The dimer model

Richard Kenyon

Contents

1 Overview **2**
1.1 General lattices 2

2 Dimer definitions **3**
2.1 Electrical field 5
2.2 Random tilings 5
2.3 Facets 9
2.4 Measures 9
2.4.1 Phases 10

3 Gibbs measures **10**
3.1 Definition 10
3.2 Ergodic Gibbs measures 12

4 Kasteleyn theory **12**
4.1 Kasteleyn weighting 12
4.2 Kasteleyn matrix 13
4.3 Local statistics 15

5 Partition function **16**
5.1 Rectangle 16
5.2 Torus 17
5.3 Partition function 19

6 General graphs **19**
6.1 The amoeba of P 21
6.2 Phases of EGMs 22

6.3 Harnack curves 26
6.4 Example 26

1 Overview

Our goal is to study the planar dimer model and the associated random interface model. There are a number of references to the dimer model, including a few short survey articles, which can be found in the bibliography.

The dimer model is a classical statistical mechanics model, the first results on which were obtained by Kasteleyn and Temperley/Fisher in the 1960's who computed the partition function. Later works by Fisher, Stephenson, Temperley, Blöte, Hilhorst, Percus and many others contributed to understanding correlations and other properties.

Essentially all of these authors considered the model on either $\mathbb{Z}^2$ or the hexagonal lattice. While these cases already contain much of interest, only when one generalizes to other lattices (with larger fundamental domains) does one see the "complete" picture. For example gaseous phases and semi-frozen phases (defined below) only occur in this more general setting.

Our goal here is to describe recent results on the planar, periodic, bipartite dimer model, obtained through joint work with Andrei Okounkov and Scott Sheffield [6, 4]. We solve the dimer model in the general setting of a planar bipartite periodic lattice.

1.1 General lattices

The dimer model on a general periodic bipartite lattice has a surprisingly rich behavior. We can in fact achieve a reasonably complete and satisfying theory for dimers in this setting: we have a good understanding of the set of translation-invariant Gibbs states, the influence of boundary conditions, and the scaling limits of fluctuations, which are described by a massless Gaussian free field.

Other statistical mechanical models such as the 6-vertex model can be defined in similar generality but have not been solved except for very special situations such as $\mathbb{Z}^2$ with periodic weights. One would expect that a detailed study of these models on general periodic lattices would also yield a wealth of new behavior, but at the moment the tools for this study have been lacking.

Dimer models on non-bipartite lattices, of which the Ising model is an example, could also potentially be studied on general periodic lattices; Pfaffian methods exist for their solution. However at the moment no one has undertaken such a study. I believe such a project would be very enlightening about the nature of the Ising model and more generally the nonbipartite dimer model.

2 Dimer definitions

A **dimer covering**, or **perfect matching**, of a graph is a subset of edges which covers every vertex exactly once, that is, every vertex is the endpoint of exactly one edge. See Figure 1. Le $M(\mathcal{G})$ be the set of dimer covers of the graph $\mathcal{G}$.

In these lectures we will deal only with **bipartite planar** graphs. A graph is bipartite when the vertices can be colored black and white in such a way that each edge connects vertices of different colors. Alternatively, this is equivalent to each cycle having even length.

Let $\mathcal{E}$ be a function on the edges of a finite graph $\mathcal{G}$, defining the energy associated to having a dimer on that edge. To a dimer cover m we associate a total energy $\mathcal{E}(m)$ to be the sum of the energies the edges covered by dimers. The partition function for a finite graph $\mathcal{G}$ is then

$$Z = Z(\mathcal{G}, \mathcal{E}) = \sum_{m \in M(\mathcal{G})} e^{-\mathcal{E}(m)}.$$

The corresponding Boltzmann measure is a probability measure assigning a dimer cover m a probability $\mu(m) = \frac{1}{Z} e^{-\mathcal{E}(m)}$.

Note that if we add a constant energy $\mathcal{E}_0$ to each edge containing a given vertex, then the energy of any dimer cover changes by $\mathcal{E}_0$ and so the associated Boltzmann measure on $M(\mathcal{G})$ does not change. Therefore we define two energy functions $\mathcal{E}, \mathcal{E}'$ to be equivalent, $\mathcal{E} \sim \mathcal{E}'$ if one can be obtained from the other by a sequence of such operations. It is not hard to show that $\mathcal{E} \sim \mathcal{E}'$ if and only if the **alternating sums** along faces are equal: given a face with edges $e_1, e_2, \ldots, e_{2k}$ in cyclic order, the sum $\mathcal{E}_1 - \mathcal{E}_2 + \mathcal{E}_3 \cdots - \mathcal{E}_{2k}$ and $\mathcal{E}'_1 - \mathcal{E}'_2 + \mathcal{E}'_3 \cdots - \mathcal{E}'_{2k}$ (which we call alternating sums) must be equal. It is possible to interpret these alternating sums as magnetic fluxes through the faces but we will not take this point of view here.

Kasteleyn showed how to enumerate the dimer covers of any planar graph, and in fact compute the partition function for any $\mathcal{E}$: the partition function is a Pfaffian of an associated signed adjacency matrix. The random interface interpretation we will discuss is valid only for bipartite graphs, so we will only be concerned here with bipartite graphs. In this case one can replace the Pfaffian with a determinant. There are many open problems involving dimer coverings of non-bipartite planar graphs, which at present we do not have tools to attack. More on this matter later.

Our prototypical examples are the dimer models on $\mathbb{Z}^2$ and the honeycomb graph. These are equivalent to, respectively, the domino tiling model (tilings with 2×1 rectangles) and the "lozenge tiling" model (tilings with 60° rhombi). The case when $\mathcal{E}$ is identically 0 on edges is already quite interesting. See Figures 1 and 2. A less simple case is the case of $\mathbb{Z}^2$ in which $\mathcal{E}$

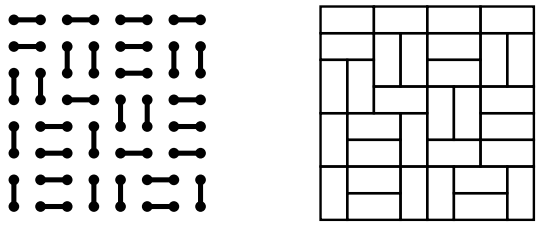

Figure 1:

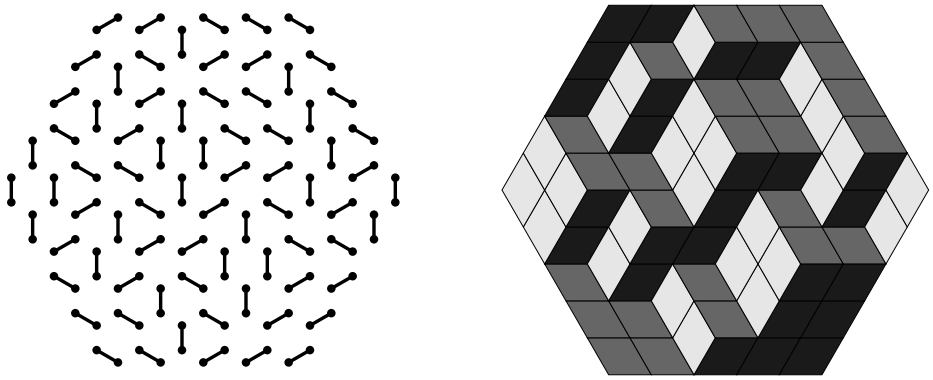

Figure 2:

is periodic with period n, that is, translations in $n\mathbb{Z}^2$ leave $\mathcal{E}$ invariant, but no larger group does.

