

A GENERALISATION OF THE HONEYCOMB DIMER MODEL TO HIGHER DIMENSIONS

BY PIET LAMMERS

*Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,
p.g.lammers@statslab.cam.ac.uk*

Linde, Moore and Nordahl introduced a generalisation of the honeycomb dimer model to higher dimensions. The purpose of this article is to describe a number of structural properties of this generalised model. First, it is shown that the samples of the model are in one-to-one correspondence with the perfect matchings of a hypergraph. This leads to a generalised Kasteleyn theory: the partition function of the model equals the Cayley hyperdeterminant of the adjacency hypermatrix of the hypergraph. Second, we prove an identity which relates the covariance matrix of the random height function directly to the random geometrical structure of the model. This identity is known in the planar case but is new for higher dimensions. It relies on a more explicit formulation of Sheffield's *cluster swap* which is made possible by the structure of the honeycomb dimer model. Finally, we use the special properties of this explicit cluster swap to give a new and simplified proof of strict convexity of the surface tension in this case.

1. Introduction.

1.1. *Background.* Random models on shift-invariant Euclidean graphs such as the square lattice and the hexagonal lattice form a well-known subject of study in both combinatorics and statistical physics [7]. There are several integrable models which allow for a quantitative analysis. In the integrable setting, the focus is on deriving quantitative results concerning (asymptotics of) partition functions and correlation functions. Examples of such models are the Ising model [20, 25], ice-type models [18, 22, 26], and the dimer model [8, 24], see also [1]. These quantitative estimates in turn imply qualitative results, for example (non)uniqueness of shift-invariant Gibbs measures, and (when the samples are height functions) strict convexity of the surface tension. Sometimes it is possible to derive qualitative results even in the absence of quantitative estimates. Georgii [6] provides an excellent overview for the theory of Gibbs measures for general (nonintegrable) models, and Sheffield [21] derives many key results for models of height functions in the gradient setting, including strict convexity of the surface tension in any dimension.

The focus of this paper is a natural generalisation of the hexagonal dimer model (Figure 1(c)) to arbitrary dimension. The generalised model first appeared in the work of Linde, Moore and Nordahl [19]. It also belongs to the category of models under consideration in the thesis of Sheffield [21]. A great deal is known about the original two-dimensional dimer model: we mention three pivotal developments. The mathematical study of dimers was initiated by Kasteleyn, Temperley, and Fisher. Kasteleyn [8] and Temperley and Fisher [24] independently calculated the number of perfect matchings of an $n \times m$ grid or, equivalently, the number of domino tilings of an $n \times m$ rectangle. Kasteleyn [9, 10] later showed that the

Received June 2019; revised May 2020.

MSC2020 subject classifications. Primary 82B41, 60K35; secondary 60F10.

Key words and phrases. Dimer model, random tilings, height functions, variational principle, limit shapes, surface tension strict convexity.

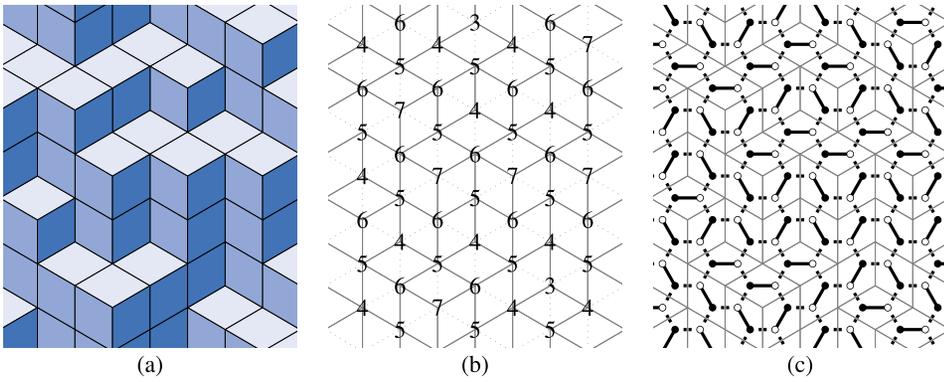


FIG. 1. Dimension $d = 2$; projections onto H of several representations: (a) as a stepped surface or lozenge tiling, (b) as a height function on the simplicial lattice (X^d, E^d) , (c) as a dimer cover or perfect matching of the dual hexagonal lattice.

number of perfect matchings of any bipartite planar graph equals the determinant of a matrix that is closely related to the adjacency matrix of the concerned graph. Cohn, Kenyon and Propp [3] proved the variational principle for domino tilings; the scope of their article includes the hexagonal dimer model. Remarkably, a closed-form solution for the surface tension is found, something that is not to be expected in higher dimensions. Their derivation of the closed-form formula relies on a bijection between dimer configurations and height functions, together with an original application of the Kasteleyn theory. Kenyon, Okounkov, and Sheffield [14] establish a bijection between the set of accessible slopes and the set of ergodic Gibbs measures. They furthermore classify the ergodic Gibbs measures into three categories (frozen, liquid and gaseous) which describe qualitatively the behaviour of the random surface. Their paper contains many more qualitative and quantitative results.

1.2. *The honeycomb dimer model in dimension $d \geq 2$.* Let (X^d, E^d) denote the graph obtained from the square lattice \mathbb{Z}^{d+1} by identifying vertices which differ by an integer multiple of the vector $\mathbf{n} := \mathbf{e}_1 + \dots + \mathbf{e}_{d+1}$. This graph is called the *simplicial lattice*; its vertices are equivalence classes of vertices of the square lattice. Let Ω denote the set of functions $f : X^d \rightarrow \mathbb{Z}$ which have the property that $f(\mathbf{0}) \in (d+1)\mathbb{Z}$ and $f([\mathbf{x} + \mathbf{e}_i]) - f([\mathbf{x}]) \in \{-d, 1\}$ for any $\mathbf{x} \in \mathbb{Z}^{d+1}$ and $1 \leq i \leq d+1$. Functions in Ω are called *height functions*—see Figure 1(b). The set Ω is in bijection with the set of *stepped surfaces* in \mathbb{R}^{d+1} . Informally, a stepped surface is a union of unit hypercubes with integer coordinates, such that each hypercube is well supported, and such that there is no overhang. Stepped surfaces are related directly to the three-dimensional interpretation of the familiar picture of lozenge tilings for $d = 2$; see Figure 1(a). Each stepped surface is furthermore associated with a *tiling* of the hyperplane orthogonal to \mathbf{n} ; this tiling is essentially obtained by projecting the exposed faces of the hypercubes of the stepped surface onto this hyperplane. These bijections are all introduced in the work of Linde, Moore and Nordahl [19].

In this article, we are interested in the model of uniformly random height functions whenever fixed or periodic boundary conditions are enforced. For fixed boundary conditions, it is also possible to consider more general Boltzmann measures, in the spirit of the classical planar dimer model. The purpose of this article is to point out a number of structural properties of the generalised model which lead to new results.

1.3. *The double dimer model and the cluster swap.* For the double dimer model, one superimposes two dimer configurations of the same graph. The union of two such dimer covers

decomposes into a number of isolated edges which appear in both dimer covers, called *double edges*, and a number of closed loops of even length, where each edge of the loop is contained in exactly one of the original dimer covers. If the distribution of the two dimer configurations is uniformly random, subject say to fixed boundary conditions, then the orientation of each loop is uniformly random in its two states. More precisely, this means that for each loop, one can flip a fair coin to decide if each dimer should change the configuration that it belongs to, without changing the distribution of the product measure. Many results have been obtained for the double dimer model: see the work of Kenyon and Pemantle [11, 15] for the relevant literature.

Sheffield introduces *cluster swapping* in the seminal monograph *Random surfaces* [21]. The technique employs the same idea of considering two configurations—height functions, in this case—at once, then identifying and resampling independent structures. The cluster swap applies to *simply attractive potentials*, that is, models of height functions which are induced by a convex nearest-neighbour potential. The setup is much more intricate than for the original double dimer model. To compensate for the possible change in potential, one first identifies a ferromagnetic Ising model, then achieves independence through the Edwards–Sokal representation [5] which in turn derives from the Swendsen–Wang update [23]. The cluster swap applies directly to the heights of the two height functions; it is not an operation on their gradients.

1.4. *Main results.* First, we develop a new construction, the *cluster boundary swap*, for the generalised model. This cluster boundary swap is entirely analogous to the resampling operation in the double dimer model. In particular, the difference of two uniformly random height functions decomposes as a geometrical structure consisting of boundaries—which generalise loops—and double edges, and, conditional on this structure, a number of fair coin flips, one for the orientation of each boundary. Moreover, the operation directly manipulates the gradients of the two height functions, which works to our advantage. The name *cluster boundary swap* is chosen intentionally, as it should be considered a special case of the cluster swap, adapted and optimised for the special geometrical structure that is present in the generalisation of the honeycomb dimer model to higher dimensions. The author is not aware of a similar construction for any other model in dimension $d > 2$; in particular, no generalisation is known for the dimer model on the two-dimensional square lattice.

Second, we use the cluster boundary swap to obtain an identity which relates the covariance matrix of the random height function f to the geometrical structure of the model. We prove that the variance of f at a vertex $\mathbf{x} \in X^d$ is exactly $\frac{1}{2}(d+1)^2$ times the expectation of the number of boundaries separating \mathbf{x} from infinity in the product measure. The identity makes sense for fixed boundary conditions only, but it does apply to general Boltzmann measures (which includes the uniform probability measure). Similarly, for $\mathbf{x}, \mathbf{y} \in X^d$, we prove that the covariance between f at \mathbf{x} and \mathbf{y} equals $\frac{1}{2}(d+1)^2$ times the number of boundaries that separate both \mathbf{x} and \mathbf{y} from infinity. This identity is new and was previously known, to our knowledge, only for two-dimensional dimer models.

Third, we prove that the set of tilings, as introduced in [19], are in one-to-one correspondence with the perfect matchings of a hypergraph which is the natural dual of the simplicial lattice (X^d, E^d) . This hypergraph is $d!$ -regular and $d!$ -partite, in the sense that there is a partition of its vertex set into $d!$ -parts, such that each hyperedge contains exactly one vertex in each part. We derive a generalised Kasteleyn theory: we prove that the partition function of any Boltzmann measure equals the Cayley hyperdeterminant of the adjacency hypermatrix of this hypergraph. It is well known that no Kasteleyn weighting is necessary for the Kasteleyn theory on the hexagonal lattice. We prove the same here: we may take the adjacency hypermatrix without modifications as our Kasteleyn hypermatrix. Unfortunately, this does not by

the knowledge of the author lead to an explicit formula for the surface tension, nor to an expression for dimer-dimer correlations in terms of an inverse of the Kasteleyn hypermatrix.

Finally, we use the special properties of the cluster boundary swap—relative to the more general cluster swap—to greatly streamline the proof in [21] for strict convexity of the surface tension. Note that strict convexity of the surface tension for $d = 2$ was first derived in [3] by means of an explicit calculation. Strict convexity of the surface tension is important, because it implies that the model is stable on a macroscopic scale. The macroscopic behaviour of the model is described by a large deviations principle, which in turn implies a variational principle—these are due to generic arguments which do not require that the surface tension is strictly convex. If the surface tension is strictly convex, however, then the rate function in the large deviations principle has a unique minimizer, and consequently the random height function concentrates around a unique limit shape. See [3, 12, 13] for the variational principle in the original dimer setting.

1.5. Structure of the article. Section 2 formally introduces the model, and gives an overview of the several representations of each sample. These constructions derive directly from [19]. Section 3 describes the probability measures under consideration, by enforcing fixed or periodic boundary conditions. We construct the cluster boundary swap in Section 4. In Section 5, we state and prove the identity for the covariance structure of the random height function, and Section 6 contains the generalised Kasteleyn theory. The purpose of the remainder of the article is to motivate and prove the result of strict convexity of the surface tension. In Section 7, we introduce gradient Gibbs measures, which play an important role in the proof, but which are also interesting in their own right. Section 8 motivates the result: it introduces the surface tension, and describes how it is related to the large deviations principle and the variational principle. Strict convexity of the surface tension is finally proven in Section 9.

2. Stepped surfaces, tilings, height functions. Linde, Moore and Nordahl [19] observed that each sample of the model has three natural representations. The purpose of this section is to give an overview of these representations, and to state or derive some basic properties. For details, refer to [19]. Throughout this article, $d \geq 2$ denotes the fixed dimension that we work in.

2.1. Stepped surfaces. Informally, a stepped surface is a union of unit hypercubes with corners in \mathbb{Z}^{d+1} . The hypercubes must be properly stacked without overhang—recall for this notion the familiar picture of lozenge tilings, which corresponds to $d = 2$. If a hypercube is present at some coordinate $\mathbf{x} \in \mathbb{Z}^{d+1}$, then we require the presence of a hypercube at $\mathbf{x} - \mathbf{e}_i$ for every $1 \leq i \leq d + 1$; this indeed ensures that every hypercube is well supported, and excludes overhang.

The formal definition is as follows. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}$, then write $\mathbf{x} \leq \mathbf{y}$ whenever $x_i \leq y_i$ for all i . A set $A \subset \mathbb{R}^{d+1}$ is called *closed under \leq* whenever $\mathbf{x} \in A$ and $\mathbf{y} \leq \mathbf{x}$ implies $\mathbf{y} \in A$. Write $L(A)$ for the closure of a set $A \subset \mathbb{R}^{d+1}$ under \leq , that is,

$$L(A) := \{\mathbf{y} \in \mathbb{R}^{d+1} : \mathbf{y} \leq \mathbf{x} \text{ for some } \mathbf{x} \in A\}.$$

A *stepped surface* is a strict nonempty subset of \mathbb{R}^{d+1} of the form $L(A)$ for some $A \subset \mathbb{Z}^{d+1}$. Let Ψ denote the set of stepped surfaces. If F is a stepped surface, then write $V(F)$ for the set $\partial F \cap \mathbb{Z}^{d+1}$. The set $V(F)$ should be thought of as the discrete boundary of F . It is a simple exercise to work out that $F = L(V(F))$. In particular, each stepped surface is characterised by this discrete boundary.

2.2. *The height function of a stepped surface.* Consider a stepped surface F . The height function associated to F is essentially a function that has the discrete boundary $V(F)$ of F as its graph. The value of this function at each vertex $\mathbf{x} \in V(F)$ is given by $\mathbf{x}_1 + \dots + \mathbf{x}_{d+1} = (\mathbf{x}, \mathbf{n})$, where \mathbf{n} is the vector $\mathbf{e}_1 + \dots + \mathbf{e}_{d+1}$ and (\cdot, \cdot) the natural inner product. This value is also called the *height* of the vertex \mathbf{x} .

Write (X^d, E^d) for the graph obtained from the square lattice \mathbb{Z}^{d+1} after identifying all vertices which differ by an integer multiple of \mathbf{n} . In particular, each vertex $[\mathbf{x}] \in X^d$ is an equivalence class of the form $[\mathbf{x}] := \mathbf{x} + \mathbb{Z}\mathbf{n}$ for some $\mathbf{x} \in \mathbb{Z}^{d+1}$. The $2d + 2$ neighbours of $[\mathbf{x}]$ are of the form $[\mathbf{x} \pm \mathbf{e}_i]$. The graph (X^d, E^d) is called the *simplicial lattice*. It is a simple exercise to derive from the definition of a stepped surface that $|[\mathbf{x}] \cap V(F)| = 1$ for any $[\mathbf{x}] \in X^d$. To derive this, observe first that $\mathbf{x} + \mathbb{R}\mathbf{n}$ intersects ∂F once because F is closed under \leq , then note that this point of intersection has integral coordinates whenever \mathbf{x} does.

The *height function* associated to F is given by the function

$$f : X^d \rightarrow \mathbb{Z}, [\mathbf{x}] \mapsto (\mathbf{y}, \mathbf{n}) \quad \text{where } \{\mathbf{y}\} = [\mathbf{x}] \cap V(F).$$

This is the desired parametrisation of $V(F)$; the set $V(F)$ is equal to

$$V(f) := \{\mathbf{x} \in \mathbb{Z}^{d+1} : (\mathbf{x}, \mathbf{n}) = f([\mathbf{x}])\}.$$

The assignment $F \mapsto f$ is injective because $F \mapsto V(F)$ is injective.

Let us now identify the image of the map $F \mapsto f$. The function f satisfies $f([\mathbf{x}]) \equiv (\mathbf{x}, \mathbf{n}) \pmod{d + 1}$; we call this the *parity condition*. Assert furthermore that $f([\mathbf{x} + \mathbf{e}_i]) \leq f([\mathbf{x}]) + 1$. This assertion is called the *Lipschitz condition*.

Suppose that the assertion is false, that is, that instead $f([\mathbf{x} + \mathbf{e}_i]) > f([\mathbf{x}]) + 1$. Write \mathbf{y} for the unique vertex in $[\mathbf{x}] \cap V(F)$, and write \mathbf{z} for the unique vertex in $[\mathbf{x} + \mathbf{e}_i] \cap V(F)$. Then $\mathbf{z}_j > (\mathbf{y} + \mathbf{e}_i)_j \geq \mathbf{y}_j$ for all j . In this case, it is impossible that both \mathbf{z} and \mathbf{y} are contained in ∂F , because F is closed under \leq . This proves the assertion.

The Lipschitz condition also implies that $f([\mathbf{x} + \mathbf{e}_i]) \geq f([\mathbf{x}]) - d$, since $[\mathbf{e}_1 + \dots + \mathbf{e}_{d+1}] = [\mathbf{n}] = [\mathbf{0}]$. The Lipschitz condition, this new inequality, and the parity condition imply jointly that the gradient ∇f of f satisfies

$$\nabla f([\mathbf{x}], [\mathbf{x} + \mathbf{e}_i]) := f([\mathbf{x} + \mathbf{e}_i]) - f([\mathbf{x}]) \in \{-d, 1\}.$$

On the other hand, if the flow ∇f satisfies this equation for any \mathbf{x} and i and if $f([\mathbf{0}])$ is an integer multiple of $d + 1$, then f also clearly satisfies the parity condition.

