Commun. math. Phys. 51, 15—40 (1976)

Communications in
Mathematical
Physics

Statistical Mechanics Models and the Modular Group

Fernando Lund*,**

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540, USA

Mario Rasetti*** and Tullio Regge†

Institute for Advanced Study, Princeton, New Jersey 08540, USA

Abstract. A lattice homogeneous under the modular group Γ of fractional linear transformations is constructed. The generating function for close packed dimer configurations on this infinite lattice is found directly, without doing it first for a finite lattice, using the Pfaffian method. This requires orienting the lattice. The group $SL(2, \mathbb{Z})$ is used to this end. Computation of the generating function is reduced to a particular case of the problem of finding the number of words reducible to the identity for a group which is the free product of two cyclic groups. Solution of this problem gives the dimer generating function as the solution of an algebraic equation. Considered as a function of the activities, the free energy has a logarithmic singularity.

Next an Ising model is built on the same lattice. The free energy per spin is evaluated by solving a dimer problem on an associated lattice following the general prescription of Fisher. It is a rational function of the solution of a system of two algebraic equations.

I. Introduction

In the solution of the dimer problem for rectangular lattices by the Pfaffian method [1, 2] use is made of the invariance of the lattice under a group of translations. In fact, the generating function in this case is given in terms of the determinant of a cyclic matrix which is easily diagonalized by means of a Fourier transformation. The question naturally arises then, whether it is possible to construct lattices which are homogeneous under a group other than the translations, say, a non-abelian one, and solve there the dimer problem. Since the Ising problem for a two-dimensional lattice is rather simply related to a dimer problem [3], one would also hope to find the Ising partition function for this new homogeneous lattice.

* Research sponsored in part by National Science Foundation Grant No. GP 30799 X to Princeton University

** Present address: Insitute for Advanced Study, Princeton, New Jersey 08540, USA

*** On leave from the Politecnico of Turin, Italy

† Research sponsored in part by National Science Foundation Grant No. 40768X

It is the purpose of this work to construct a lattice which is homogeneous under the modular group of fractional linear transformations and solve there the dimer and Ising problems by the Pfaffian method developed by Kasteleyn [1] and Fisher [3].

Consider then the Poincaré model of the hyperbolic plane [4]: it consists of the upper half $\mathbb{H}$ of the complex plane $\mathbb{C}$. Its geodesics are arcs of circles and straight lines orthogonal to the real axis. If we let $z \in \mathbb{C}$, $z = x + iy$, then the metric of this plane is

$$ds^2 = \frac{dx^2 + dy^2}{y^2}. \tag{1.1}$$

Notice that with this metric, any point on the real axis is infinitely far away from any point on the upper half plane. The fractional linear transformations

$$z \to \frac{az+b}{cz+d}; \qquad ad - bc = 1; \qquad a, b, c, d \in \mathbb{R} \tag{1.2}$$

are orientation preserving isometries of this geometry, and it is easy to see that they form a group. The modular group Γ is the subgroup of (1.2) in which a, b, c, d are integers. It is a discrete subgroup. As such, it has a fundamental domain $\mathbb{D}$ which is known to be

$$\mathbb{D} = \{z \in \mathbb{H} \,|\, \mathrm{Re}\, z < \tfrac{1}{2}, |z| > 1\}. \tag{1.3}$$

Also, Γ is isomorphic to $\mathrm{SL}(2, \mathbb{Z})/\{\pm \mathbb{1}\}$, the group of 2×2 matrices with integer coefficients and unit determinant divided by its center. In fact, $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ determine the same fractional transformation. When no confusion arises, we shall denote by the same symbol the elements of Γ and $\mathrm{SL}(2, \mathbb{Z})$. Let now

$$S: z \to -\frac{1}{z},$$

$$T: z \to z + 1. \tag{1.4}$$

These transformations generate Γ. They satisfy the identities $S^2 = (TS)^3 = \mathbb{I}$, which are essentially unique, as all others are a consequence of them. More precisely, a presentation of Γ is given by $\langle S, T; S^2, (TS)^3 \rangle$. In other words, Γ is the free product of the cyclic group of order 2 generated by S and the cyclic group of order 3 generated by TS.

The transformation S leaves $i = \sqrt{-1}$ fixed, TS leaves $\varrho = e^{\frac{i\pi}{3}}$ fixed and ST leaves ϱ^2 fixed. These are the only points of $\mathbb{D}$ which are left fixed by elements of Γ.

By acting on $\mathbb{D}$ with Γ one obtains a tesselation [5] of the hyperbolic plane. It is here that we construct a lattice as follows: Take an arbitrary point inside $\mathbb{D}$, call it $\boldsymbol{a}$, and act on it with $g \in \Gamma$. That is, consider all its images under the modular group. Then connect with bonds the points $g(\boldsymbol{a})$ with $gS(\boldsymbol{a})$, $gT(\boldsymbol{a})$ and $gT^{-1}(\boldsymbol{a})$ and associate an activity x with the T and T^{-1} bonds, and ℓ with the S bonds. In this way the lattice of Figure 1 is formed. This may be described in an alternative way: it is well known that the upper half plane may be conformally mapped inside the unit circle and that if we consider straight lines as limiting cases of circles, this map is such that circles go into circles. In particular, the new geodesics are

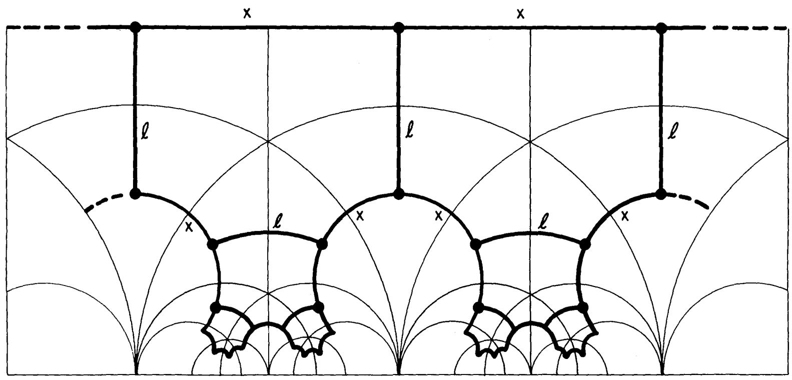


Fig. 1. Lattice homogeneous under the modular group, drawn on the upper half plane

circular arcs orthogonal to the unit circle. In this way, another description of the hyperbolic plane is obtained. The tesselation induced by the modular group as well as the lattice we have just constructed are illustrated for this model in Figure 2. We see that the lattice consists of hexagons, each one connected to three others along alternating faces in such a way as to form an infinite tree.

By construction, this lattice is homogeneous under Γ. Its generating function per site for close packed dimer configurations will be computed directly, without finding it first for a finite lattice and taking the thermodynamic limit afterwards. This is done as follows: according to Kasteleyn [1], the dimer generating function Z_s for a lattice with N vertices is given by the determinant of a certain operator A associated with a prescribed orientation of the lattice:

$$\log Z_s = \frac{1}{2N} \log \det A = \frac{1}{2N} \operatorname{tr} \log A . \tag{1.5}$$

Now, A acts on a vector space of dimension N, so that

$$\log Z_s = \tfrac{1}{2}(\operatorname{tr} \mathbb{I})^{-1} \operatorname{tr} \log A , \tag{1.6}$$

where $\mathbb{I}$ is the identity. This expression then holds for arbitrary N, and it is possible to take it as the *definition* of $\log Z_s$ in the thermodynamic limit, when the vector space becomes infinite dimensional. That this definition makes sense is shown by actually performing calculations with it. In particular, it has been used [6] to recover the generating function of Kasteleyn [1] and the free energy first found by Onsager [7] for the Ising model on a rectangular lattice.

In what follows we shall omit the factor $(\operatorname{tr}\mathbb{I})^{-1}$ from our expressions, so that one should read $(\operatorname{tr}\mathbb{I})^{-1} \operatorname{tr} A$ whenever $\operatorname{tr} A$ appears. In particular, the trace of the identity will be one.

In Section II the lattice is oriented according to Kasteleyn's rules. In so doing the invariance under Γ is lost. However, it is possible to retain invariance under the larger group $\mathrm{SL}(2, \mathbb{Z})$. The operator whose determinant gives the generating function is expressed in terms of the regular representation of certain elements of this group. Use of (1.5) reduces the problem then to the combinatorial question of counting a certain class of elements in $\mathrm{SL}(2, \mathbb{Z})$, which is a particular case of the

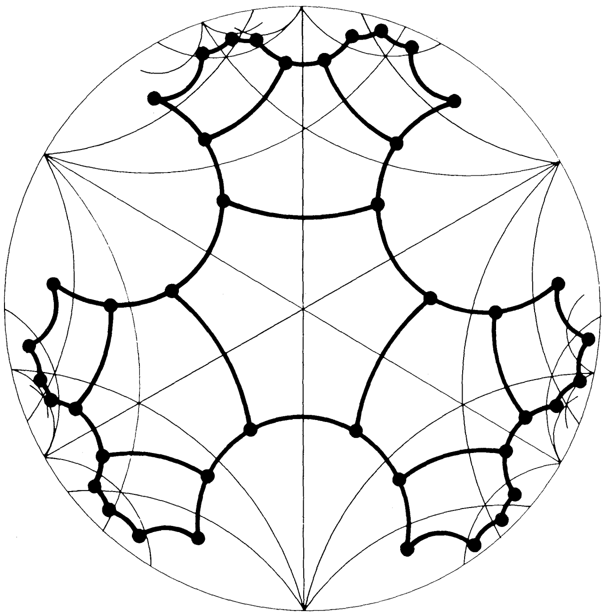


Fig. 2. Lattice homogeneous under the modular group, drawn inside the unit circle

problem [8] of finding the number of words reducible to the identity for a group which is the free product of two cyclic groups. This is solved by means of generating functions and gives $\log Z_s$ as the integral of a solution of an algebraic equation.

The Ising problem for the lattice is considered in Section III. The corresponding dimer counting lattice is constructed following Fisher [3]. After this, the steps of II are repeated. They become more complicated, especially the counting problem. The free energy turns out to be a rational function of the solution of a system of two algebraic equations. Section IV contains some final remarks.

II. Dimer Generating Function

The first step is to orient the lattice such that every hexagon has odd parity (i.e., such that the number of bonds oriented in either direction is odd). A "fundamental hexagon" is oriented first and then the rest. Take then the original point ($\boldsymbol{a}$) from which the lattice was constructed and the points $T(\boldsymbol{a})$, $TS(\boldsymbol{a})$, $TST(\boldsymbol{a})$, $(TS)^2(\boldsymbol{a})$, $(TS)^2T(\boldsymbol{a})$. They are the vertices of the fundamental hexagon. Draw arrows now in the following way:

$$\boldsymbol{a} \to T(\boldsymbol{a}) \to TS(\boldsymbol{a}) \to TST(\boldsymbol{a}) \to (TS)^2(\boldsymbol{a}) \to (TS)^2T(\boldsymbol{a}) \leftarrow (TS)^3(\boldsymbol{a}) = \boldsymbol{a} .$$

Clearly this fundamental hexagon has odd parity. Draw also an arrow from $T^{-1}(\boldsymbol{a})$ to $\boldsymbol{a}$. This orientation, however, breaks the symmetry of the lattice, as it is incompatible with the identity $S^2 = \mathbb{I}$.