2.1 Electrical field

Let $\mathcal{G}$ be a planar, bipartite, periodic graph. This means that $\mathcal{G}$ is a planar bipartite weighted graph on which translations in $\mathbb{Z}^2$ (or some other rank-2 lattice) act by energy-preserving and color-preserving isomorphisms. Here by **color-preserving** isomorphisms, we mean, isomorphism which maps white vertices to white and black vertices to black. Note for example that for the graph $\mathcal{G} = \mathbb{Z}^2$ with nearest neighbor edges, the lattice generated by $(2, 0)$ and $(1, 1)$ acts by color-preserving isomorphisms, but $\mathbb{Z}^2$ itself does not act by color-preserving isomorphisms. So the fundamental domain (in this case) contains two vertices, one white and one black.

For simplicity we will assume our periodic graphs are embedded so that the lattice of energy- and color-preserving isomorphisms is $\mathbb{Z}^2$, so that we can describe a translation using a pair of integers.

Let $\vec{E} = (E_x, E_y)$ be a constant electric field in $\mathbb{R}^2$. If we regard each dimer as a dipole, for example if we assign white vertices charge $+1$ and black vertices a charge -1, then each dimer gets an additional energy due to its interaction with $\vec{E}$; this energy is just $\vec{E} \cdot (\mathrm{b} - \mathrm{w})$ for a dimer with vertices w and b.

On a finite graph $\mathcal{G}$, the Boltzmann measure on $M(\mathcal{G})$ is unaffected by the presence of $\vec{E}$, because for every dimer cover the energy due to $\vec{E}$ is

$$\sum_m \vec{E} \cdot (\mathrm{b} - \mathrm{w}) = \sum_{\mathrm{b}} \vec{E} \cdot \mathrm{b} - \sum_{\mathrm{w}} \vec{E} \cdot \mathrm{w}$$

where the sums on the right are over all the black and all the white vertices, respectively. On (nonplanar) graphs with periodic boundary conditions, or on infinite graphs, $\vec{E}$ will influence the measure in a nontrivial way.

2.2 Random tilings

Look at a larger domino picture and the lozenge picture, Figures 3 and 4. These are both uniform random tilings of the corresponding regions, that is, they are chosen from the distribution in which all tilings are equally weighted. In the first case there are about e^{455} possible domino tilings and

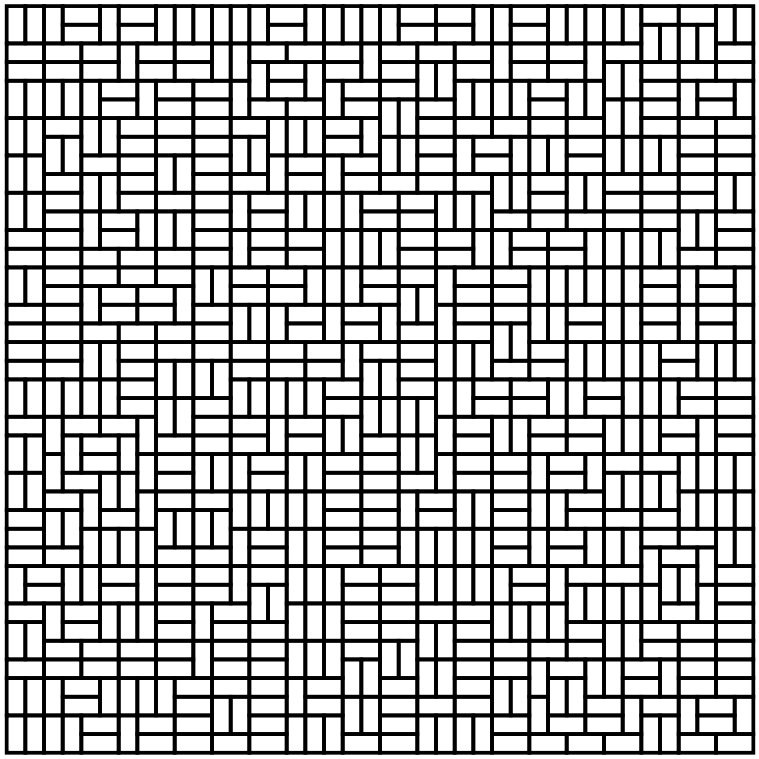

Figure 3:

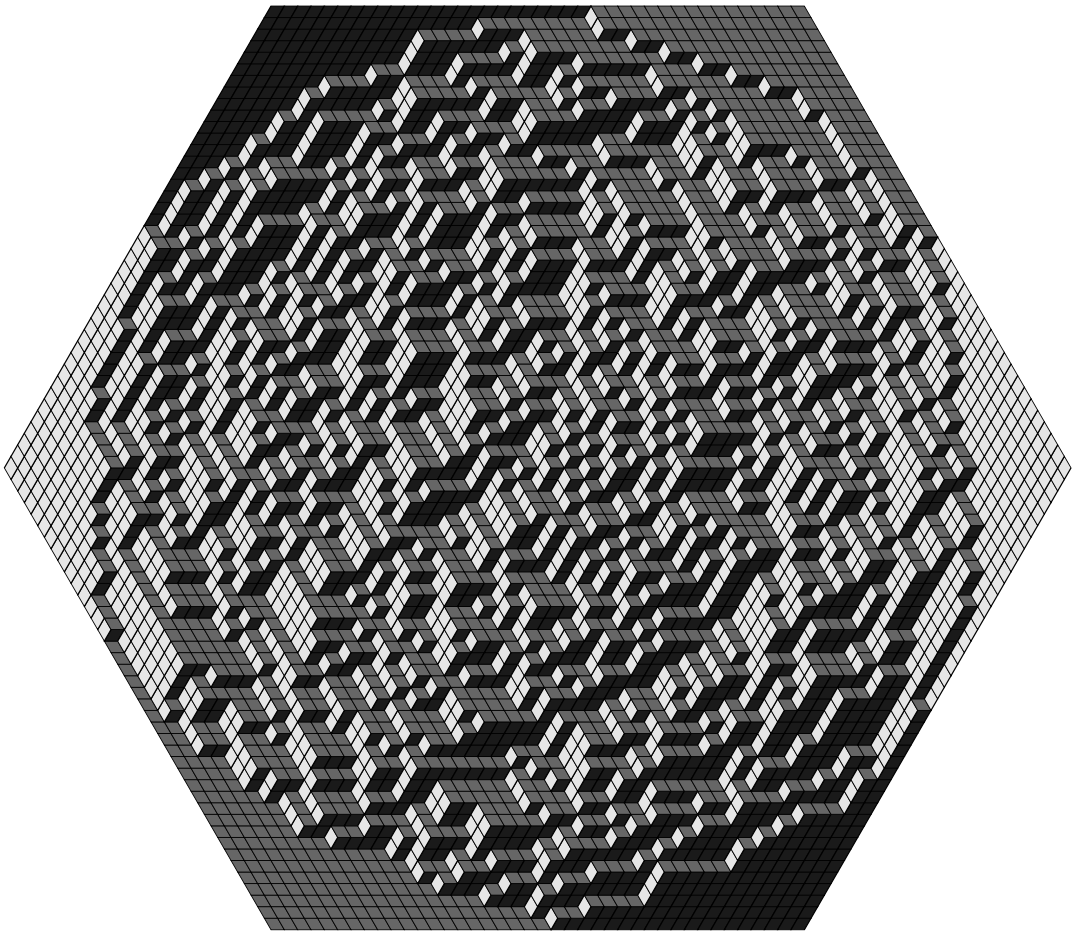

Figure 4:

in the second, about e^{1255} lozenge tilings [1] These two pictures clearly display some very different behavior. The first picture appears homogeneous while in the second, the densities of the tiles of a given orientation vary throughout the region. The reason for this behavior is the boundary height function.

