similar to the bulk case. In this section we briefly outline a few but important differences.

The allowed conformal mappings in BCFT must keep invariant both the boundary itself and the boundary conditions imposed along it. For the global conformal transformations (10.44) the invariance of the real axis forces $a_{ij} \in \mathbb{R}$, i.e., they form the group $SL(2, \mathbb{R})$ and the number of parameters is halved from 6 to 3. For an infinitesimal local conformal transformation $z \rightarrow w(z) = z + \varepsilon(z)$ the requirement reads $\varepsilon(\bar{z}) = \bar{\varepsilon}(z)$. This property can be used to eliminate the $\bar{\varepsilon}(z)$ part altogether, since it is just the analytic continuation

³¹This should not (as is sometimes seen in the literature) be confused with imposing *fixed* boundary conditions, which would rather correspond to an infinite symmetry-breaking field applied on the boundary.

of $\varepsilon(\bar{z})$ into the lower half plane. It follows that $\bar{L}_n = L_{-n}$, and so one half of the conformal generators has been eliminated.

At the level of the stress tensor, the requirement is $T(\bar{z}) = \bar{T}(z)$. In Cartesian coordinates this reads $T_{xy} = 0$ on the real axis, the so-called *conformal boundary condition*. Its physical meaning is that there is no energy-momentum flow across \mathbb{R} . This has important consequences on the conformal Ward identity (10.57) where $T(z)$ is applied to a product of primary fields $X = \prod_j \phi_j(z_j, \bar{z}_j)$ situated in the upper half plane. The contour C surrounding all z_j can then be taken as a large semicircle with the diameter parallel to the real axis. However, writing the same identity for $\bar{T}(\bar{z})$ yields another Ward identity involving the conjugate semicircle contour \bar{C} , and since $\bar{T} = T$ when $z \in \mathbb{R}$, the two contours can be fused into a complete circle surrounding both z_j and \bar{z}_j . The end result, cf. (10.58), is thus

$$T(z)X = \sum_j \left(\frac{h_j}{(z - z_j)^2} + \frac{\partial_{z_j}}{z - z_j} + \frac{\bar{h}_j}{(\bar{z} - \bar{z}_j)^2} + \frac{\partial_{\bar{z}_j}}{\bar{z} - \bar{z}_j} \right) X. \quad (12.1)$$

In conclusion, everything happens as if each primary field in the upper half plane were accompanied by a mirror field in the lower half plane. This means that computations in the BCFT can be done using a *method of images* similar to that used in electrostatics when solving the Laplace equation with boundary conditions. Correlation functions are computed as if the theory were defined on the whole complex plane, and governed by a single Virasoro algebra (10.69): the physical fields are then situated in the upper half plane, and their unphysical mirror images in the lower half plane. The simplification of getting rid of \bar{L}_n has thus been achieved at the price of doubling the number of points in correlation functions. In practice, the former simplification largely outweighs the latter complication.

In particular, the n -point boundary correlation functions satisfy the very same differential equations (10.86) as $2n$ -point bulk correlation functions, but with different boundary conditions. The most interesting cases are $n = 1$ and $n = 2$, both tractable in the bulk picture in several situations of practical importance. As examples of the physical information which can be extracted from these cases we should mention, for $n = 1$, the probability profile of finding a monomer of a loop at a certain distance from the boundary, and for $n = 2$, the probability that a polymer comes close to the boundary at two prescribed points [BE94]. A particularly celebrated application of the

$n = 2$ case is Cardy's computation [Ca92] of the *crossing probability* that a percolation cluster traverses a large rectangle, as a function of the aspect ratio of the latter.

The radial quantisation scheme of section 10.5 still makes sense in BCFT. The associated conformal mapping

$$w(z) = \frac{L}{\pi} \log z \quad (12.2)$$

transforms the upper half plane into a semi-infinite strip of width L with non-periodic transverse boundary conditions. The two rims of the strip are then the images of the positive and the negative real axis, and the time (resp. space) direction is parallel (resp. perpendicular) to the axis of the strip. The dilatation operator reads $\mathcal{D} = L_0$ and the Hamiltonian $\mathcal{H} = (\pi/L)(L_0 - c/24)$. Non-trivial eigenstates of \mathcal{H} are formed by a *boundary operator* $\phi_j(0)$ acting on the vacuum state, $|h\rangle = \phi_j(0)|0\rangle$.

In general, we expect boundary operators to have different scaling dimensions than bulk operators. This can be understood from the method of images: when a primary field approaches the boundary it interacts with its mirror image and, by the OPE (10.74), produces a series of other primaries which then describe the boundary critical behaviour.

Likewise, a field $\phi_{(r,s)}$ with a given interpretation in the bulk will typically have a different interpretation when situated on the boundary. Examples pertinent to loop models will be given in section 12.5.

The finite-size formulae (10.63) and (10.65) can be adapted to the case of a strip of width L . For this, one uses the method of images and the mapping (12.2). The end results read:

$$\begin{aligned} f_0(L) &= f_0(\infty) + \frac{f_0^S}{L} - \frac{\pi c}{24L^2} + o(L^{-2}), \\ f_\phi(L) - f_0(L) &= \frac{f_\phi^S - f_0^S}{L} + \frac{\pi\Delta}{L^2} + o(L^{-2}), \end{aligned} \quad (12.3)$$

where there is now a non-universal $1/L$ dependence due to the presence of surface free energies f^S . For some (but not all) choices of excited levels $f_\phi(L)$ it can be argued that $f_\phi^S = f_0^S$, thus simplifying the second of these formulae.

Note that (10.45) applied to a boundary operator is the reason why we have not discussed *finite* Dirichlet boundary conditions at the beginning of

this section. More generally, any uniform boundary condition is expected to flow under the renormalisation group towards a *conformally invariant boundary condition*. It is one of the goals of BCFT to classify such boundary conditions. One of the main results obtained is the following [Ca05]: For diagonal models (i.e., $n_{h,\bar{h}} = \delta_{h,\bar{h}}$ in (10.75)) there is a bijection between the primary fields in the bulk CFT and the conformally invariant boundary conditions in the BCFT.