We consider then two vector spaces: one spanned by vectors $\boldsymbol{a}_g$, $g \in \Gamma$, which are in one-to-one correspondence with the points $g(\boldsymbol{a})$, and the other spanned by $\boldsymbol{a}_h$, $h \in \mathrm{SL}(2, \mathbb{Z})$, in which the labels are 2×2 matrices. The bonds incident on the point $\boldsymbol{a}$ define the action of an operator $O^{\sim}$ on $\boldsymbol{a}_{\mathrm{II}}$.

$$O^{\sim}\boldsymbol{a}_{\mathrm{II}}=\ell\boldsymbol{a}_S+x\boldsymbol{a}_T-x\boldsymbol{a}_{T^{-1}}\,. \tag{2.1}$$

We want to know what is $O^{\sim}\boldsymbol{a}_g$, $g\in\Gamma$. The first guess, namely $O^{\sim}\boldsymbol{a}_g=\ell\boldsymbol{a}_{gS}+x\boldsymbol{a}_{gT}-x\boldsymbol{a}_{gT^{-1}}$ is wrong, as it is inconsistent with $S^2=\mathbb{I}$. We can define however an operator O acting on $\boldsymbol{a}_h$, $h\in\mathrm{SL}(2,\mathbb{Z})$, in a similar way without contradiction:

$$O\boldsymbol{a}_h=\ell\boldsymbol{a}_{h\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)}+x\boldsymbol{a}_{h\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)}-x\boldsymbol{a}_{h\left(\begin{smallmatrix}1&-1\\0&1\end{smallmatrix}\right)}\,. \tag{2.2}$$

Now, to each $g\in\Gamma$ there correspond two elements in $\mathrm{SL}(2,\mathbb{Z})$ say $\pm\begin{pmatrix}g_{11}&g_{12}\\g_{21}&g_{22}\end{pmatrix}$. There are then infinitely many injections of Γ into $\mathrm{SL}(2,\mathbb{Z})$. Consider a particular one, say the one that associates to each g a 2×2 matrix such that $g_{11}>0$, and when $g_{11}=0$, then $g_{21}>0$. Notice that g_{11} and g_{21} cannot vanish simultaneously. Such a matrix will be called *positive*. A non-positive matrix will be called *negative*.

The operator (2.2) joins vectors labeled both by positive and negative matrices. If $\boldsymbol{a}_h$ is a vector labeled by a negative matrix h, we transform it into a vector labeled by the positive matrix $-h$ in the following way:

$$\boldsymbol{a}_h=-\boldsymbol{a}_{-h},\ h\in\mathrm{SL}(2,\mathbb{Z})\,. \tag{2.3}$$

We thus obtain an operator which acts on the subspace spanned by basis vectors labeled by positive matrices only. As these vectors are in one-to-one correspondence with the $\boldsymbol{a}_g$, $g\in\Gamma$, we *define* this operator to be $O^{\sim}$. In this way we obtain an orientation of the modular lattice. This is seen as follows: Take the point $g(\boldsymbol{a})$, $g\in\Gamma$. Associate with g the positive matrix $\begin{pmatrix}g_{11}&g_{12}\\g_{21}&g_{22}\end{pmatrix}$. Take next the product $\begin{pmatrix}g_{11}&g_{12}\\g_{21}&g_{22}\end{pmatrix}\begin{pmatrix}0&-1\\1&0\end{pmatrix}=\begin{pmatrix}g_{12}&-g_{11}\\g_{22}&-g_{21}\end{pmatrix}$. If this product is positive, we draw an arrow from $gS(\boldsymbol{a})$ to $g(\boldsymbol{a})$.

Next, take $\begin{pmatrix}g_{11}&g_{12}\\g_{21}&g_{22}\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}=\begin{pmatrix}g_{11}&g_{11}+g_{12}\\g_{21}&g_{21}+g_{22}\end{pmatrix}$. The signs of g_{11} and g_{21} are preserved. The rule then is to draw an arrow from $g(\boldsymbol{a})$ to $gT(\boldsymbol{a})$. Similarly, an arrow must be drawn from $gT^{-1}(\boldsymbol{a})$ to $g(\boldsymbol{a})$.

First, we show that this rule does give the orientation we have imposed on the fundamental hexagon.

i) $g(\boldsymbol{a})=\boldsymbol{a}$, the matrix associated with g is $\begin{pmatrix}1&0\\0&1\end{pmatrix}$.

Since $\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}0&-1\\1&0\end{pmatrix}=\begin{pmatrix}0&-1\\1&0\end{pmatrix}$, there is an arrow from $\boldsymbol{a}$ to $S(\boldsymbol{a})=(TS)^2T(\boldsymbol{a})$, as required. An arrow also goes to $T(\boldsymbol{a})$.

ii) $g(\boldsymbol{a})=T(\boldsymbol{a})$, $T\to\begin{pmatrix}1&1\\0&1\end{pmatrix}$.

Since $\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}0&-1\\1&0\end{pmatrix}=\begin{pmatrix}1&-1\\1&0\end{pmatrix}$, an arrow goes from $T(\boldsymbol{a})$ to $TS(\boldsymbol{a})$. An arrow from $TT^{-1}(\boldsymbol{a})=\boldsymbol{a}$ to $T(\boldsymbol{a})$ is also required, and this is consistent with the prescription of i).

One sees similarly that the orientation of the remaining four bonds is correctly given.

Now we show that every hexagon $g(\boldsymbol{a})$, $gT(\boldsymbol{a})$, $gTS(\boldsymbol{a})$, $gTST(\boldsymbol{a})$, $gST^{-1}(\boldsymbol{a})$, $gS(\boldsymbol{a})$ has odd parity. There are arrows from $g(\boldsymbol{a})$ to $gT(\boldsymbol{a})$, from $gTS(\boldsymbol{a})$ to $gTST(\boldsymbol{a})$ and from $gST^{-1}(\boldsymbol{a})$ to $gS(\boldsymbol{a})$. Suppose now that $\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is positive. In this case an arrow goes from $g(\boldsymbol{a})$ to $gS(\boldsymbol{a})$. To gST^{-1} we associate the matrix

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} g_{12} & -g_{11}-g_{12} \\ g_{22} & -g_{21}-g_{22} \end{pmatrix}.$$

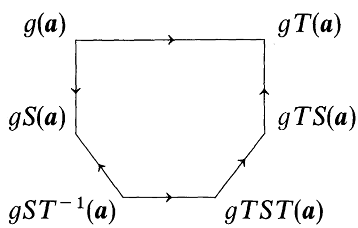
There are three possibilities: g_{12} larger, equal to, or smaller than zero. Take $g_{12}>0$ first. Now,

$$\begin{pmatrix} g_{12} & -g_{11}-g_{12} \\ g_{22} & -g_{21}-g_{22} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}=\begin{pmatrix} -g_{11}-g_{12} & -g_{12} \\ -g_{21}-g_{22} & -g_{22} \end{pmatrix}.$$

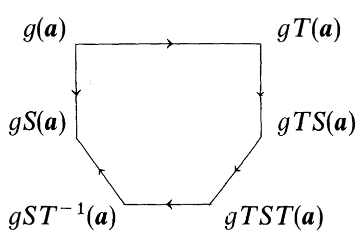
On the other hand, to gTS there corresponds

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}=\begin{pmatrix} g_{11}+g_{12} & -g_{11} \\ g_{21}+g_{22} & -g_{21} \end{pmatrix}.$$

If $-g_{11}-g_{12}>0$, an arrow goes from $gST^{-1}(\boldsymbol{a})$ to $gST^{-1}S(\boldsymbol{a})$. But this means $g_{11}+g_{12}<0$, and an arrow goes then from $gTS(\boldsymbol{a})$ to $gT(\boldsymbol{a})$, and the hexagon is odd:



If $g_{11}+g_{12}>0$, the arrows go from $gT(\boldsymbol{a})$ to $gTS(\boldsymbol{a})$ and from $gST^{-1}S(\boldsymbol{a})$ to $gST^{-1}(\boldsymbol{a})$ and the hexagon again is odd:



If $g_{11}+g_{12}=0$, we have nevertheless that $(gST^{-1}S)_{21}=-(gTS)_{21}$ and the same considerations hold.

The cases $g_{12}<0$ and $g_{12}=0$ are similarly treated, again obtaining odd hexagons.

We have then oriented the lattice using (2.2) and (2.3) and a particular injection of Γ into $\mathrm{SL}(2,\mathbb{Z})$. The question arises whether different recipes could also give good orientations while leading to different results. The only possibility for this to happen is to have, instead of O, an operator O' in which the relative sign of the activities is changed. An investigation of this possibility [6] leads to the conclusion that the induced orientation of the lattice gives hexagons of even parity and is thus inadmissable.

We must then compute the determinant of the operator resulting from (2.2) with the constraint (2.3). The operator O may be written as

$$O=\ell\mathscr{R}\begin{pmatrix}0&-1\\1&0\end{pmatrix}+x\mathscr{R}\begin{pmatrix}1&-1\\0&1\end{pmatrix}-x\mathscr{R}\begin{pmatrix}1&-1\\0&1\end{pmatrix}\tag{2.4}$$

where $\mathscr{R}$ is the regular representation of $\mathrm{SL}(2,\mathbb{Z})$.

Notice that if instead of (2.3) we had imposed the constraint $\boldsymbol{a}_h=\boldsymbol{a}_{-h}$, we would obtain exactly the regular representation for Γ. In fact, this constraint is saying that we identify h and $-h$, which is precisely what one does when dividing $\mathrm{SL}(2,\mathbb{Z})$ by $\{\pm\mathbb{1}\}$, and in particular $\mathscr{R}(-\mathbb{I})$ becomes the identity. Constraint (2.3) says that instead of this, we want $\mathscr{R}(-\mathbb{I})$ to become *minus* the identity.

More precisely, if we call $\mathscr{E}$ the vector space for the regular representation of $\mathrm{SL}(2,\mathbb{Z})$, then the regular representation of Γ is given by $\mathscr{E}/\ker\mathscr{A}$, where $\mathscr{A}=\mathscr{R}(\mathbb{I})+\mathscr{R}(-\mathbb{I})$. In fact, $(\boldsymbol{a}_{-h}-\boldsymbol{a}_h)\in\ker\mathscr{A}$, which means that, in $\mathscr{E}/\ker\mathscr{A}$, $\boldsymbol{a}_h$ and $\boldsymbol{a}_{-h}$ are identified. Analogously, we construct $\mathscr{E}/\ker\mathscr{B}$, where $\mathscr{B}=\mathscr{R}(\mathbb{I})-\mathscr{R}(-\mathbb{I})$. This time, one has $(\boldsymbol{a}_h+\boldsymbol{a}_{-h})\in\ker\mathscr{B}$ so that $-\boldsymbol{a}_h$ and $\boldsymbol{a}_{-h}$ are identified as required by (2.3). Consequently, $O^{\sim}$ is the endomorphism induced in $\mathscr{E}/\ker\mathscr{B}$ by O.