A lozenge tiling of a simply connected region is the projection along the $(1,1,1)$-direction of a piecewise linear surface in $\mathbb{R}^3$. This surface has pieces which are integer translates of the sides of the unit cube. Such a surface which projects injectively in the $(1,1,1)$-direction is called a **stepped surface** or **skyscraper surface**. Random lozenge tilings are therefore random stepped surfaces. A random lozenge tiling of a fixed region as in Figure 4 is a random stepped surface spanning a fixed boundary curve in $\mathbb{R}^3$.

Domino tilings can also be interpreted as random tilings. Here the third coordinate is harder to visualize, but see Figure 5 for a definition. It is not

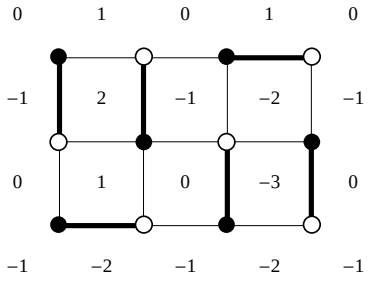

Figure 5: The height is integer-valued on the faces and changes by ± 1 across an unoccupied edge; the change is $+1$ when crossing so that a white vertex is on the left.

hard to see that dimers on any bipartite graph can be interpreted as random surfaces.

When thought of as random surfaces, these models have a **limit shape phenomenon**. This says that, for a fixed boundary curve in $\mathbb{R}^3$, or sequence of converging boundary curves in $\mathbb{R}^3$, if we take a random stepped surface on finer and finer lattices, then with probability tending to 1 the random surface will lie close to a fixed non-random surface, the so-called **limit shape**. So this limit shape surface is not just the average surface but the only surface you will see if you take an extremely fine mesh...the measure is concentrating as the mesh size tends to zero on the delta-measure at this surface. Not

[1] How do you pick a random sample from such a large space?

surprisingly, this limit shape is described by a variational principle: it is the surface which minimizes a "surface tension" integral.

In Figure 3, the height function along the boundary is (with small variation) linear, and a random surface corresponding to a random domino tiling of the square is flat (with fluctuations of smaller order). For Figure 4, however, the boundary height function is not flat: it consists of 6 sides of a large cube. Consequently the stepped surface spanning it is forced to bend; this bending is responsible for the varying densities of tiles throughout the figure (indeed, the tile densities determine the slope of the surface locally).

2.3 Facets

One thing to notice about lozenge tiling in figure 4 is the presence of regions near the vertices of the hexagon where the lozenges are aligned. This phenomenon persists in the limit of small mesh size and in fact, in the limit shape surface there is a **facet** near each corner, where the limit shape is planar. This is a phenomenon which does not occur in one dimension. The limit shapes for dimers generally contain facets and smooth (in fact analytic) curved regions separating these facets. It is remarkable that one can solve through analytic means, for reasonably general boundary conditions, for the entire limit shape, including the locations of the facets. We won't discuss this computation here but refer the reader to [4].

2.4 Measures

What do we see if we zoom in to a point in figure 4? That is, consider a sequence of such figures with the same fixed boundary but decreasing mesh size. Pick a point in the hexagon and consider the configuration restricted to a small window around that point, window which gets smaller as the mesh size goes to zero. One can imagine for example a window of side $\sqrt{\epsilon}$ when the mesh size is ϵ. This gives a sequence of random tilings of (relative to the mesh size) larger and larger domains, and in the limit (assuming that a limit of these "local measures" exists) we will get a random tiling of the plane.

We will see different types of behaviors, depending on which point we zoom in on. If we zoom in to a point in the facet, we will see in the limit a boring figure in which all tiles are aligned. This is an example of a measure on lozenge tilings of the plane which consists of a delta measure at a single tiling. This measure is an (uninteresting) example of an ergodic Gibbs measure (see

definition below). If we zoom into a point in the non-frozen region, one can again ask what limiting measure on tilings of the plane is obtained. One of the important open problems is to understand this limiting measure, in particular to prove that the limit exists. Conjecturally it exists and only depends on the slope of the average surface at that point and not on any other property of the boundary conditions. For each possible slope (s, t) we will define below a measure $\mu_{s,t}$, and the **local statistics conjecture** states that, for any fixed boundary, $\mu_{s,t}$ is the measure which occurs in the limit at any point where the limit shape has slope (s, t). For certain boundary conditions this has been proved [3].

2.4.1 Phases

Gibbs measures on $M(\mathcal{G})$ come in three types, or **phases**, depending on the fluctuations of a typical (for that measure) surface. Suppose we fix the height at a face near the origin to be zero. A measure is said to be a **frozen phase** if the height fluctuations are finite almost surely, that is, the fluctuation of the surface away from its mean value is almost surely bounded, no matter how far away from the origin you are. A measure is said to be in a **liquid phase** if the fluctuations have variance increasing with increasing distance, that is, the variance of the height at a point tends to infinity almost surely for points farther and farther from the origin. Finally a measure is said to be a **gaseous phase** if the height fluctuations are unbounded but the variance of the height at a point is bounded independently of the distance from the origin.

We'll see that for uniform lozenge or domino tilings we can have both liquid and frozen phases but not gaseous phases. An example of a graph for which we have a all three phases is the uniform square octagon dimer model, see Figure 6. In general the classification of phases depends on algebraic properties of the underlying graph and edge weights, as we'll see.

3 Gibbs measures

3.1 Definition

Let $X = X(\mathcal{G})$ be the set of dimer coverings of a graph $\mathcal{G}$, possibly infinite, with edge weight function w. Recall the definition of the Boltzmann probability measure on $X(\mathcal{G})$ when $\mathcal{G}$ is finite: a dimer covering has probability

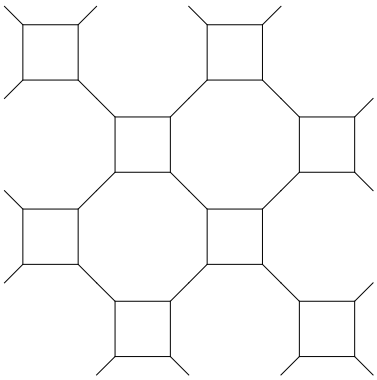

Figure 6:

proportional to the product of its edge weights. When $\mathcal{G}$ is infinite, this definition will of course not work. For an infinite graph $\mathcal{G}$, a probability measure on X is a **Gibbs measure** if it is a weak limit of Boltzmann measures on a sequence of finite subgraphs of $\mathcal{G}$ filling out $\mathcal{G}$. By this we mean, for any finite subset of $\mathcal{G}$, the probability of any particular configuration occurring on this finite set converges. That is, the probabilities of **cylinder sets** converge. Here a cylinder set is a subset of $X(\mathcal{G})$ consisting of all coverings containing a given finite subset of edges.

For a sequence of Boltzmann measures on increasing graphs, the limiting measure may not exist, but subsequential limits will always exist. The limit may not be unique, however; that is, it may depend on the approximating sequence of finite graphs. This will be the case for dimers and it is this non-uniqueness which makes the dimer model interesting.

The important property of Gibbs measures is the following. Let A be a cylinder set defined by the presence of a finite set of edges. Let B be another cylinder set defined by a different set of edges but using the same set of vertices. Then, for the approximating Boltzmann measures, the ratio of the measures of these two cylinder sets is equal to the ratio of the product of the edge weights in A and the product of the edge weights in B. This is true along the finite growing sequence of graphs and so in particular the same is true for the limiting Gibbs measures. In fact this property *characterizes* Gibbs measures: given a finite set of vertices, the measure on this set of vertices conditioned on the exterior (that is, integrating over the configuration in the exterior) is just the Boltzmann measure on this finite set.