A *height function* is a function $f : X^d \rightarrow \mathbb{Z}$ which satisfies $f([\mathbf{0}]) \in (d + 1)\mathbb{Z}$ and

$$\nabla(f([\mathbf{x}]), f([\mathbf{x} + \mathbf{e}_i])) = f([\mathbf{x} + \mathbf{e}_i]) - f([\mathbf{x}]) \in \{1, -d\}$$

for all \mathbf{x} and i ; write Ω for the set of height functions. If F is a stepped surface, then the associated height function f is indeed an element of Ω . The injective map $\Psi \rightarrow \Omega, F \mapsto f$ is in fact a bijection; its inverse is given by the map $\Xi : \Omega \rightarrow \Psi, f \mapsto L(V(f))$. For details, refer to [19].

2.3. *The tiling associated to a stepped surface.* If f is a height function, then write

$$T(f) := \{\{\mathbf{x}, \mathbf{y}\} \in E^d : \nabla f(\mathbf{x}, \mathbf{y}) = -d\}.$$

The set $T(f)$ characterises ∇f , and therefore it characterises the function f up to constants. The first goal is to characterise the image of the map $f \mapsto T(f)$ over Ω .

A path $(\mathbf{s}_k)_{0 \leq k \leq n} \subset X^d$ of length $n = d + 1$ is called a *rooted simplicial loop* or simply a *simplicial loop* if there exists a permutation $\xi \in S_{d+1}$ such that $\mathbf{s}_k = \mathbf{s}_{k-1} + \mathbf{e}_{\xi(k)}$ for any $1 \leq k \leq d + 1$. This implies that \mathbf{s} is closed because $[\mathbf{x}] + \mathbf{e}_1 + \dots + \mathbf{e}_{d+1} = [\mathbf{x} + \mathbf{n}] = [\mathbf{x}]$. Write R^d for the set of rooted simplicial loops. Sometimes we are not concerned with the

starting points of the loops. In those cases, two loops are considered equal if they are equal up to indexation—this is equivalent to requiring that the two loops traverse the same set of edges. Write U^d for the set of *unrooted simplicial loops*.

Let \mathbf{s} denote a simplicial loop. Knowing that ∇f must integrate to zero along this simplicial loop, it is immediate that exactly one of the edges of \mathbf{s} is contained in $T(f)$. Write Θ for the set of *tilings*, that is, the set of subsets $T \subset E^d$ with the property that $|T \cap \mathbf{s}| = 1$ for any simplicial loop \mathbf{s} . To see that the map $\Omega \rightarrow \Theta, f \mapsto T(f)$ is surjective, let $T \in \Theta$, and write α_T for the unique flow on (X^d, E^d) such that

$$(2.1) \quad \alpha_T([\mathbf{x}], [\mathbf{x} + \mathbf{e}_i]) = \begin{cases} 1 & \text{if } \{[\mathbf{x}], [\mathbf{x} + \mathbf{e}_i]\} \notin T, \\ -d & \text{if } \{[\mathbf{x}], [\mathbf{x} + \mathbf{e}_i]\} \in T. \end{cases}$$

Then $\alpha_T = \nabla f$ for some height function $f \in \Omega$ if and only if α_T is conservative. The flow α_T integrates to zero along any simplicial loop, by definition of a tiling. It is a simple exercise in group theory to see that this implies that α_T integrates to zero along any closed path, since the graph (X^d, E^d) can be written as the Cayley graph on the generators $\mathbf{e}_1, \dots, \mathbf{e}_{d+1}$ subject to the relators which are given by all possible permutations of these $d + 1$ elements—these relators correspond exactly to the simplicial loops. This proves that $T = T(f)$ for some height function $f \in \Omega$, and therefore the map $\Omega \rightarrow \Theta, f \mapsto T(f)$ is surjective. Since $T(f)$ characterises f up to constants, this also implies that the map $\Phi : \Omega \rightarrow (d + 1)\mathbb{Z} \times \Theta, f \mapsto (f(0), T(f))$ is a bijection.

2.4. *The geometrical intuition behind tilings.* Let us connect the combinatorial tilings $T \in \Theta$ to the familiar geometric picture of lozenge tilings. If $d = 2$, then (X^d, E^d) is the triangular lattice, and the set $T \subset E^d$ has the property that T contains exactly one edge of every triangle. Thus, if we remove the set T from this triangular lattice, then we are left with a collection of lozenges.

The boundary $\partial F \subset \mathbb{R}^{d+1}$ of a stepped surface is a union of hypercubes of codimension one with integral coordinates. Write H for the orthogonal complement of \mathbf{n} , and write $P : \mathbb{R}^{d+1} \rightarrow H$ for orthogonal projection onto H . For each stepped surface F , the projection map P restricts to a bijection from ∂F to H . If $d = 2$, then P maps each two-dimensional square contained in ∂F , to a lozenge embedded in H , and jointly these lozenges partition H —we ignore here the fact that the topological boundaries of the lozenges overlap. We observed in the previous paragraph that each edge in T encodes exactly one of these lozenges. The same reasoning applies to higher dimensions: the map P maps the hypercubes of codimension 1 which make up ∂F to H , and these projected hypercubes partition the d -dimensional real vector space H up to overlapping boundaries. Finally, each edge in T encodes exactly one of these projected hypercubes.

2.5. *Lemmas for analysing stepped surfaces.*

2.5.1. *The map Ξ preserves the lattice structure.*

LEMMA 2.1. *Let f_1 and f_2 be height functions and let $F_1 = \Xi(f_1)$ and $F_2 = \Xi(f_2)$. Then:*

1. $f_1 \leq f_2$ if and only if $F_1 \subset F_2$,
2. $f_1 \vee f_2$ is a height function, $F_1 \cup F_2$ is a stepped surface, and $\Xi(f_1 \vee f_2) = F_1 \cup F_2$,
3. $f_1 \wedge f_2$ is a height function, $F_1 \cap F_2$ is a stepped surface, and $\Xi(f_1 \wedge f_2) = F_1 \cap F_2$.

Of course, 2 and 3 extend to finite unions, intersections, maxima and minima.

See Lemma 2 in [19] for a proof.

2.5.2. *Local moves.* If two height functions agree at all but one vertex, then it is said that they differ by a *local move*. A local move is equivalent to adding or removing a single unit hypercube to the corresponding stepped surface.

LEMMA 2.2. *Suppose that $R \subset X^d$ is finite and that f and g are height functions that are equal outside R and satisfy $f \leq g$ on R . Then there exists a sequence of height functions $(f_k)_{0 \leq k \leq n} \subset \Omega$ with:*

1. $f_0 = f$ and $f_n = g$,
2. for every $0 \leq k < n$, there is a unique $\mathbf{x} \in R$ such that $f_{k+1} = f_k + (d + 1) \cdot 1_{\mathbf{x}}$.

The sequence is increasing and all functions f_k agree to f and g on $X^d \setminus R$.

See Lemma 3 in [19] for a proof. If f and g agree outside R but neither $f \leq g$ nor $f \geq g$, then one can first go down from f to $f \wedge g$ and then up from $f \wedge g$ to g . Lemma 2.1 ensures that $f \wedge g$ is a height function.

2.5.3. *The Kirszbraun theorem.* Consider a vertex $[\mathbf{x}] \in X^d$. Since all elements in $[\mathbf{x}]$ differ from one another by multiples of \mathbf{n} , and since $\mathbf{n} \in \text{Ker } P$, there is a unique element $\mathbf{y} \in H$ such that $\{\mathbf{y}\} = P[\mathbf{x}]$. Let us identify each vertex $[\mathbf{x}] \in X^d$ with this corresponding point $\mathbf{y} \in H$. Write $\mathbf{g}_i := P\mathbf{e}_i = \mathbf{e}_i - \mathbf{n}/(d + 1)$. The neighbours of $\mathbf{x} \in X^d \subset H$ are then given by $\mathbf{x} \pm \mathbf{g}_i$.

Define the asymmetric norm $\|\cdot\|_+$ on H by $\|\mathbf{x}\|_+ := -(d + 1) \min_i x_i$. Remark that $\|\cdot\|_+$ is the largest asymmetric norm on H subject to $\|\mathbf{g}_i\|_+ \leq 1$ for all i . In other words, a function $f : X^d \rightarrow \mathbb{R}$ satisfies the Lipschitz condition introduced in Section 2.2 if and only if

$$(2.2) \quad f(\mathbf{y}) - f(\mathbf{x}) \leq \|\mathbf{y} - \mathbf{x}\|_+$$

for all $\mathbf{x}, \mathbf{y} \in X^d$. A function $f : D \rightarrow \mathbb{R}$ defined on a subset $D \subset H$ is called *Lipschitz* whenever f satisfies (2.2) for all $\mathbf{x}, \mathbf{y} \in D$. This definition is consistent with the previous definition of the Lipschitz property for height functions.

LEMMA 2.3. *If a function $f : H \rightarrow \mathbb{R}$ is Lipschitz, then there exists a unique largest height function g subject to $g \leq f|_{X^d}$; the value of g at each $\mathbf{x} \in X^d$ is given by $g(\mathbf{x}) := k$, where k is the largest integer which makes g satisfy the parity condition at \mathbf{x} , and which does not exceed $f(\mathbf{x})$.*

If $f : H \rightarrow \mathbb{R}$ is a Lipschitz function, then write $\lfloor f \rfloor$ for the largest height function subject to $\lfloor f \rfloor \leq f|_{X^d}$; its function values are given by the previous lemma. We leave its proof to the reader. The previous lemma results in a discrete analogue of the Kirszbraun theorem for the current setting.

LEMMA 2.4. *If $R \subset H$ and $f : R \rightarrow \mathbb{R}$ is Lipschitz, then f extends to a Lipschitz function $\bar{f} : H \rightarrow \mathbb{R}$. If $R \subset X^d$ and $f : R \rightarrow \mathbb{Z}$ is Lipschitz and satisfies the parity condition for every $\mathbf{x} \in R$, then f extends to a height function $\bar{f} \in \Omega$.*

The first assertion is the original Kirszbraun theorem. For the second assertion, let $R \subset X^d$ and $f : R \rightarrow \mathbb{Z}$ be as in the lemma. The function f extends to some Lipschitz function $g : H \rightarrow \mathbb{R}$ by the Kirszbraun theorem. The previous lemma states that $\lfloor g \rfloor$ is a height function that equals f on R . This proves the second assertion.

3. Random height functions. In this section, we introduce boundary conditions (fixed or periodic) in order to reduce Ω to a finite set, and define and study probability measures on these finite sets. We assert that the model is monotone in boundary conditions, and state an immediate corollary of this fact by employing the Azuma–Hoeffding inequality.

3.1. *Fixed boundary conditions.* We reduce Ω to a finite set by applying fixed boundary conditions. One can study the uniform probability measure on this finite set. One can also define more general Boltzmann measures. If $R \subset X^d$, then write R^c for $X^d \setminus R$. Write $E^d(R)$ for the set of edges in E^d that are incident to at least one vertex in R .

DEFINITION 3.1. Define, for any height function f and for any tiling T ,

$$\Omega(R, f) := \{g \in \Omega : g|_{R^c} = f|_{R^c}\},$$

$$\Theta(R, T) := \{Y \in \Theta : Y \setminus E^d(R) = T \setminus E^d(R)\}.$$

Call a set $R \subset X^d$ a *region* if R is finite and if R^c is connected.

LEMMA 3.2. Let $R \subset X^d$ be a finite set, f a height function and $T := T(f)$. Then:

1. $\Omega(R, f)$ and $\Theta(R, T)$ are finite sets,
2. The map $g \mapsto T(g)$ restricts to an injection from $\Omega(R, f)$ to $\Theta(R, T)$,
3. If R is a region, then the restricted map in the previous statement is a bijection.

PROOF. Without loss of generality, R does not contain $\mathbf{0}$. If $g \in \Omega(R, f)$ then we must have $g(\mathbf{0}) = f(\mathbf{0})$ and $T(g) \setminus E^d(R) = T(f) \setminus E^d(R)$. The map $g \mapsto T(g)$ restricts to an injection because the gradient ∇g can be reconstructed from $T(g)$, which is sufficient for reconstructing g as the constant is determined from $g(\mathbf{0}) = f(\mathbf{0})$. The logarithm of $|\Theta(R, T)|$ is bounded by $|E^d(R)| \log 2 < \infty$. We have now proven the first two assertions of the lemma.

Next, we prove that the same restriction map is also surjective whenever R is a region. Fix $Y \in \Theta(R, T)$ and define $g := \Phi^{-1}(f(\mathbf{0}), Y)$. It suffices to show that $g \in \Omega(R, f)$, that is, that $g(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in R^c$. Let \mathbf{p} denote a path from $\mathbf{0}$ to \mathbf{x} through $(X^d \setminus R, E^d \setminus E^d(R))$; such a path exists by definition of a region. Then $g(\mathbf{x}) - f(\mathbf{0})$ and $f(\mathbf{x}) - f(\mathbf{0})$ can both be calculated in terms of integrals of ∇g and ∇f , respectively over the path \mathbf{p} . Since $Y \setminus E^d(R) = T \setminus E^d(R)$, these gradients are equal on the edges of \mathbf{p} , which proves that $g(\mathbf{x}) = f(\mathbf{x})$. \square

Fix a finite set $R \subset X^d$ and a height function $f \in \Omega$. Write f^\pm for the pointwise minimum and maximum over all height functions in the finite set $\Omega(R, f)$. These are also height function by virtue of Lemma 2.1, and clearly $\Omega(R, f) = \{g \in \Omega : f^- \leq g \leq f^+\}$. The same lemma implies the following result.

LEMMA 3.3. Fix a finite set $R \subset X^d$ and a height function f . Write f^\pm as in the preceding discussion and define $F^\pm := \Xi(f^\pm)$. Then Ξ restricts to a bijection from $\Omega(R, f)$ to $\{F \in \Psi : F^- \subset F \subset F^+\}$.

Next, we define Boltzmann measures. The uniform probability measures on $\Omega(R, f)$ and $\Theta(R, T)$ are examples of Boltzmann measures. The introduction of Boltzmann measures allows us to increase the relative probability of tilings containing certain edges.

DEFINITION 3.4. Let R be a region, f a height function and $T := T(f)$ a tiling. A positive real function w on $E^d(R)$ is called a *weight function*. Let \mathbb{P}_w be the probability

measure on the set $\Theta(R, T)$ such that $\mathbb{P}_w(Y) \propto \prod_{e \in Y \cap E^d(R)} w(e)$ for any $Y \in \Theta(R, T)$, that is,

$$\mathbb{P}_w(Y) := \frac{1}{Z_w} \prod_{e \in Y \cap E^d(R)} w(e) \quad \text{where } Z_w := \sum_{Y \in \Theta(R, T)} \prod_{e \in Y \cap E^d(R)} w(e).$$

The probability measure \mathbb{P}_w is called a *Boltzmann measure* and the normalising constant Z_w is called the *partition function*. The measure \mathbb{P}_w is also considered a probability measure on the sample space $\Omega(R, f)$ by defining $\mathbb{P}_w(g) := \mathbb{P}_w(T(g))$. Write \mathbb{P} for \mathbb{P}_w with w identically equal to 1, and write Z for the corresponding partition function. Observe that $Z = |\Omega(R, f)| = |\Theta(R, T)|$. The definition of Z_w makes sense also when w takes complex values.

We prove that Z_w equals the Cayley hyperdeterminant of a suitable hypermatrix in Section 6.

3.2. *The periodic setting.* Write H^* for the natural dual space of the d -dimensional real vector space H . A linear form $s \in H^*$ is called a *slope*.

DEFINITION 3.5. Periodic boundary conditions are characterised by a pair (L, s) , where $L \subset X^d$ is a full-rank sublattice and $s \in H^*$ a slope. A function $f : X^d \rightarrow \mathbb{R}$ is called (L, s) -*periodic* if, for any $\mathbf{x} \in X^d$ and $\mathbf{y} \in L$,

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + s(\mathbf{y}).$$

Write $\Omega(L, s)$ for the set of (L, s) -periodic height functions that map $\mathbf{0}$ to 0. Call a pair (L, s) *valid* if $\Omega(L, s)$ is nonempty.

It is not a priori clear which periodic boundary conditions (L, s) are valid. If f is a height function, then write $f|_L$ and $s|_L$ for the restrictions of f and s to $L \subset X^d \subset H$ respectively. First, every function $f \in \Omega(L, s)$ must satisfy $f|_L = s|_L$, so if $s|_L$ does not extend to a height function then $\Omega(L, s)$ is empty. On the other hand, if $s|_L$ extends to a height function, then the minimum amongst all possible extensions is (L, s) -periodic. Therefore, (L, s) is valid if and only if $s|_L$ extends to a height function. Lemma 2.4 imposes a Lipschitz condition and a parity condition on $s|_L$. Clearly $s|_L$ is Lipschitz if and only if s is Lipschitz.

DEFINITION 3.6. Write \mathcal{S} for the set of slopes in H^* which are Lipschitz, that is,

$$\mathcal{S} := \left\{ s \in H^* : \max_i s(\mathbf{g}_i) \leq 1 \right\}.$$

PROPOSITION 3.7. *The set $\mathcal{S} \subset H^*$ is a closed d -simplex with extreme points $(s^i)_{1 \leq i \leq d+1}$, where each slope s^i satisfies, for any j ,*

$$s^i(\mathbf{g}_j) := \begin{cases} -d & \text{if } i = j, \\ 1 & \text{if } i \neq j. \end{cases}$$

Introduce also the parity condition for the following result.

LEMMA 3.8. *Suppose given periodic boundary conditions (L, s) . Then (L, s) is valid if and only if $s \in \mathcal{S}$ and $s([\mathbf{x}]) \equiv (\mathbf{x}, \mathbf{n}) \pmod{d+1}$ for every $[\mathbf{x}] \in L \subset X^d$. In particular, if $L \subset (d+1)X^d$, then $(\mathbf{x}, \mathbf{n}) \in (d+1)\mathbb{Z}$ for every $[\mathbf{x}] \in L$, and under this extra condition, (L, s) is valid if and only if $s \in \mathcal{S}$ and $s(\mathbf{x}) \in (d+1)\mathbb{Z}$ for every $\mathbf{x} \in L$.*

If $L = n(d + 1)X^d$ for some $n \in \mathbb{N}$, then (L, s) is valid if and only if $s \in \mathcal{S}$ and

$$s(n(d + 1)\mathbf{g}_i) = n(d + 1)s(\mathbf{g}_i) \in (d + 1)\mathbb{Z}$$

for every $1 \leq i \leq d + 1$, that is, $s(\mathbf{g}_i) \in \mathbb{Z}/n$.