We now know that the generating function Z_s we are looking for is given by

$$2\log Z_s=\log\det O^{\sim}=\operatorname{tr}\log O^{\sim}.\tag{2.5}$$

We shall want then to compute traces of operators acting on $\mathscr{E}/\ker\mathscr{B}$ induced by the regular representation of some element of $\mathrm{SL}(2,\mathbb{Z})$ which acts on $\mathscr{E}$. We know that $\operatorname{tr}\mathscr{R}(\mathbb{I})=1$, $\operatorname{tr}\mathscr{R}(h)=0$, $h\neq\mathbb{I}\in\mathrm{SL}(2,\mathbb{Z})$. Let $h^{\sim}$ be the operator induced by $\mathscr{R}(h)$. What is $\operatorname{tr}h^{\sim}$? The answer is

$$\operatorname{tr}(h^{\sim})=\begin{cases}+1 & \text{if}\quad h=\mathbb{I}\\ -1 & \text{if}\quad h=-\mathbb{I}\\ 0 & \text{otherwise}.\end{cases}\tag{2.6}$$

That $\operatorname{tr}(\mathbb{I}^{\sim})=1$ is clear since the identity induces the identity. Also, $(-\mathbb{I})^{\sim}=-(\mathbb{I}^{\sim})$, and more generally, $(-h)^{\sim}=-(h^{\sim})$. This is just another way of saying that $\mathscr{R}(-\mathbb{I})\boldsymbol{a}_h=\boldsymbol{a}_{-h}$ and $-\boldsymbol{a}_h$ must be identified. Only $\pm\mathbb{I}$ induces operators proportional to the identity in $\mathscr{E}/\ker\mathscr{B}$. In fact, suppose $k^{\sim}=\alpha\mathbb{1}$, $k\neq\pm\mathbb{I}$, α a scalar and $\mathbb{1}$ the identity of $\mathscr{E}/\ker\mathscr{B}$. In this case we should identify $\boldsymbol{a}_k$ and $\alpha\boldsymbol{a}_{\mathbb{I}}$. But $\boldsymbol{a}_{\mathbb{I}}$ is identified *only* with $-\boldsymbol{a}_{-\mathbb{I}}$ (and with itself, of course). Consequently $\alpha=-1$ and $k=-\mathbb{I}$ or $\alpha=1$ and $k=\mathbb{I}$.

We are finally in a position to compute (2.5):

$$2\log Z_s=\operatorname{tr}\log(\ell S^{\sim}+xT^{\sim}-xT^{\sim-1}),\tag{2.7}$$

where we abuse somewhat the language by calling $\mathscr{R}\begin{pmatrix}0&-1\\1&0\end{pmatrix}^{\sim}=S^{\sim}$, $\mathscr{R}\begin{pmatrix}1&1\\0&1\end{pmatrix}^{\sim}=T^{\sim}$ and $\mathscr{R}\begin{pmatrix}1&-1\\0&1\end{pmatrix}^{\sim}=T^{\sim-1}$. Factoring out the $\ell S^{\sim}$ term and developing the logarithm in a power series, we obtain

$$2\log Z_s = \operatorname{tr}\log \ell S^{\sim} + \sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{x}{\ell}\right)^k \operatorname{tr}(S^{\sim}T^{\sim} - S^{\sim}T^{\sim-1})^k . \tag{2.8}$$

The problem is now to find how many words of length k in the two letters $S^{\sim}T^{\sim}$ and $-S^{\sim}T^{\sim-1}$ can be reduced to the identity, using $(-S^{\sim}T^{\sim})^3 = +\mathbb{I} = (S^{\sim}T^{\sim-1})^3$. These two letters generate cyclic groups of order three, and words in them are elements of a group which is the free product of the two groups. We consider then the slightly more general case of a group which is the free product of two cyclic groups of order n and m. Its presentation is $\langle R, U; R^n, U^m\rangle$. What we shall do is to compute the generating function

$$H(x, y) = \sum_{p,q} N(p, q)x^p y^q , \tag{2.9}$$

where $N(p, q)$ is the number of words in R and U only (that is R^{-1} and U^{-1} do not appear) with R appearing p times and U q times which are reducible to the identity. Notice that

$$\operatorname{tr}(R+U)^k = \sum_{p+q=k} N(p, q) . \tag{2.10}$$

At this point we must introduce some definitions:

a) The *product* of two words $\mathscr{W}_1$ and $\mathscr{W}_2$ is the word $\mathscr{W}_1\mathscr{W}_2$.

b) A word reducible to the identity will be called simply an identity word.

c) An identity word which cannot be factored as the product of two identity words will be called a *monoblock*. If it can be decomposed as the product of two monoblocks, but not of three, it will be called a *diblock*. Similarly we shall have *3-blocks* and in general *n-blocks*. We shall also refer to identity words which are not monoblocks as *multiblocks*.

d) A *UR-word* is a word whose first letter is U and its last letter is R. We have then UR-, RU-, RR-, and UU-words.

e) $N_k^{UR}(p, q)$ is the number of UR-k-blocks with R appearing p times and U appearing q times.

f) $N^{UR}(p, q)$ is the number of UR-identity words with R appearing p times and U appearing q times.

A word in R and U is reduced to the identity [8] by deleting the letters R^n and U^m. In order to find all the identity words of a given length we shall employ a constructive procedure inverse to the reduction: we start with the empty word, that is, 1. Then we insert repeatedly R^n and U^m until we get to the desired length. It is clear that in fact all the identity words can be constructed following exactly the reduction steps in reverse order. We want to count how many *different* words we can obtain. Notice, for instance, that R^{2n} counts as only one word, although it can be obtained by inserting R^n into R^n in $(n+1)$ ways.

First, we establish the following

Lemma 2.1. *There are no UR or RU monoblocks.*

Proof. Since $R^n = 1$ and $U^m = 1$, the minimum length of an identity word where both R and U appear is $m+n$, and the only UR-identity word of this length is U^mR^n which is obviously not a monoblock. Larger UR-words are formed by inserting identity words in this basic UR-word, so that they have the form

$$U\mathscr{W}_1U\mathscr{W}_2U\ldots\mathscr{W}_{m-1}U\mathscr{W}_mR\mathscr{W}_{m+1}R\ldots\mathscr{W}_{m+n-1}R ,$$

which is clearly the product of the three identity words $(U\mathscr{W}_1 \dots \mathscr{W}_{m-1} U)$, $(\mathscr{W}_m)$, and $(R\mathscr{W}_{m+1} \dots R)$. Adjoining identity words to the right or left of $U^m R^n$ makes things even worse as there is one more factor in the product. Thus, there are no UR-monoblocks. The same reasoning holds when R and U are interchanged which shows that there are no RU-monoblocks and the proof is complete.

Consider now the most general UU-monoblock. It has the form $U\mathscr{W}_1 U \dots U\mathscr{W}_{m-1} U$, where the $\mathscr{W}_i$'s are identity words which may, in particular, be the empty word. Then the following property holds:

Lemma 2.2. *If $\mathscr{M} = U\mathscr{W}_1 U \dots U\mathscr{W}_{m-1} U$ is a UU-monoblock, then the $\mathscr{W}_i$, $i = 1, 2, \dots, m-1$ are products of RR-monoblocks only. Similarly, if $R\mathscr{W}_1 R \dots R\mathscr{W}_{n-1} R$ is a RR-monoblock, then the $\mathscr{W}_i$, $i = 1, \dots, n-1$ are products of UU-monoblocks only.*

Proof. It is clear that any identity word is the product of monoblocks. Also, we have seen there are no UR- or RU-monoblocks. Suppose then that one of the words inside the UU-monoblock, say $\mathscr{W}_j$, contains as a factor another UU-monoblock: $W_j = \mathscr{A}(U\mathscr{B}_1 U \dots U\mathscr{B}_{m-1} U)\mathscr{C}$, where $\mathscr{A}$, $\mathscr{C}$ and $\mathscr{B}_i$ $1 \leqq i \leqq m-1$ are identity words. Substituting into $\mathscr{M}$ we have

$$\begin{aligned}\mathscr{M} &= U\mathscr{W}_1 U \dots U\mathscr{W}_{j-1} U\mathscr{A} U\mathscr{B}_1 U \dots U\mathscr{B}_{m-1} U\mathscr{C} U\mathscr{W}_{j+1} U \dots U\mathscr{W}_{m-1} U \\ &= \mathscr{M}_1 \mathscr{B}_{m-j} \mathscr{M}_2\end{aligned}$$

where

$$\mathscr{M}_1 = U\mathscr{W}_1 U \dots U\mathscr{W}_{j-1} U\mathscr{A} U\mathscr{B}_1 U \dots U\mathscr{B}_{m-j-1} U\,,$$

and

$$\mathscr{M}_2 = U\mathscr{B}_{m-j+1} U \dots U\mathscr{B}_{m-1} U\mathscr{C} U\mathscr{W}_{j+1} U \dots U\mathscr{W}_{m-1} U$$

are identity words, contradicting the fact that $\mathscr{M}$ is a monoblock. Interchanging U with R and m with n proves the lemma for RR-monoblocks, completing the proof.

As a consequence of this lemma we have

$$N_1^{UU}(p,q) = \sum_{\substack{p_1 + \dots + p_{m-1} = 1 \\ q_1 + \dots + q_{m-1} = q - m}} \mathscr{N}^{RR}(p_1, q_1)\mathscr{N}^{RR}(p_2, q_2) \dots \mathscr{N}^{RR}(p_{m-1}, q_{m-1}) \tag{2.11}$$

and

$$N_1^{RR}(p,q) = \sum_{\substack{p_1 + \dots + p_{n-1} = p - n \\ q_1 + \dots + q_{n-1} = q}} \mathscr{N}^{UU}(p_1, q_1)\mathscr{N}^{UU}(p_2, q_2) \dots \mathscr{N}^{UU}(p_{n-1}, q_{n-1})\,, \tag{2.12}$$

where $\mathscr{N}^{RR}(p,q)$ is the number of RR-identity words which are the product of RR-monoblocks only, and in which R appears p times and U appears q times. If we call $\mathscr{N}_j^{UU}(p,q)$ the corresponding number of UU-identity words which are the product of a number j of UU-monoblocks (notice $\mathscr{N}_1 = N_1$) we have

$$\mathscr{N}_j^{UU}(p,q) = \sum_{\substack{p_1 + \dots + p_j = p \\ q_1 + \dots + q_j = q}} N_1^{UU}(p_1, q_1) N_1^{UU}(p_2, q_2) \dots N_1^{UU}(p_j, q_j)\,, \tag{2.13}$$

$$\mathscr{N}_j^{RR}(p,q) = \sum_{\substack{p_1 + \dots + p_j = p \\ q_1 + \dots + q_j = q}} N_1^{RR}(p_1, q_1) N_1^{RR}(p_2, q_2) \dots N_1^{RR}(p_j, q_j)\,. \tag{2.14}$$

Notice how (2.13) says in particular that there is only one word U^{3k}. In fact, $N_1^{UU}(p,q) = 0$ except for $p = 0$, $q = 3$. In this case, $N_1^{UU}(0,3) = 1$, namely U^3.