3.2 Ergodic Gibbs measures

For a periodic graph $\mathcal{G}$, a **translation-invariant measure** on $X(\mathcal{G})$ is simply one for which the measure of a subset of $X(\mathcal{G})$ in invariant under the translation-isomorphism action.

The **slope** (s, t) of a translation-invariant measure is the expected height change in the $(1, 0)$ and $(0, 1)$ directions, that is, s is the expected height change between a face and its translate by $(1, 0)$ and t is the expected height change between a face and its translate by $(0, 1)$.

A **ergodic** Gibbs measure, or **EGM**, is one in which translation-invariant sets have measure 0 or 1. Typical examples of translation-invariant sets are: the set of coverings which contain a translate of a particular pattern.

Theorem 1 (Sheffield [10])**.** *For the dimer model on a periodic planar bipartite periodically edge-weighted graph, for each slope (s, t) for which there exists a translation-invariant measure, there exists a unique EGM $\mu_{s,t}$.*

In particular we can classify EGMs by their slopes. The existence is not hard to establish by taking limits of Boltzmann measures on larger and larger tori while restricted height changes (h_x, h_y), see below. The uniqueness is much harder; we won't discuss this here.

4 Kasteleyn theory

We show how to compute the number of dimer coverings of any bipartite planar graph using the KTF (Kasteleyn-Temperley-Fisher) technique. While this technique extends to nonbipartite planar graphs, we will have no use for this generality here.

4.1 Kasteleyn weighting

A **Kasteleyn weighting** of a planar bipartite graph is a choice of sign for each undirected edge with the property that each face with 0 mod 4 edges has an odd number of $-$ signs and each face with 2 mod 4 edges has an even number of $-$ signs.

In certain circumstances it will be convenient to use complex numbers of modulus 1 rather than signs ± 1. In this case the condition is that the alternating product of edge weights (the first, divided by the second, times

the third, and so on) around a face is negative real or positive real depending on whether the face has 0 or 2 mod 4 edges.

This condition appears mysterious at first but we'll see why it is important below[2]. It is not hard to see (using a homological argument) that any two Kasteleyn weightings are gauge equivalent: they can be obtained one from the other by a sequence of operations consisting of multiplying all edges at a vertex by a constant.

The existence of a Kasteleyn weighting is also easily established using spanning trees (put signs +1 on the edges of a spanning tree; the remaining edges have determined sign). We leave this fact to the reader, as well as the proof of the following (easily proved by induction)

Lemma 2. *Given a cycle of length $2k$ enclosing ℓ points, the alternating product of signs around this cycle is $(-1)^{1+k+\ell}$.*

Note finally that for the (edge-weighted) honeycomb graph, all faces have 2 mod 4 edges and so no signs are necessary in the Kasteleyn weighting.

4.2 Kasteleyn matrix

A **Kasteleyn matrix** is a weighted, signed adjacency matrix of the graph $\mathcal{G}$. Given a Kasteleyn weighting of $\mathcal{G}$, define a $|B|\times|W|$ matrix K by $K(\mathrm{b},\mathrm{w}) = 0$ if there is no edge from w to b, otherwise $K(\mathrm{b},\mathrm{w})$ is the Kasteleyn weighting times the edge weight $w(\mathrm{bw}) = e^{-\mathcal{E}(\mathrm{b},\mathrm{w})}$.

For the graph in Figure 7 with Kasteleyn weighting indicated, the Kasteleyn matrix is

$$\begin{pmatrix} a & 1 & 0 \\ 1 & -b & 1 \\ 0 & 1 & c \end{pmatrix}.$$

Note that gauge transformation corresponds to pre- or post-multiplication

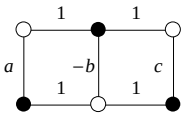

Figure 7:

[2]The condition might appear more natural if we note that the alternating product is required to be $e^{\pi i N/2}$ where N is the number of triangles in a triangulation of the face

of K by a diagonal matrix.

Theorem 3 ([1, 11])**.** $Z = |\det K|$.

In the example, the determinant is $-a - c - abc$.

Proof. If K is not square the determinant is zero and there are no dimer coverings (each dimer covers one white and one black vertex). If K is a square $n \times n$ matrix, We expand

$$\det K = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) K(\mathrm{b}_1, \mathrm{w}_{\sigma(1)}) \dots K(\mathrm{b}_n, \mathrm{w}_{\sigma(1)}). \tag{1}$$

Each term is zero unless it pairs each black vertex with a unique neighboring white vertex. So there is one term for each dimer covering, and the modulus of this term is the product of its edge weights. We need only check that the signs of the nonzero terms are all equal.

Let us compare the signs of two different nonzero terms. Given two dimer coverings, we can draw them simultaneously on $\mathcal{G}$. We get a set of doubled edges and loops. To convert one dimer covering into the other, we can take a loop and move every second dimer (that is, dimer from the first covering) cyclically around by one edge so that they match the dimers from the second covering. When we do this operation for a single loop of length $2k$, we are changing the permutation σ by a k-cycle. Note that by Lemma 2 above the sign change of the edge weights in the corresponding term in (1) is ± 1 depending on whether $2k$ is 2 mod 4 or 0 mod 4 (since ℓ is even there), exactly the same sign change as occurs in $\operatorname{sgn}(\sigma)$. These two sign changes cancel, showing that these two coverings (and hence any two coverings) have the same sign. □

An alternate proof of this theorem for honeycomb graphs—which avoids using Lemma 2—goes as follows: if two dimer coverings differ only on a single face, that is, an operation of the type in Figure 8 converts one cover into the other, then these coverings have the same sign in the expansion of the determinant, because the hexagon flip changes σ by a 3-cycle which is an even permutation. Thus it suffices to notice that any two dimer coverings can be obtained from one another by a sequence of hexagon flips. This can be seen using the lozenge tiling picture since applying a hexagon flip is equivalent to adding or subtracting a cube from the stepped surface. Any

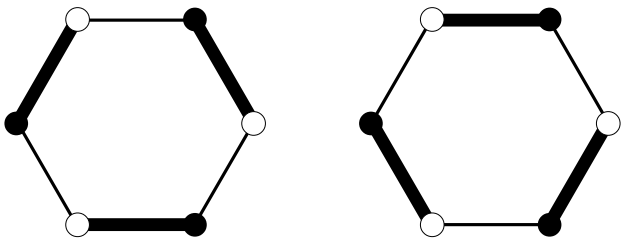

Figure 8:

two surfaces with the same connected boundary can be obtained from one another by adding and/or subtracting cubes.

While there is a version of Theorem 3 (using Pfaffians) for non-bipartite planar graphs, there is no corresponding sign trick for nonplanar graphs in general (the exact condition is that a graph has a Kasteleyn weighting if and only if it does not have $K_{3,3}$ as minor [7]).

4.3 Local statistics

There is an important corollary to Theorem 3:

Corollary 4 ([2]). *Given a set of edges $X = \{\mathrm{w}_1\mathrm{b}_1, \ldots, \mathrm{w}_k\mathrm{b}_k\}$, the probability that all edges in X occur in a dimer cover is*

$$\left(\prod_{i=1}^{k} K(\mathrm{b}_i, \mathrm{w}_i)\right) \det(K^{-1}(\mathrm{w}_i, \mathrm{b}_j))_{1\le i,j\le k}.$$

The proof uses the Jacobi Lemma that says that a minor of a matrix A is $\det A$ times the complementary minor of A^{-1}.

The advantage of this result is that the probability of a set of k edges being present is only a $k \times k$ determinant, independently of the size of the graph. One needs only be able to compute K^{-1}. In fact the corollary is valid even for infinite graphs, once K^{-1} has been appropriately defined.