For example, for the Ising model ($m = 3$ and $m' = 4$ in (10.87)) the three different bulk primary operators (the identity $I = \phi_{(1,1)}$, the spin $\sigma = \phi_{(1,2)}$, and the energy $\varepsilon = \phi_{(2,1)}$) correspond to three types of uniform boundary conditions in the lattice model of spins (fixed $s = +1$ and $s = -1$, and free boundary conditions).

To this point we have discussed only uniform boundary conditions. It is important to realise that the radial quantisation picture with a boundary operator $\phi_j(0)$ situated at the origin is compatible also with mixed boundary conditions, i.e., one boundary condition on the negative real half-axis and another on the positive half-axis. In this case, $\phi_j(0)$ is called a *boundary condition changing operator*. One then needs a second operator $\phi_j(\infty)$ situated at infinity to change back the boundary condition. A more symmetric picture is obtained by mapping the upper half plane to the strip, through (12.2). There are then different boundary conditions on the two sides of the strip, and a boundary condition changing operators is located at either end of the strip. More generally, one may study a BCFT on any simply connected domain with a variety of different boundary conditions along the boundary, each separated by a boundary condition changing operator.

For bulk CFT, crucial insight was gained by considering the theory on a torus. The analogous tool for BCFT is to consider the theory on an annulus.³² In analogy with the torus case, we denote by $\omega_1 \in \mathbb{R}$ the width of the annulus and by $\omega_2 \in i\mathbb{R}$ its length (in the periodic direction), defining $\tau = \omega_2/\omega_1 \in i\mathbb{R}$. The boundary conditions on the two rims are denoted, symbolically, a and

³²It makes sense to think of this in the radial quantisation, or transfer matrix, picture. The theories are initially considered on a semi-infinite cylinder (resp. a strip) with specified transverse boundary conditions (periodic, resp. non-periodic) and unspecified longitudinal boundary conditions. This gives access to the transfer matrix eigenvalues. To access the fine structure, such as amplitudes of the eigenvalues, one must impose periodic longitudinal boundary conditions and take the length of the cylinder (resp. strip) to be *finite*.

b. Then

$$Z_{ab}(\tau) = \text{Tr} \left(q^{L_0 - c/24} \right) \quad (12.4)$$

with $q = \exp(\pi i \tau)$. This should be compared with (10.93). The analogue of (10.94),

$$Z_{ab}(\tau) = \sum_h n_h^{(ab)} \chi_{(c,h)}(\tau), \quad (12.5)$$

then becomes *linear* in the characters. Equivalently, one might exchange the space and time direction and view the annulus as a cylinder of circumference ω_2 and finite length ω_1 , with boundary conditions a (resp. b) in the initial (resp. final) state. This leads to

$$Z_{ab}(\tau) = \left\langle b \left| e^{\tau^{-1} \mathcal{H}_{\text{bulk}}} \right| a \right\rangle, \quad (12.6)$$

where now $\mathcal{H}_{\text{bulk}}$ is the Hamiltonian of the *bulk* CFT propagating between boundary states $|a\rangle$ and $\langle b|$. The links between bulk and boundary CFT result from a detailed study of the equivalence between (12.4) and (12.6).

12.3 Coulomb gas on an annulus

Consider now instead the loop model defined on an annulus which we shall take as an $L \times M$ rectangle with coordinates $x \in [0, L]$ and $y \in [0, M]$. The boundary conditions are free (f) in the x -direction and periodic in the y -direction. Cardy [Ca06] has shown how to impose the correct marginality requirement for this geometry.

Consider first the continuum-limit partition function $Z = Z_{\text{ff}}(\tau)$ from (12.4) in the limit $M/L \gg 1$ of a very long and narrow annulus. The modular parameters $\tau = iM/L$ and $q = \exp(i\pi\tau) = \exp(-\pi M/L)$. We expect in this limit that only the identity operator contributes to Z , and so

$$Z \sim q^{-c/24} \sim \exp\left(\frac{\pi c M}{24L}\right). \quad (12.7)$$

The central charge c is (11.12) from the bulk theory, and in particular is known to vary with the coupling constant g .

The question then arises how (12.7) is compatible with the continuum-limit action (11.3). According to Cardy [Ca06] the answer is that there is a background magnetic flux m_0 , a sort of electromagnetic dual of the background electric charge e_0 present in the cylinder geometry. Thus, in the

continuum limit there is effectively a number (in general fractional) m_0 of oriented loops running along the rims of the annulus, giving rise to a height difference between the left and the right rim. Accepting this hypothesis, we can write

$$\phi(x, y) = \tilde{\phi}(x, y) + \frac{\pi m_0 x}{L}, \quad (12.8)$$

where $\tilde{\phi}$ is a “gauged” height field that still contains the elastic fluctuations but obeys identical Dirichlet boundary conditions on both rims, say $\tilde{\phi}(0, y) = \tilde{\phi}(L, y) = 0$.

According to the functional integrations in section 10.7, the field $\tilde{\phi}$ contributes $q^{-1/24}$ to Z , corresponding to $c = 1$. The last term in (12.8) modifies the action (11.3) by $\Delta S = \frac{g}{4\pi}(\pi m_0)^2 \frac{M}{L}$ and thus multiplies Z by a factor $e^{-\Delta S} = q^{gm_0^2/4}$, which correctly reproduces the contribution of the last term in (11.12) to (12.7) provided that we set

$$m_0 = \pm \frac{(1 - g)}{g}. \quad (12.9)$$

This value of m_0 can be retrieved from a marginality requirement which has the double advantage of being more physically appealing and of not invoking the formula (11.12) for c . Indeed, if m_0 is too large a pair of oriented loop strands will shed from the rims, corresponding to a vortex pair of strength $m = \pm 2$ situated at the top and the bottom of the annulus. This vortex pair can then annihilate in order to reduce the free energy. And if m_0 is too small the opposite will occur. The equilibrium requirement is then that inserting such a vortex pair must be an exactly marginal perturbation in the RG sense, i.e., the corresponding *boundary* scaling dimension is $\Delta_v = 1$.