We also have

$$\mathcal{N}^{UU}(p,q)=\sum_{j=1}^{\infty}\mathcal{N}_j^{UU}(p,q)\,, \tag{2.15}$$

$$\mathcal{N}^{RR}(p,q)=\sum_{j=1}^{\infty}\mathcal{N}_j^{RR}(p,q)\,. \tag{2.16}$$

We see that (2.11)–(2.14) give recursive formulae for N_1^{UU} and N_1^{RR}. Make now the following definitions:

$$f_j(x,y)=\sum_{p,q}\mathcal{N}_j^{RR}(p,q)x^py^q\,, \tag{2.17}$$

$$g_j(x,y)=\sum_{p,q}\mathcal{N}_j^{UU}(p,q)x^py^q\,, \tag{2.18}$$

$$f(x,y)=\sum_{p,q}\mathcal{N}^{RR}(p,q)x^py^q\,, \tag{2.19}$$

$$g(x,y)=\sum_{p,q}\mathcal{N}^{UU}(p,q)x^py^q\,. \tag{2.20}$$

Then, multiplying (2.11) and (2.12) by x^py^q and summing over p and q we get

$$g_1(x,y)=y^mf^{m-1}(x,y)\,, \tag{2.21}$$

$$f_1(x,y)=x^ng^{n-1}(x,y)\,. \tag{2.22}$$

Similarly, from (2.15) and (2.16) we get

$$f(x,y)=\sum_{j=0}^{\infty}f_j(x,y)\,, \tag{2.23}$$

$$g(x,y)=\sum_{j=0}^{\infty}g_j(x,y)\,, \tag{2.24}$$

and from (2.13) and (2.14):

$$f_j(x,y)=f_1^j(x,y)\,, \tag{2.25}$$

$$g_j(x,y)=g_1^j(x,y)\,. \tag{2.26}$$

Substituting (2.25) into (2.23) and then in (2.21) we have

$$g_1(x,y)=y^m(1-f_1(x,y))^{1-m}\,. \tag{2.27}$$

Similarly, from (2.26), (2.24), and (2.22) we get

$$f_1(x,y)=x^n(1-g_1(x,y))^{1-n}\,. \tag{2.28}$$

Expressions (2.27) and (2.28) give the generating functions for UU and RR monoblocks.

Consider now the total number of identity words with R appearing p times and U q times $N(p,q)$ and the corresponding total number of k-blocks $N_k(p,q)$. One has

$$N(p,q)=\sum_{k=1}^{\infty}N_k(p,q) \tag{2.29}$$

and

$$N_k(p,q)=\sum_{\substack{p_1+\ldots+p_k=p\\ q_1+\ldots+q_k=q}} N_1(p_1,q_1)\ldots N_1(p_k,q_k)\,. \tag{2.30}$$

Since there are no UR- or RU-monoblocks,

$$N_1(p,q)=N_1^{RR}(p,q)+N_1^{UU}(p,q)\,. \tag{2.31}$$

If we call

$$H_k(x,y)=\sum_{p,q} N_k(p,q)x^p y^q\,, \tag{2.32}$$

we see that (2.31) implies

$$H_1(x,y)=f_1(x,y)+g_1(x,y)\,. \tag{2.33}$$

Also, (2.30) implies

$$H_k(x,y)=H_1^k(x,y)\,, \tag{2.34}$$

and (2.29) implies that the generating function (2.9) is then given by

$$H(x,y)=\sum_k H_k(x,y)\,. \tag{2.35}$$

Substituting (2.33) into (2.34) and into (2.35) we finally get

$$H(x,y)=\frac{f_1(x,y)+g_1(x,y)}{1-(f_1(x,y)+g_1(x,y))}\,. \tag{2.36}$$

That is, the generating function for words in R and U is given in terms of the solution of two algebraic equations, namely, (2.27) and (2.28).

Going back now to (2.8)–(2.10), we see that our dimer generating function is given by

$$2\log Z_s=\operatorname{tr}\log \ell S^{\sim}+\int^z \frac{dz'}{z'}\,G(z'),\qquad z=-x\ell^{-1}\,, \tag{2.37}$$

where $G(z)=H(z,z)$, $m=n=3$ and $R=S^{\sim}T^{\sim -1}$, $U=-S^{\sim}T^{\sim}$. In this case we have $\mathcal{N}_j^{RR}(p,q)=\mathcal{N}_j^{UU}(q,p)$ and consequently $f_1(z,z)=g_1(z,z)=f_1(z)$. Equations (2.27) and (2.28) carry in this case the same information, namely

$$(1-f_1)^2 f_1=z^3\,. \tag{2.38}$$

Substitution of (2.36) leads then to

$$G(G+2)^2(G+1)^{-3}=8z^3\,. \tag{2.39}$$

The function G defined by this equation has a pole at $z^3=\frac{1}{8}$ and a branch point at $z^3=\frac{4}{27}$. The coefficients a_n in its power series

$$G(y)=\sum_n a_n y^n \tag{2.40}$$

may be found explicitly by substituting (2.39) into Cauchy's formula

$$a_n = \frac{1}{2\pi i}\oint dy \frac{G(y)}{y^{n+1}}, \tag{2.41}$$

to obtain

$$a_n = \frac{1}{2\pi i}\oint dG \frac{(2-G)(G+1)^{3n-1}}{G^n(G+2)^{2n+1}}. \tag{2.42}$$

Making the change of variables $\zeta = G(G+2)^{-2}$, developing the resulting integrand in a power series and integrating term by term one finds

$$a_n = 2^{-3n}\left\{\binom{3n}{n} - \sum_{p=0}^{n-1}\binom{3n}{p}\right\}. \tag{2.43}$$

Substituting (2.43) in (2.37) we get the series development for the dimer generating function:

$$2\log Z_s = \log\ell + \sum_{n=1}^{\infty}\frac{1}{3n}\left(-\frac{x}{\ell}\right)^{3n}\left[\binom{3n}{n} - \sum_{p=0}^{n-1}\binom{3n}{p}\right], \tag{2.44}$$

where we have used $\operatorname{tr}\log\ell S^{\sim} = \log\ell$.

We have found then a series expansion for the dimer generating function per site on a modular lattice. This series converges for $\left|\frac{x}{\ell}\right| < \frac{1}{2}$. Moreover, it has a singularity at $\frac{x}{\ell} = -\frac{1}{2}$, where its behavior is found to be from (2.37) and (2.39)

$$2\log Z_{s\, z\sim\frac{1}{2}} \sim \tfrac{1}{12}\log(1-2z), \qquad z = -\frac{x}{\ell}. \tag{2.45}$$

That is, the free energy $F = -\beta^{-1}\log Z_s$ has a logarithmic singularity, and consequently the specific heat will have a polar singularity with critical exponent $\alpha = 1$.

III. Ising Model

Now we place spin variables σ_g which can take the values ± 1 at each site of our lattice. Sites interact with an energy $-E_1(E_1 > 0)$ if we can get from one to the other by operating on the right with S, and with an energy $-E_2(E_2 > 0)$ if we can get from one to the other by similarly operating with T or T^{-1}. We have then an Ising model for a two dimensional lattice described by the Hamiltonian

$$H = -\tfrac{1}{2}E_1\sum_{g\in\Gamma}\sigma_g\sigma_{gS} - E_2\sum_{g\in\Gamma}\sigma_g\sigma_{gT} \tag{3.1}$$

where the $\frac{1}{2}$ in front of E_1 comes from the fact that the "$g-gS$" interaction is counted twice. Our aim is to compute the free energy per spin for the model described by this Hamiltonian. Fisher [3] has given a general prescription that relates the partition function per spin Z for an Ising model on any planar lattice to the dimer generating function per site Z_s on an associated lattice. In our case, this relation is given by

$$\log Z = \log 2\sinh\tfrac{1}{2}\beta E_1 \sinh\beta E_2 + \log Z_s \tag{3.2}$$

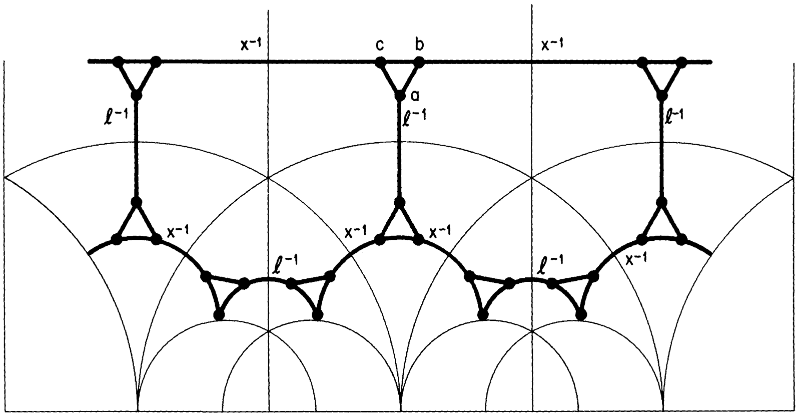


Fig. 3. Lattice whose dimer generating function gives the free energy for an Ising model built on the lattice of Figure 1

and the lattice is drawn on Figure 3. The activities of this lattice are related to the interaction energies by

$$\begin{aligned} \ell &= \tanh \tfrac{1}{2}\beta E_1 \\ x &= \tanh \beta E_2 \,. \end{aligned} \tag{3.3}$$

The problem now is to compute $\log Z_s$. We do this again by applying Kasteleyn's theorem [1].

First, we see that the construction of the lattice may be described as follows: take three non-collinear points inside a fundamental region for Γ, call them $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and connect them by bonds. Act then on this triangle with $g \in \Gamma$. Finally, connect by bonds $g(\boldsymbol{b})$ with $gT(\boldsymbol{c})$ and $g(\boldsymbol{a})$ with $gS(\boldsymbol{a})$. This lattice must now be oriented and the Pfaffian of the corresponding operator computed.

We follow the same procedure of Section II. We start by orienting the fundamental decorated hexagon of Figure 4, which defines the action of an operator $O^{\sim}$ on vectors $\boldsymbol{a}_{\mathbb{1}}$, $\boldsymbol{b}_{\mathbb{1}}$, $\boldsymbol{c}_{\mathbb{1}}$:

$$\begin{aligned} O^{\sim}(\boldsymbol{a}_{\mathbb{1}}) &= \ell^{-1}\boldsymbol{a}_S + \boldsymbol{b}_{\mathbb{1}} + \boldsymbol{c}_{\mathbb{1}} \\ O^{\sim}(\boldsymbol{b}_{\mathbb{1}}) &= -\boldsymbol{a}_{\mathbb{1}} - \boldsymbol{c}_{\mathbb{1}} - x^{-1}\boldsymbol{c}_T \\ O^{\sim}(\boldsymbol{c}_{\mathbb{1}}) &= -\boldsymbol{a}_{\mathbb{1}} + \boldsymbol{b}_{\mathbb{1}} + x^{-1}\boldsymbol{b}_{T^{-1}} \,. \end{aligned} \tag{3.4}$$

This induces us to define an operator O which acts on $\mathscr{E} \otimes \mathbb{R}^3$, where $\mathscr{E}$ is the space where the regular representation of $\mathrm{SL}(2, \mathbb{Z})$ is realized:

$$\begin{aligned} O(\boldsymbol{a}_h) &= \ell^{-1}\boldsymbol{a}_{h\left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right)} + \boldsymbol{b}_h + \boldsymbol{c}_h \\ O(\boldsymbol{b}_h) &= -\boldsymbol{a}_h - \boldsymbol{c}_h - x^{-1}\boldsymbol{c}_{h\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)} \qquad h \in \mathrm{SL}(2, \mathbb{Z}) \\ O(\boldsymbol{c}_h) &= -\boldsymbol{a}_h + \boldsymbol{b}_h + x^{-1}\boldsymbol{b}_{h\left(\begin{smallmatrix}1 & -1\\ 0 & 1\end{smallmatrix}\right)} \,. \end{aligned} \tag{3.5}$$

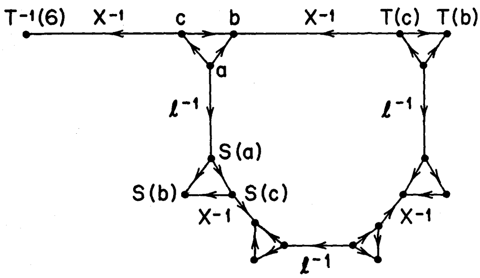


Fig. 4. Orientation of the fundamental decorated hexagon that induces an orientation of the lattice of Figure 3