5 Partition function

5.1 Rectangle

Here is the simplest example. Assume mn is even and let $\mathcal{G}_{m,n}$ be the $m \times n$ square grid. Its vertices are $V = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ and edges connect nearest neighbors. Let $Z_{m,n}$ be the partition function for dimers with edge weights 1. This is just the number of dimer coverings of $\mathcal{G}_{m,n}$.

A Kasteleyn weighting is obtained by putting weight 1 on horizontal edges and $i = \sqrt{-1}$ on vertical edges. Since each face has four edges, the condition in (4.1) is satisfied.[3]

The corresponding Kasteleyn matrix K is an $mn/2 \times mn/2$ matrix (recall that K is a $|W| \times |B|$ matrix). The eigenvalues of the matrix $\tilde{K} = \begin{pmatrix} 0 & K \\ K^t & 0 \end{pmatrix}$ are in fact simpler to compute. Let $z = e^{i\pi j/(m+1)}$ and $w = e^{i\pi k/(n+1)}$. Then the function

$$f_{j,k}(x, y) = (z^x - z^{-x})(w^y - w^{-y}) = -4 \sin(\frac{\pi j x}{m+1}) \sin(\frac{\pi k y}{n+1})$$

is an eigenvector of $\tilde{K}$ with eigenvalue $z + \frac{1}{z} + i(w + \frac{1}{w})$. To see this, check that

$$\lambda f(x, y) = f(x+1, y) + f(x-1, y) + if(x, y+1) + if(x, y-1)$$

when (x, y) is not on the boundary of $\mathcal{G}$, and also true when f is on the boundary assuming we extend f to be zero just outside the boundary, i.e. when $x = 0$ or $y = 0$ or $x = m+1$ or $y = n+1$.

As j, k vary in $(1, m) \times (1, n)$ the eigenfunctions $f_{j,k}$ are independent (a well-known fact from Fourier series). Therefore we have a complete diagonalization of the matrix $\tilde{K}$, leading to

$$Z_{m,n} = \left(\prod_{j=1}^{m} \prod_{k=1}^{n} 2 \cos \frac{\pi j}{m+1} + 2i \cos \frac{\pi k}{n+1} \right)^{1/2}. \tag{2}$$

Here the square root comes from the fact that $\det \tilde{K} = (\det K)^2$.

[3] A weighting gauge equivalent to this one, and using only weights ± 1, is to weight alternate columns of vertical edges by -1 and all other edge $+1$. This was the weighting originally used by Kasteleyn [1]; our current weighting (introduced by Percus [9]) is slightly easier for our purposes.

Note that if m, n are both odd then this expression is zero because of the term $j = (m+1)/2$ and $k = (n+1)/2$ in (2).

For example $Z_{8,8} = 12988816$. For large m, n we have

$$\lim_{m,n\to\infty} \frac{1}{mn} \log Z_{m,n} = \frac{1}{2\pi^2} \int_0^\pi \int_0^\pi \log(2\cos\theta + 2i\cos\phi) d\theta\, d\phi$$

which can be shown to be equal to G/π, where G is Catalan's constant $G = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \ldots$.

We can also write an explicit expression for the inverse Kasteleyn matrix and its limit. See below, Theorem 6.

5.2 Torus

A graph on a torus does not in general have a Kasteleyn weighting. However we can still make a "local" Kasteleyn matrix whose determinant can be used to count dimer covers.

Rather that show this in general, let us work out a detailed example. Let H_n be the honeycomb lattice on a torus, as in Figure 9, which shows H_3. It has n^2 black vertices and n^2 white vertices, and $3n^2$ edges. Weight edges

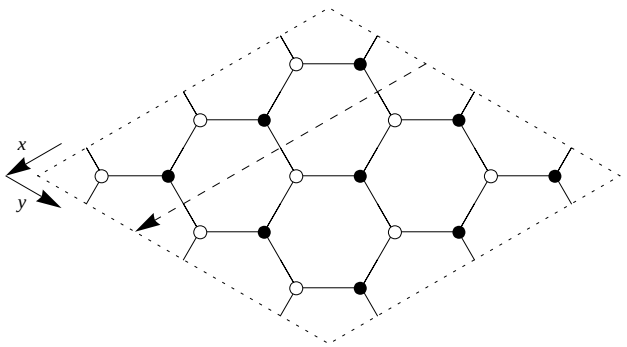

Figure 9: Honeycomb graph on a torus.

a, b, c according to direction: a the horizontal, b the NW-SE edges and c the NE-SW edges. Let $\hat{x}$ and $\hat{y}$ be the directions indicated. Given a dimer cover m of H_n, the number N_a, N_b, N_c of a, b and c type edges is a multiple of n; for example there are the same number of b-type edges crossing any NE-SW dashed line as in the Figure. Let $h_x(m)$ and $h_y(m)$ be $1/n$ times the number of

b-, respectively c-type edges, respectively. That is $h_x = N_b/n$ and $h_y = N_c/n$. These quantities measure the height change of m on a path winding around the torus, that is, thinking of a dimer cover as a stepped surface, there is a "locally defined" height function which changes by an additive constant on a path which winds around the torus; this additive constant is h_x in the $\hat{x}$ direction and h_y in the $\hat{y}$ direction.

Let K_n be the weighted adjacency matrix of H_n. That is $K(\mathrm{b}, \mathrm{w}) = 0$ if there is no edge from b to w and otherwise $K(\mathrm{b}, \mathrm{w}) = a, b$ or c according to direction.

From the proof of Theorem 3 we can see that $\det K$ is a weighted, signed sum of dimer coverings. Our next goal is to determine the signs.

Lemma 5. *The sign of a dimer covering in* $\det K$ *depends only on its height change* (h_x, h_y) *modulo* 2. *Three of the four parity classes gives the same sign and the fourth has the opposite sign.*

Proof. Let N_b, N_c be the number of b and c type edges in a cover. If we take the union of a covering with the covering consisting of all a-type edges, we can compute the sign of the cover by the product of the sign changes when shifting along each loop. The number of loops is $q = \mathrm{GCF}(h_x, h_y)$ and each of these has homology class $(h_y/q, h_x/q)$ (note that the b edges contribute to h_x but to the $(0, 1)$ homology class). The length of each loop is $\frac{2n}{q}(h_x + h_y)$ and so each loop contributes sign $(-1)^{1+\frac{n}{q}(h_x+h_y)}$, for a total of $(-1)^{q+n(h_x+h_y)}$. Note that q is even if and only if h_x and h_y are both even. So if n is even then the sign is -1 unless $(h_x, h_y) \equiv (0, 0) \bmod 2$. If n is odd the sign is $+1$ unless $(h_x, h_y) \equiv (1, 1) \bmod 2$. □

In particular $\det K$, when expanded as a polynomial in a, b and c, has coefficients which count coverings with particular height changes. For example for n odd we can write

$$\det K = \sum (-1)^{h_x h_y} C_{h_x,h_y} a^{n(n-h_x-h_y)} b^{nh_x} c^{nh_y}$$

since $h_x h_y$ is odd exactly when h_x, h_y are both odd.

Define $Z_{00} = Z_{00}(a, b, c)$ to be this expression. Define

$$Z_{10}(a, b, c) = Z_{00}(a, be^{\pi i/n}, c),$$

$$Z_{01}(a, b, c) = Z_{00}(a, b, ce^{\pi i/n}),$$

and

$$Z_{11}(a,b,c) = Z_{00}(a, be^{\pi i/n}, ce^{\pi i/n}).$$

Then one can verify that, when n is odd,

$$Z = \frac{1}{2}(Z_{00} + Z_{10} + Z_{01} - Z_{11}) \tag{3}$$

which is equivalent to Kasteleyn's expression [1] for the partition function. The case n even is similar and left to the reader.