DEFINITION 3.9. Write $L_n := n(d + 1)X^d$ for every $n \in \mathbb{N}$. Define

$$\mathcal{S}_n := \{s \in \mathcal{S} : s(\mathbf{g}_i) \in \mathbb{Z}/n \text{ for every } 1 \leq i \leq d + 1\}.$$

In other words, \mathcal{S}_n is precisely the set of slopes s such that (L_n, s) is valid.

Remark that \mathcal{S}_n converges to \mathcal{S} in the Hausdorff metric, in the sense that every slope $s \in \mathcal{S}$ can be approximated by a sequence of slopes $(s_n)_{n \in \mathbb{N}}$ where $s_n \in \mathcal{S}_n$ for each n .

3.3. *Symmetries of the periodic setting.* For $\mathbf{x} \in X^d$, write $\theta_{\mathbf{x}}$ for the map $H \rightarrow H, \mathbf{y} \mapsto \mathbf{y} + \mathbf{x}$. This map is called a *shift*, and it is also clearly a symmetry of X^d . Write $\Theta(L)$ for the group $\{\theta_{\mathbf{x}} : \mathbf{x} \in L\}$ for any sublattice L of X^d , and write $\Theta := \Theta(X^d)$.

If f is a function defined on either H or X^d , then write θf for the function defined by $\theta f(\mathbf{x}) := f(\theta\mathbf{x})$. If f is a height function and $\theta \in \Theta$, then define the height function $\tilde{\theta}f$ by $\tilde{\theta}f := \theta f - f(\theta\mathbf{0}) + f(\mathbf{0})$. In other words, $\tilde{\theta}f$ is the unique height function such that $(\tilde{\theta}f)(\mathbf{0}) = f(\mathbf{0})$ and $\theta T(\theta f) = T(f)$. The map $\tilde{\theta} : \Omega \rightarrow \Omega$ is bijective, and $\tilde{\theta}$ restricts to a bijection from $\Omega(L, s)$ to $\Omega(L, s)$ for any periodic boundary conditions (L, s) .

Each height function $f \in \Omega(L, s)$ is characterised by the $\Theta(L)$ -invariant set $T(f) \subset E^d$. This implies in particular that the set $\Omega(L, s)$ is finite, because $|\Omega(L, s)| \leq 2^{|E^d/\Theta(L)|} < \infty$.

LEMMA 3.10. *Pick valid periodic boundary conditions (L, s) , write \mathbb{P} for the uniform probability measure on $\Omega(L, s)$, and write f for the random function in $\Omega(L, s)$. Then $\tilde{\theta}f \sim f$ in \mathbb{P} for any $\theta \in \Theta$. Moreover, $\mathbb{E}f(\mathbf{x}) = s(\mathbf{x})$ for every $\mathbf{x} \in X^d$.*

PROOF. The first assertion is obvious as $\tilde{\theta} : \Omega(L, s) \rightarrow \Omega(L, s)$ is a bijection and \mathbb{P} is uniform on this set. The first assertion implies that the map $\mathbb{E}(f(\cdot)) : X^d \rightarrow \mathbb{R}$ is additive. Therefore, it must extend to a linear form in H^* . Now L is full-rank and $\mathbb{P}(f(\mathbf{x}) = s(\mathbf{x})) = 1$ for every $\mathbf{x} \in L$, and therefore $\mathbb{E}(f(\cdot))$ must extend to the linear form $s \in \mathcal{S} \subset H^*$. \square

3.4. *Monotonicity of the random function.* Monotonicity in boundary conditions is often an essential property for understanding the macroscopic behaviour of the system. See Lemmas 3 and 4 in [19] for a proof of the following theorem—note that these lemmas apply both to fixed boundary conditions in the general Boltzmann setting, as well as to periodic boundary conditions. The proof simply says that the Glauber dynamics preserve the monotonicity and mix to the correct distribution.

THEOREM 3.11 (Monotonicity). *Let R be a region, let $b_1, b_2 \in \Omega$, and fix a weight function w . Write \mathbb{P}_1 and \mathbb{P}_2 for the Boltzmann measures with weight w on $\Omega(R, b_1)$ and $\Omega(R, b_2)$ respectively. Write a_- and a_+ for the infimum and supremum of $(b_1 - b_2)|_{R^c}$, respectively. Then there exists a probability distribution \mathbb{P} on the pair $(f_1, f_2) \in \Omega(R, b_1) \times \Omega(R, b_2)$ with marginals \mathbb{P}_1 and \mathbb{P}_2 such that $a_- \leq f_1 - f_2 \leq a_+$ almost surely.*

Now let (L, s) denote a valid pair of periodic boundary conditions, let $b_1, b_2 \in \Omega(L, s)$, and fix $R \subset X^d$. Write \mathbb{P}_1 and \mathbb{P}_2 for the uniform probability measure on $\Omega(L, s)$ conditioned on $f|_R = b_1|_R$ and $f|_R = b_2|_R$ respectively. Write a_- and a_+ for the infimum and supremum of $(b_1 - b_2)|_R$ respectively. Then there exists a probability distribution \mathbb{P} on the pair $(f_1, f_2) \in \Omega(L, s)^2$ with marginals \mathbb{P}_1 and \mathbb{P}_2 such that $a_- \leq f_1 - f_2 \leq a_+$ almost surely.

3.5. *Pointwise concentration inequalities.*

THEOREM 3.12. *Let $R \subset X^d$ denote a finite set, let $b \in \Omega$, and fix a weight function w . Write \mathbb{P} for the Boltzmann measure with weight w on $\Omega(R, b)$. Fix $\mathbf{x} \in X^d$. Then the following inequalities hold true:*

1. $\text{Var } f(\mathbf{x}) \leq (d + 1)^2 n,$
2. $\mathbb{P}(f(\mathbf{x}) - \mu \geq (d + 1)a) \leq \exp -\frac{a^2}{2n}$ for all $a \geq 0$ whenever $n > 0,$
3. $\mathbb{P}(f(\mathbf{x}) - \mu \leq (d + 1)a) \leq \exp -\frac{a^2}{2n}$ for all $a \leq 0$ whenever $n > 0,$

where $n = d_{(X^d, E^d)}(\mathbf{x}, R^c)$ and $\mu = \mathbb{E} f(\mathbf{x}).$

Now let (L, s) denote a valid pair of periodic boundary conditions, and write \mathbb{P} for the uniform probability measure on $\Omega(L, s)$. Then (1)–(3) remain true for the choices $n = d_{(X^d, E^d)}(\mathbf{x}, L) \leq d_{(X^d, E^d)}(\mathbf{x}, \mathbf{0})$ and $\mu = \mathbb{E} f(\mathbf{x}) = s(\mathbf{x}).$

PROOF. This is a simple consequence of monotonicity (Theorem 3.11) and the Azuma–Hoeffding inequality. Focus on fixed boundary conditions; the proof for periodic boundary conditions is entirely analogous. Let $\mathbf{p} = (\mathbf{p}_k)_{0 \leq k \leq n}$ denote a path of shortest length from R^c to \mathbf{x} . Write $(f_k)_{0 \leq k \leq n}$ for the martingale such that f_k equals the expectation of $f(\mathbf{x})$ conditional on the values of f on the vertices in $\{\mathbf{p}_0, \dots, \mathbf{p}_k\}$. By Theorem 3.11, we have $|f_k - f_{k+1}| \leq d + 1$. The theorem now follows from the Azuma–Hoeffding inequality. \square

4. The cluster boundary swap. In the seminal work [21], Sheffield introduces *cluster swapping*. This technique is related to the double dimer model, where for uniform probability measures, the orientation of each loop is uniformly random in the two states, independently of any other structure that is present. In this section, we introduce the *cluster boundary swap*. The cluster boundary swap is a special case of the cluster swap, adapted and optimised for the special nature of the model under consideration. Its properties are reminiscent of the double dimer model, because, conditional on the geometrical structure involving boundaries and double edges, the orientation of each boundary is uniformly random in its two states.

4.1. *The boundary graph and the level set decomposition.* In Section 2.3, it was proven that every height function f is characterised by the pair $\Phi(f) = (f(\mathbf{0}), T(f))$. In this section, f_1 and f_2 denote height functions, and we write $(a_i, T_i) := \Phi(f_i)$ for $i \in \{1, 2\}$. The difference function $f_1 - f_2$ is denoted by g . The goal of this subsection is to understand the level set structure of g .

LEMMA 4.1. *Let $\mathbf{s} \in R^d$ be a rooted simplicial loop. As one walks along \mathbf{s} :*

1. *The function f_1 moves up by 1 exactly d times,*
2. *The function f_1 moves down by d exactly once,*
3. *Either g remains constant, or it changes value twice,*
4. *If g is not constant, then the difference between its two values is $d + 1$.*

The lemma follows immediately from the observations in Section 2.3. Write $A \ominus B$ for the symmetric difference of arbitrary sets A and B .

DEFINITION 4.2. Define the graph $G_g = (V_g, E_g)$ as follows. Its vertex set V_g is given by

$$V_g := T_1 \ominus T_2 = \{e \in E^d : g \text{ is not constant on } e\} \subset E^d,$$

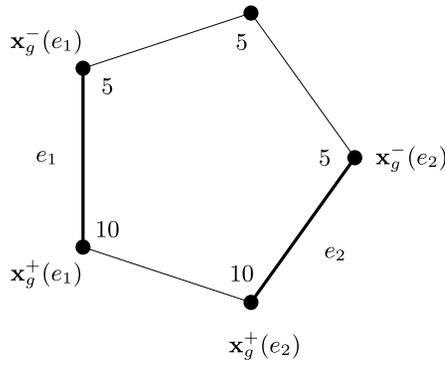


FIG. 2. The values of $g = f_1 - f_2$ along a simplicial loop; $d = 4$.

and two vertices $e_1, e_2 \in V_g \subset E^d$ are neighbours if some simplicial loop travels through both e_1 and e_2 . The graph G_g is called the *boundary graph*. For $e \in V_g$, write $\mathbf{x}_g^-(e), \mathbf{x}_g^+(e) \in X^d$ for the vertices contained in $e \subset X^d$ on which g takes the smaller value and the larger value, respectively; see Figure 2. For any $C \subset V_g$, we write $\mathbf{x}_g^\pm(C) := \{\mathbf{x}_g^\pm(e) : e \in C\}$.

For example, if $d = 2$, then one can identify each edge of the triangular lattice with the lozenge obtained by removing that edge. The set V_g is then precisely the set of edges of lozenges which appear in exactly one of the two configurations. Two edges in V_g are neighbours if they belong to the same triangle of the triangular lattice. Each connected component of G_g corresponds to the set of edges of the triangular lattice crossed by a nontrivial loop of the double dimer model.

LEMMA 4.3. *Let $C \subset V_g$ be a connected component of G_g . Then $(X^d, E^d \setminus C)$ consists of two connected components, one containing $\mathbf{x}_g^-(C)$, and the other containing $\mathbf{x}_g^+(C)$. Moreover, each of $\mathbf{x}_g^-(C)$ and $\mathbf{x}_g^+(C)$ is contained in a connected component of the graph $(X^d, E^d \setminus V_g)$.*

PROOF. Suppose that the G_g -vertices e_1 and e_2 are neighbours in the graph G_g ; write s for a simplicial loop passing through both e_1 and e_2 . Then s contains no other edges in V_g by Proposition 4.1, 3, and therefore $\mathbf{x}_g^-(e_1)$ and $\mathbf{x}_g^-(e_2)$ are connected in the graph $(X^d, E^d \setminus V_g)$; see also Figure 2. Induct on this argument to see that $\mathbf{x}_g^-(C)$ is contained in a connected component of $(X^d, E^d \setminus V_g)$. Identical reasoning applies to the set $\mathbf{x}_g^+(C)$, and we also learn that each of $\mathbf{x}_g^\pm(C)$ is contained in a connected component of $(X^d, E^d \setminus C)$.

The sets $\mathbf{x}_g^\pm(C)$ cover all the endpoints of edges in C , and therefore two possibilities remain: either the graph $(X^d, E^d \setminus C)$ is connected, or it consists of two connected components, with one containing $\mathbf{x}_g^-(C)$, and the other containing $\mathbf{x}_g^+(C)$. To establish the lemma we must exclude the first possibility. Every simplicial loop intersects C an even number of times. The group theory arguments that proved that the flow in Section 2.3 was conservative, imply here that any closed walk through (X^d, E^d) intersects C an even number of times. This proves that $(X^d, E^d \setminus C)$ is not connected. \square

DEFINITION 4.4. A *g-level set* is a connected component of the graph $(X^d, E^d \setminus V_g)$. A *g-boundary* is a connected component of the graph $G_g = (V_g, E_g)$. The *g-level sets* are considered subsets of X^d , and the *g-boundaries* are considered subsets of $V_g \subset E^d$. If E is a *g-boundary*, then write $X_g^\pm(E)$ for the *g-level set* containing $\mathbf{x}_g^\pm(E)$. The *level set decomposition of g* or *LSD(g)* is an undirected graph, where the vertices are the *g-level sets* and the

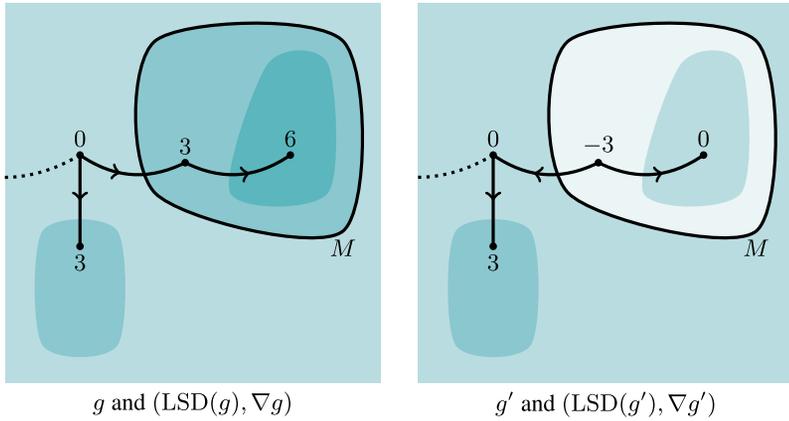


FIG. 3. The level set decomposition and a cluster boundary swap by M ; $d = 2$.

edges are the g -boundaries. The g -boundary E connects the g -level sets $X_g^-(E)$ and $X_g^+(E)$. Write g for the graph homomorphism $g : \text{LSD}(g) \rightarrow (d + 1)\mathbb{Z}$ that assigns the value $g(X)$ to a g -level set X . The vector field ∇g directs the edges in $\text{LSD}(g)$: it orients each g -boundary E from $X_g^-(E)$ to $X_g^+(E)$. Write $(\text{LSD}(g), \nabla g)$ for this directed graph.

If $d = 2$, then the g -boundaries correspond exactly to the loops of the double dimer model. The g -level sets correspond to the connected components of \mathbb{R}^2 with all loops of the double dimer model removed. In Figure 3 we see an example of this graph. Each g -level set contracts into a single $\text{LSD}(g)$ -vertex. The $\text{LSD}(g)$ -edges are comprised of the g -boundaries separating the g -level sets.

LEMMA 4.5. *The graph $\text{LSD}(g)$ is well defined and a tree.*

PROOF. It follows from Lemma 4.3 that every ∇g -directed $\text{LSD}(g)$ -edge has a well-defined startpoint and endpoint, and that removing an edge disconnects the graph. \square

4.2. The cluster boundary swap.

LEMMA 4.6. *Let $M \subset V_g \subset E^d$ be a union of g -boundaries. Then the sets $T'_i := T_i \ominus M$ are tilings for $i \in \{1, 2\}$. Write f'_1 and f'_2 for the unique height functions such that $\Phi(f'_i) = (a_i, T'_i)$ and define $g' = f'_1 - f'_2$. Then $G_{g'} = G_g$ and $\text{LSD}(g') = \text{LSD}(g)$. Moreover, $\nabla g'$ and ∇g are the same except that the g -boundaries contained in M have reversed orientation, that is, $\nabla g' = (-1)^{1_M} \cdot \nabla g$.*

PROOF. First, claim that T'_1 and T'_2 are tilings. We focus on T'_1 . Let s be a simplicial loop, and abuse notation by writing s also for the set of edges crossed by this loop. It suffices to prove that $|T'_1 \cap s| = 1$. Now either $s \cap M$ is empty, or contains two edges, one from T_1 and one from T_2 . In the former case we have $T'_1 \cap s = T_1 \cap s$ and consequently $|T'_1 \cap s| = 1$. In the latter case, we have $T'_1 \cap s = T_2 \cap s$ and consequently $|T'_1 \cap s| = 1$, as desired. This proves the claim. The appropriate functions f'_1 and f'_2 exist because Φ is a bijection from Ω to $(d + 1)\mathbb{Z} \times \Theta$. Next,

$$V_{g'} := T'_1 \ominus T'_2 = (T_1 \ominus M) \ominus (T_2 \ominus M) = T_1 \ominus T_2 = V_g,$$

and consequently $G_{g'} = G_g$ and $\text{LSD}(g') = \text{LSD}(g)$. Recall the definition of α_T in terms of T in (2.1). We have

$$\nabla g' = \alpha_{T'_1} - \alpha_{T'_2} = \alpha_{T_1 \ominus M} - \alpha_{T_2 \ominus M} = (-1)^{1_M} \cdot (\alpha_{T_1} - \alpha_{T_2}) = (-1)^{1_M} \cdot \nabla g;$$

this follows directly from the fact that $M \subset T_1 \ominus T_2$ and from the definition of α_T . \square

DEFINITION 4.7. Define

$$(T_1, T_2) \ominus M := (T'_1, T'_2) = (T_1 \ominus M, T_2 \ominus M),$$

$$(f_1, f_2) \ominus M := (f'_1, f'_2) = (\Phi^{-1}(a_1, T_1 \ominus M), \Phi^{-1}(a_2, T_2 \ominus M)),$$

whenever these are related as in the previous lemma. Write $(f_1, f_2) \sim (f'_1, f'_2)$ whenever $(f'_1, f'_2) = (f_1, f_2) \ominus M$ for some union of g -level sets M , in which case we say that the two pairs *differ by a cluster boundary swap*. The relation \sim is an equivalence relation; write $[(f_1, f_2)]$ for the equivalence class of (f_1, f_2) .