Imposing the constraints

$$\begin{aligned} &\boldsymbol{a}_h = -\boldsymbol{a}_{-h} \\ &\boldsymbol{b}_h = -\boldsymbol{b}_{-h} \\ &\boldsymbol{c}_h = -\boldsymbol{c}_{-h}, \end{aligned} \tag{3.6}$$

and choosing a particular injection of Γ into $\mathrm{SL}(2,\mathbb{Z})$ consistent with (3.4) orients then the whole lattice by defining an operator $O^{\sim}$ which is the operator induced by O on $\mathscr{E}\otimes\mathbb{R}^3/\ker(\mathscr{B}\otimes\mathbb{I}_3)$, where $\mathscr{B}=\mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\mathscr{R}\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$, and $\mathbb{I}_3$ is the identity in $\mathbb{R}^3$. $\mathscr{R}$ is the regular representation of $\mathrm{SL}(2,\mathbb{Z})$. Expression (3.5) may then be rewritten as

$$O=\begin{pmatrix} \ell^{-1}\mathscr{R}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} & \mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} & \mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \\ -\mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} & 0 & -\mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-x^{-1}\mathscr{R}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} \\ -\mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} & \mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+x^{-1}\mathscr{R}\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} & 0 \end{pmatrix}. \tag{3.7}$$

The dimer generating function is given by

$$2\log Z_s = \mathrm{tr}\,\log O^{\sim}. \tag{3.8}$$

From (3.7) we see that tr $\log O$ may be written as $\left(\text{here we put } 1=\mathscr{R}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$

$$\mathrm{tr}\,\log O^{\sim} = \mathrm{tr}\,\log\left\{\begin{pmatrix} -\ell^{-1}S^{\sim} & 0 & 0 \\ 0 & -x^{-1}T^{\sim} & 0 \\ 0 & 0 & x^{-1}T^{\sim-1}\end{pmatrix}+\begin{pmatrix} 0 & +1 & +1 \\ -1 & -1 & 0 \\ -1 & 0 & 1\end{pmatrix}\right\} \tag{3.9}$$

where we have interchanged the second and third row and changed the sign of the first one. Factoring out the first term in the right hand side of (3.9) and developing the logarithm we have

$$\operatorname{tr}\log O^{\sim}=\operatorname{tr}\log\begin{pmatrix}-\ell^{-1}S^{\sim} & 0 & 0\\ 0 & -x^{-1}T^{\sim} & 0\\ 0 & 0 & x^{-1}T^{\sim-1}\end{pmatrix}$$

$$+\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\operatorname{tr}(-\ell S^{\sim}\otimes \boldsymbol{a}+xT^{\sim-1}\otimes \boldsymbol{b}+xT^{\sim}\otimes \boldsymbol{c})^k \tag{3.10}$$

where

$$\boldsymbol{a}=\begin{pmatrix}0 & 1 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$$

$$\boldsymbol{b}=\begin{pmatrix}0 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}\quad \boldsymbol{c}=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ -1 & 0 & 1\end{pmatrix}. \tag{3.11}$$

These matrices satisfy the following multiplication rules:

$$\boldsymbol{a}^2=\boldsymbol{bc}=\boldsymbol{cb}=0\,, \tag{3.12a}$$

$$\boldsymbol{b}^2=\boldsymbol{b},\, \boldsymbol{c}^2=\boldsymbol{c}\,, \tag{3.12b}$$

$$(\boldsymbol{ab})^2=\boldsymbol{ab},\, (\boldsymbol{ba})^2=\boldsymbol{ba},\, (\boldsymbol{ac})^2=-\boldsymbol{ac},\, (\boldsymbol{ca})^2=-\boldsymbol{ca}\,, \tag{3.12c}$$

$$\boldsymbol{abac}=\boldsymbol{ac},\, \boldsymbol{acab}=-\boldsymbol{ab},\, \boldsymbol{baca}=-\boldsymbol{ba},\, \boldsymbol{caba}=\boldsymbol{ca} \tag{3.12d}$$

and their traces are

$$\operatorname{tr}(\boldsymbol{ab})=-\operatorname{tr}(\boldsymbol{ac})=1\,. \tag{3.13}$$

From now on, we shall omit the twiddle in $S^{\sim}$, $T^{\sim}$ and $T^{\sim-1}$.

We are now faced with the problem of finding the trace of an operator of the form $\mathscr{W}_1\otimes\mathscr{W}_2$, with $\mathscr{W}_2$ being a word in the letters $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\mathscr{W}_1$ a word in S, T and T^{-1}. We see that only words $\mathscr{W}_1$ reducible to the identity contribute, so that counting these words is equivalent to counting closed loops on the modular lattice. The condition $\operatorname{tr}\mathscr{W}_2\neq 0$ gives different weights to different polygons. In particular, (3.12a) says that no step can be immediately retreated. That is, in $\mathscr{W}_1$ the letter S never appears raised to a power other than one or zero, and T never appears immediately preceding or following T^{-1}. This means in particular that in the reduction of $\mathscr{W}_1$ the identity $S^2=-\mathbb{I}$ is not used. Also, if $\mathscr{W}_1$ starts with S, it cannot end with it, as in this case one would have that $\mathscr{W}_2$ is of the form $\mathscr{W}_2=\boldsymbol{a}\mathscr{W}\boldsymbol{a}$ implying $\operatorname{tr}\mathscr{W}_2=\operatorname{tr}\mathscr{W}\boldsymbol{a}^2=0$. Similarly, if $\mathscr{W}_1$ starts with T (resp. T^{-1}), it cannot end with T^{-1} (resp. T). We shall call such a word *admissible*.

Make now the following definitions:

$$\begin{aligned}&R=-ST,\, U=ST^{-1}\,,\\ &R^{-1}=T^{-1}S,\, U^{-1}=-TS\,,\end{aligned} \tag{3.14a}$$

and we have

$$R^3=U^3=\mathbb{I}\,. \tag{3.14b}$$

Now the fact that we are interested in counting words in S, T, T^{-1} which cannot be reduced by using $S^2=-\mathbb{I}$ or $TT^{-1}=\mathbb{I}$ says that they have to be reduced using (3.14b) and/or their inverses, and we have then that admissible words are words in R, U, their inverses, T and T^{-1}. The following lemma shows that it is enough to consider a class of words in R, U and their inverses.

Lemma 3.1. *There is a one-to-one correspondence between admissible words in S, T and T^{-1} and words in R, R^{-1}, U and U^{-1} in which R and R^{-1} are not contiguous and neither are U and U^{-1}. In addition, if R (resp. U) is the first letter, then the last one cannot be U^{-1} (resp. R^{-1}). Such a word will also be called admissible.* (In this context, the first and last letters of a word are considered to be contiguous.)

Proof. In two parts. First we start with a word in R, R^{-1}, U, U^{-1} and get to a word in S, T, T^{-1}. After that we perform the inverse step.

a) Consider an admissible word in R, R^{-1}, U and U^{-1}. It is a fortiori a word in S, T and T^{-1}. The letters T and T^{-1} are not contiguous since this is possible only by having either R and R^{-1} contiguous, or U and U^{-1}. The letter S, however, appears squared whenever there appears the combinations $R^{-1}U$ or $U^{-1}R$. Notice that S does not appear raised to a power higher than two. We obtain then the wanted admissible word in S, T and T^{-1} simply by deleting all S^2. Notice that such deletion gives rise to T^{-2} if it is done on $R^{-1}U$, and T^2 if done on $U^{-1}R$ so that no new prohibited combinations appear. An overall minus sign is added if the number of such deletions is odd.

b) Consider now an admissible word $\mathscr{W}_1$ in S, T, and T^{-1}. We distinguish two cases: (i) The first letter is S. In this case, there follows a power of $T^{\pm 1}$, say $T^{\pm n}(n\neq 0)$. If T^{+n}, write $\mathscr{W}_1$ as $\mathscr{W}_1=(ST)T^{n-1}\ldots=(-R)T^{n-1}\ldots$. If n odd, write T^{n-1} as $(-)^{\frac{1}{2}(n-1)}(TSST)^{\frac{1}{2}(n-1)}=(-)^{\frac{1}{2}(n-1)}(U^{-1}R)^{\frac{1}{2}(n-1)}$. If n even, write $T^{n-1}=(-)^{\frac{1}{2}(n-2)}(U^{-1}R)^{\frac{1}{2}(n-2)}T$. The next letter is S. For n odd, the process is repeated. For n even, $\mathscr{W}_1$ is

$$\mathscr{W}_1=(-)^{\frac{1}{2}(n)}R(U^{-1}R)^{\frac{1}{2}(n-2)}TS\ldots=(-)^{\frac{1}{2}(n-2)}R(U^{-1}R)^{\frac{1}{2}(n-2)}U^{-1}\ldots .$$

After U^{-1}, a power of $T^{\pm 1}$ comes again, and we have a word whose first letter is not S. This is the second case. Before going into it, notice that if the second letter of $\mathscr{W}_1$ is T^{-n} instead of T^{+n}, the process of intercalating $(-S^2)$ yields

$$\mathscr{W}_1=ST^{-n}S\ldots=UT^{-n+1}S\ldots=\begin{cases}(-)^{\frac{1}{2}(n-1)}U(R^{-1}U)^{\frac{1}{2}(n-1)}S\ldots(\text{odd}\,n)\\(-)^{\frac{1}{2}(n-2)}U(R^{-1}U)^{\frac{1}{2}(n-2)}R^{-1}\ldots(\text{even}\,n)\,.\end{cases}$$

(ii) The first letter is T^n. If n even, we write $T^n=(-)^{\frac{1}{2}n}(U^{-1}R)^n$. The letter that follows is S, and we are back in case (i). If n odd, $T^n=(-)^{\frac{1}{2}(n-1)}(U^{-1}R)^{\frac{1}{2}(n-1)}T$, and $\mathscr{W}_1=(-)^{\frac{1}{2}(n-1)}(U^{-1}R)TS\ldots=(-)^{\frac{1}{2}(n-1)}(U^{-1}R)^{\frac{1}{2}(n-1)}(-U)^{-1}\ldots$ and the procedure is repeated. The case in which the first letter of $\mathscr{W}_1$ is T^{-n} is obtained from this one by interchanging $R\leftrightarrow U$. We have then constructed a way of uniquely getting from one type of word to the other, and the proof is complete.

Go back now to the series of (3.10). It is of the form

$$\sum_{m+n=k}\frac{(-1)^k}{k}(-\ell)^m x^n\mathscr{A}_{mn}\,. \tag{3.15}$$

It is easy to see that k must be even and that $m\leqq n$. Also, from (3.13)–(3.15) one sees that all words contributing to $\mathscr{A}_{mn}$ do so with the same sign, namely $(-1)^{\frac{1}{2}(n-m)}$.