5.3 Partition function

Dealing with a torus makes computing the determinant much easier, since the graph now has many translational symmetries. The matrix K commutes with translation operators and so can be simultaneously diagonalized with them. Simultaneous eigenfunctions of the horizontal and vertical translations (and K) are the exponential functions (also called Bloch-Floquet eigenfunctions) $f_{z,w}(x,y) = z^{-x}w^{-y}$ where $z^n = 1 = w^n$. There are n choices for z and n for w leading to a complete set of eigenfunctions. The corresponding eigenvalue for K is $a + bz + cw$.

In particular

$$\det K = Z_{00} = \prod_{z^n=1} \prod_{w^n=1} a + bz + cw.$$

This leads to

$$Z_{10} = \prod_{z^n=-1} \prod_{w^n=1} a + bz + cw.$$

$$Z_{01} = \prod_{z^n=1} \prod_{w^n=-1} a + bz + cw.$$

$$Z_{11} = \prod_{z^n=-1} \prod_{w^n=-1} a + bz + cw.$$

6 General graphs

As one might expect, this can be generalized to other planar, periodic, bipartite graphs. Representative examples are the square grid and the square-octagon grid (Figure 6).

For simplicity, we're going to deal mainly with honeycomb dimers with periodic energy functions. Since it is possible, after simple modifications, to embed any other periodic bipartite planar graph in a honeycomb graph (possibly increasing the size of the period), we're actually not losing any generality. We'll also illustrate our calculations in an example in section 6.4.

So let's start with the honeycomb with a periodic weight function ν on the edges, periodic with period ℓ in directions $\hat{x}$ and $\hat{y}$. As in section 5.2 for any n we are led to an expression for the partition function for the $n\ell \times n\ell$ torus:

$$Z(H_{n\ell}) = \frac{1}{2}(Z_{00} + Z_{10} + Z_{01} - Z_{11})$$

where

$$Z_{\tau_1,\tau_2} = \prod_{z^n=(-1)^{\tau_1}} \prod_{w^n=(-1)^{\tau_2}} P(z,w),$$

and where $P(z,w)$ is a polynomial, the **characteristic polynomial**, with coefficients depending on the energy function $\mathcal{E}$. The polynomial $P(z,w)$ is the determinant of $K(z,w)$, the Kasteleyn matrix for the $\ell \times \ell$ torus (with appropriate extra weights z and w on edges crossing fundamental domains); $K(z,w)$ is just the restriction of K to the space of Bloch-Floquet eigenfunctions (eigenfunctions for the operators of translation by $\ell\hat{x}, \ell\hat{y}$) with multipliers z, w. We can also consider $P(z,w)$ to be the signed weighted sum of dimer covers of the $\ell \times \ell$ torus consisting of a single $\ell \times \ell$ fundamental domain (with the appropriate extra weight $(-1)^{h_x h_y} z^{h_x} w^{h_y}$).

The algebraic curve $P(z,w) = 0$ is called the **spectral curve** of the dimer model, since it describes the spectrum of the K operator on the whole weighted honeycomb graph.

Many of the physical properties of the dimer model are encoded in the polynomial P.

Theorem 6. *The partition function per fundamental domain Z satisfies*

$$\log Z := \lim_{n\to\infty} \frac{1}{n^2} \log Z(H_{n\ell}) = \frac{1}{(2\pi i)^2} \int_{|z|=|w|=1} \log P(z,w) \frac{dz}{z} \frac{dw}{w}.$$

The inverse Kasteleyn matrix limit satisfies

$$K^{-1}(\mathrm{b}, \mathrm{w}) = \frac{1}{(2\pi i)^2} \int_{|z|=|w|=1} \frac{z^x w^y \mathbf{Q}(z,w)_{b,w}}{P(z,w)} \frac{dz}{z} \frac{dw}{w}$$

where $\mathbf{Q}(z,w)/P(z,w) = K^{-1}(z,w)$*,* (x,y) *is the translation between the fundamental domain containing* b *to that containing* w*, and* $(b,w) \equiv (\mathrm{b},\mathrm{w}) \bmod \mathbb{Z}^2$*.*

Note that the values of K^{-1} in the limit are linear combinations of Fourier coefficients of $1/P$ (since $\mathbf{Q}$ is a matrix of polynomials in z, w).

Recall the definition of h_x, h_y for a dimer cover of H_n. Any translation-invariant Gibbs state has an **expected slope**, which is the expected amount that the height function changes under translation by one fundamental domain in the $\hat{x}$- or $\hat{y}$- direction. It is a theorem of Sheffield that there is a unique translation-invariant ergodic Gibbs measure on dimer covers for each possible slope (s,t). The set of possible slopes for translation-invariant Gibbs measures can be naturally identified with the Newton polygon $\mathcal{N}(P)$ of P, that is, the convex hull in $\mathbb{R}^2$ of the set of integer exponents:

$$\mathcal{N}(P) = \mathrm{cvxhull}(\{(i,j) : z^i w^i \text{ is a monomial of } P\}).$$

Note that there is a **periodic** dimer cover with slope $(h_x, h_y)/n$ for every $(h_x, h_y) \in \mathcal{N}(p)$.

6.1 The amoeba of P

The **amoeba** of an algebraic curve $P(z,w) = 0$ is the set

$$\mathbb{A}(P) = \{(\log|z|, \log|w|) \in \mathbb{R}^2 \ : \ P(z,w) = 0\}.$$

In other words, it is a projection to $\mathbb{R}^2$ of the zero set of P in $\mathbb{C}^2$, sending (z,w) to $(\log|z|, \log|w|)$. Note that for each point $(X,Y) \in \mathbb{R}^2$, the amoeba contains (X,Y) if and only if the torus $\{(z,w) \in \mathbb{C}^2 \ : \ |z| = e^X, |w| = e^Y\}$ intersects $P(z,w) = 0$.

The amoeba has "tentacles" which are regions where $z \to 0, \infty$, or $w \to 0, \infty$. Each tentacle is asymptotic to a line $\alpha \log|z| + \beta \log|w| + \gamma = 0$. These tentacles divide the complement of the amoeba into a certain number of unbounded complementary components. There may be bounded complementary components as well.

The **Ronkin function** of P is the function on $\mathbb{R}^2$ given by

$$R(X,Y) = \frac{1}{(2\pi i)^2} \int_{|z|=e^X} \int_{|w|=e^Y} \log P(z,w) \frac{dz}{z} \frac{dw}{w}.$$

The following facts are standard; see [8]. The Ronkin function of P is convex in $\mathbb{R}^2$, and linear on each component of the complement of $\mathbb{A}(P)$[4].

The gradient $\nabla R(X,Y)$ takes values in $\mathcal{N}(P)$. The Ronkin function is in fact Legendre dual to a function $\sigma(s,t)$ defined on $\mathcal{N}(P)$. That is, we define

$$\sigma(s,t) = \max_{X,Y}(-R(X,Y) + sX + tY).$$

This σ is the **surface tension** function for the dimer model.

This duality $(X,Y) \mapsto (s,t)$ maps each component of the complement of $\mathbb{A}(P)$ to a single point of the Newton polygon $\mathcal{N}(P)$. This is a point with integer coordinates. Unbounded complementary components correspond to integer points on the boundary of $\mathcal{N}(P)$; bounded complementary components correspond to integer points in the interior of $\mathcal{N}(P)$[5].

See Figure 13 for an example of an amoeba of a spectral curve.

See Figures 10 and 11 for plots of σ and R in the case of uniform honeycomb dimers ($P(z,w) = 1 + z + w$).