The free energy increase for creating the vortex pair is, by the same gauge argument as before,

$$\Delta S = \frac{g}{4\pi}((m_0 \pm 2)^2 - m_0^2) \left(\frac{\pi}{L}\right)^2 ML, \quad (12.10)$$

and noting the factor of 24 between c and the scaling dimension Δ_v in (12.3), we now have $e^{-\Delta S} = q^{-\Delta_v}$ from (12.7), so that

$$\Delta_v = \frac{g}{4}((m_0 \pm 2)^2 - m_0^2) = 1 \quad (12.11)$$

and we recover (12.9). The ambiguity on the sign in (12.9) will be lifted in section 12.6 below.

12.4 $O(n)$ model

The loop model originating from the Potts model does not allow for the phenomenology described in section 12.1. By construction, the loops cover all edges of the (medial) lattice, hence cannot be further attracted to the boundary by enhancing the coupling at the surface. Clearly we need a model in which loop coexist with some empty regions of space.

Notwithstanding these comments, the Potts model allows for a very rich surface critical behaviour. This is obtained by assigning a weight n_b to loops touching (once or more) the boundary, different from the weight of bulk loops [JS08a, JS08b, DJS09]. Discussing this in detail requires in particular the introduction of boundary extensions of the Temperley-Lieb algebra.

The $O(n)$ model is defined initially by associated with each vertex $i \in V$ of a regular planar lattice $G = (V, E)$ a vector spin $\mathbf{S}_i \in \mathbb{R}^n$ of unit length, $|\mathbf{S}_i|^2 = 1$. It turns out convenient to absorb in the integration measure a factor n/Ω_n , where Ω_n is the surface area of the unit sphere in \mathbb{R}^n . Thus, if S_i^α and S_i^β are components of a vector spin \mathbf{S}_i , we have the basic integration rule

$$\int d\mathbf{S}_i S_i^\alpha S_i^\beta = \delta(\alpha, \beta). \quad (12.12)$$

The partition function of the $O(n)$ model is defined by

$$Z = \int_{\mathbf{s}} \prod_{(ij) \in E} e^{-V(\mathbf{S}_i, \mathbf{S}_j)}, \quad (12.13)$$

where we have introduced a short-hand notation for the integration over all spins

$$\int_{\mathbf{s}} \equiv \prod_{i \in V} \left(\int d\mathbf{S}_i \right) \quad (12.14)$$

and $V(\mathbf{S}_i, \mathbf{S}_j)$ is some scalar potential describing the interaction between \mathbf{S}_i and \mathbf{S}_j . In most texts on the $O(n)$ model in general dimension d , one takes

$$V(\mathbf{S}_i, \mathbf{S}_j) = -K \mathbf{S}_i \cdot \mathbf{S}_j. \quad (12.15)$$

In $d = 2$ it is however much more convenient to define

$$e^{-V(\mathbf{S}_i, \mathbf{S}_j)} = 1 + K \mathbf{S}_i \cdot \mathbf{S}_j, \quad (12.16)$$



Figure 26: Allowed vertices in the $O(n)$ model on the hexagonal lattice with their corresponding Boltzmann weights.

where K is a dimensionless coupling constant.

The high-temperature ($K \ll 1$) expansion of (12.13) with potential (12.16) parallels the Fortuin-Kasteleyn cluster expansion of the Potts model partition function. To each term in the expansion we associate an edge subset $E' \subseteq E$, with $e = (ij) \in E$ if we take the term $K\mathbf{S}_i \cdot \mathbf{S}_j$ in (12.16). For each $i \in V$, by the symmetry $\mathbf{S}_i \rightarrow -\mathbf{S}_i$, the contribution to Z of a term associated with E' vanishes unless i is incident on an *even* number of edges in E' .

As a further simplification we now take G to be the hexagonal lattice. Since each vertex $i \in V$ has degree three, the only edge sets E' contributing to the expansion of Z are those where the vertices of $G' = (V, E')$ all have degree zero or two, as shown in Figure 26. In other words, G' is a set of self-avoiding and mutually avoiding loops. The contribution to Z of a loop of length p edges is

$$Z_p = K^p \int d\mathbf{S}_1 \cdots \int d\mathbf{S}_p \sum_{\alpha_1, \dots, \alpha_p} S_1^{\alpha_1} S_2^{\alpha_1} S_2^{\alpha_2} S_3^{\alpha_2} \cdots S_p^{\alpha_p} S_1^{\alpha_p} \quad (12.17)$$

$$= K^p \sum_{\alpha_1, \dots, \alpha_p} \delta(\alpha_1, \alpha_2) \delta(\alpha_2, \alpha_3) \cdots \delta(\alpha_p, \alpha_1) \quad (12.18)$$

$$= K^p n, \quad (12.19)$$

where we have used (12.12) repeatedly. We have then finally

$$Z = \sum'_{E' \subseteq E} K^{|E'|} n^{l(E')}, \quad (12.20)$$

where $l(E')$ is the number of cycles (loops), and the prime on the summation reminds us that the summation is constrained to edge subsets E' such that each vertex $i \in V$ is incident on zero or two edges of E' .

The $O(n)$ model partition function in the form (12.20) is quite similar to the loop representation (8.24) of the Potts model, except that the loop weight \sqrt{Q} has been replaced by n , and that vertices are now allowed to be empty of loops with a relative Boltzmann weight K^{-1} proportional to the temperature.

At infinite temperature ($K = 0$), we have thus $E' = \emptyset$ and $Z = 1$. As the temperature is lowered, loops will start appearing, and one would expect that there exists some critical coupling K_c such that the average length of a loop diverges. Obviously, this means that the correlation length will diverge as well, and so K_c could be expected to be the locus of a second order phase transition. The exact solution of the $O(n)$ model however shows that these hypotheses are only fulfilled for $-2 \leq n \leq 2$. Assuming this to be the case, if K_c is small enough, one could hope that the critical behaviour is identical to that of the generic $O(n)$ model, since the two potentials (12.15)–(12.16) agree to first order in K .

The exact solution implies that one has in fact

$$K_c = (2 \pm \sqrt{2-n})^{-1/2} . \quad (12.21)$$

The plus (resp. minus) sign is referred to as the *dilute* (resp. *dense*) phase.