We obtain then

$$\operatorname{tr}\log O^{\sim} = -\log \ell x^2 + \sum_{m \leqq n} \frac{(-1)^{\frac{1}{2}(n+m)}}{n+m} \ell^m x^n \begin{pmatrix} \text{number of admissible} \\ \text{identity words in } S, \\ T \text{ and } T^{-1} \text{ with } S \text{ ap-} \\ \text{pearing } m \text{ times, } T \text{ and} \\ T^{-1} \text{ a total of } n \text{ times} \end{pmatrix}. \quad (3.16)$$

Notice that in this context, "identity word" means a word which is reducible to $\pm\mathbb{I}$. From Lemma 3.1 we have that the number of words referred to in (3.16) is the same as the number of admissible identity words in R, R^{-1}, U, U^{-1} of length n, and in which there appear the combinations $R^{-1}U$ and $U^{-1}R$ a total of $\frac{1}{2}(n-m)$ times. Consider then the generating function

$$G(z, \lambda) = \sum_{k,n} z^k \lambda^n N(k, n) \quad (3.17)$$

where $N(k, n)$ is the number of admissible identity words of length k in R, R^{-1}, U and U^{-1} with the combinations $U^{-1}R$ and $R^{-1}U$ appearing a total of n times [notice the change in notation for $N(k, n)$ with respect to Sect. II]. The relation between (3.17) and (3.16) is given by

$$\operatorname{tr}\log O^{\sim} = -\log \ell x^2 + \int^z \frac{dz'}{z'} G(z', \lambda) \Bigg|_{\substack{z = -x\ell \\ \lambda = -\ell^{-2}}} . \quad (3.18)$$

In the following we shall speak of identity words only, it being understood that they are admissible. We shall also use the nomenclature of Section II, speaking of monoblocks, multiblocks, UR-words, etc., with the obvious generalizations to include the new letters U^{-1} and R^{-1}. The problem now is to keep track of the number of times the combinations $R^{-1}U$ and $U^{-1}R$ appear in a word of a given length.

In Section II, we found the generic monoblock by inserting multiblocks of a certain kind inside a "basic" monoblock. These multiblocks were themselves formed as products of monoblocks, so that the number of monoblocks of a given length was expressed in terms of the number of shorter monoblocks. Moreover, as the generic identity word is a product of monoblocks, the corresponding generating function could then be expressed in terms of the corresponding monoblock function. It is this process which we want to follow here again.

The generic identity word will here also be a product of monoblocks. However, some monoblocks cannot appear side by side, i.e. one ending in R cannot be followed by one starting with R^{-1}. Also, the power of $(-\ell)^2$ is raised by one whenever a word ending in R^{-1} (resp. U^{-1}) is followed by a word beginning by U (resp. R). More precisely, a word which is the product of a word ending in R^{-1} contributing to $N(k_1, n_1)$ and another starting with U which contributes to $N(k_2, n_2)$, will itself contribute to $N(k_1+k_2, n_1+n_2+1)$.

We now introduce some notation: we associate indexes 1, 2, 3, 4 with R, R^{-1}, U, U^{-1} respectively. Notice that we don't worry about minus signs as we know that all words with fixed k, n contribute with the same overall sign which has already been determined, giving (3.16). The generating function for monoblocks starting with the letter corresponding to i and ending with the letter corresponding to $j\,(1 \leqq i, j \leqq 4)$ will be called f_{ij}:

$$f_{ij}(z, \lambda) = \sum_{k,n} z^k \lambda^n N_{1,ij}(k, n) \tag{3.19}$$

where $N_{1,ij}(k, n)$ is the number of monoblocks starting with the letter corresponding to i and ending with the letter corresponding to j [we shall call them *(i, j)-monoblocks*, (i, j)-words and so on] of length k and with n combinations $R^{-1}U$ or $U^{-1}R$.

Let us call $\boldsymbol{F} = (f_{ij})_{1 \leqq i, j \leqq 4}$ the 4×4 matrix formed by the f_{ij}. The following lemma generalizes Lemma 2.1:

Lemma 3.2.

$$\boldsymbol{F} = \begin{pmatrix} f_{11} & f_{12} & 0 & 0 \\ \lambda f_{12} & f_{11} & 0 & 0 \\ 0 & 0 & f_{11} & f_{12} \\ 0 & 0 & \lambda f_{12} & f_{11} \end{pmatrix}. \tag{3.20}$$

Proof (Sketch. For details see Ref. [6]). In three parts.

a) The vanishing of f_{13}, f_{14}, etc. is proved with the same method that was used in Lemma 2.1.

b) By changing R with U one establishes a one-to-one correspondence between different types of monoblocks that proves $f_{11} = f_{13}$, $f_{22} = f_{44}$, $f_{12} = f_{34}$, $f_{21} = f_{43}$. Taking the inverse of a $(1, 1)$-monoblock gives a $(2, 2)$-monoblock, proving $f_{11} = f_{22}$.

c) A one-to-one relation between $(1, 2)$- and $(2, 1)$-monoblocks is established by reading a word from right to left instead of from left to right. This gives $N_{1,12}(k, n) = N_{1,21}(k, n+1)$ and consequently $f_{21} = \lambda f_{12}$, completing the proof.

The following four lemmas generalize lemma 2.2 and may be proved with the same technique used there.

Lemma 3.3. *If $\mathcal{M} = R\mathcal{W}_1 R\mathcal{W}_2 R$ is a $(1, 1)$-monoblock, then both $\mathcal{W}_1$ and $\mathcal{W}_2$ (which are identity words) are products of monoblocks which cannot have as factors $(1, 1)$- or $(2, 2)$-monoblocks. Moreover, $\mathcal{W}_1$ cannot have as factors $(2, 1)$-monoblocks, and $\mathcal{W}_2$ cannot have as factors $(1, 2)$-monoblocks.*

Lemma 3.4. *If $\mathcal{M} = R\mathcal{W}R^{-1}$ is a $(1, 2)$-monoblock, the identity word $\mathcal{W}$ cannot contain as factors either $(1, 1)$-, $(2, 2)$- or $(2, 1)$-monoblocks.*

Lemma 3.5. *Let $\mathcal{M} = R\mathcal{W}_1 R\mathcal{W}_2 R$ be a $(1, 1)$-monoblock such that $\mathcal{W}_1 = \mathcal{A}R\mathcal{B}R^{-1}\mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are identity words. That is, $\mathcal{W}_1$ has a $(1,2)$-monoblock as a factor. Then, the word $\mathcal{B}$ cannot have another $(1, 2)$-monoblock as a factor.*

Lemma 3.6. *If $\mathcal{M} = R\mathcal{W}R^{-1}$ is a $(1, 2)$-monoblock and $\mathcal{W} = \mathcal{A}R\mathcal{B}R^{-1}\mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are identity words, then $\mathcal{B}$ cannot have a $(1, 2)$-monoblock as a factor.*

Let us now go back to the generating function (3.17). Clearly,

$$N(k, n) = \sum_j N_j(k, n), \tag{3.21}$$

where $N_j(k, n)$ is the number of admissible j-blocks, and therefore

$$G(z, \lambda) = \sum_j G_j(z, \lambda), \tag{3.22}$$

where $G_j(z,\lambda)$ is the generating function for admissible j-blocks. Consider $G_1(z,\lambda)$. At first sight, one would be tempted to write $N_1(k,n)=\sum_{ij} N_{1,ij}(k,n)$ and hence $G_1=\sum_{ij} f_{ij}$. Not all the monoblocks are admissible, however, as they must satisfy the hypotheses of Lemma 3.1. That is, neither R and R^{-1} nor U and U^{-1} can be contiguous. Thus, f_{12}, f_{21}, f_{34}, and f_{43} are not admissible, and $G_1=\mathrm{tr}(\boldsymbol{F})$.

Consider now G_2. This time there are two types of restrictions: on the one hand, the initial letter of the diblock restricts the choice for last letter. On the other hand, not just any two monoblocks can be put side by side. We tackle the second problem first. Consider then an (i,j)-diblock. We want to find $N_{2,ij}(k,n)$. The number of (i,j)-diblocks we get is the number of ways we can multiply two monoblocks in such a way that the initial letter is i, the final letter j, the total length comes out k, and the total number of $U^{-1}R$ and $R^{-1}U$ combinations comes out n. That is:

$$\begin{aligned} N_{2,ij}(k,n)= \sum_{\substack{k_1+k_2=k\\ n_1+n_2=n}} & N_{1,i1}(k_1,n_1)[N_{1,1j}(k_2,n_2)+N_{1,3j}(k_2,n_2)+N_{1,4j}(k_2,n_2)] \\ & +N_{1,i2}(k_1,n_1)[N_{1,2j}(k_2,n_2)+N_{1,3j}(k_2,n_2-1)+N_{1,4j}(k_2,n_2)] \\ & +N_{1,i3}(k_1,n_1)[N_{1,1j}(k_2,n_2)+N_{1,2j}(k_2,n_2)+N_{1,3j}(k_2,n_2)] \\ & +N_{1,i4}(k_1,n_1)[N_{1,1j}(k_2,n_2-1)+N_{1,2j}(k_2,n_2)+N_{1,4j}(k_2,n_2)]\,. \end{aligned} \tag{3.23}$$

Certain monoblocks cannot be put side by side. For example, a (1, 1)-monoblock followed by a (2, 1)-monoblock would mean a combination of RR^{-1}, which is forbidden. Notice also that a $(i,2)$- [resp. a $(i,4)$-] monoblock followed by a $(3,j)$- [resp. a $(1,j)$-] monoblock means an additional $R^{-1}U$(resp. a $U^{-1}R$) combination, a fact which is reflected in $N_{1,31}$ and $N_{1,21}$ entering (3.23) with n_2-1 instead of n_2. If we now call $G_{2,ij}(z,\lambda)=\sum_{k,n} z^k\lambda^n N_{2,ij}(k,n)$, we have, after multiplying (3.23) by $z^k\lambda^n$ and summing over k and n

$$\begin{aligned} G_{2,ij}(z,\lambda)= & f_{i1}(f_{1j}+f_{3j}+f_{4j}) \\ & +f_{i2}(f_{2j}+\lambda f_{3j}+f_{4j}) \\ & +f_{i3}(f_{1j}+f_{2j}+f_{3j}) \\ & +f_{i4}(\lambda f_{1j}+f_{2j}+f_{4j}) \end{aligned} \tag{3.24}$$

A compact way of writing this is

$$G_{2,ij}=(\boldsymbol{FTF})_{ij}\,, \tag{3.25}$$

where $\boldsymbol{F}$ is defined by (3.20) and

$$\boldsymbol{T}\equiv\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & \lambda & 1 \\ 1 & 1 & 1 & 0 \\ \lambda & 1 & 0 & 1 \end{pmatrix}. \tag{3.26}$$

Now for the problem of incompatibility of certain initial and final letters in the words contributing to G_2. If the initial letter is R^{-1} (resp. U^{-1}) the final letter cannot be R (resp. U). If the initial letter is R or U the final letter cannot be either R^{-1} or U^{-1}. Thus

$$\begin{aligned} G_2 = G_{2,11} & \qquad\quad + G_{2,13} \\ & + G_{2,22} + G_{2,23} + G_{2,24} \\ + G_{2,31} & \qquad\quad + G_{2,33} \\ + G_{2,41} & + G_{2,42} \qquad\quad + G_{2,44} \end{aligned} \tag{3.27}$$

If we introduce

$$\Omega = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \tag{3.28}$$

then Equation (3.27) can be written, using (3.25), as

$$G_2 = \Omega_{ij} G_{2,ji} = \mathrm{tr}(\Omega \boldsymbol{F} \boldsymbol{T} \boldsymbol{F}) . \tag{3.29}$$

Here and in what follows we adopt the convention of summing over repeated indices.

The reasoning leading to (3.29) leads also to

$$G_\ell = \Omega_{ij} G_{\ell,ij} \tag{3.30}$$

and

$$G_{\ell,ij} = (\boldsymbol{F}(\boldsymbol{T}\boldsymbol{F})^{\ell-1})_{ij} , \tag{3.31}$$

giving finally

$$G = \mathrm{tr}[\Omega \boldsymbol{F}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}] . \tag{3.32}$$

This expresses then the generating function (3.17) in terms of the generating functions for monoblocks. We now turn our attention to them.