REMARK 4.8.

1. If $(f'_1, f'_2) = (f_1, f_2) \ominus M$, then $f'_1 + f'_2 = f_1 + f_2$; a cluster boundary swap does not change the sum of the two involved height functions. To see that this is the case, observe that M is a subset of $T_1 \ominus T_2$, and therefore $1_{T_1 \ominus M} + 1_{T_2 \ominus M} = 1_{T_1} + 1_{T_2}$ and

$$\nabla f_1 + \nabla f_2 = \alpha_{T_1} + \alpha_{T_2} = \alpha_{T_1 \ominus M} + \alpha_{T_2 \ominus M} = \nabla f'_1 + \nabla f'_2.$$

2. The cluster boundary swap was formalised in terms of the height functions f_1 and f_2 . The cluster boundary swap should however be understood as an operation on the gradients ∇f_1 and ∇f_2 of these height functions. This gradient operation is made into an operation on the nongradient height functions by choosing the vertex $\mathbf{0} \in X^d$ as a reference vertex at which the height is held constant.

Figure 3 illustrates a cluster boundary swap. The thick contour is the set M , and the two difference functions g and g' are related by $g = f_1 - f_2$ and $g' = f'_1 - f'_2$ where $(f'_1, f'_2) := (f_1, f_2) \ominus M$. Swapping by M effectively inverts the orientation of the corresponding g -boundary. One can swap any union of g -boundaries. Therefore, one can direct the edges of $\text{LSD}(g)$ in any desired way. We obtain the following theorem.

THEOREM 4.9. *The relation \sim is an equivalence relation on Ω^2 . The elements in the equivalence class of (f_1, f_2) correspond naturally to the graph homomorphisms from the tree $\text{LSD}(g)$ to $(d + 1)\mathbb{Z}$ that map $\mathbf{0}$ to $g(\mathbf{0})$.*

5. The variance and covariance structure. This section is dedicated to a straightforward application of Theorem 4.9 in the fixed boundary setting.

THEOREM 5.1. *Let R be a region not containing $\mathbf{0}$, let b be a height function, and let $w : E^d(R) \rightarrow (0, \infty)$ be a weight function. Denote the Boltzmann measure on $\Omega(R, b)$ with weight w by \mathbb{P}_w . Abuse notation by writing \mathbb{P}_w for $\mathbb{P}_w \times \mathbb{P}_w$; write (f_1, f_2) for the pair of random functions in this measure, and write $g := f_1 - f_2$. Also write f for f_1 . Then for any $\mathbf{x} \in X^d$, we have*

$$\text{Var}_w f(\mathbf{x}) = \frac{1}{2}(d + 1)^2 \mathbb{E}_w d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x}).$$

In other words, the variance of $f(\mathbf{x})$ in \mathbb{P}_w equals $\frac{1}{2}(d + 1)^2$ times the \mathbb{P}_w -expectation of the number of g -boundaries that separate \mathbf{x} from $\mathbf{0}$.

PROOF. The random variables $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are i.i.d., and therefore

$$\text{Var}_w f(\mathbf{x}) = \frac{1}{2} \text{Var}_w(f_1(\mathbf{x}) - f_2(\mathbf{x})) = \frac{1}{2} \mathbb{E}_w(f_1(\mathbf{x}) - f_2(\mathbf{x}))^2 = \frac{1}{2} \mathbb{E}_w g(\mathbf{x})^2.$$

It suffices to prove that $\mathbb{E}_w g(\mathbf{x})^2 = (d + 1)^2 \mathbb{E}_w d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x})$. In fact, we make the stronger claim that

$$\mathbb{E}_w (g(\mathbf{x})^2 | [(f_1, f_2)]) = (d + 1)^2 d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x}).$$

The left-hand side is $\sigma([(f_1, f_2)])$ -measurable by definition. For the right-hand side, observe that the graph $\text{LSD}(g)$ is constant on each equivalence class $[(f_1, f_2)]$, which means that $d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x})$ is also $\sigma([(f_1, f_2)])$ -measurable. The proof of the claim relies on Theorem 4.9.

Assert first that $[(f_1, f_2)] \subset \Omega(R, b)^2$ whenever $(f_1, f_2) \in \Omega(R, b)^2$. The set R^c is connected by the definition of a region, and it contains $\mathbf{0}$. Therefore, R^c is contained in the g -level set containing $\mathbf{0}$. A cluster boundary swap does not alter the values of f_1 and f_2 on this g -level set, and therefore f_1, f_2, f'_1, f'_2 , and b all assume the same values on R^c provided that $(f'_1, f'_2) \sim (f_1, f_2)$ and $(f_1, f_2) \in \Omega(R, b)^2$. This proves the assertion. Next, assert that \mathbb{P}_w conditioned on $[(f_1, f_2)]$ is uniform on $[(f_1, f_2)]$. To see that this is the case, observe that

$$\mathbb{P}_w((f_1, f_2)) \propto \prod_{e \in E^d(R)} w(e)^{(1_{T(f_1)} + 1_{T(f_2)})(e)}.$$

Now $1_{T(f_1)} + 1_{T(f_2)} = 1_{T(f'_1)} + 1_{T(f'_2)}$ whenever $(f'_1, f'_2) \sim (f_1, f_2)$, which proves the assertion.

Theorem 4.9 now provides the distribution of the function g in the measure \mathbb{P}_w conditioned on $[(f_1, f_2)]$. In particular, as $\text{LSD}(g)$ is a tree, the distribution of $g(\mathbf{x})$ is given by summing the outcomes of $d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x})$ fair coin flips, each worth $\pm(d + 1)$. It is well known that the expectation of the square of this random variable is $(d + 1)^2 d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{x})$, which proves the claim. \square

In fact, the exact same calculation works for the covariance of $f(\mathbf{x})$ with $f(\mathbf{y})$.

THEOREM 5.2. *Work in the setting of the previous theorem. Then for any $\mathbf{x}, \mathbf{y} \in X^d$, we have*

$$\text{Cov}_w(f(\mathbf{x}), f(\mathbf{y})) = \frac{1}{2}(d + 1)^2 \mathbb{E}_w d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{z}),$$

where \mathbf{z} is the last $\text{LSD}(g)$ -vertex of the $\text{LSD}(g)$ -path from $\mathbf{0}$ to \mathbf{x} that also appears in the $\text{LSD}(g)$ -path from $\mathbf{0}$ to \mathbf{y} . In other words, the covariance of $(f(\mathbf{x}), f(\mathbf{y}))$ in \mathbb{P}_w equals $\frac{1}{2}(d + 1)^2$ times the expectation of the number of g -boundaries that separate both \mathbf{x} and \mathbf{y} from $\mathbf{0}$.

PROOF. Again, we have $\text{Cov}_w(f(\mathbf{x}), f(\mathbf{y})) = \frac{1}{2} \text{Cov}_w(g(\mathbf{x}), g(\mathbf{y}))$, and we prove that

$$\mathbb{E}_w (g(\mathbf{x})g(\mathbf{y}) | [(f_1, f_2)]) = (d + 1)^2 d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{z}).$$

The conditioned measure \mathbb{P}_w directs the edges of $\text{LSD}(g)$ independently and uniformly at random, as in the previous theorem. Thus, under the conditioned measure \mathbb{P}_w , we have

$$(g(\mathbf{x}), g(\mathbf{y})) \sim (A + X, A + Y),$$

where A, X and Y are independent, where A is determined by summing the outcome of $d_{\text{LSD}(g)}(\mathbf{0}, \mathbf{z})$ fair independent coin flips each valued $\pm(d + 1)$, where X is determined by flipping $d_{\text{LSD}(g)}(\mathbf{z}, \mathbf{x})$ coins, and where Y is determined by flipping $d_{\text{LSD}(g)}(\mathbf{z}, \mathbf{y})$ coins. This proves the assertion. \square

6. Generalisation of the Kasteleyn theory. Consider fixed boundary conditions (R, f) and (R, T) with R a region and $T = T(f)$. The goal of this section is to show that $Z = |\Omega(R, f)| = |\Theta(R, T)|$ equals the Cayley hyperdeterminant of the adjacency hypermatrix of a suitably defined hypergraph. In fact, we have no trouble in generalising to Boltzmann measures; we show that one can insert the weights w into the adjacency hypermatrix so that the Cayley hyperdeterminant equals Z_w . The hypergraph, which we shall denote by (U^d, H^d) , is dual to the simplicial lattice (X^d, E^d) . Recall that U^d is the set of unrooted simplicial loops that was introduced earlier. In dimension $d = 2$ we recover exactly the theory of the dimer model on the hexagonal lattice. (See Figure 1(c).)

6.1. *The dual of the simplicial lattice.* In this subsection, we define the hypergraph (U^d, H^d) . Consider first the collection of simplicial loops. If $\mathbf{s} = (\mathbf{s}_k)_{0 \leq k \leq d+1} \in R^d$ is a rooted simplicial loop, then \mathbf{s} is characterised by its starting point $\mathbf{s}_0 \in X^d$ and the permutation $\xi \in S_{d+1}$ which describes in which order the increments $(\mathbf{g}_i)_i$ appear. This automatically gives a bijection from R^d to $X^d \times S_{d+1}$. Let us agree to index each unrooted loop $\mathbf{s} \in U^d$ (by default) such that $\mathbf{s}_1 = \mathbf{s}_0 + \mathbf{g}_{d+1}$. There is a unique way of doing so, because the increment \mathbf{g}_{d+1} appears exactly once in each loop. With this convention, each unrooted loop $\mathbf{s} \in U^d$ is characterised by its starting point \mathbf{s}_0 and the order $\xi \in S_d$ in which the remaining increments $\{\mathbf{g}_1, \dots, \mathbf{g}_d\}$ appear in the path after the first increment. By adopting the convention, we obtain a bijection from U^d to $X^d \times S_d$. We identify the unrooted loop \mathbf{s} with its image under the bijection, so that every pair $(\mathbf{x}, \xi) \in X^d \times S_d$ denotes also an unrooted simplicial loop.

DEFINITION 6.1. For any $e \in E^d$, write $h(e)$ for the set of unrooted simplicial loops that traverse e .

Write $e = \{\mathbf{x}, \mathbf{x} + \mathbf{g}_j\} \in E^d$ and let us make a number of observations about the set $h(e)$. First, the assignment $e \mapsto h(e)$ is injective, because the edge e is the only edge that is traversed by all loops in $h(e)$. Second, there are precisely $d!$ unrooted simplicial loops that traverse e , since they correspond to the $d!$ ways that we can order the d increments $(\mathbf{g}_i)_{i \neq j}$ that we need to walk back to \mathbf{x} from $\mathbf{x} + \mathbf{g}_j$. Therefore $h(e)$ contains $d!$ unrooted loops. Finally, if $\mathbf{s}^1, \mathbf{s}^2 \in h(e)$ are distinct loops identified with the pairs $(\mathbf{x}^1, \xi^1), (\mathbf{x}^2, \xi^2) \in X^d \times S_d$, then the permutations $\xi^1, \xi^2 \in S_d$ must be distinct. Conclude that for every $\xi \in S_d$, there is a unique $\mathbf{x} \in X^d$ such that $(\mathbf{x}, \xi) \in h(e)$.

The reason that we introduced the map h is the following. A set $T \subset E^d$ is a tiling if and only if $h(T)$ is a partition of U^d , the set of simplicial loops. One could rephrase this statement by saying that $h(T)$ is a perfect matching of the hypergraph $(U^d, h(E^d))$.

DEFINITION 6.2. Write H^d for the set $\{h(e) : e \in E^d\}$. The hypergraph (U^d, H^d) is called the *dual hyperlattice* or simply the *dual* (of the simplicial lattice).

LEMMA 6.3. *The map $h : T \mapsto \{h(e) : e \in T\}$ is a bijection from Θ to the set of perfect matchings of the hypergraph (U^d, H^d) .*

Note that (U^d, H^d) is really dual to (X^d, E^d) because the map h is a bijection from E^d to H^d . The hyperlattice is $d!$ -uniform, because every hyperedge $h(e)$ contains $d!$ elements. It is also $d!$ -partite with the partition $\{X^d \times \{\xi\} : \xi \in S_d\}$, because every hyperedge $h(e)$ contains one loop in each member $X^d \times \{\xi\}$. The $d!$ -partite structure of the dual of the simplicial lattice is special and it is a feature that distinguishes the simplicial lattice from other lattices (in particular, the author is not aware of a similar construction for the square lattice in dimension larger than two). The $d!$ -partite structure enables us to generalise the Kasteleyn theory.

6.2. *The approach suggested by the classical dimer theory.* The purpose of this section is to demonstrate that the size of $Z = |\Omega(R, f)| = |\Theta(R, T)|$ equals the Cayley hyperdeterminant of a suitable adjacency hypermatrix. First, recall how this works in the Kasteleyn theory for the dimer model on the hexagonal lattice. If $d = 2$, then $d! = 2$, that is, (U^d, H^d) is a regular bipartite graph. In fact, it is really the planar dual of the triangular lattice: the hexagonal lattice. The vertex set U^d is split into its two parts: a set of black and a set of white vertices. A dimer cover (i.e., a perfect matching of the graph) matches each black vertex to one white vertex. This is illustrated by Figure 1(c). The dimer cover is thus encoded by a bijective map σ from the set of black vertices to the set of white vertices; each dimer is of the form $\{b, \sigma(b)\}$ with b ranging over the set of black vertices. To calculate the number of dimer covers, one needs to count the bijections σ from the black vertices to the white that produce a dimer cover. If K is an $n \times n$ matrix, then $\text{Det } K$ is defined as (this is the Leibniz formula)

$$(6.1) \quad \text{Det } K = \sum_{\sigma \in S_n} \left[\text{Sign } \sigma \prod_{k=1}^n K(k, \sigma(k)) \right].$$

If the matrix K is suitably chosen, then the term in the square brackets reduces to the indicator function of the event that σ encodes a dimer cover, in which case $\text{Det } K$ equals the number of dimer covers. This is the Kasteleyn theory for dimer models. These observations suggest the following approach, consisting of four steps:

1. First, encode each tiling as a tuple of bijections. It turns out that we need $d! - 1$ bijections in each tuple, because the graph (U^d, H^d) is $d!$ -partite, and because we need one bijection for each colour beyond the first. This is Lemma 6.4.
2. Second, we show that applying fixed boundary conditions fixes the bijections at all but a finite number of points. Each tiling $Y \in \Theta(R, T)$ is thus encoded as a $(d! - 1)$ -tuple of bijections between finite sets. This is Lemma 6.6.
3. Third, we define a rank $d!$ adjacency hypermatrix K and construct the Cayley hyperdeterminant for this hypermatrix, such that the each nonzero term in the sum in the definition of $\text{Det } K$ corresponds to a tiling $Y \in \Theta(R, T)$. These are Definitions 6.7 and 6.8.
4. Finally, each nonzero term in this sum takes value 1 or -1 . This is due to the signs that appear in the determinant formula (note that the sign also appears in (6.1)). It takes some effort to show that all nonzero terms have the same sign. This is Lemma 6.9.

Once this has all been done, it is clear that $Z = |\Omega(R, f)| = |\Theta(R, T)| = \pm \text{Det } K$. The dual hyperlattice (U^d, H^d) plays a crucial role in the analysis. Fix, throughout the remainder of this section, an enumeration $\{\xi^1, \dots, \xi^{d!}\} = S_d$.

6.3. *The Kasteleyn theory in dimension $d \geq 2$.* We start with Step 1 of the proposed approach. Let $Y \in \Theta$ be a tiling of (X^d, E^d) , so that $h(Y)$ is a perfect matching of (U^d, H^d) . Each hyperedge $h(e) \in h(Y)$ contains one simplicial loop in each of the $d!$ parts of the partition of U^d . The bijections corresponding to Y are the unique maps

$$(6.2) \quad \sigma_i : X^d \times \{\xi^1\} \rightarrow X^d \times \{\xi^i\}$$

such that

$$(6.3) \quad \{\mathbf{s}, \sigma_2(\mathbf{s}), \sigma_3(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})\} \in h(Y)$$

for every unrooted simplicial loop $\mathbf{s} \in X^d \times \{\xi^1\}$. All elements of $h(Y)$ are given by ranging \mathbf{s} over $X^d \times \{\xi^1\}$ in (6.3). This is completely analogous to the dimer model.

Suppose given arbitrary bijections $(\sigma_i)_{2 \leq i \leq d!}$ as in (6.2). Then the set of sets of simplicial loops

$$(6.4) \quad \{\{\mathbf{s}, \sigma_2(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})\} : \mathbf{s} \in X^d \times \{\xi^1\}\}$$

is automatically a partition of $U^d = X^d \times S_d$, because each map σ_i is a bijection and therefore each loop (\mathbf{x}, ξ^i) appears precisely once. Conclude that (6.4) is a perfect matching of (U^d, H^d) if and only if (6.4) is a subset of the hyperedge set H^d . This yields the following result: Step 1 of the suggested approach.

LEMMA 6.4. *The set of $(d! - 1)$ -tuples of bijections*

$$(\sigma_i : X^d \times \{\xi^1\} \rightarrow X^d \times \{\xi^i\})_{2 \leq i \leq d!} \text{ satisfying } \{\mathbf{s}, \sigma_2(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})\} \in H^d$$

for every loop $\mathbf{s} \in X^d \times \{\xi^1\}$ is in bijection with the perfect matchings of (U^d, H^d) . The perfect matching of a tuple (under this bijection) is given by ranging \mathbf{s} over $X^d \times \{\xi^1\}$; this is precisely the set in (6.4).

Now consider Step 2 of the suggested approach. Suppose given a region R and a tiling T , and consider a tiling $Y \in \Theta(R, T)$. By definition, $Y \in \Theta(R, T)$ if and only if $Y \setminus E^d(R) = T \setminus E^d(R)$. Therefore all loops traversing $T \setminus E^d(R)$ must be matched in the same way as in T , and the loops traversing $T \cap E^d(R)$ can be matched differently. However, the loops that are matched differently are not allowed to produce new hyperedges outside the set $h(E^d(R))$, since we want $Y \setminus E^d(R) = T \setminus E^d(R)$. We first need to identify, for each part of the partition $\{X^d \times \{\xi^i\} : 1 \leq i \leq d!\}$ of U^d , the set of loops traversing $T \cap E^d(R)$, that is, the loops that are allowed to be matched differently. This motivates the following definition.