The Coulomb gas treatment of the $O(n)$ model is almost identical to that of the Potts model. The only subtle difference is at the level of critical exponents. Indeed, the $O(n)$ model allows for watermelon configurations with any number ℓ of legs, whereas for the Potts model only even ℓ is allowed. Apart from that the central charge and the ℓ -leg exponents are identical.

The dense (resp. dilute) phase of the $O(n)$ model has the same Coulomb gas coupling—hence belongs to the same universality class—as the critical (resp. tricritical) Potts model.

The fact that $\Delta_4 > 2$ in the dilute phase means that loop crossings are irrelevant in the RG sense. This justifies a posteriori the truncation made when going from (12.15) to (12.16). In particular for $d = 2$, the dilute phase of the $O(n)$ model on the hexagonal lattice correctly describes the continuum limit of the generic $O(n)$ model.

12.5 Surface critical behaviour of loop models

The $O(n)$ model with suitably modified surface couplings permits one to realise the ordinary, special, and extraordinary surface transitions described

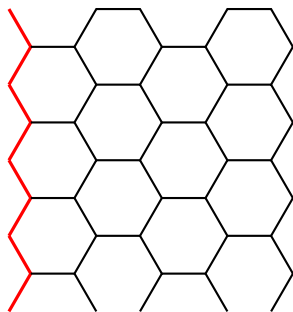


Figure 27: Hexagonal lattice in an annular geometry. The top and the bottom of the figure are identified. Boundary edges on the left are shown in red colour.

qualitatively in section 12.1. To this end, one studies the model defined in the annular geometry of section 12.3.

To be precise, the special transition requires the loops to be in the dilute phase, and so we shall assume this to be the case throughout section 12.5. The results for the ordinary and extraordinary transitions hold true in the dense phase as well.

A well-studied case is the hexagonal-lattice loop model (12.20). The lattice is oriented such that one third of the lattice bonds are parallel to the x -axis, as shown in Fig. 27. The fugacity of a monomer is still denoted K in the bulk, but we now take a different weight K_s for a monomer touching the *left* rim of the annulus, $x = 0$. In contrast, the right rim of the annulus, $x = L$, enjoys free boundary conditions, meaning that its surface monomers still carry the usual weight K .

In this section we wish to limit the discussion to the case where only the left boundary sustains particular (\neq free) boundary conditions; this is sometimes referred to as *mixed* boundary conditions. The case where both boundaries are distinguished is also of interest [DJS09].

The loop model described above has been thoroughly studied by Batchelor and coworkers [BS93, YB95, BY95, BC97], in particular using Bethe Ansatz analysis. They find in particular that when $K_s = K$ the model is integrable and belongs to the universality class of the *ordinary* transition, while for

$$K_s = K_s^S \equiv (2 - n)^{-1/4} \quad (12.22)$$

it is also integrable and describes the special transition.³³ This is consistent with a boundary RG scenario, where K_s^S is a repulsive fixed point that flows towards either of the attractive fixed points $K_s^O < K$ and $K_s^E = \infty$, the former (resp. latter) point describing the ordinary (resp. the extraordinary) transition.

12.6 Watermelon exponents

Surface watermelon exponents can be defined as in section 11.4.1, the only difference being that the ℓ legs are inserted at the boundary. We shall denote these exponents by Δ_ℓ^O , Δ_ℓ^S , Δ_ℓ^E at the ordinary, special, extraordinary surface transition respectively. Whenever a result applies to any of these transitions, we use the generic notation Δ'_ℓ , where the prime indicates a surface rather than a bulk exponent.

For the ordinary transition, Δ_ℓ^O can be derived by a slight refinement of the marginality argument given in section 12.3. First recall that in the continuum limit there is a background flux m_0 given by (12.9), corresponding to a (fractional) number of oriented loop strands running along the rims of the annulus. Suppose now that we wish to evaluate the scaling dimension Δ_ℓ^O corresponding to having $\ell > 0$ non-contractible oriented loop strands running around the periodic direction of the annulus. This can be done by evaluating the free energy increase $\Delta S = S_\ell - S_0$ due to these strands, as in (12.10)

$$\Delta S = \frac{g}{4\pi} ((\ell + m_0)^2 - m_0^2) \left(\frac{\pi}{L}\right)^2 ML \quad (12.23)$$

and using $e^{-\Delta S} = q^{-\Delta_\ell^O}$ from (12.7).

The question now arises which sign for m_0 to pick in (12.9). With the plus sign we would have $\Delta_2 = 1$ independently of g , in clear contradiction with numerical results [DS86]. Taking therefore the minus sign leads to the result

$$\Delta_\ell^O = \frac{1}{4}g\ell^2 - \frac{1}{2}(1 - g)\ell. \quad (12.24)$$

The derivation just presented follows the argument of Cardy [Ca06], but in fact (12.24) was found a long time before by other means. Duplantier and Saleur [DS86] were the first to propose (12.24) for any ℓ , by noting that

³³Technically speaking this is the mixed ordinary-special transition, but we have simplified the terminology according to the above remarks.

their numerical transfer matrix results were in excellent agreement with the following locations in the Kac table (10.81)

$$\Delta_\ell^{\text{O}} = \begin{cases} h_{1,1+\ell} & \text{for the dense } \text{O}(n) \text{ model} \\ h_{1+\ell,1} & \text{for the dilute } \text{O}(n) \text{ model} \end{cases} \quad (12.25)$$

from which (12.24) follows by the identification (11.13). On a more rigorous level, (12.24) has been established by Bethe Ansatz (BA) techniques [SB89, BS93, YB95].

For the special transition, Δ_ℓ^{S} does not seem to permit a CG derivation. It is however known from the BA analysis [BY95, YB95] that one has

$$\begin{aligned} \Delta_\ell^{\text{S}} &= \frac{1}{4}g(1+\ell)^2 - (1+\ell) + \frac{4 - (1-g)^2}{4g} \\ &= h_{1+\ell,2} \quad \text{for the dilute } \text{O}(n) \text{ model} \end{aligned} \quad (12.26)$$

in this case.