Let us call $f_{11} \equiv f$ and $f_{12} \equiv g$. Consider f first. It is the generating function for (1, 1)-monoblocks, whose general form is $\mathcal{M} = R\mathcal{W}_1 R\mathcal{W}_2 R$, where $\mathcal{W}_1$ and $\mathcal{W}_2$ are identity words subject to the restrictions imposed by Lemmas 3.3 and 3.5. Notice now the following: suppose $\mathcal{W}_1 = \mathcal{A}R\mathcal{B}R^{-1}\mathcal{C}$ and $\mathcal{W}_2 = \mathcal{D}$ where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are identity words. Then, $\mathcal{M} = R\mathcal{A}R\mathcal{B}R^{-1}\mathcal{C}R\mathcal{D}R$. This same monoblock can be formed by taking $\mathcal{W}_1 = \mathcal{A}$ and $\mathcal{W}_2 = \mathcal{B}R^{-1}\mathcal{C}R\mathcal{D}$. Since we need a unique way of forming a monoblock, we shall exclude words $\mathcal{W}_2$ which have a (2, 1)-monoblock as a factor. In view of Lemma 3.3, this means that only (i, j)-monoblocks with $i, j = 3, 4$ are allowed as factors of $\mathcal{W}_2$. The number of (1, 1)-monoblocks $N_{1,11}(k, n)$ is then the product of the number of $\mathcal{W}_1$ and $\mathcal{W}_2$ words which satisfy the restrictions imposed by compatibility, the Lemmas 3.3 and 3.5 and the one just imposed on $\mathcal{W}_2$, and such that the total length is k and the number of $U^{-1}R$ and $R^{-1}U$ combinations is n. Let us call $\mathcal{N}^1_{ij}(k_1, n_1)$ the number of (i, j)-identity words of length k_1 and with n_1 $U^{-1}R$ and $R^{-1}U$ combinations which satisfy the restrictions imposed in $\mathcal{W}_1$, and $\mathcal{N}^2_{ij}(k_2, n_2)$ the corresponding number of (i, j)-identity words satisfying the restrictions imposed on $\mathcal{W}_2$. We have then

$$N_{1,11}(k, n) = \sum_{\substack{k_1 + k_2 = k \\ n_1 + n_2 = n}} \mathcal{N}^1(k_1, n_1) \mathcal{N}^2(k_2 - 3, n_2) \tag{3.33}$$

where

$$\begin{aligned}\mathcal{N}^1(k_1,n_1)=&\mathcal{N}^1_{13}(k_1,n_1)+\mathcal{N}^1_{14}(k_1,n_1-1)\\&+\mathcal{N}^1_{33}(k_1,n_1)+\mathcal{N}^1_{34}(k_1,n_1-1)\\&+\mathcal{N}^1_{43}(k_1,n_1)+\mathcal{N}^1_{44}(k_1,n_1-1)\,,\end{aligned}\tag{3.34}$$

$$\begin{aligned}\mathcal{N}^2(k_2,n_2)=&\mathcal{N}^2_{33}(k_2,n_2)+\mathcal{N}^2_{34}(k_2,n_2-1)\\&+\mathcal{N}^2_{43}(k_2,n_2)+\mathcal{N}^2_{44}(k_2,n_2-1)\,.\end{aligned}\tag{3.35}$$

We see that $\mathcal{N}^1_{2i}$ and $\mathcal{N}^1_{i2}$ do not contribute to $\mathcal{N}^1$, since $\mathcal{W}_1$ cannot begin or end with R^{-1}. Neither does $\mathcal{N}^1_{i1}$ contribute, as $\mathcal{W}_1$ cannot have as factors monoblocks ending with R (more precisely, it cannot have (1, 1)- or (2, 1)-monoblocks as factors. But we have already seen in Lemma 3.2 that there are no (3,1)- or (4,1)-monoblocks). The structure of (3.35) reflects the fact that only (ij)-monoblocks with $i,j=3,4$ contribute in $\mathcal{W}_2$.

Let us take $\mathcal{N}^1_{ij}(k,n)$. It is the number of admissible identity (i,j)-words satisfying the restriction imposed on $\mathcal{W}_1$. If we call $\mathcal{N}^1_{\ell,ij}(k,n)$ the corresponding number of ℓ-blocks, we have

$$\mathcal{N}^1_{ij}(k,n)=\sum_{\ell}\mathcal{N}^1_{\ell,ij}(k,n)\,.\tag{3.36}$$

The multiblocks contributing to $\mathcal{N}^1_{\ell,ij}$ are formed by multiplying (ij)-monoblocks with $i,j=3,4$ and also (1, 2)-monoblocks. Not all the (1, 2)-monoblocks contribute, however. They are restricted by Lemma 3.5. We have then

$$\mathcal{N}^1_{1,ij}=N_{1,ij}\quad\text{for}\quad i,j=3,4\tag{3.37}$$

but

$$\mathcal{N}^1_{1,12}\neq N_{1,12}\,,\tag{3.38}$$

and

$$\mathcal{N}^1_{1,ij}=0\quad\text{otherwise}\,.\tag{3.39}$$

Let us call h the generating function for (1, 2)-monoblocks contributing to $\mathcal{W}_1$:

$$h(z,\lambda)=\sum_{k,n}z^k\lambda^n\mathcal{N}^1_{1,12}(k,n)\,.\tag{3.40}$$

At this point we need some more notation, so we make the following definitions:

$$A^a(z,\lambda)\equiv\sum_{k,n}\mathcal{N}^a(k,n)z^k\lambda^n,\qquad a=1,2\,,\tag{3.41}$$

$$A^a_{ij}(z,\lambda)\equiv\sum_{k,n}\mathcal{N}^a_{ij}(k,n)z^k\lambda^n\,,\tag{3.42}$$

$$A^a_{\ell,ij}(z,\lambda)\equiv\sum_{k,n}\mathcal{N}^a_{\ell,ij}(k,n)z^k\lambda^n\,.\tag{3.43}$$

In particular, $A^1_{1,12}=h$ and $A^1_{1,ij}=f_{ij}$ for $i,j=3,4$. All other $A^1_{1,ij}$ vanish. Consider now $\mathcal{N}^1_{2,ij}(k,n)$. From what has been said before we have

$$\begin{aligned}\mathcal{N}^1_{2,ij}(k,n)=\sum_{\substack{k_1+k_2=k\\n_1+n_2=n}}&\mathcal{N}^1_{1,i2}(k_1,n_1)[\mathcal{N}^1_{1,3j}(k_2,n_2-1)+\mathcal{N}^1_{1,4j}(k_2,n_2)]\\&+\mathcal{N}^1_{1,i3}(k_1,n_1)[\mathcal{N}^1_{1,1j}(k_2,n_2)+\mathcal{N}^1_{1,3j}(k_2,n_2)]\\&+\mathcal{N}^1_{1,i4}(k_1,n_1)[\mathcal{N}^1_{1,1j}(k_2,n_2-1)+\mathcal{N}^1_{1,4j}(k_2,n_2)]\,.\end{aligned}\tag{3.44}$$

Notice the similarity with (3.23). In fact, it is that same equation plus the restrictions imposed on the monoblocks making up $\mathscr{W}_1$. Multiplying by $z^k\lambda^n$ and summing over k and n we get

$$\begin{aligned} A^1_{2,ij} = A^1_{1,i2}(& \qquad \lambda A^1_{1,3j} + A^1_{1,4j}) \\ + A^1_{1,i3}(& A^1_{1,1j} + A^1_{1,3j} \qquad) \\ + A^1_{1,i4}(& \lambda A^1_{1,1j} \qquad + A^1_{1,4j}) \end{aligned} \tag{3.45}$$

which is similar to (3.24). A relation analogous to (3.25) can be obtained if we introduce

$$\boldsymbol{F}' = \begin{pmatrix} 0 & h & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & f_{33} & f_{34} \\ 0 & 0 & f_{43} & f_{44} \end{pmatrix}. \tag{3.46}$$

That is, $(\boldsymbol{F}')_{ij} = A^1_{1,ij}$.

In this case,

$$A^1_{2,ij} = (\boldsymbol{F}'\boldsymbol{T}\boldsymbol{F}')_{ij}, \tag{3.47}$$

with $\boldsymbol{T}$ given by (3.26). Clearly now, $A^1_{\ell,ij} = A^1_{\ell,ip}(\boldsymbol{T})_{pq}A^1_{\ell,qj}$ and consequently

$$A^1_{\ell,ij} = (\boldsymbol{F}'(\boldsymbol{T}\boldsymbol{F}')^{\ell-1})_{ij}. \tag{3.48}$$

Notice that $A^1_{\ell,i1} = A^1_{\ell,2i} = 0$, $1 \leqq i \leqq 4$.

If we multiply (3.34) by $z^{k_1}\lambda^{n_1}$ and sum, we shall have

$$\begin{aligned} A^1(z,\lambda) &= A^1_{13} + \lambda A^1_{14} + A^1_{33} + \lambda A^1_{34} + A^1_{43} + A^1_{44} \\ &= (\boldsymbol{T})_{1p} A^1_{pq} (\boldsymbol{T})_{q1}. \end{aligned} \tag{3.49}$$

Using (3.36) we have $A^1_{pq} = \sum\limits_{\ell=0} A^1_{\ell,pq}$, where $A^1_{0,pq} = 1$ is the contribution of the empty word, which is not forbidden by the restrictions imposed on $\mathscr{W}_1$. Introducing here (3.48) and then the result into (3.49) we get

$$A^1 = (\boldsymbol{T}(\mathbb{1} - \boldsymbol{F}'\boldsymbol{T})^{-1})_{11}. \tag{3.50}$$

We have now A^1. Next, we get A^2. From (3.35) we have

$$A^2 = A^2_{33} + \lambda A^2_{34} + A^2_{43} + \lambda A^2_{44}. \tag{3.51}$$

Since (3.36) also holds for $\mathscr{N}^2$, we again have $A^2_{ij} = \sum\limits_{\ell} A^2_{\ell,ij}$. Computing $A^2_{\ell,ij}$ is now simpler, as $i, j = 3, 4$ and the only rule to be remembered is that U and U^{-1} cannot be neighbors. In fact, since A^2 is obtained from A^1 just by forbidding the (1, 2)-monoblocks, it is enough to put $h = 0$ in (3.50) to get A^2. Explicitly,

$$A^2 = (\boldsymbol{T}(\mathbb{1} - \boldsymbol{F}''\boldsymbol{T})^{-1})_{11}, \tag{3.52}$$

where

$$\boldsymbol{F}'' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & f_{33} & f_{34} \\ 0 & 0 & f_{43} & f_{44} \end{pmatrix}. \tag{3.53}$$

Note that also here there is a contribution from the empty word.

Finally, from (3.33) we have

$$f(z,\lambda)=z^3A^1(z,\lambda)A^2(z,\lambda)\,. \tag{3.54}$$

This gives an equation for f in terms of g and h. We now look for similar equations for g and for h.

Now, g is the generating function for (1, 2)-monoblocks, whose general form is $\mathscr{M}=R\mathscr{W}R^{-1}$, where $\mathscr{W}$ is an identity word subject to the restriction of Lemmas 3.4 and 3.6. Notice that these are the same restrictions that were imposed on $\mathscr{W}_1$ when finding the generating function f. The only difference is that this time there is no additional $U^{-1}R$ combination when $\mathscr{W}$ ends with U^{-1}. Formula (3.34) holds then with the difference that instead of the terms $\mathscr{N}^1_{i4}(k,n-1)$ there are terms $\mathscr{N}^1_{i4}(k,n)$. The reasoning that led to (3.49) now leads to

$$g(z,\lambda)=z^2(A^1_{13}+A^1_{14}+A^1_{33}+A^1_{34}+A^1_{43}+A^1_{44}) \tag{3.55}$$

which may be written as

$$g=z^2(\boldsymbol{T}(\mathbb{1}-\boldsymbol{F}'\boldsymbol{T})^{-1})_{12}\,. \tag{3.56}$$

Notice that here there is no contribution from the empty word as $(\boldsymbol{T})_{12}=0$. $\mathscr{W}=1$ would mean $\mathscr{M}=RR^{-1}$, which is forbidden.