DEFINITION 6.5. Define, for a fixed region R and a fixed tiling T , and for $1 \leq i \leq d!$,

$$\begin{aligned} X_i &:= \{(\mathbf{x}, \xi^i) : \text{the loop } (\mathbf{x}, \xi^i) \text{ intersects } T \cap E^d(R)\} \\ &= \{(\mathbf{x}, \xi^i) : \text{the loop } (\mathbf{x}, \xi^i) \text{ does not intersect } T \setminus E^d(R)\} \subset X^d \times \{\xi^i\}. \end{aligned}$$

Observe that $|X_i| = |T \cap E^d(R)|$, and therefore the sets X_i all have the same, finite size. The sets X_i contain the loops that are allowed to match differently. This is Step 2 of the suggested approach.

LEMMA 6.6. *Let R be a region and T a tiling. The set of $(d! - 1)$ -tuples of bijections*

$$(\sigma_i : X_1 \rightarrow X_i)_{2 \leq i \leq d!} \text{ satisfying } \{\mathbf{s}, \sigma_2(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})\} \in h(E^d(R))$$

for every loop $\mathbf{s} \in X_1$ is in bijection with the set of perfect matchings $h(Y)$ corresponding to tilings $Y \in \Theta(R, T)$. The perfect matching of a tuple (under this bijection) is

$$\{\{\mathbf{s}, \sigma_2(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})\} : \mathbf{s} \in X_1\} \cup \{h(e) : e \in T \setminus E^d(R)\}.$$

The Kasteleyn hypermatrix and its determinant are now straightforwardly defined.

DEFINITION 6.7. Let R be a region and T a tiling. Define

$$K : X_1 \times \dots \times X_{d!} \rightarrow \{0, 1\}, (\mathbf{s}_1, \dots, \mathbf{s}_{d!}) \mapsto 1(\{\mathbf{s}_1, \dots, \mathbf{s}_{d!}\} \in h(E^d(R))),$$

where $1(\cdot)$ equals one if the statement inside holds true and zero otherwise. The map K is called the *Kasteleyn hypermatrix*.

From this definition and the previous lemma it follows that

$$|\Theta(R, T)| = \sum_{\sigma_2: X_1 \rightarrow X_2, \dots, \sigma_{d!}: X_1 \rightarrow X_{d!}} \left[\prod_{\mathbf{s} \in X_1} K(\mathbf{s}, \sigma_2(\mathbf{s}), \dots, \sigma_{d!}(\mathbf{s})) \right],$$

where the sum is over bijective maps only. This because the product produces a 1 if the tuple $(\sigma_i)_{2 \leq i \leq d!}$ corresponds to an element of $\Theta(R, T)$ and zero otherwise. Recall that $|X_1| = \dots = |X_{d!}| = |T \cap E^d(R)|$ and write n for this finite number. To simplify notation, we identify each set X_i with $[n] := \{1, \dots, n\}$, so that the previous equality is written

$$(6.5) \quad |\Theta(R, T)| = \sum_{\sigma_2, \dots, \sigma_{d!} \in S_n} \left[\prod_{k=1}^n K(k, \sigma_2(k), \dots, \sigma_{d!}(k)) \right].$$

The expression on the right looks similar to the definition of the determinant of a matrix, and if we insert the signs of the permutations then we obtain precisely the Cayley hyperdeterminant.

DEFINITION 6.8. Suppose given a map $A : [n]^m \rightarrow \mathbb{C}$ for some $n \in \mathbb{N}, m \in 2\mathbb{N}$. Define

$$(6.6) \quad \text{Det } A := \sum_{\sigma_2, \dots, \sigma_m \in S_n} \left(\left[\prod_{i=2}^m \text{Sign } \sigma_i \right] \left[\prod_{k=1}^n A(k, \sigma_2(k), \dots, \sigma_m(k)) \right] \right)$$

$$(6.7) \quad = \frac{1}{n!} \sum_{\sigma_1, \dots, \sigma_m \in S_n} \left(\left[\prod_{i=1}^m \text{Sign } \sigma_i \right] \left[\prod_{k=1}^n A(\sigma_1(k), \sigma_2(k), \dots, \sigma_m(k)) \right] \right).$$

This expression is called the *Cayley hyperdeterminant* of A .

The equality in the definition is straightforwardly verified, and it requires m to be even. If we replace A by K in (6.6) then (6.5) and (6.6) are the same, except that some signs appear in (6.6) that do not appear in (6.5). We conclude that the nonzero terms of the sum in (6.6) correspond precisely to the elements of $\Theta(R, T)$. This is Step 3 of the proposed approach. In order to prove that $|\Theta(R, T)| = \pm \text{Det } K$, it suffices to show that all terms of the sum in the definition of $\text{Det } K$ have the same sign (this is Step 4).

LEMMA 6.9. Let R be a region and let T be a tiling. Write K for the Kasteleyn hypermatrix. Then all nonzero terms in the sum in the definition of $\text{Det } K$ have the same sign.

PROOF. Let R, T and K be as in the lemma. We want to show that all terms of the sum in (6.6) (with A replaced with K) have the same sign. The idea is to show that the sign is invariant under making a local move as defined in Section 2.5.2. Write f for the unique height function such that $\Phi(f) = (0, T)$. The nonzero terms in (6.6) correspond bijectively (through the bijections that we have set up in Lemma 3.2, 3 and in Lemma 6.6) to the height functions in $\Omega(R, f)$. We pick two height functions $f', f'' \in \Omega(R, f)$ and prove that the corresponding terms in (6.6) have the same sign. By Lemma 2.2, we may assume, without loss of generality, that $f'' = f' + (d + 1) \cdot 1_{\mathbf{x}}$ for some $\mathbf{x} \in R$. Let $T', T'' \in \Theta(R, T)$ be the tilings corresponding to f', f'' respectively. Recall that

$$T' = \{ \{\mathbf{y}, \mathbf{y} + \mathbf{g}_i\} \in E^d : \nabla f'(\mathbf{y}, \mathbf{y} + \mathbf{g}_i) = -d \},$$

and for T'' we have an identical expression in terms of f'' . Remark that $f'' = f'$ except at the point \mathbf{x} , and therefore $\nabla f = \nabla f'$ except at the edges incident to \mathbf{x} . Since $f'' = f' + (d + 1) \cdot 1_{\mathbf{x}}$ and since both f' and f'' are height functions, we must have

$$\begin{aligned} \nabla f'(\mathbf{x}, \mathbf{x} + \mathbf{g}_i) &= \nabla f''(\mathbf{x} - \mathbf{g}_i, \mathbf{x}) = 1, \\ \nabla f'(\mathbf{x} - \mathbf{g}_i, \mathbf{x}) &= \nabla f''(\mathbf{x}, \mathbf{x} + \mathbf{g}_i) = -d, \end{aligned}$$

and therefore

$$(6.8) \quad \begin{aligned} T' \setminus T'' &= \{\{\mathbf{x}, \mathbf{x} - \mathbf{g}_i\} : 1 \leq i \leq d + 1\}, \\ T'' \setminus T' &= \{\{\mathbf{x}, \mathbf{x} + \mathbf{g}_i\} : 1 \leq i \leq d + 1\}. \end{aligned}$$

This means that all loops are matched the same (in the matchings $h(T')$ and $h(T'')$), except for the loops traversing the vertex \mathbf{x} . In order to prove the lemma, we work out the effect of this difference on the signs in (6.6). Let $(\sigma'_i)_{2 \leq i \leq d!}$ denote the bijections from Lemma 6.6 corresponding to T' . This means that

$$\{\{\mathbf{s}, \sigma'_2(\mathbf{s}), \dots, \sigma'_{d!}(\mathbf{s})\} : \mathbf{s} \in X_1\} = h(T' \cap E^d(R)).$$

Define, for each $1 \leq i \leq d!$, the bijections

$$\delta_i : X_i \rightarrow X_i, (\mathbf{x}, \xi^i) \mapsto \begin{cases} (\mathbf{x} + \mathbf{g}_j, \xi^i) & \text{if } (\mathbf{x}, \xi^i) \text{ traverses } \{\mathbf{x} - \mathbf{g}_j, \mathbf{x}\} \text{ for some } j, \\ (\mathbf{x}, \xi^i) & \text{if } (\mathbf{x}, \xi^i) \text{ does not traverse } \mathbf{x}. \end{cases}$$

Note that δ_i is a permutation consisting of one cycle of length $d + 1$. Claim that

$$(6.9) \quad \{\{\delta_1(\mathbf{s}), \delta_2 \circ \sigma'_2(\mathbf{s}), \dots, \delta_{d!} \circ \sigma'_{d!}(\mathbf{s})\} : \mathbf{s} \in X_1\} = h(T'' \cap E^d(R)).$$

To support the claim, recall (6.8) and observe simply that

$$\begin{aligned} &\{\delta_1(\mathbf{s}), \delta_2 \circ \sigma'_2(\mathbf{s}), \dots, \delta_{d!} \circ \sigma'_{d!}(\mathbf{s})\} \\ &= \begin{cases} h(\{\mathbf{x}, \mathbf{x} + \mathbf{g}_i\}) & \text{if } \{\mathbf{s}, \sigma'_2(\mathbf{s}), \dots, \sigma'_{d!}(\mathbf{s})\} = h(\{\mathbf{x}, \mathbf{x} - \mathbf{g}_i\}) \text{ for some } i, \\ \{\mathbf{s}, \sigma'_2(\mathbf{s}), \dots, \sigma'_{d!}(\mathbf{s})\} & \text{otherwise.} \end{cases} \end{aligned}$$

This proves the claim. Since $\delta_1 : X_1 \rightarrow X_1$ is a bijection, the sets in (6.9) are equal to

$$\{\{\mathbf{s}, \delta_2 \circ \sigma'_2 \circ \delta_1^{-1}(\mathbf{s}), \dots, \delta_{d!} \circ \sigma'_{d!} \circ \delta_1^{-1}(\mathbf{s})\} : \mathbf{s} \in X_1\}$$

This implies that the bijections from Lemma 6.6 corresponding to T'' are, for $2 \leq i \leq d!$,

$$\sigma''_i = \delta_i \circ \sigma'_i \circ \delta_1^{-1} : X_1 \rightarrow X_i.$$

Now note that $\text{Sign } \delta_i = (-1)^d$ (since δ_i is a cycle of length $d + 1$). Conclude that $\text{Sign } \delta_i \cdot \text{Sign } \delta_1^{-1} = 1$, and therefore σ'_i and σ''_i have the same sign in (6.6), for all i . \square

We have completed the final step of the approach that was suggested by the Kasteleyn theory for dimer covers. This yields the following theorem.

THEOREM 6.10. *Let R be a region, let f be a height function, and let $T = T(f)$. Write K for the Kasteleyn hypermatrix. Then $Z = |\Omega(R, f)| = |\Theta(R, T)| = \pm \text{Det } K$.*

6.4. Boltzmann measures. Recall the definition of a Boltzmann measure in Section 3.1. The number $|\Omega(R, f)| = |\Theta(R, T)|$ equals the partition function Z of the uniform probability measures on $\Omega(R, f)$ and $\Theta(R, T)$. The Kasteleyn theory is easily generalised to Boltzmann measures by inserting the weights into the Kasteleyn hypermatrix.

DEFINITION 6.11. Let R be a region, T a tiling, and $w : E^d(R) \rightarrow \mathbb{C}$ any (complex-valued) weight function. Define

$$\begin{aligned} K_w &: X_1 \times \dots \times X_{d!} \rightarrow \mathbb{C}, \\ (\mathbf{s}_1, \dots, \mathbf{s}_{d!}) &\mapsto 1(\{\mathbf{s}_1, \dots, \mathbf{s}_{d!}\} \in h(E^d(R))) \cdot w(h^{-1}(\{\mathbf{s}_1, \dots, \mathbf{s}_{d!}\})). \end{aligned}$$

The map K_w is called the *weighted Kasteleyn hypermatrix*.

By comparing the definition of Z_w with the definition of the Cayley hyperdeterminant, and taking into account Lemma 6.9, it is readily verified that $Z_w = \pm \text{Det } K_w$.

6.5. *Explicit calculations.* Two natural questions arise, which unfortunately we are not able to answer. First, one may ask if it is possible to find a simple expression for the partition function in the case of the torus, in order to derive a closed-form formula for the surface tension. This does not appear to be the case. The Cayley hyperdeterminant is invariant under a change of basis, and it is therefore meaningful to perform a Fourier transform on K . The transformed hypermatrix is “diagonal”, in the sense that its rank- $d!$ hyperdeterminant reduces to the permanent of a hypermatrix of rank $d! - 1$. This leads to the desired formula only in dimension two. Second, one may ask if it is possible to derive results on dimer-dimer correlations. The absence of a natural inverse for K leads the author to believe that also this second question is more difficult in the general case.

7. Gradient Gibbs measures. In previous sections, we introduced fixed boundary conditions and periodic boundary conditions, which enabled us to study probability measures on finite subsets of Ω . This section introduces *shift-invariant gradient Gibbs measures*, which prove to be an effective tool for studying the large-scale behaviour of the model. While gradient Gibbs measures are interesting in their own right, their main purpose here are their use in the proof of strict convexity of the surface tension in Section 9.

7.1. *Definition.* Write f for the random function in Ω . Define for any $R \subset X^d$,

$$\begin{aligned} \mathcal{F} &:= \sigma(f(\mathbf{x}) : \mathbf{x} \in X^d), & \mathcal{F}_R &:= \sigma(f(\mathbf{x}) : \mathbf{x} \in R), \\ \mathcal{F}^\nabla &:= \sigma(f(\mathbf{x}) - f(\mathbf{y}) : \mathbf{x}, \mathbf{y} \in X^d), & \mathcal{F}_R^\nabla &:= \sigma(f(\mathbf{x}) - f(\mathbf{y}) : \mathbf{x}, \mathbf{y} \in R). \end{aligned}$$

Note that $\mathcal{F}_R^\nabla = \mathcal{F}^\nabla \cap \mathcal{F}_R$ is finite whenever R is finite because it is generated by finitely many random variables, each taking finitely many values. Write $\mathcal{P}(\Omega, \mathcal{X})$ for the collection of probability measures on the measurable space (Ω, \mathcal{X}) for any σ -algebra \mathcal{X} on Ω . Probability measures in $\mathcal{P}(\Omega, \mathcal{F}^\nabla)$ are called *gradient measures*.

A gradient measure $\mu \in \mathcal{P}(\Omega, \mathcal{F}^\nabla)$ is called *shift-invariant* whenever $\mu(\tilde{\theta}A) = \mu(A)$ for any $A \in \mathcal{F}^\nabla$ and $\theta \in \Theta$, where $\tilde{\theta}A := \{\tilde{\theta}f : f \in A\}$. In other words, a gradient measure μ is shift-invariant whenever ∇f and $\theta\nabla f$ have the same law in μ for every $\theta \in \Theta$. The set of shift-invariant gradient measures is denoted by $\mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$. If $\mu \in \mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$, then it follows from shift-invariance that the map $\mu(f(\cdot) - f(\mathbf{0})) : X^d \rightarrow \mathbb{R}$ is additive over X^d . Therefore there exists a unique $s \in H^*$ such that $s(\mathbf{x}) = \mu(f(\mathbf{x}) - f(\mathbf{0}))$ for every $\mathbf{x} \in X^d$, and we must have $s \in \mathcal{S}$ because $s(\mathbf{g}_i) = \mu(f(\mathbf{g}_i) - f(\mathbf{0})) \leq 1$ for every $1 \leq i \leq d + 1$. Write $s(\mu)$ for s , the *slope* of $\mu \in \mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$.

Let (L, s) denote valid periodic boundary conditions and let μ denote the probability measure that is uniformly random in the finite set $\Omega(L, s)$. Lemma 3.10 implies that μ restricts to a shift-invariant gradient measure of slope $s(\mu) = s$.

Let us now introduce the notion of a Gibbs measure. Fix a measure $\mu \in \mathcal{P}(\Omega, \mathcal{F})$. The measure μ is called a *Gibbs measure* if for every finite $R \subset X^d$, the distribution of f in μ is the same as the distribution of a sample f obtained by first sampling g from μ , then sampling f from $\Omega(R, g)$ uniformly at random. The definition is formalised in terms of specifications and the Dobrushin–Lanford–Ruelle (DLR) equations. For each finite $R \subset X^d$, let γ_R denote the probability kernel from $(\Omega, \mathcal{F}_{R^c})$ to (Ω, \mathcal{F}) such that for any $f \in \Omega$, the probability measure $\gamma_R(\cdot, f)$ is uniform in $\Omega(R, f)$. It is obvious from the definition that $\Omega(R, f)$ is invariant under changing the values of f on R , so that $\gamma_R(A, \cdot)$ is indeed \mathcal{F}_{R^c} -measurable for every $A \in \mathcal{F}$. The kernels γ_R satisfy the consistency condition; if $S \subset R$, then $\gamma_R\gamma_S = \gamma_R$. The collection of probability kernels γ_R is called a *specification*, and a measure $\mu \in \mathcal{P}(\Omega, \mathcal{F})$ is called a *Gibbs measure* if μ satisfies the DLR equation

$$(7.1) \quad \mu = \mu\gamma_R$$

for each finite $R \subset X^d$. This is equivalent to our previous, informal description. By the consistency condition it is sufficient to check the DLR equations for an increasing exhaustive sequence of finite subsets of X^d . Each kernel γ_R restricts to a kernel from $(\Omega, \mathcal{F}_{R^c}^\nabla)$ to $(\Omega, \mathcal{F}^\nabla)$. We shall write γ_R^∇ for this restriction. A gradient measure $\mu \in \mathcal{P}(\Omega, \mathcal{F}^\nabla)$ is called a *gradient Gibbs measure* if

$$\mu = \mu\gamma_R^\nabla$$

for each finite subset R of X^d .