6.2 Phases of EGMs

Note that the Ronkin function can also be written

$$R(X,Y) = \frac{1}{(2\pi i)^2} \int_{|z|=1} \int_{|w|=1} \log P(e^X z, e^Y w) \frac{dz}{z} \frac{dw}{w}.$$

As such the Ronkin function is the negative of the free energy as a function of the electric field:

$$\log Z(E_x, E_y) = -R(E_x, E_y).$$

In particular the amoeba of P can be thought of as the phase diagram for the dimer model as a function of the electric field $\vec{E} = (E_x, E_y)$. The amoeba boundaries are places where the partition function is not analytic as a function of $\vec{E}$. It remains to see what the different phases "mean".

Sheffield's theorem says that to every point (s,t) in the Newton polygon of P there is a unique ergodic Gibbs measure $\mu_{s,t}$ with that average slope (and these are all the ergodic Gibbs measures).

[4]Thus the complementary components of $\mathbb{A}(P)$ are convex.

[5]Not every integer point in $\mathcal{N}(P)$ may correspond to a complementary component of $\mathbb{A}$.

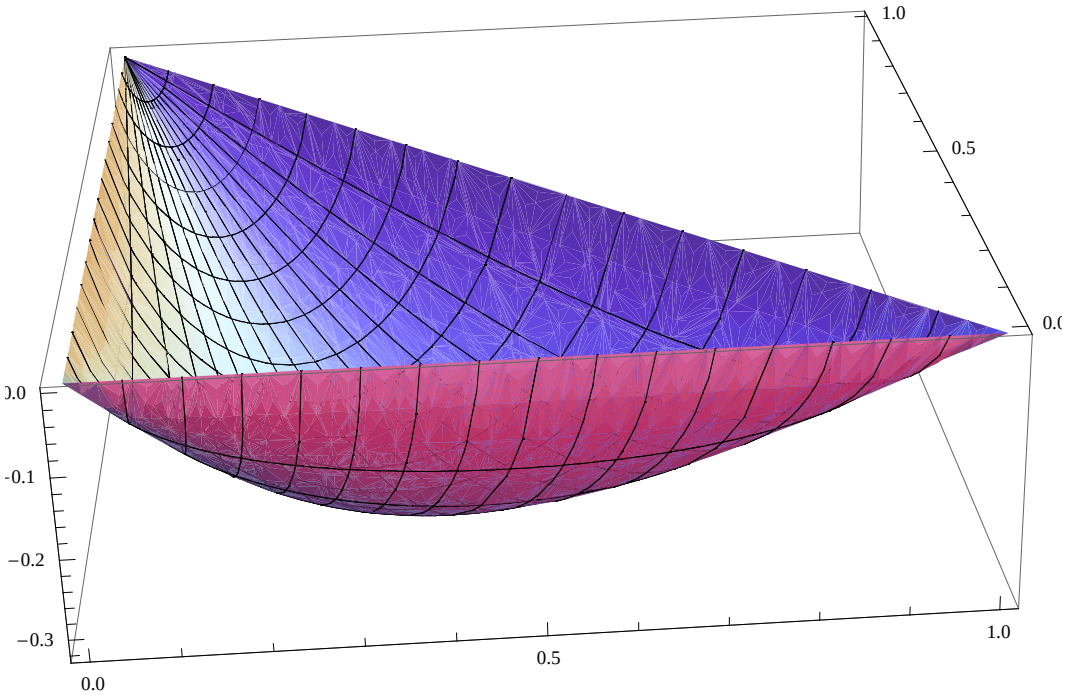

Figure 10:

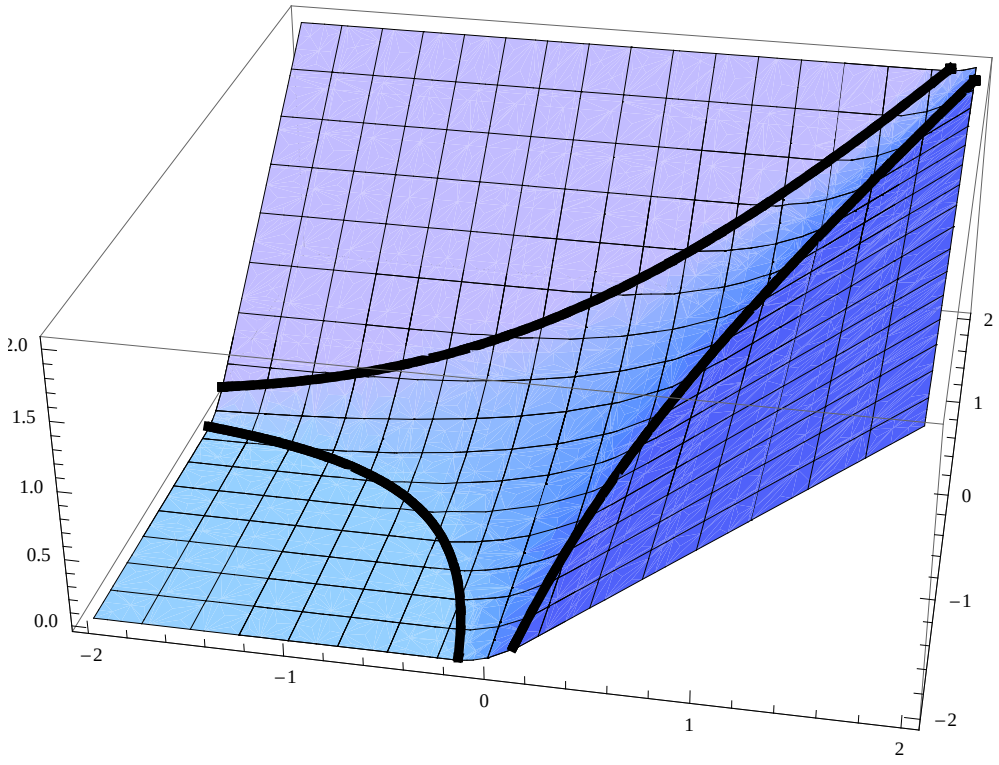

Figure 11:

To study the different phases, we study the decay of correlations for a fixed measure, that is, for a given $\mu_{s,t}$, how does the probability of two edges in a dimer cover compare with the product of their probabilities? The covariance of two edges is by definition

$$\mathbf{Cov}(e_1, e_2) = \mathbf{Pr}(e_1 \& e_2) - \mathbf{Pr}(e_1)\mathbf{Pr}(e_2).$$

This can be written in terms of the inverse Kasteleyn matrix as (see section 4.3)

$$\begin{aligned}\mathbf{Cov}(e_1, e_2) &= K_{\mathrm{w}_1\mathrm{b}_1} K_{\mathrm{w}_2\mathrm{b}_2} \det \begin{pmatrix} (K^{-1})_{\mathrm{b}_1\mathrm{w}_1} & (K^{-1})_{\mathrm{b}_1,\mathrm{w}_2} \\ (K^{-1})_{\mathrm{b}_2,\mathrm{w}_1} & (K^{-1})_{\mathrm{b}_2,\mathrm{w}_2} \end{pmatrix} - K_{\mathrm{w}_1\mathrm{b}_1} K_{\mathrm{w}_2\mathrm{b}_2} (K^{-1})_{\mathrm{b}_1,\mathrm{w}_1} (K^{-1})_{\mathrm{b}_2,\mathrm{w}_2} \\ &= w(\mathrm{w}_1\mathrm{b}_1) w(\mathrm{w}_2\mathrm{b}_2) |K^{-1}(\mathrm{b}_1, \mathrm{w}_2) K^{-1}(\mathrm{b}_2, \mathrm{w}_1)|.\end{aligned}$$

The local statistics for a measure $\mu_{s,t}$ are determined by the inverse Kasteleyn matrix $K_{\vec{E}}^{-1}$, where $\vec{E} = (E_x, E_y)$ is related to (s,t) via Legendre duality, $\nabla R(E_x, E_y) = (s,t)$. As discussed in Theorem 6, values of K^{-1} are (linear combinations of) Fourier coefficients of $1/P(z,w)$, or, more precisely, Fourier coefficients of $1/P(e^{E_x}z, e^{E_y}w)$. In particular, if $P(e^{E_x}z, e^{E_y}w)$ has no zeroes on the unit torus $\{|z| = |w| = 1\}$, then $1/P$ is analytic and so its Fourier coefficients decay exponentially fast. The corresponding covariance will decay exponentially fast in the separation between the edges. On the other hand if $P(z,w)$ has simple zeroes on the unit torus, its Fourier coefficients decay linearly, and the covariance of two edges will decay quadratically in the separation.