Alternatively, one may imagine producing the special ℓ -leg operator $\mathcal{O}_\ell^{\text{S}}$ by fusion of the ordinary ℓ -leg operator $\mathcal{O}_\ell^{\text{O}}$ and an ordinary-to-special boundary condition changing operator ϕ_{OS} . The scaling dimension (12.26) pertains to the insertion of this composite operator at either strip end. Comparing the Kac indices in (12.25) and (12.26), and using the CFT fusion rules (10.88), immediately leads to the identification $\phi_{\text{OS}} = \phi_{1,2}$. If one wants special boundary conditions on both the left and the right rim, two insertions of ϕ_{OS} are needed (to change from special to ordinary and back again). One would then expect $h_{1+\ell,3}$, as is indeed confirmed by the BA analysis [BY95, YB95].

Finally, the extraordinary transition is rather trivially related to the ordinary transition. Indeed, for $K_{\text{s}} = \infty$ the entire left rim of the annulus will be coated by a straight polymer strand, so that the remaining system (of width $L - 1$) effectively sees free boundary conditions—this is dubbed the *teflon effect* in [BC97]. Thus, for $\ell = 0$ the coating strand will be the left half of a long stretched-out loop, whose right half will act as a one-leg operator, and one effectively observes the exponent Δ_1^{O} . For $\ell > 0$, one of the legs will act as the coating strand, and one observes $\Delta_{\ell-1}^{\text{O}}$.

12.6.1 Physical interpretation

We have seen in section 11.4.2 that the 2-leg operator marks a point on a loop. Therefore the fractal dimension d_{b} of the points on the boundary

covered by loop segments is related to Δ'_2 . Since we are now in a boundary theory, the relation is

$$d_b^O = 1 - \Delta'_2. \quad (12.27)$$

At the ordinary transition we thus find from (12.24) that $d_b = 2(1 - g)$. But since $1 \leq g \leq 2$ in the dilute phase this implies formally $d_b \leq 0$. In other words, at the ordinary transition loops are repelled from the boundary in the continuum limit. This is consistent with the qualitative discussion of section 12.1, and in particular the use of Dirichlet boundary conditions $\phi|_{\mathbb{R}} = 0$ to describe the ordinary transition.

At the special transition

$$d_b^S = 1 - \Delta_2^S = \frac{7}{2} - \frac{3}{4g} - 2g. \quad (12.28)$$

Thus d_b decreases from $\frac{3}{4}$ to 0 as n decreases from 2 to 0 (and g increases from 1 to $\frac{3}{2}$). In other words, monomers are critically attracted towards the boundary, and d_b is a non-trivial number.

The limit $n \rightarrow 0$ of the dilute $O(n)$ model produces a self-avoiding walk. Its *bulk* fractal dimension is $d_h = 2 - \Delta_2 = \frac{4}{3}$, and since $d_b^S = 0$ it is insensitive to the special boundary condition.

The goal of this last part is to exhibit various applications, obtained by combining elements of the previous parts. There is a very large number of possibilities for such combinations. Accordingly I hope to develop this part over the years! Some ideas: critical exponents for the Potts model obtained from Wiener-Hopf calculations, indecomposability and logarithmic CFT, four-point functions from differential equations, crossing formulae in percolation, . . .

13 Exact CFT partition functions

By combining some key results of CFT (chapter 10) with the decomposition of the Markov trace (chapter 9) it is possible to construct the exact continuum-limit partition functions of the loop models defined on an annulus.

Consider the Potts loop model on an annulus of size $L \times M$. The periodic direction is that of size M . We recall that this model is defined by giving a weight $n = -2 \cos(\pi g)$ to each contractible loop, and (possibly different) weight $n_1 = 2 \cos(\pi e_0)$ to each non-contractible loop.

According to (12.5) we have

$$Z \equiv Z_{\text{ff}}(q) = \sum_h n_h \chi_{(c,h)}(q), \tag{13.1}$$

where the sum is over the boundary scaling dimensions h . Here $\chi_{(c,h)}(q)$ is the generic character (10.77). We recall that the modular parameter for the boundary theory is $q = \exp(i\pi\tau) = \exp(-\pi M/L)$. The degeneracy factor n_h states how many times a given character appears in the partition function, and as usual for non-minimal theories it needs not in general be an integer. We omit in the following the subscript ff which reminds us that the boundary conditions on both rims of the annulus are free.

The CFT partition function is then

$$Z[g, e_0] = \frac{q^{-c/24}}{P(q)} \sum_{\ell \in \mathbb{Z}} \frac{\sin((1 + \ell)\pi e_0)}{\sin(\pi e_0)} q^{\frac{g\ell^2}{4} - \frac{(1-g)\ell}{2}} \tag{13.2}$$

This expression was first obtained by Saleur and Bauer [SB89], using techniques of integrability and quantum groups. It was later rederived and discussed by Cardy [Ca06] from a Coulomb gas point of view. We hold by now all the necessary ingredients to prove this relation:

- The front factor $\frac{q^{-c/24}}{P(q)}$ is the usual contribution from the free boson, viz., the “gauged” height field $\tilde{\phi}$ of (12.8).
- The $q^{\Delta_\ell^O}$ factor codes the critical exponents (12.24) of the ℓ -leg (watermelon) operators at the ordinary surface transition.
- The degeneracy factor

$$n_\ell = U_\ell(\bar{n}/2) = \frac{\sin((1+\ell)\pi e_0)}{\sin(\pi e_0)} \quad (13.3)$$

comes from the algebraic decomposition (9.55) of the TL (Markov) trace over ordinary matrix traces.

- The sum $\sum_{\ell \in \mathbb{Z}}$ is over the number of non-contractible lines on the annulus.

The attentive reader may object that

1. Since the Kac labels (r, s) of $\Delta_\ell^O = h_{1,1+\ell}$ are integers, the expansion (13.1) should not be over generic characters $\chi_{(c,h)}(\tau)$ of (10.77) but over the irreducibles $K_{r,s}(\tau)$ of (10.90).
2. The sum in (13.2) should be over $\ell \in \mathbb{N}$ and not $\ell \in \mathbb{Z}$.

While these observations are certainly correct, a little analysis shows that taking both of them into account leads to exactly the same result (13.2).

13.1 A percolation crossing formula

The result (13.2) contains a lot of precious information in a very compact form. To illustrate the scope of this expression, we consider the limit $n \rightarrow 1$, which corresponds to bond percolation on the square lattice. In this case $c = 0$.