7.2. *Existence and concentration.*

THEOREM 7.1. *For each slope $s \in \mathcal{S}$, there is a shift-invariant gradient Gibbs measure $\mu \in \mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$ of slope s such that, for any $\mathbf{x}, \mathbf{y} \in X^d$, we have the bounds:*

1. $\text{Var}_\mu(f(\mathbf{y}) - f(\mathbf{x})) \leq (d + 1)^2n$,
2. $\mu(f(\mathbf{y}) - f(\mathbf{x}) - s(\mathbf{y} - \mathbf{x})) \geq (d + 1)a \leq \exp -\frac{a^2}{2n}$ for all $a \geq 0$ whenever $n > 0$,
3. $\mu(f(\mathbf{y}) - f(\mathbf{x}) - s(\mathbf{y} - \mathbf{x})) \leq (d + 1)a \leq \exp -\frac{a^2}{2n}$ for all $a \leq 0$ whenever $n > 0$,

where $n = d_{(X^d, E^d)}(\mathbf{x}, \mathbf{y})$.

The *topology of local convergence* or \mathcal{L} -*topology* on $\mathcal{P}(\Omega, \mathcal{X})$ is the coarsest topology that makes the evaluation map $\mu \mapsto \mu(A)$ continuous for every finite $R \subset X^d$ and for any $A \in \mathcal{X} \cap \mathcal{F}_R$. Constructing (gradient) Gibbs measures on $\mathcal{P}(\Omega, \mathcal{X})$ is much easier whenever choosing $\mathcal{X} = \mathcal{F}^\nabla$ and not $\mathcal{X} = \mathcal{F}$, because \mathcal{F}_R^∇ is finite for finite $R \subset X^d$ —see the following lemma.

LEMMA 7.2. *The set $\mathcal{P}(\Omega, \mathcal{F}^\nabla)$ is compact in the topology of local convergence.*

PROOF. The proof is entirely straightforward. Let $(\mu_n)_{n \in \mathbb{N}}$ denote a sequence of measures in $\mathcal{P}(\Omega, \mathcal{F}^\nabla)$ and let $(\Gamma_m)_{m \in \mathbb{N}}$ denote an increasing exhaustive sequence of finite subsets of X^d . Fix $m \in \mathbb{N}$. The σ -algebra $\mathcal{F}_{\Gamma_m}^\nabla$ is finite and therefore there exists a subsequence $(k_n)_{n \in \mathbb{N}} \subset \mathbb{N}$ such that μ_{k_n} converges on $\mathcal{F}_{\Gamma_m}^\nabla$ as $n \rightarrow \infty$. By a standard diagonalisation argument, we may assume that convergence occurs for all $m \in \mathbb{N}$. The limiting measure exists by the Kolmogorov extension theorem. \square

PROOF OF THEOREM 7.1. Let $s \in \mathcal{S}$ be the slope of interest. Let $(s_n)_{n \in \mathbb{N}}$ be a sequence of slopes converging to s with $s_n \in \mathcal{S}_n$ for every n . Write μ_n for the uniform probability measure on $\Omega(L_n, s_n)$, for every $n \in \mathbb{N}$. Each measure μ_n restricts to a shift-invariant gradient measure in $\mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$, and $s(\mu_n) = s_n$.

Now apply the previous lemma to obtain a subsequence $(k_n)_{n \in \mathbb{N}}$ along which the sequence of gradient measures $(\mu_n)_{n \in \mathbb{N}}$ converges in the topology of local convergence, say to $\mu \in \mathcal{P}(\Omega, \mathcal{F}^\nabla)$. The limit μ must be shift-invariant as all measures $(\mu_n)_{n \in \mathbb{N}}$ are shift-invariant. At each vertex $\mathbf{x} \in X^d$ we have

$$\mu(f(\mathbf{x}) - f(\mathbf{0})) = \lim_{n \rightarrow \infty} \mu_{k_n}(f(\mathbf{x}) - f(\mathbf{0})) = \lim_{n \rightarrow \infty} s_{k_n}(\mathbf{x}) = s(\mathbf{x}),$$

which means that $s(\mu) = s$. One shows similarly that (1)–(3) follow from Theorem 3.12.

It suffices to prove that the gradient measure μ is a Gibbs measure, that is, that $\mu\gamma_R^\nabla = \mu$ for every finite $R \subset X^d$. Fix a finite subset $R \subset X^d$. Now suppose that $\mu\gamma_R^\nabla$ equals μ on \mathcal{F}_S^∇ for any finite $S \subset X^d$. Then the two measures must be the same, by the uniqueness statement of the Kolmogorov extension theorem. It thus suffices to prove that $\mu\gamma_R^\nabla$ equals μ on \mathcal{F}_S^∇ for

any finite $S \subset X^d$. We may assume that $R \subset S$ and $\partial R \subset S$ by expanding S if necessary. By using shift-invariance, we may finally assume that $\mathbf{0} \notin S$.

We make the stronger claim that already in the nongradient setting and before taking limits, we have

$$(7.2) \quad \mu_n \upharpoonright_R |_{\mathcal{F}_S} = \mu_n |_{\mathcal{F}_S}$$

for n sufficiently large. The distribution μ_n is not invariant under resampling f on R . However, if $R + \mathbf{x}$ and $R + \mathbf{y}$ are disjoint and not adjacent for any $\mathbf{x}, \mathbf{y} \in L_n$ distinct, then μ_n is invariant under resampling f on R , then translating this modification to $R + \mathbf{x}$ for each $\mathbf{x} \in L_n \setminus \{\mathbf{0}\}$. Thus, if n is so large that S and $R + \mathbf{x}$ are disjoint for any $\mathbf{x} \in L_n \setminus \{\mathbf{0}\}$, then (7.2) holds true because the additional modifications do not affect the values of μ_n on the σ -algebra \mathcal{F}_S . This proves the claim. \square

For any $n \in \mathbb{N}$, let Π_n denote a *centred box* of sides $2n$, that is,

$$\Pi_n := \{a_1 \mathbf{g}_1 + \dots + a_d \mathbf{g}_d : -n \leq a_1, \dots, a_d < n\} \subset X^d.$$

Note that $|\Pi_n| = (2n)^d$.

PROPOSITION 7.3. *Let μ denote a measure of Theorem 7.1 of slope $s \in \mathcal{S}$. Then μ -almost surely*

$$(7.3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(f - f(\mathbf{0}))|_{\Pi_n} - s|_{\Pi_n}\|_\infty = 0.$$

This follows from a union bound and the inequalities in Theorem 7.1.

8. The surface tension and the variational principle. The purpose of this section is to give an overview of three closely related concepts which describe the macroscopic behaviour of the model. These motivate the study of strict convexity of the surface tension in Section 9. First, there is indeed the surface tension, which describes the asymptotic number of height functions approximating a certain slope. Second, there is the large deviations principle, which describes the asymptotic number of height functions approximating an arbitrary continuous profile. The rate function is the integral of the surface tension over the gradient of the continuous profile of interest. Third, there is the variational principle, which is a direct corollary of the large deviations principle, and describes the typical macroscopic behaviour of the random height function. The surface tension is usually convex, making that the rate function in the large deviations principle is also convex. In the next section, we shall also prove that the surface tension is strictly convex, which implies that the rate function has a unique minimiser, which in turn implies concentration around a single continuous profile in the variational principle. For the results in this section, we refer to [21] and [16].

8.1. *The surface tension.*

DEFINITION 8.1. Recall the definition of $\lfloor f \rfloor$ for Lipschitz functions $f : H \rightarrow \mathbb{R}$ on Page 1039. Recall also the definition of $\Pi_n \subset X^d$ at the end of the previous section (Page 1055). The *surface tension* is the function $\sigma : \mathcal{S} \rightarrow \mathbb{R}$ defined by

$$\sigma(s) := \lim_{n \rightarrow \infty} -\frac{1}{|\Pi_n|} \log |\Omega(\Pi_n, \lfloor s \rfloor)|.$$

For convergence of the limit in the definition of $\sigma(s)$, we refer to Section 4 in [16]. The argument is effectively a supermultiplicativity argument: if $A, B \subset X^d$ are finite and disjoint with no vertex of A adjacent to a vertex of B , then $|\Omega(A \cup B, f)| = |\Omega(A, f)| \cdot |\Omega(B, f)|$, and if $A \subset B$, then $|\Omega(A, f)| \leq |\Omega(B, f)|$. In fact, the definition of $\sigma(s)$ is stable under modifications of order $o(n)$ to $\lfloor s \rfloor$ as $n \rightarrow \infty$; see Lemma 4.5 in [16] for the following result.

THEOREM 8.2. *If $s \in \mathcal{S}$, and if $(f_n)_{n \in \mathbb{N}} \subset \Omega$ satisfies $\|f_n|_{\Pi_n} - s|_{\Pi_n}\|_\infty = o(n)$ as $n \rightarrow \infty$, then*

$$\sigma(s) = \lim_{n \rightarrow \infty} -\frac{1}{|\Pi_n|} \log |\Omega(\Pi_n, f_n)|.$$

The previous result implies immediately that σ is continuous; see also Lemma 4.3 in [16].

THEOREM 8.3. *The surface tension $\sigma : \mathcal{S} \rightarrow \mathbb{R}$ is continuous.*

8.2. The large deviations principle. Before stating the large deviations principle, we must introduce a suitable topological space to work in, and we must specify how a sequence of fixed boundary conditions converges to a continuous boundary profile. This is the purpose of the following definition.

DEFINITION 8.4. Write $\text{Lip}(D)$ for the collection of real-valued Lipschitz functions on D for any $D \subset H$. A *domain* is a bounded open set $D \subset H$ such that ∂D has zero Lebesgue measure. A *boundary profile* is a pair (D, b) where D is a domain and $b \in \text{Lip}(\partial D)$. An *approximation* of (D, b) is a sequence of pairs $((D_n, b_n))_{n \in \mathbb{N}}$ such that $D_n \subset X^d$ is finite and $b_n \in \Omega$ for any $n \in \mathbb{N}$, and such that

$$\frac{1}{n} D_n \rightarrow D, \quad \frac{1}{n} \text{Graph } b_n|_{\partial D_n} \rightarrow \text{Graph } b$$

in the Hausdorff topologies on H and $H \times \mathbb{R}$ respectively as $n \rightarrow \infty$.

If (D, b) is a boundary profile with approximation $((D_n, b_n))_{n \in \mathbb{N}}$, then write $(\gamma_n)_{n \in \mathbb{N}}$ for the sequence of measures defined by $\gamma_n := \gamma_{D_n}(\cdot, b_n)$, the uniform probability measure in the finite set $\Omega(D_n, b_n)$. The topological space for the large deviations principle associated with this sequence is the set $\text{Lip}(\bar{D})$ endowed with the topology of uniform convergence—which is equivalent to the topology of pointwise convergence as \bar{D} is compact. We must bring all samples from each measure γ_n to the space $\text{Lip}(\bar{D})$ for the large deviations principle to make sense. For each $n \in \mathbb{N}$, define the map $K_n : \Omega \rightarrow \text{Lip}(\bar{D})$ as follows. First, for each $f \in \Omega$, define \tilde{f} to be the smallest Lipschitz extension of f to H . Define each map K_n by

$$K_n(f) : \bar{D} \rightarrow \mathbb{R}, x \mapsto \frac{1}{n} \tilde{f}(nx).$$

Finally, let λ denote the unique translation-invariant measure on H for which

$$\{a_1 \mathbf{g}_1 + \dots + a_d \mathbf{g}_d : a_1, \dots, a_d \in [0, 1]\} \subset H$$

has measure one.

THEOREM 8.5. *Let (D, b) denote a boundary profile with approximation $((D_n, b_n))_{n \in \mathbb{N}}$ and associated measures $\gamma_n := \gamma_{D_n}(\cdot, b_n)$. Write γ_n^* for the pushforward of γ_n along K_n . Then the sequence of measures $(\gamma_n^*)_{n \in \mathbb{N}}$ satisfies a large deviations principle in the topological space $\text{Lip}(\bar{D})$ with speed n^d and rate function*

$$I(f) := -P(D, b) + \begin{cases} \int_D \sigma(\nabla f) d\lambda & \text{if } f|_{\partial D} = b, \\ \infty & \text{otherwise,} \end{cases}$$

where $P(D, b)$ is called the pressure of the boundary profile (D, b) , defined to be the unique constant such that the minimum of I is 0, and equal to

$$P(D, b) = \lim_{n \rightarrow \infty} -\frac{1}{n^d} \log |\Omega(D_n, b_n)|.$$

We shall continue using the definitions of I and $P(D, b)$ in the sequel. The large deviations principle was proven in a much more general setting by Sheffield in [21]. The large deviations principle with boundary conditions is stated in Section 7.3.2. The large deviations principle in [21] does not only address the macroscopic profile of each height functions but also its “local statistics” within macroscopic regions, something we are not concerned with here. In [21] it is required that the boundary profile is “not taut”. This requirement is however only necessary to understand the local statistics, and may be omitted when one is interested in the macroscopic profile only. For a more recent and elementary proof of the large deviations principle for the macroscopic profile only, we refer to Theorem 2.17 in [16].

8.3. *The variational principle.* The variational principle is a direct corollary of the large deviations principle. Note that the set of minimisers of I in Theorem 8.5 is exactly the set of minimisers of

$$\int_D \sigma(\nabla f) d\lambda$$

over all functions $f \in \text{Lip}(\bar{D})$ which restrict to b on ∂D .

THEOREM 8.6. *Assume the setting of Theorem 8.5. Let A denote an open neighbourhood of $\{I = 0\} \subset \text{Lip}(\bar{D})$. Then $\gamma_n^*(A) \rightarrow 1$ as $n \rightarrow \infty$. In particular, if σ is strictly convex, then I has a unique minimiser $f^* \in \text{Lip}(\bar{D})$, and in that case $\gamma_n^*(A) \rightarrow 1$ as $n \rightarrow \infty$ for any open neighbourhood A of f^* .*

9. Strict convexity of the surface tension. In [3], the authors find an explicit formula for σ for the case $d = 2$ by appealing to the integrable nature of the model. A direct corollary is that the surface tension is strictly convex. In [21], Sheffield proves that the surface tension related to any simply attractive model is strictly convex. In particular, this implies the following theorem.

THEOREM 9.1. *For any $d \geq 2$, the surface tension is strictly convex on the interior of S .*

The proof of Sheffield relies crucially on cluster swapping. The purpose of the section is to give an alternative proof of Theorem 9.1, which is simpler than the proof in [21] due to the special nature of the cluster swap in the particular setting of this article.

9.1. *The specific entropy.* First, we give an alternative characterisation of $\sigma(s)$ in terms of the shift-invariant gradient Gibbs measure of slope s whose existence is guaranteed by Theorem 7.1. For this, we require the notions of *entropy* and *specific entropy*. Let (X, \mathcal{X}) be an arbitrary measurable space endowed with a probability measure μ and a nonzero finite measure ν . Then the *relative entropy of μ with respect to ν* , denoted $\mathcal{H}(\mu, \nu)$, is defined by

$$\mathcal{H}(\mu, \nu) := \begin{cases} \nu(h \log h) = \mu(\log h) & \text{if } \mu \ll \nu \text{ and } h = d\mu/d\nu, \\ \infty & \text{if } \mu \not\ll \nu. \end{cases}$$

If \mathcal{A} is a sub- σ -algebra of \mathcal{X} , then write $\mathcal{H}_{\mathcal{A}}(\mu, \nu)$ for $\mathcal{H}(\mu|_{\mathcal{A}}, \nu|_{\mathcal{A}})$. It is well known that μ minimises $\mathcal{H}(\cdot, \nu)$ over all probability measures if and only if μ is the normalised version of

ν , in which case $\mathcal{H}(\mu, \nu) = -\log \nu(X)$. Also, if ν is a counting measure, then $h \leq 1$, and in that case $\mathcal{H}(\mu, \nu) \leq 0$.

If R is a finite subset of X^d , then write $D_R : \Omega \rightarrow \mathbb{Z}^{R \times R}$ for the map satisfying

$$(D_R f)(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}) - f(\mathbf{x})$$

for every $f \in \Omega$, $\mathbf{x}, \mathbf{y} \in R$. Call D_R the *differences map*. Note that $\text{Im } D_R$ is finite, and that $\mathcal{F}_R^\nabla = \sigma(D_R)$. Stronger: D_R may be seen as a bijection from \mathcal{F}_R^∇ to the powerset of $\text{Im } D_R$. Write λ^R for the pullback of the counting measure on $\text{Im } D_R$ along the map $D_R \rightarrow \lambda^R$ is a measure on $(\Omega, \mathcal{F}_R^\nabla)$ of size $\lambda^R(\Omega) = |\text{Im } D_R| \in \mathbb{Z}_{>0}$.

REMARK 9.2. If R is connected and $f \in \Omega$, then the values of $D_R f$ can be recovered from the values of ∇f on the edges of E^d which are contained in R , by integrating ∇f along the appropriate paths through R . Thus, for connected sets $R \subset X^d$, one should think of the map D_R as an alternative for the map

$$f \mapsto (\nabla f)|_{E^d \cap (R \times R)}$$

in the above construction. This also implies that $|\text{Im } D_R| \leq 2^{|R|}$ whenever R is connected.

Let $\mu \in \mathcal{P}(\Omega, \mathcal{F}^\nabla)$ and $R \subset X^d$ finite. Then the *entropy of μ in R* , denoted $\mathcal{H}_R(\mu)$, is defined by

$$\mathcal{H}_R(\mu) := \mathcal{H}_{\mathcal{F}_R^\nabla}(\mu, \lambda^R) = \sum_{x \in \text{Im } D_R} \mu(D_R f = x) \log \mu(D_R f = x) \in [-\log |\text{Im } D_R|, 0].$$

The *specific entropy* of μ , denoted $\mathcal{H}(\mu)$, is defined to be the limit

$$\mathcal{H}(\mu) := \lim_{n \rightarrow \infty} \frac{1}{|\Pi_n|} \mathcal{H}_{\Pi_n}(\mu) = \lim_{n \rightarrow \infty} \frac{1}{|\Pi_n|} \mathcal{H}_{\mathcal{F}_{\Pi_n}^\nabla}(\mu, \lambda^{\Pi_n})$$

whenever the sequence is convergent. Otherwise simply replace the limit by the limit inferior to obtain a well-defined limit. It can in fact be shown that the sequence is always convergent (see, e.g., [21], Chapter 2), but we shall not rely on this fact.