This then is the condition which separates the different phases of the dimer model. If a slope (s,t) is chosen so that (E_x, E_y) is in (the closure of) an unbounded component of the complement of the amoeba, then certain Fourier coefficients of $1/P$ (those contained in the appropriate dual cone) will vanish. This is enough to ensure that $\mu_{s,t}$ is in a frozen phase: covariances of edges more than one fundamental domain apart are identically zero (this requires some argument which we are not going to give here). For slopes (s,t) for which (E_x, E_y) is in (the closure of) a bounded component of the complement of the amoeba, the edge-edge covariances decay exponentially fast (in all directions). This is enough to show that the height fluctuations have bounded variance, and we are in a gaseous (but not frozen, since the correlations are nonzero) phase.

In the remaining case, (E_x, E_y) is in the interior of the amoeba, and P has zeroes on a torus. It is a beautiful and deep fact that the spectral

curves arising in the dimer model are special in that P has either two zeros, both simple, or a single node[6] over each point in the interior of $\mathbb{A}(P)$. As a consequence[7] in this case the edge-edge covariances decay quadratically (quadratically in generic directions—there may be directions where the decay is faster). It is not hard to show that this implies that the height variance between distant points is unbounded, and we are in a liquid phase.

6.3 Harnack curves

Plane curves with the property described above, that they have at most two zeros (both simple) or a single node on each torus $|z| = constant, |w| = constant$ are called **Harnack curves**, or simple Harnack curves. They were studied classically by Harnack and more recently by Mikhalkin and others [8].

The simplest definition is that a Harnack curve is a curve $P(z, w) = 0$ with the property that the map from the zero set to the amoeba $\mathbb{A}(P)$ is at most 2 to 1 over $\mathbb{A}(P)$. It will be 2 to 1 with a finite number of possible exceptions (the integer points of $\mathcal{N}(P)$) on which the map may be 1 to 1.

Theorem 7 ([6, 5])**.** *The spectral curve of a dimer model is a Harnack curve. Conversely, every Harnack curve arises as the spectral curve of some periodic bipartite weighted dimer model.*

In [5] it was also shown, using dimer techniques, that the areas of complementary components of $\mathbb{A}(P)$ and distances between tentacles are global coordinates for the space of Harnack curves with a given Newton polygon.

6.4 Example

Let's work out a detailed example illustrating the above theory. We'll take dimers on the square grid with 3×2 fundamental domain (invariant under the lattice generated by $(0, 2)$ and $(3, 1)$). Take fundamental domain with vertices labelled as in Figure 12—we chose those weights to give us enough parameters (5) to describe all possible gauge equivalence classes of weights on the 3×2 fundamental domain. Letting z be the eigenvalue of translation in

[6] A node is point where $P = 0$ looks locally like the product of two lines, e.g. $P(x, y) = x^2 - y^2 + O(x, y)^3$ near $(0, 0)$.

[7] We showed what happens in the case of a simple pole already. The case of a node is fairly hard.

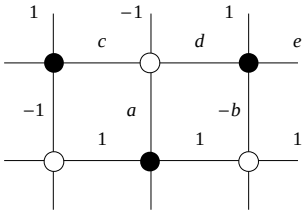

Figure 12:

direction $(3,1)$ and w be the eigenvalue of translation by $(0,2)$, the Kasteleyn matrix (white white vertices corresponding to rows and black to columns) is

$$K = \begin{pmatrix} -1+\frac{1}{w} & 1 & \frac{e}{z} \\ c & a-w & d \\ \frac{z}{W} & 1 & -b+\frac{1}{w} \end{pmatrix}.$$

We have

$$P(z,w) = \det K(z,w) = 1+b+ab+bc+d+e-\frac{1+a+ab+c+d+ae}{w}+\frac{a}{w^2}-bw+\frac{ce}{z}+d\frac{z}{w}.$$

This can of course be obtained by just counting dimer covers of $\mathbb{Z}^2/\{(0,2),(3,1)\}$ with these weights, and an appropriate factor $(-1)^{ij+j}z^i w^j$ when there are edges going across fundamental domains. Let's specialize to $b=2$ and all other edges of weight 1. The

$$P(z,w) = 9 - 2w + \frac{1}{w^2} - \frac{7}{w} + \frac{1}{z} + \frac{z}{w}.$$

The amoeba is shown in Figure 13. There are two gaseous components, corresponding to EGMs with slopes $(0,0)$ and $(0,-1)$. The four frozen EGMs correspond to slopes $(1,-1),(0,1),(0,-2)$ and $(-1,0)$. All other slopes are liquid phases.

If we take $c=2, b=\frac{1}{2}(3\pm\sqrt{3})$ and all other weights 1 then there remains only one gaseous phase; the other gas "bubble" in the amoeba shrinks to a point and becomes a node in $P=0$. If we take all edges 1 then both gaseous phases disappear; we just have the uniform measure on dominos again

References

[1] P. Kasteleyn, *Graph theory and crystal physics*, 1967 Graph Theory and Theoretical Physics pp. 43–110 Academic Press, London

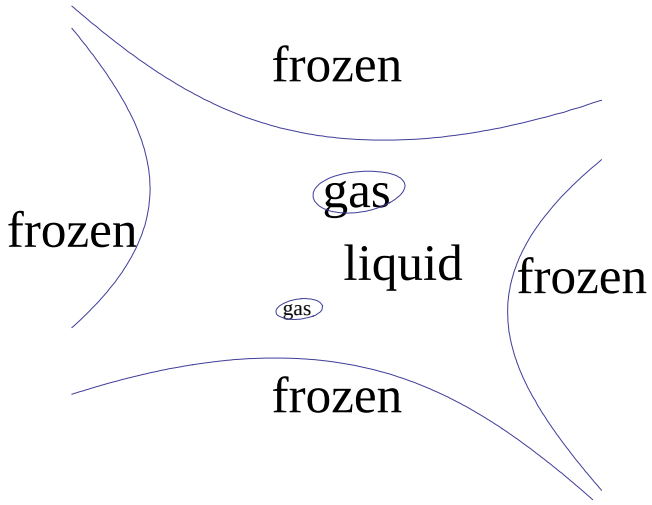

Figure 13:

[2] R. Kenyon, *Local statistics of lattice dimers*, Ann. Inst. H. Poincaré, Probabilités **33**(1997), 591–618.

[3] R. Kenyon, Height fluctuations in the honeycomb dimer model. *Comm. Math. Phys.*

[4] R. Kenyon, A. Okounkov, Dimers, Limit shapes and the complex Burgers equation

[5] R. Kenyon, A. Okounkov, Dimers and Harnack curves

[6] R. Kenyon, A. Okounkov, S. Sheffield

[7] L. Lovasz, Plummer,

[8] G. Mikhalkin,

[9] Percus,

[10] S. Sheffield,

[11] W. Temperley, M. Fisher,