The partition function itself is $Z[g = \frac{2}{3}, e_0 = \frac{1}{3}]$. The part of (13.2) under the summation is

$$\sum_{\ell \in \mathbb{Z}} \frac{\sin((1+\ell)\pi/3)}{\sin(\pi/3)} q^{\frac{\ell^2}{6} - \frac{\ell}{6}}. \quad (13.4)$$

The contributions are only non-zero in the following cases

$$\begin{aligned} \ell = 6r & : q^{6r^2-r} \\ \ell = 6r - 2 & : -q^{6r^2-5r+1} \\ \ell = 6r + 1 & : q^{6r^2+r} \\ \ell = 6r + 3 & : -q^{6r^2+5r+1} \end{aligned}$$

Let us recall the Euler pentagonal number theorem:

$$P(q) = \prod_{k=1}^{\infty} (1 - q^k) = \sum_{k=-\infty}^{\infty} (-1)^k q^{k(3k-1)/2}. \quad (13.5)$$

A term with even $k = 2r$ reads q^{6r^2-r} , and a term with odd $k = 2r + 1$ reads q^{6r^2+5r+1} . Thus regrouping the contributions with $\ell = 6r$ and $\ell = 6r + 3$ the above sum is simply $P(q)$. One finds the same result by regrouping the contributions with $\ell = 6r - 2$ and $\ell = 6r + 1$.

So seemingly $Z \left[g = \frac{2}{3}, e_0 = \frac{1}{3} \right] = 2$. But taking into account that the equivalence between the TL loop model and the Potts model requires an even number of strands N —whence also ℓ is even—we have simply

$$Z \left[g = \frac{2}{3}, e_0 = \frac{1}{3} \right] = 1. \quad (13.6)$$

Consider now the probability p that a percolation cluster connects the two rims of the annulus. This happens if and only if there are no loops wrapping around the periodic direction. Such loops can be suppressed by setting $e_0 = \frac{1}{2}$. In view of the trivial normalisation ($Z = 1$) we have then

$$\begin{aligned} p &= Z \left[g = \frac{2}{3}, e_0 = \frac{1}{2} \right] \\ &= \frac{1}{P(q)} \sum_{\ell \in \mathbb{Z}} \sin \left((1 + \ell) \frac{\pi}{2} \right) q^{\frac{p(p-1)}{6}} \end{aligned}$$

The degeneracy factor is $+1$ if $\ell = 4r$ and -1 if $\ell = 4r + 2$. Thus

$$p = \frac{1}{P(q)} \sum_{r \in \mathbb{Z}} \left(q^{\frac{4r(4r-1)}{6}} - q^{\frac{(4r+2)(4r+1)}{6}} \right).$$

This can in turn be rewritten by using the Jacobi triple product formula

$$\sum_{k \in \mathbb{Z}} (-1)^k a^k q^{\frac{k(k-1)}{2}} = \prod_{k=1}^{\infty} (1 - aq^{k-1}) (1 - a^{-1}q^n) (1 - q^n) \quad (13.7)$$

in terms of the Dedekind function $\eta(\tau) = \frac{q^{1/24}}{P(q)}$ as

$$p = \frac{\eta(-\frac{1}{3\tau})\eta(-\frac{4}{3\tau})}{\eta(-\frac{1}{\tau})\eta(-\frac{2}{3\tau})} = \sqrt{\frac{3}{2}} \frac{\eta(3\tau)\eta(\frac{3\tau}{4})}{\eta(\tau)\eta(\frac{3\tau}{2})}. \quad (13.8)$$

For a thin annulus, $q = \exp(-\pi M/L) \rightarrow 0$, we have $1-p \sim q^{1/3}$. In terms of the conjugate modulus, $\tilde{q} = \exp(-2\pi L/M)$, a long cylinder corresponds to $\tilde{q} \rightarrow 0$. In that limit

$$p \sim \sqrt{\frac{3}{2}} \tilde{q}^{\frac{5}{48}}, \quad (13.9)$$

where we recognise the magnetic exponent of the $Q \rightarrow 1$ state Potts model. The result (13.8) can be seen as expressing all corrections to scaling for this well-known result.

References

- [Af86] I. Affleck, *Universal term in the free energy at a critical point and the conformal anomaly*, Phys. Rev. Lett. **56**, 746 (1986).
- [Al05] F. Alet, J.L. Jacobsen, G. Misguich, V. Pasquier, F. Mila and M. Troyer, *Interacting classical dimers on the square lattice*, Phys. Rev. Lett. **94**, 235702 (2005).
- [Al06] F. Alet, Y. Ikhlef, J.L. Jacobsen, G. Misguich and V. Pasquier, *Classical dimers with aligning interactions on the square lattice*, Phys. Rev. E **74**, 041124 (2006).
- [BC97] M.T. Batchelor and J. Cardy, *Extraordinary transition in the two-dimensional $O(n)$ model*, Nucl. Phys. B **506**, 553 (1997).
- [BS93] M.T. Batchelor and J. Suzuki, *Exact solution and surface critical behaviour of an $O(n)$ model on the honeycomb lattice*, J. Phys. A **26**, L729 (1993).
- [BY95] M.T. Batchelor and C.M. Yung, *Exact results for the adsorption of a flexible self-avoiding polymer chain in two dimensions*, Phys. Rev. Lett. **74**, 2026 (1995).
- [Ba72] R.J. Baxter, *Partition function of the eight-vertex lattice model*, Ann. Phys. **70**, 193–228 (1972).
- [Ba82a] R.J. Baxter, *Exactly solved models in statistical mechanics* (Academic Press, London, 1982).
- [Ba82b] R.J. Baxter, *Critical antiferromagnetic square-lattice Potts model*, Proc. Roy. Soc. London Ser. A **383**, 43 (1982).
- [BPZ84] A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov, *Infinite conformal symmetry in two-dimensional quantum field theory*, Nucl. Phys. B **241**, 333 (1984).
- [Be69] F.A. Berezin, *The plane Ising model*, Uspekhi Matem. Nauk **24**, 3–22 (1969) [English translation: Russ. Math. Surv. **24**, 1 (1969)].