THEOREM 9.3. Let μ denote a measure of Theorem 7.1 of slope $s \in \mathcal{S}$. Then $\mathcal{H}(\mu) = \sigma(s)$.

PROOF. Let μ denote any gradient Gibbs measure for now. Write h^R for the Radon–Nikodym derivative

$$h^R := \frac{d\mu|_{\mathcal{F}_R^\nabla}}{d\lambda^R}$$

for any finite $R \subset X^d$. Fix $R, S \subset X^d$ finite with $S \cup \partial S \subset R$. As μ is Gibbs, we know that μ is uniformly random in $\Omega(S, f)$ whenever μ is conditioned on the values of f on S^c . This implies immediately that

$$h^R = \frac{1}{|\Omega(S, f)|} h^{R \setminus S}.$$

The function $|\Omega(S, \cdot)|$ is $\mathcal{F}_{\partial S}^\nabla$ -measurable as the model is Markov, and consequently h^R is $\mathcal{F}_{R \setminus S}^\nabla$ -measurable.

Let now μ be a measure of Theorem 7.1 of slope $s \in \mathcal{S}$, and pick $n \in \mathbb{N}$. Then

$$\begin{aligned} (2n)^{-d} \mathcal{H}_{\Pi_n}(\mu) &= (2n)^{-d} \mu(\log h^{\Pi_n}) = (2n)^{-d} \mu(\log h^{\Pi_n \setminus \Pi_{n-1}} - \log |\Omega(\Pi_{n-1}, f)|) \\ (9.1) \qquad &= (2n)^{-d} \mathcal{H}_{\Pi_n \setminus \Pi_{n-1}}(\mu) + \mu(-(2n)^{-d} \log |\Omega(\Pi_{n-1}, f)|). \end{aligned}$$

The first term in (9.2) vanishes as $n \rightarrow \infty$ because $\Pi_n \setminus \Pi_{n-1}$ is connected as a subset of (X^d, E^d) :

$$|\mathcal{H}_{\Pi_n \setminus \Pi_{n-1}}(\mu)| \leq \log |\text{Im } D_{\Pi_n \setminus \Pi_{n-1}}| \leq |\Pi_n \setminus \Pi_{n-1}| \log 2 = O(n^{d-1}).$$

The term within the expectation in (9.2) converges to $\sigma(s)$ pointwise by Proposition 7.3 and Theorem 8.2. We may apply the dominated convergence theorem because the expression within the expectation is always absolutely bounded by $\log 2$, since $0 \leq \log |\Omega(\Pi_{n-1}, f)| \leq |\Pi_{n-1}| \log 2 \leq (2n)^d \log 2$. \square

Before proceeding, let us quote an important result from the literature.

THEOREM 9.4. *Let $\mu \in \mathcal{P}_\Theta(\Omega, \mathcal{F}^\nabla)$ denote a measure which satisfies the concentration of (7.3) for some $s \in \mathcal{S}$, but which is not a Gibbs measure. Then $\mathcal{H}(\mu) > \sigma(s)$.*

PROOF OVERVIEW. Write $\mu^{n,s}$ for the measure μ conditioned on $f(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x} \in \Pi_n \setminus \Pi_{n-1}$. Then

$$\mathcal{H}_{\Pi_n}(\mu) = \mathcal{H}_{\Pi_n \setminus \Pi_{n-1}}(\mu) + \int \mathcal{H}_{\Pi_n}(\mu^{n,s}) d\mu(g).$$

This is the same decomposition as in the proof of the previous theorem. For fixed g , the integrand in this display is clearly minimised if μ is a Gibbs measure, because $\mu^{n,s}$ is then uniformly random in all extensions of $g|_{\Pi_n \setminus \Pi_{n-1}}$ to Π_n . This proves that $\mathcal{H}(\mu) \geq \sigma(s)$ whenever μ is concentrated as in (7.3); the difficulty is in proving the strict inequality whenever μ is not Gibbs. If μ is not Gibbs, then the integral in the display will for some n be strictly larger than if μ were Gibbs, but it is nontrivial to demonstrate that this difference survives the normalisation by $|\Pi_n|$ in the definition of $\mathcal{H}(\mu)$. This follows from a standard superadditivity argument; see Lemma 4.1 in [17] or Theorem 15.37 in [6]. See Theorem 2.5.2 in [21] for a proof of the current theorem in full detail. \square

9.2. The product setting. For the double dimer model, the cluster swap and the level set decomposition developed in this article, it is essential to work in the product setting. We shall introduce some straightforward technical machinery before proceeding; essentially we must adapt the constructions and results from Section 7 and from the previous subsection to the product setting. Write

$$\begin{aligned} \mathcal{F}^{2\nabla} &:= \mathcal{F}^\nabla \times \mathcal{F}^\nabla, & \gamma_R^2 &:= \gamma_R \times \gamma_R, & \lambda_2^R &:= \lambda^R \times \lambda^R, \\ \mathcal{F}_R^{2\nabla} &:= \mathcal{F}_R^\nabla \times \mathcal{F}_R^\nabla, & \gamma_R^{2\nabla} &:= \gamma_R^\nabla \times \gamma_R^\nabla. \end{aligned}$$

Let $\mathcal{P}(\Omega^2, \mathcal{F}^{2\nabla})$ denote the collection of probability measures on $(\Omega^2, \mathcal{F}^{2\nabla})$; such measures are called *double gradient measures*. If $\mu \in \mathcal{P}(\Omega^2, \mathcal{F}^{2\nabla})$, then we shall by default write (f_1, f_2) for the pair of random height functions, and $g := f_1 - f_2$ for the random difference. Write $\mathcal{P}_\Theta(\Omega^2, \mathcal{F}^{2\nabla})$ for the collection of shift-invariant measures $\mu \in \mathcal{P}(\Omega^2, \mathcal{F}^{2\nabla})$; the measure μ is called *shift-invariant* if $\mu(\tilde{\theta}A \times \tilde{\theta}B) = \mu(A \times B)$ for every $\theta \in \Theta$ and $A, B \in \mathcal{F}^\nabla$. This is equivalent to requiring that $(\nabla f_1, \nabla f_2)$ and $(\theta \nabla f_1, \theta \nabla f_2)$ have the same law in μ for each shift $\theta \in \Theta$.

The kernel $\gamma_R^2 = \gamma_R \times \gamma_R$ is simply the kernel from $(\Omega^2, \mathcal{F}_{Rc}^2)$ to $(\Omega^2, \mathcal{F}^2)$ with the property that the probability measure $(\gamma_R \times \gamma_R)(\cdot, (f_1, f_2))$ is uniformly random in the set $\Omega(R, f_1) \times \Omega(R, f_2)$, and it restricts naturally to the kernel $\gamma_R^{2\nabla} = \gamma_R^\nabla \times \gamma_R^\nabla$ —this is a probability kernel from $(\Omega^2, \mathcal{F}_{Rc}^{2\nabla})$ to $(\Omega^2, \mathcal{F}^{2\nabla})$. A double gradient measure μ is called a *double gradient Gibbs measure* if it satisfies, for every finite $R \subset X^d$, the DLR equation

$$\mu = \mu \gamma_R^{2\nabla}.$$

If μ is the product of two gradient Gibbs measures μ_1 and μ_2 , then μ is also Gibbs as

$$\mu = \mu_1 \times \mu_2 = (\mu_1 \gamma_R^\nabla) \times (\mu_2 \gamma_R^\nabla) = (\mu_1 \times \mu_2)(\gamma_R^\nabla \times \gamma_R^\nabla) = \mu \gamma_R^{2\nabla}.$$

Now let $\mu \in \mathcal{P}(\Omega^2, \mathcal{F}^{2\nabla})$ and $R \subset X^d$ finite. The entropy of μ in R , denoted $\mathcal{H}_R^2(\mu)$, is defined by

$$\mathcal{H}_R^2(\mu) := \mathcal{H}_{\mathcal{F}_R^{2\nabla}}(\mu, \lambda_R^R).$$

The specific entropy of μ , denoted $\mathcal{H}^2(\mu)$, is defined to be the limit

$$\mathcal{H}^2(\mu) := \lim_{n \rightarrow \infty} \frac{1}{|\Pi_n|} \mathcal{H}_{\Pi_n}^2(\mu) = \lim_{n \rightarrow \infty} \frac{1}{|\Pi_n|} \mathcal{H}_{\mathcal{F}_{\Pi_n}^{2\nabla}}(\mu, \lambda_2^{\Pi_n})$$

whenever the limit is convergent, and the limit inferior otherwise.

The direct generalisation of Theorems 9.3 and 9.4 to the product setting reads as follows.

THEOREM 9.5. *Let $\mu \in \mathcal{P}_\Theta(\Omega^2, \mathcal{F}^{2\nabla})$ denote a shift-invariant product measure such that*

$$(9.2) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(f_i - f_i(\mathbf{0}))|_{\Pi_n} - s_i|_{\Pi_n}\|_\infty = 0$$

almost surely for $i \in \{1, 2\}$ and for some fixed slopes $s_1, s_2 \in \mathcal{S}$. Then $\mathcal{H}^2(\mu) \geq \sigma(s_1) + \sigma(s_2)$, with equality if and only if μ is a Gibbs measure.

9.3. Proof overview. Fix throughout this section two distinct Lipschitz slopes $s_1, s_2 \in \mathcal{S}$ such that their average $s_a := (s_1 + s_2)/2$ lies in the interior of \mathcal{S} . The ultimate goal of this section is to prove that $2\sigma(s_a) < \sigma(s_1) + \sigma(s_2)$, which implies Theorem 9.1: that σ is strictly convex on the interior of \mathcal{S} .

In the remainder of this section, let μ_i denote the shift-invariant gradient Gibbs measure of Theorem 7.1 of slope s_i for each $i \in \{1, 2\}$, and fix $\mu := \mu_1 \times \mu_2$. Then μ is a shift-invariant double gradient Gibbs measure. Moreover, μ has the concentration of (9.2), and therefore Theorem 9.5 implies that $\mathcal{H}^2(\mu) = \sigma(s_1) + \sigma(s_2)$.

The sets $T(f_1)$ and $T(f_2)$, the graph $G_g = (V_g, E_g)$, the g -level sets, the g -boundaries, and the directed graph $(\text{LSD}(g), \nabla g)$ are all invariant under adding constants to f_1 and f_2 , as each of them is characterised entirely by the gradients $\nabla f_1, \nabla f_2$ and $\nabla g := \nabla f_1 - \nabla f_2$. The gradient ∇g also determines $X_g^\pm(E)$ for any g -boundary E .

LEMMA 9.6. *It is μ -almost certain that $\text{LSD}(g)$ contains a subgraph that is graph isomorphic to \mathbb{Z} . Moreover, every g -level set and every g -boundary involved in such a subgraph of $\text{LSD}(g)$ is unbounded.*

This lemma is essential in understanding the geometry of $\text{LSD}(g)$. It is expected that the difference function $g = f_1 - f_2$ of a typical sample from μ looks somewhat like the leftmost subfigure of Figure 4.

PROOF OF LEMMA 9.6. As $s_g := s_1 - s_2 \neq 0$, there exists an index $1 \leq i \leq d + 1$ such that $s_g(\mathbf{g}_i) \neq 0$. Fix such an i , and write \mathbf{p} for the \mathbb{Z} -indexed path $\mathbf{p} := (\mathbf{p}_k)_{k \in \mathbb{Z}} := (k\mathbf{g}_i)_{k \in \mathbb{Z}}$ through (X^d, E^d) . Write \mathbf{q}_k for the g -level set containing \mathbf{p}_k for each $k \in \mathbb{Z}$. For each $k \in \mathbb{Z}$ the vertices \mathbf{p}_k and \mathbf{p}_{k+1} are either contained in the same g -level set, or in two distinct neighbouring g -level sets. We consider $\mathbf{q} := (\mathbf{q}_k)_{k \in \mathbb{Z}}$ a walk through $\text{LSD}(g)$ although \mathbf{q} is not a walk in the strict sense: it may visit the same g -level set multiple times in a row. Proposition 7.3 says that μ -almost surely $g(\mathbf{p}_k) - g(\mathbf{p}_0) = ks_g(\mathbf{g}_i) + o(k)$ as $k \rightarrow \infty$ or

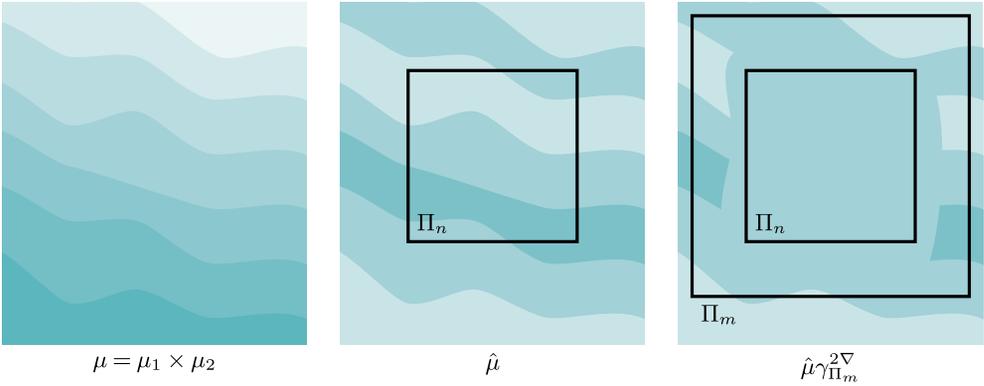


FIG. 4. Difference samples from three different measures; Π_m is a trifurcation box.

$k \rightarrow -\infty$. This implies that there is a well-defined and unique loop-erased bi-infinite version of the path \mathbf{q} up to indexation, which is the desired \mathbb{Z} -isomorphic subgraph of $\text{LSD}(g)$. This proves the first part of the lemma.

Focus on the second statement, which is deterministic in nature. Fix a g -boundary $E \subset E^d$ that is an edge of a subgraph of $\text{LSD}(g)$ that is isomorphic \mathbb{Z} . Then removing E from $\text{LSD}(g)$ disconnects $\text{LSD}(g)$ and separates the graph into two infinite components. In particular, this implies that the graph $(X^d, E^d \setminus E)$ consists of two infinite connected components. If E were finite, then one of the two connected components of $(X^d, E^d \setminus E)$ had to be finite, and therefore we conclude that E is infinite. The g -boundary E connects the two g -level sets $X_g^-(E)$ and $X_g^+(E)$ when considered an $\text{LSD}(g)$ -edge, and these must also be infinite as one of them contains the infinite set $\mathbf{x}_g^-(E)$ and the other $\mathbf{x}_g^+(E)$. This proves the second statement of the lemma. \square

We now give an overview of the remainder of the proof. The key idea is to construct a new shift-invariant double gradient measure $\hat{\mu} \in \mathcal{P}_\Theta(\Omega^2, \mathcal{F}^{2\nabla})$. Write \hat{f}_1, \hat{f}_2 and $\hat{g} := \hat{f}_1 - \hat{f}_2$ for the random functions in $\hat{\mu}$. To sample from $\hat{\mu}$, first draw a pair (f_1, f_2) from the original measure $\mu = \mu_1 \times \mu_2$. Then obtain (\hat{f}_1, \hat{f}_2) from (f_1, f_2) by flipping a fair coin for every g -boundary in order to determine whether or not to reverse the orientation of that g -boundary. In other words, we rerandomise the orientation of each g -boundary. In the measure $\hat{\mu}$, the orientations of the edges in the graph $(\text{LSD}(\hat{g}), \nabla \hat{g})$ are thus uniformly random and independent of all other structure that is present. First, we show that the resampling operation does not affect the specific entropy, that is,

$$\mathcal{H}^2(\hat{\mu}) = \mathcal{H}^2(\mu) = \sigma(s_1) + \sigma(s_2).$$

Second, we prove that for $i \in \{1, 2\}$ we $\hat{\mu}$ -almost surely have

$$(9.3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(\hat{f}_i - \hat{f}_i(\mathbf{0}))|_{\Pi_n} - s_a|_{\Pi_n}\|_\infty = 0.$$

Note that the concentration of the gradient of either function is around the average slope s_a . Third, we prove that $\hat{\mu}$ is not Gibbs. Theorem 9.5 therefore implies that $\mathcal{H}^2(\hat{\mu}) > 2\sigma(s_a)$, the desired result.

Let us elaborate on the third step, before proceeding. Suppose that $\hat{\mu}$ is Gibbs, in order to derive a contradiction. A trifurcation box of \hat{g} is a finite subset of X^d of the form $R = \theta \Pi_n$ such that, for some infinite \hat{g} -level set $X \subset X^d$, removing R from X means breaking X into at least three infinite components. We show that \hat{g} has a trifurcation box with positive

probability in the measure $\hat{\mu}\gamma_{\Pi_m}^{2\nabla}$ for m sufficiently large: see the middle and rightmost sub-figures in Figure 4. If $\hat{\mu}$ were Gibbs then $\hat{\mu} = \hat{\mu}\gamma_{\Pi_m}^{2\nabla}$, and therefore a sample \hat{g} from $\hat{\mu}$ has a trifurcation box with positive probability. Trifurcation boxes do almost surely not occur in shift-invariant measures, by a simple geometrical argument described by Burton and Keane in their celebrated paper [2]. This proves that $\hat{\mu}$ is not Gibbs.

9.4. *Detailed proof.*

LEMMA 9.7. *The specific entropy of $\hat{\mu}$ equals the specific entropy of μ .*

PROOF. We prove the stronger statement that $\mathcal{H}_{\Pi_n}^2(\hat{\mu}) = \mathcal{H}_{\Pi_n}^2(\mu) + O(n^{d-1})$ as $n \rightarrow \infty$. Fix $n \in \mathbb{N}$ large. The measure μ is Gibbs and therefore satisfies the DLR equation

$$\mu = \mu\gamma_{\Pi_n}^{2\nabla}.$$

This implies in particular that the distribution of a sample (f_1, f_2) from μ is invariant under subsequently rerandomising the orientation of each g -boundary that is contained in $E^d(\Pi_n)$. As this is true for all n , we derive that the distribution of the sample is in fact invariant under rerandomising the orientation of any finite g -boundary. Thus, to sample (\hat{f}_1, \hat{f}_2) from $\hat{\mu}$, one may first sample a pair (f_1, f_2) from μ , then rerandomise the orientation of only the g -boundaries which are infinite. Write $\tilde{\mu}$ for the corresponding coupling of μ and $\hat{\mu}$, and write $(f_1, f_2, \hat{f}_1, \hat{f}_2)$ for the random 4-tuple in $\tilde{\mu}$.