Finally, we find h. It is the generating function for monoblocks $R\mathscr{W}R^{-1}$ where $\mathscr{W}$ has all the restrictions that applied when finding g plus the additional one of not having (1, 2)-monoblocks as factors. We see that h is related to g in the same way as A^2 was related to A^1. We got from A^1 to A^2 by letting $\boldsymbol{F}'\to\boldsymbol{F}''$. It is very easy to see that this is also the step that must be taken to go from g to h. Consequently,

$$h=z^2(\boldsymbol{T}(\mathbb{1}-\boldsymbol{F}''\boldsymbol{T})^{-1})_{12}\,. \tag{3.57}$$

As was the case for g, the empty word does not contribute.

Since $\boldsymbol{F}''$ is essentially a 2×2 matrix, it is easy to compute h as well as A^2. The results are

$$h=z^2\frac{2f-2f^2+g+\lambda g+2\lambda g^2}{(1-f)^2-\lambda g^2}\,, \tag{3.58}$$

$$A^2=\frac{1-f+\lambda f-\lambda f^2+2\lambda g+\lambda^2g^2}{(1-f)^2-\lambda g^2}\,. \tag{3.59}$$

The computation of g and A^1 is much more tedious as it involves the inversion of a 4×4 matrix. The procedure, however, is straightforward and one obtains

$$f=z^3\frac{(1+2\lambda g+\lambda^2g^2-f+\lambda f-\lambda f^2)^2}{((1-f)^2-\lambda g^2)^2-z^2\lambda(2f-2f^2+g+\lambda g+2\lambda g^2)^2}\,, \tag{3.60}$$

$$g=z^2\cdot\frac{(2\lambda g^2+2f^2+\lambda g+g)[(1-f)^2-\lambda g^2+z^2((1+\lambda f)^2-4\lambda f^2+f\lambda g+4\lambda^2g^2-\lambda^3g^2)]}{((1-f)^2-\lambda g^2)^2-z^2\lambda(2f-2f^2+g+\lambda g+2\lambda g^2)^2}\,. \tag{3.61}$$

To obtain the last two expressions we have made use of (3.58) and (3.59).

This expression for g can be considerably simplified. This is achieved by noticing that, if we define the determinant

$$D = \det \begin{vmatrix} 1-f & h & \lambda h & h \\ 1 & 0 & 1 & 1 \\ f+\lambda g & f+g & 1-f & g \\ \lambda(f+g) & f+\lambda g & \lambda g & 1-f \end{vmatrix}, \tag{3.62}$$

then

$$f = z^3 \frac{(D_{12})^2}{D_{22}D_{12,12}}, \tag{3.63}$$

$$g = z^2 \frac{D}{D_{22}}, \tag{3.64}$$

$$h = z^2 \frac{D_{11}}{D_{12,12}}, \tag{3.65}$$

where D_{ab} denotes the minor obtained by deleting the a^{th} row and the b^{th} column, and $D_{ab,cd}$ denotes the minor obtained by deleting rows a, b, and columns c, d. Using now the identity

$$D_{11}D_{22} - D_{12}D_{12} = DD_{12,12} \tag{3.66}$$

we obtain

$$fh + zh = zg \tag{3.67}$$

which, upon substitution of (3.58) and (3.59) leads to

$$g = z(z+f)[(1-f)^2 - \lambda g^2]^{-1}[2f - 2f^2 + g + \lambda g + 2\lambda g^2]. \tag{3.68}$$

We have then two algebraic equations, (3.60) and (3.68), in two unknowns, f and g, the monoblock generating functions. Insertion of the solution to this set of equations in (3.32) gives the generating function in terms of which the partition function is given by (3.18). The equations, however, are complicated enough so that an explicit solution to them has not yet been found. The expression for G in terms of f and g can be given, after the matrix inverse to $(\mathbb{1} - \boldsymbol{TF})$ has been found. The result is

$$G = N\Delta^{-1}, \tag{3.69}$$

where

$$\begin{aligned} N = {} & 4f - 2(4-\lambda)f^2 - 8\lambda f^3 - 2\lambda(\lambda-6)f^4 + 12\lambda fg + 2\lambda(\lambda^2+5\lambda+3)fg^2 \\ & - 14\lambda f^2 g + 4\lambda^2(1-\lambda)fg^3 + 2\lambda^2(\lambda-12)f^2g^2 + 2\lambda(\lambda+4)g^2 \\ & + 2\lambda^2(\lambda^2 - 2\lambda + 13)g^3 - 4\lambda^3(\lambda-4)g^4, \end{aligned} \tag{3.70}$$

$$\Delta = [(1+\lambda g)^2 - \lambda f^2][(2f+\lambda g - 1)^2 - \lambda(f+2g)^2]. \tag{3.71}$$

It is also possible to find the first several terms of the series expansions for f and g by solving (3.60) and (3.68) iteratively. This gives

$$f=z^3\{1+2z^3(1+\lambda)+8z^5\lambda+z^6(7+12\lambda+5\lambda^2)$$
$$+z^8(68\lambda+40\lambda^2)+z^{10}(192\lambda+280\lambda^2+40\lambda^3)\dots\}\,, \tag{3.71}$$
$$g=z^5\{2+2z^2(2+\lambda)+z^3(6+4\lambda)+z^4(6+8\lambda+2\lambda^2)$$
$$+z^5(20+38\lambda+4\lambda^2)+\dots\}\,. \tag{3.72}$$

In (3.71) the coefficient of z^9 does not vanish, but we have not computed it. The coefficients of z, z^2, z^4, and z^7 do vanish however. Substitution of (3.71) and (3.72) into (3.69) then gives

$$G=4z^3+z^6(16+10\lambda)+56z^8\lambda+z^9(76+96\lambda+28\lambda^2)$$
$$+z^{10}(144\lambda+64\lambda^2)+z^{11}(640\lambda+288\lambda^2)$$
$$+z^{12}(392+1056\lambda+770\lambda^2+156\lambda^3)$$
$$+z^{13}(1920\lambda+2464\lambda^2+320\lambda^3)+\dots\,. \tag{3.73}$$

The coefficients in (3.73) have been checked by computing directly the trace of the relevant power of $(-\ell S\otimes \boldsymbol{a}+xT^{-1}\otimes \boldsymbol{b}+xT\otimes \boldsymbol{c})$.

IV. Concluding Remarks

We have constructed a lattice which is homogeneous under the modular group of fractional linear transformations Γ. This was done by taking a point on the upper half of the complex plane, acting on it with Γ to obtain an infinite set of points, and drawing bonds between every point and its nearest neighbors. The result was a cactus lattice with hexagonal leaves joined along bonds (not by the vertices).

The generating function for close-packed dimer configurations on this lattice was computed by the Pfaffian method. For this it was necessary to orient the lattice, thus breaking its symmetry. Invariance under a group was restored by considering $SL(2,\mathbb{Z})$, which is related to Γ through $\Gamma\approx SL(2,\mathbb{Z})/\{\pm\mathbb{1}\}$. This idea may be used also to recover the generating function for dimers on a finite toroidal lattice [6] first found by Kasteleyn [9]. There, one has to orient the toroidal lattice, which is invariant under $\mathbb{Z}/N\mathbb{Z}\times\mathbb{Z}/M\mathbb{Z}$. The orientation breaks the symmetry but one recovers it by going to a larger group like $\mathbb{Z}/2N\mathbb{Z}\times\mathbb{Z}/M\mathbb{Z}$. The advantage of having symmetry under a group lies in the fact that it is then possible to express the operator whose determinant gives the generating function in terms of the regular representation of certain elements of the group. A renormalization of the trace allowed then to compute the generating function directly in the thermodynamic limit, without the intermediate step of computing it first for a finite lattice and taking the limit afterwards. The logarithm of the generating function was then developed in a power series whose coefficients were the number of a certain class of words reucible to the identity. They were found by solving the problem of counting the number of words of a fixed length in the generators of a group which is the free product of two cyclic groups. The series converges for $\left|\frac{x}{\ell}\right| < \frac{1}{2}$, and for $\frac{x}{\ell} = -\frac{1}{2}$ there is a singularity.

One may envisage two further problems that may be attacked with the approach given here. One is the dimer problem on lattices homogeneous under the triangle groups, whose presentation is [5] $\langle A, B; A^n, B^m, (AB)^p \rangle$. The question of counting the number of words reducible to the identity, however, is likely to become extremely involved. The other is the dimer problem on finite lattices homogeneous under the finite group Γ/Γ_p, where Γ_p is the principal congruence subgroup of level p. These lattices however are embedded on surfaces whose genus g grows as the lattice grows. Since we need 4^g Pfaffians for the calculation of the dimer generating function of such a lattice, this will be a serious difficulty.

Finally, an Ising model was constructed on the original lattice. Evaluation of the partition function was reduced to a dimer problem on an associated lattice following the general prescription of Fisher. The procedure was the same as for the simpler lattice of Section II, although the details were considerably more involved, especially the evaluation of the coefficients of the series development for the free energy. This function was then given as a rational function of the solutions of a system of two algebraic equations. A preliminary study of the series expansion (3.73) seems to indicate that this free energy is analytic for $|z| < \frac{1}{2}$. We hope to come back to this point in a future paper.

Acknowledgements. We wish to thank J. Lascoux, G. Dell'Antonio and E. Lieb for their critical comments. One of us (F.L.) gratefully acknowledges the hospitality of the Seattle Summer Institute in Theoretical Physics, where part of this work was done.

References

1. Kasteleyn, P. W.: J. Math. Phys. **4**, 1332 (1963)
2. McCoy, B., Wu, T. T.: The two-dimensional Ising Model. Cambridge, Mass.: Harvard University Press (1973)
3. Fisher, M. E.: J. Math. Phys. **7**, 1776 (1966)
4. The standard reference concerning the modular group is Klein, F., Fricke, R.: Vorlesungen über die Theorie der elliptischen Modulfunktionen. Leipzig: Teubner 1890. A concise summary of its main properties is given in Chapter I of Gunning, R. C.: Lectures on modular forms. Princeton: Princeton University Press 1962. Facts concerning fractional linear transformations, hyperbolic geometry and discontinuous groups may be found in Chapter 3 of Siegel, C. L.: Topics in complex function theory. London: Wiley 1971
5. See Magnus, W.: Non-euclidean tesselations and their groups. New York-London: Academic Press 1973
6. Lund, F.: Ph. D. Thesis. Princeton University (unpublished)
7. Onsager, L.: Phys. Rev. **65**, 117 (1944)
8. For this and other topics concerning combinatorial group theory, see Magnus, W., Karrass, A., Solitar, D.: Combinatorial group theory. London: Wiley 1966. In particular the fundamental problem of deciding in a finite number of steps whether a word in a group defines the identity element of the group itself, was formulated by Dehn, M.: Math. Ann. **71**, 116 (1911)
9. Kasteleyn, P. W.: Physica **27**, 1209 (1961)

Communicated by G. Gallavotti

Received May 5, 1976