- [Be31] H.A. Bethe, *Zur Theorie der Metalle I. Eigenwerte und Eigenfunktionen der linearen Atomkette*, Zeitschrift für Physik A **71**, 205–226 (1931).
- [Bi83] K. Binder, *Critical behaviour at surfaces*, in C. Domb and J.L. Lebowitz (eds.), *Phase transitions and critical phenomena* **8**, 1 (Academic Press, London, 1983).
- [BCN86] H.W.J. Blöte, J.L. Cardy and M.P. Nightingale, *Conformal invariance, the central charge, and universal finite-size amplitudes at criticality*, Phys. Rev. Lett. **56**, 742 (1986).
- [BE94] T.W. Burkhardt and E. Eisenriegler, *Conformal theory of the two-dimensional $O(N)$ model with ordinary, extraordinary, and special boundary conditions*, Nucl. Phys. B **424**, 487 (1994).
- [Ca84a] J.L. Cardy, *Conformal invariance and universality in finite-size scaling*, J. Phys. A **17**, L385 (1984).
- [Ca84b] J.L. Cardy, *Conformal invariance and surface critical behavior*, Nucl. Phys. B **240**, 514 (1984).
- [Ca86] J. Cardy, *Operator content of two-dimensional conformally invariant theories*, Nucl. Phys. B **270**, 186 (1986).
- [Ca92] J.L. Cardy, *Critical percolation in finite geometries*, J. Phys. A **25**, L201 (1992).
- [Ca05] J. Cardy, *Boundary conformal field theory*, in J.-P. Francoise, G. Naber and T.S. Tsun (eds.), *Encyclopedia of mathematical physics* (Elsevier, 2005).
- [Ca06] J. Cardy, *The $O(n)$ model on the annulus*, J. Stat. Phys. **125**, 1 (2006).
- [DSZ87a] P. Di Francesco, H. Saleur and J.B. Zuber, *Modular invariance in non-minimal two-dimensional conformal theories*, Nucl. Phys. B **285**, 454 (1987).
- [DSZ87b] P. Di Francesco, H. Saleur and J.B. Zuber, *Relations between the Coulomb gas picture and conformal invariance of two-dimensional critical models*, J. Stat. Phys. **49**, 57 (1987).

- [DMS87] P. Di Francesco, P. Mathieu and D. Sénéchal, *Conformal field theory* (Springer Verlag, New York, 1987).
- [DJS09] J. Dubail, J.L. Jacobsen and H. Saleur, *Conformal two-boundary loop model on the annulus*, Nucl. Phys. B **813**, 430–459 (2009).
- [DS86] B. Duplantier and H. Saleur, *Exact surface and wedge exponents for polymers in two dimensions*, Phys. Rev. Lett. **57**, 3179 (1986).
- [DS87] B. Duplantier and H. Saleur, *Exact critical properties of two-dimensional dense self-avoiding walks*, Nucl. Phys. B **290**, 291 (1987).
- [En80] I.G. Enting, *Generating functions for enumerating self-avoiding rings on the square lattice*, J. Phys. A: Math. Gen. **13**, 3713 (1980).
- [En96] I.G. Enting, *Series expansions from the finite lattice method*, Nucl. Phys. B (Proc. Suppl.) **47**, 180–187 (1996).
- [FZ87] V.A. Fateev and A.B. Zamolodchikov, *Conformal quantum field theory models in two dimensions having Z_3 symmetry*, Nucl. Phys. B **280**, 644 (1987).
- [Fi61] M.E. Fisher, *Statistical mechanics of dimers on a plane lattice*, Phys. Rev. **124**, 1664–1672 (1961).
- [FS63] M.E. Fisher and J. Stephenson, *Statistical Mechanics of dimers on a plane lattice: II. Dimer correlations and monomers*, Phys. Rev. **132**, 1411–1431 (1963).
- [FQS84] D. Friedan, Z. Qiu and S. Shenker, *Conformal invariance, unitarity and critical exponents in two dimensions*, Phys. Rev. Lett. **52**, 1575 (1984).
- [Ge79] P.-G. de Gennes, *Scaling concepts in polymer physics* (Cornell University Press, New York, 1979).
- [GP93] U. Grimm and P.A. Pearce, *Multi-colour braid-monoid algebras*, J. Phys. A **26**, 7435–7460 (1993).
- [Gu09] A.J. Guttmann (ed.), *Polygons, polyominoes and polycubes* (Springer, Heidelberg, 2009).

- [HR05] T. Halverson and A. Ram, *Partition algebras*, Eur. J. Comb. **26**, 869–921 (2005).
- [IJS08] Y. Ikhlef, J.L. Jacobsen and H. Saleur, *A staggered six-vertex model with non-compact continuum limit*, Nucl. Phys. B **789**, 483–524 (2008).
- [Is25] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift für Physik A **31**, 253–258 (1925).
- [IZ86] C. Itzykson and J.-B. Zuber, *Two-dimensional conformal invariant theories on a torus*, Nucl. Phys. B **275**, 580 (1986).
- [JS06] J.L. Jacobsen and H. Saleur, *The antiferromagnetic transition for the square-lattice Potts model*, Nucl. Phys. B **743**, 207 (2006).
- [JS08a] J.L. Jacobsen and H. Saleur, *Conformal boundary loop models*, Nucl. Phys. B **788**, 137–166 (2008).
- [JS08b] J.L. Jacobsen and H. Saleur, *Combinatorial aspects of conformal boundary loop models*, J. Stat. Mech. P01021 (2008).
- [Je03] I. Jensen, *A parallel algorithm for the enumeration of self-avoiding polygons on the square lattice*, J. Phys. A: Math. Gen. **36**, 5731 (2003).
- [JG99] I. Jensen and A.J. Guttmann, *Self-avoiding polygons on the square lattice*, J. Phys. A: Math. Gen. **32**, 4867 (1999).
- [Jo77] J.V. José, L.P. Kadanoff, S. Kirkpatrick and D.R. Nelson, *Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model*, Phys. Rev. B **16**, 1217 (1977).
- [Ka78] L.P. Kadanoff, *Lattice Coulomb gas representations of two-dimensional problems*, J. Phys. A **11**, 1399 (1978).
- [KB79] L.P. Kadanoff and A.C. Brown, *Correlation functions on the critical lines of the Baxter and Ashkin-Teller models*, Ann. Phys. **121**, 318 (1979).

- [Ka61] P.W. Kasteleyn, *The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice*, Physica **27**, 1209–1225 (1961).
- [Ka63] P.W. Kasteleyn, *Dimer statistics and phase transitions*, J. Math. Phys. **4**, 287–293 (1963).
- [Ke99] A.E. Kennelly, *Equivalence of triangles and stars in conducting networks*, Electrical World and Engineer **34**, 413-414 (1899).
- [KV92] P. Kleban and I. Vassileva, *Domain boundary energies in finite regions at 2D criticality via conformal field theory*, J. Phys. A: Math. Gen. **25** 5779 (1992).
- [Kn81] H.J.F. Knops, *Renormalization connection between the eight-vertex model and the Gaussian model*, Ann. Phys. **128**, 448 (1981).
- [Ko97] J. Kondev, *Liouville field theory of fluctuating loops*, Phys. Rev. Lett. **78**, 4320 (1997).
- [KH95] J. Kondev and C.L. Henley, *Four-coloring model on the square lattice: A critical ground state*, Phys. Rev. B **52**, 6628 (1995).
- [KW41] H.A. Kramers and G.H. Wannier, *Statistics of the two-dimensional ferromagnet. Part II*, Phys. Rev. **60**, 263-276 (1941).
- [Le20] W. Lenz, *Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körper*, Physikalische Zeitschrift **21**, 613–615 (1920).
- [LH89] J.L. Cardy, *Conformal invariance and statistical mechanics*, and P. Ginsparg, *Applied conformal field theory*, both in *Fields, strings and critical phenomena* (Les Houches, session XLIX), eds. E. Brézin and J. Zinn-Justin (Elsevier, New York, 1989).
- [Li67] E.H. Lieb, *Residual entropy of square ice*, Phys. Rev. **162**, 162-172 (1967).
- [LP75] A. Luther and I. Peschel, *Calculation of critical exponents in two dimensions from quantum field theory in one dimension*, Phys. Rev. B **12**, 3908 (1975).

- [Ma91] P. Martin, *Potts models and related problems in statistical mechanics* (World Scientific Publishing, Singapore, 1991).
- [MW73] B.M. McCoy and T.T. Wu, *The two-dimensional Ising model* (Harvard University Press, Cambridge MA, 1973).
- [MPW63] E.W. Montroll, R.B. Potts and J.C. Ward, *Correlations and spontaneous magnetization of the two-dimensional Ising model*, J. Math. Phys. **4**, 308-322 (1963).
- [Nh82] B. Nienhuis, *Exact critical point and critical exponents of $O(n)$ models in two dimensions*, Phys. Rev. Lett. **49**, 1062 (1982).
- [Nh84] B. Nienhuis, *Critical behavior of two-dimensional spin models and charge asymmetry in the Coulomb gas*, J. Stat. Phys. **34**, 731 (1984).
- [Nh87] B. Nienhuis, *Coulomb gas formulations of two-dimensional phase transitions*, in C. Domb and J.L. Lebowitz (eds.), *Phase transitions and critical phenomena* **11**, 1–53 (Academic Press, London, 1987).
- [Ni83] M. den Nijs, *Extended scaling relations for the magnetic critical exponents of the Potts model*, Phys. Rev. B **27**, 1674 (1983).
- [Ni84] M. den Nijs, *Extended scaling relations for the chiral and cubic crossover exponents*, J. Phys. A **17**, L295 (1984).
- [On44] L. Onsager, *Crystal statistics. I. A two-dimensional model with an order-disorder transition*, Phys. Rev. **65**, 117–149 (1944).
- [On49] L. Onsager, *Discussion*, Nuovo Cimento (suppl.) **6**, 261 (1949).
- [Pl88] V.N. Plechko, *Grassmann path-integral solution for a class of triangular type decorated Ising models*, Physica A **152**, 51–97 (1988).
- [Po70] A.M. Polyakov, *Conformal symmetry of critical fluctuations*, JETP Lett. **12**, 381 (1970).
- [RK88] D.S. Rokhsar and S.A. Kivelson, *Superconductivity and the quantum hard-core dimer gas*, Phys. Rev. Lett. **61**, 2376–2379 (1988).
- [RE59] G.S. Rushbrooke and J. Eve, *On noncrossing lattice polygons*, J. Chem. Phys. **31**, 1333–1334 (1959).

- [SB89] H. Saleur and M. Bauer, *On some relations between local height probabilities and conformal invariance*, Nucl. Phys. B **320**, 591 (1989).
- [SML64] T.D. Schultz, D.C. Mattis and E.H. Lieb, *Two-dimensional Ising model as a soluble problem of many fermions*, Rev. Mod. Phys. **36**, 856-871 (1964).
- [TF61] H.N.V. Temperley and M.E. Fisher, *Dimer problem in statistical mechanics—An exact result*, Phil. Mag. **6**, 1061–1063 (1961).
- [YY66] C.N. Yang and C.P. Yang, *One-dimensional chain of anisotropic spin-spin interactions. I. Proof of Bethe’s hypothesis for ground state in a finite system*, Phys. Rev. **150**, 321–27 (1966).
- [Wi69] K.G. Wilson, *Non Lagrangian models of current algebra*, Phys. Rev. **179**, 1499 (1969).
- [Wu06] F.Y. Wu, *Dimers on two-dimensional lattices*, Int. J. Mod. Phys. B **20**, 5357–5371 (2006).
- [Ya52] C.N. Yang, *The spontaneous magnetization of a two-dimensional Ising model*, Phys. Rev. **85**, 808-816 (1952).
- [YB95] C.M. Yung and M.T. Batchelor, *$O(n)$ model on the honeycomb lattice via reflection matrices: Surface critical behaviour*, Nucl. Phys. B **453**, 552 (1995).
- [Za86] A.B. Zamolodchikov, *Conformal symmetry and multicritical points in two-dimensional quantum field theory*, Sov. J. Nucl. Phys. **44**, 530 (1986).