Write $\mathcal{H}_{A,B}^2(\tilde{\mu})$ for $\mathcal{H}_{\mathcal{F}_A^{2\nabla} \times \mathcal{F}_B^{2\nabla}}(\tilde{\mu}, \lambda_2^A \times \lambda_2^B)$ for $A, B \subset X^d$ finite and claim that:

1. $\mathcal{H}_{\Pi_n}^2(\mu) = \mathcal{H}_{\Pi_n, \Pi_n \setminus \Pi_{n-1}}^2(\tilde{\mu}) + O(n^{d-1})$,
2. $\mathcal{H}_{\Pi_n}^2(\hat{\mu}) = \mathcal{H}_{\Pi_n \setminus \Pi_{n-1}, \Pi_n}^2(\tilde{\mu}) + O(n^{d-1})$,
3. $\mathcal{H}_{\Pi_n, \Pi_n \setminus \Pi_{n-1}}^2(\tilde{\mu}) = \mathcal{H}_{\Pi_n, \Pi_n}^2(\tilde{\mu}) = \mathcal{H}_{\Pi_n \setminus \Pi_{n-1}, \Pi_n}^2(\tilde{\mu})$.

It is clear that these three claims jointly imply the lemma.

Focus on the first claim, and consider thus the measures

$$\mu|_{\mathcal{F}_{\Pi_n}^{2\nabla}}, \quad \tilde{\mu}|_{\mathcal{F}_{\Pi_n}^{2\nabla} \times \mathcal{F}_{\Pi_n \setminus \Pi_{n-1}}^{2\nabla}}.$$

The first claim is intuitive: the restriction of μ records the values of $D_{\Pi_n}(f_1)$ and $D_{\Pi_n}(f_2)$, and the restriction of $\tilde{\mu}$ records also the values of $D_{\Pi_n \setminus \Pi_{n-1}}(\hat{f}_1)$ and $D_{\Pi_n \setminus \Pi_{n-1}}(\hat{f}_2)$. Informally, the extra information that the restriction of $\tilde{\mu}$ records is of order n^{d-1} , because $\log |\text{Im } D_{\Pi_n \setminus \Pi_{n-1}}|^2 = O(n^{d-1})$. We now formalise this idea. For $x \in (\text{Im } D_{\Pi_n})^2$, we write $\hat{\mu}^x$ for the measure $\tilde{\mu}$ conditioned on the event $\{(D_{\Pi_n} f_1, D_{\Pi_n} f_2) = x\}$ and projected onto the product of the third and fourth component of the product measurable space $(\Omega, \mathcal{F}^\nabla)^4$. Then a standard entropy calculation implies that

$$\mathcal{H}_{\Pi_n, \Pi_n \setminus \Pi_{n-1}}^2(\tilde{\mu}) = \mathcal{H}_{\Pi_n}^2(\mu) + \int \mathcal{H}_{\Pi_n \setminus \Pi_{n-1}}^2(\hat{\mu}^x) d\mu((D_{\Pi_n} f_1, D_{\Pi_n} f_2) = x),$$

see Theorems D.3 and D.13 of [4] or Lemma 2.1.3 of [21]. As in the proof of Theorem 9.3, we have

$$|\mathcal{H}_{\Pi_n \setminus \Pi_{n-1}}^2(\hat{\mu}^x)| \leq \log |\text{Im } D_{\Pi_n \setminus \Pi_{n-1}}|^2 = O(n^{d-1}).$$

This proves the first claim. The second claim follows by identical reasoning.

Focus on the third claim, in particular on the equality on the left—the equality on the right shall follow by the same arguments. For the equality on the left it suffices to demonstrate that, with $\tilde{\mu}$ -probability one, the tuple

$$(9.4) \quad (D_{\Pi_n} f_1, D_{\Pi_n} f_2, D_{\Pi_n} \hat{f}_1, D_{\Pi_n} \hat{f}_2)$$

can be reconstructed almost surely from

$$(9.5) \quad (D_{\Pi_n} f_1, D_{\Pi_n} f_2, D_{\Pi_n \setminus \Pi_{n-1}} \hat{f}_1, D_{\Pi_n \setminus \Pi_{n-1}} \hat{f}_2).$$

We know that (f_1, f_2) and (\hat{f}_1, \hat{f}_2) differ by cluster boundary swaps, where all boundaries which are swapped, are infinite. To recover (9.4) from (9.5), we must therefore understand whether or not each infinite boundary which intersects Π_{n-1} , should be swapped or not. However, as each such boundary is infinite, it must intersect $\Pi_n \setminus \Pi_{n-1}$, and its orientation for (\hat{f}_1, \hat{f}_2) can therefore be read off from (9.5). \square

LEMMA 9.8. *Equation (9.3) holds true $\hat{\mu}$ -almost surely for $i \in \{1, 2\}$.*

PROOF. It suffices to prove that $\hat{\mu}$ -almost surely

$$(9.6) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n} + (\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_n} - 2s_a|_{\Pi_n}\|_{\infty} = 0,$$

$$(9.7) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n} - (\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_n}\|_{\infty} = 0.$$

First focus on (9.6). Recall the definition of μ ; Proposition 7.3 implies that almost surely

$$\lim_{n \rightarrow \infty} \frac{1}{n} \|(f_1 - f_1(\mathbf{0}))|_{\Pi_n} + (f_2 - f_2(\mathbf{0}))|_{\Pi_n} - (s_1 + s_2)|_{\Pi_n}\|_{\infty} = 0.$$

Recall that the sum of two height functions is invariant under a cluster boundary swap, and that $2s_a = s_1 + s_2$. The previous display therefore implies (9.6). For (9.7), we must show that $\hat{\mu}$ -almost surely

$$(9.8) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \|(\hat{g} - \hat{g}(\mathbf{0}))|_{\Pi_n}\|_{\infty} = 0$$

Fix some $\mathbf{x} \in X^d$; we are interested in the distribution of $\hat{g}(\mathbf{x}) - \hat{g}(\mathbf{0})$. The distribution of $\hat{g}(\mathbf{x}) - \hat{g}(\mathbf{0})$ conditional on $\text{LSD}(\hat{g})$ is given by summing $d_{\text{LSD}(\hat{g})}(\mathbf{0}, \mathbf{x}) \leq d_{(X^d, E^d)}(\mathbf{0}, \mathbf{x})$ fair coins each valued $\pm(d + 1)$. This follows immediately from the definition of the measure $\hat{\mu}$. Application of the Azuma–Hoeffding inequality yields, for $a \geq 0$ and $\mathbf{x} \neq \mathbf{0}$,

$$\hat{\mu}(|\hat{g}(\mathbf{x}) - \hat{g}(\mathbf{0})| \geq (d + 1)a) \leq 2 \exp - \frac{a^2}{2d_{(X^d, E^d)}(\mathbf{0}, \mathbf{x})}.$$

A union bound now implies (9.8). \square

LEMMA 9.9. *The double gradient measure $\hat{\mu}$ is not Gibbs.*

PROOF. Define the $\mathcal{F}^{2\vee}$ -measurable event

$$I(n) := \left\{ \hat{g} \text{ takes the same value on three distinct infinite } \hat{g}\text{-level sets} \right. \\ \left. \text{that all intersect } \Pi_n, \text{ one of which contains } \mathbf{0} \right\} \subset \Omega^2,$$

and claim that $\hat{\mu}(I(n)) > 0$ for $n \in \mathbb{N}$ sufficiently large. First, a cluster boundary swap leaves $\text{LSD}(g)$ invariant, and therefore Lemma 9.6 holds true for g replaced with \hat{g} and μ with $\hat{\mu}$. Therefore, it is $\hat{\mu}$ -almost certain that $\text{LSD}(\hat{g})$ contains a \mathbb{Z} -indexed self-avoiding walk $\mathbf{p} = (\mathbf{p}_k)_{k \in \mathbb{Z}}$. Let \mathbf{p} be chosen deterministically in terms of $\text{LSD}(\hat{g})$, so that $(\hat{g}(\mathbf{p}_{k+1}) - \hat{g}(\mathbf{p}_k))_{k \in \mathbb{Z}}$ is a sequence of i.i.d. random variables each distributed uniformly in $\pm(d + 1)$, independent of $\text{LSD}(\hat{g})$. In particular, the event $\{\hat{g}(\mathbf{p}_{k+2}) - \hat{g}(\mathbf{p}_k) = 0\}$ has probability $\frac{1}{4}$ for each fixed k . As $\hat{\mu}$ is shift-invariant, we may choose \mathbf{p} such that $\hat{\mu}(\mathbf{0} \in \mathbf{p}_0) > 0$. Choose $n \in \mathbb{N}$ sufficiently large such that, conditional on $\{\mathbf{0} \in \mathbf{p}_0\}$, the set Π_n intersects $\mathbf{p}_{\pm 2}$ with positive probability.

Note that $\hat{g}(\mathbf{p}_{\pm 2}) - \hat{g}(\mathbf{p}_0) = 0$ with probability $\frac{1}{4}$ independently of the occurrence of both previous events, and therefore the original event $I(n)$ has positive probability. This is the claim. Fix $n \in \mathbb{N}$ such that $\varepsilon := \frac{1}{2}\hat{\mu}(I(n)) > 0$. See the middle display in Figure 4 for the level set decomposition of the difference function \hat{g} corresponding to a sample from the event $I(n)$.

If h and h' are real-valued functions defined on two disjoint subsets A and A' of X^d respectively, then write hh' for the unique function on $A \cup A'$ which equals h on A and h' on A' . Next, define for $m \geq n$ the $\mathcal{F}^{2\nabla}$ -measurable event

$$L(m) := \{ \text{the function } (\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n} (\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_m^c} \text{ extends to a height function} \} \\ = \{ \text{the function } (\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n} (\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_m^c} \text{ is Lipschitz} \} \subset \Omega^2,$$

and claim that $\hat{\mu}(L(m)) \rightarrow 1$ as $m \rightarrow \infty$. Recall that (9.3) holds true for $i = 2$ with $\hat{\mu}$ -probability one. Therefore it suffices to show that the function

$$(\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n} (\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_m^c}$$

is Lipschitz for m sufficiently large whenever \hat{f}_2 satisfies (9.3). To see this, write h_m^\pm for the largest and smallest functions in $\Omega(\Pi_m, \hat{f}_2 - \hat{f}_2(\mathbf{0}))$, respectively. In other words, the functions h_m^\pm are the largest and smallest extensions of $(\hat{f}_2 - \hat{f}_2(\mathbf{0}))|_{\Pi_m^c}$ to X^d which are height functions. Note that there exist constants $\alpha^\pm > 0$ such that $h_m^\pm(\mathbf{0}) = \pm\alpha^\pm m + o(m)$ as $m \rightarrow \infty$ due to (9.3) and because s_a is in the interior of the set of Lipschitz slopes. In particular,

$$h_m^+|_{\Pi_n} \geq (\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n}, \quad h_m^-|_{\Pi_n} \leq (\hat{f}_1 - \hat{f}_1(\mathbf{0}))|_{\Pi_n}$$

for m sufficiently large. This proves the claim. Fix m so large that $\hat{\mu}(L(m)) \geq 1 - \varepsilon$, which implies that $\hat{\mu}(I(n) \cap L(m)) \geq \varepsilon > 0$.

Define for $\mathbf{x} \in X^d$ the $\mathcal{F}^{2\nabla}$ -measurable event

$$T(\mathbf{x}) := \left\{ \begin{array}{l} (X^d, E^d \setminus (V_{\hat{g}} \cup E^d(\Pi_m + \mathbf{x}))) \text{ has three infinite connected} \\ \text{components that are contained in a single } \hat{g}\text{-level set} \end{array} \right\} \\ = \left\{ \begin{array}{l} (X^d, E^d \setminus (V_{\hat{g}} \cup E^d(\Pi_m + \mathbf{x}))) \text{ has three infinite connected} \\ \text{components that are contained in one connected component of} \\ (X^d, E^d \setminus V_{\hat{g}}) \end{array} \right\} \subset \Omega^2;$$

we shall first focus on $T := T(\mathbf{0})$. See the rightmost subfigure in Figure 4 for the level set decomposition of the difference function \hat{g} corresponding to a sample from T . We claim that $\Omega(\Pi_m, f_1) \times \Omega(\Pi_m, f_2)$ intersects T whenever $(f_1, f_2) \in I(n) \cap L(m)$. Fix such a pair (f_1, f_2) and assume, without loss of generality, that $f_1(\mathbf{0}) = f_2(\mathbf{0}) = 0$. Assert first that there exists a height function f'' which equals f_1 on Π_n and which equals f_2 on Π_m^c , and which at any vertex \mathbf{x} takes values between $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$. The following construction works: write first f' for a height function which extends $f_1|_{\Pi_n} f_2|_{\Pi_m^c}$ to X^d , then define f'' by

$$f'' := (f' \vee (f_1 \wedge f_2)) \wedge (f_1 \vee f_2).$$

This proves the assertion. Now $f'' \in \Omega(\Pi_m, f_2)$ as f'' equals f_2 on Π_m^c . Moreover, as at each vertex the value of f'' is in between the values of f_1 and f_2 , we have $\{f_1 = f_2\} \subset \{f_1 = f''\} \subset X^d$, and we also know that $\Pi_n \subset \{f_1 = f''\}$. This implies that $(f_1, f'') \in T$, since the three connected components of $\{f_1 = f_2\}$ are contained in a single connected component of $\{f_1 = f''\}$, and because $\{f_1 = f_2\} \setminus \Pi_m = \{f_1 = f''\} \setminus \Pi_m$, that is, removing Π_m disconnects these three components again.

Suppose that $\hat{\mu}$ is a Gibbs measure, in order to derive a contradiction. Then the event T must occur with positive probability for $\hat{\mu}$, since $\hat{\mu}(I(n) \cap L(m)) > 0$, and since T has positive probability for

$$\gamma_{\Gamma_m}^{2\nabla}(\cdot, (\hat{f}_1, \hat{f}_2))$$

whenever $(\hat{f}_1, \hat{f}_2) \in I(n) \cap L(m)$. The argument of Burton and Keane [2] dictates that trifurcation boxes do almost surely not occur in shift-invariant probability measures. In other words, $\hat{\mu}(T) = 0$. This is the desired contradiction. To see that $\hat{\mu}(T) = 0$, observe that $\hat{\mu}(T(\mathbf{x}))$ is independent of \mathbf{x} , implying that trifurcation boxes must occur with positive density whenever $\hat{\mu}(T) > 0$. This is impossible due to geometrical constraints of the amenable graph (X^d, E^d) . \square

This establishes Theorem 9.1.

Acknowledgement. The author is grateful to Nathanaël Berestycki and James Norris for their enthusiastic supervision of the writing of this paper. The author thanks Amir Dembo, Richard Kenyon, Scott Sheffield and Martin Tassy for helpful discussions. Finally, the author would like to thank the anonymous referees for their insightful feedback.

The author was supported by the Department of Pure Mathematics and Mathematical Statistics, University of Cambridge and the UK Engineering and Physical Sciences Research Council grant EP/L016516/1.

REFERENCES

- [1] BAXTER, R. J. (1982). *Exactly Solved Models in Statistical Mechanics*. Academic Press, London. [MR0690578](#)
- [2] BURTON, R. M. and KEANE, M. (1989). Density and uniqueness in percolation. *Comm. Math. Phys.* **121** 501–505. [MR0990777](#)
- [3] COHN, H., KENYON, R. and PROPP, J. (2001). A variational principle for domino tilings. *J. Amer. Math. Soc.* **14** 297–346. [MR1815214](#) <https://doi.org/10.1090/S0894-0347-00-00355-6>
- [4] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications. Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. [MR2571413](#) <https://doi.org/10.1007/978-3-642-03311-7>
- [5] EDWARDS, R. G. and SOKAL, A. D. (1988). Generalization of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm. *Phys. Rev. D* (3) **38** 2009–2012. [MR0965465](#) <https://doi.org/10.1103/PhysRevD.38.2009>
- [6] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. [MR2807681](#) <https://doi.org/10.1515/9783110250329>
- [7] GRIMMETT, G. (2018). *Probability on Graphs: Random Processes on Graphs and Lattices*, 2nd ed. *Institute of Mathematical Statistics Textbooks* **8**. Cambridge Univ. Press, Cambridge. [MR3751350](#) <https://doi.org/10.1017/9781108528986>
- [8] KASTELEYN, P. W. (1961). The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice. *Physica* **27** 1209–1225.
- [9] KASTELEYN, P. W. (1963). Dimer statistics and phase transitions. *J. Math. Phys.* **4** 287–293. [MR0153427](#) <https://doi.org/10.1063/1.1703953>
- [10] KASTELEYN, P. W. (1967). Graph theory and crystal physics. In *Graph Theory and Theoretical Physics* 43–110. Academic Press, London. [MR0253689](#)
- [11] KENYON, R. (2014). Conformal invariance of loops in the double-dimer model. *Comm. Math. Phys.* **326** 477–497. [MR3165463](#) <https://doi.org/10.1007/s00220-013-1881-0>
- [12] KENYON, R. (2018). Limit shapes and their analytic parameterizations. *Proc. Int. Cong. Math.* **3** 3123–3138.
- [13] KENYON, R. and OKOUNKOV, A. (2007). Limit shapes and the complex Burgers equation. *Acta Math.* **199** 263–302. [MR2358053](#) <https://doi.org/10.1007/s11511-007-0021-0>
- [14] KENYON, R., OKOUNKOV, A. and SHEFFIELD, S. (2006). Dimers and amoebae. *Ann. of Math. (2)* **163** 1019–1056. [MR2215138](#) <https://doi.org/10.4007/annals.2006.163.1019>
- [15] KENYON, R. and PEMANTLE, R. (2016). Double-dimers, the Ising model and the hexahedron recurrence. *J. Combin. Theory Ser. A* **137** 27–63. [MR3403514](#) <https://doi.org/10.1016/j.jcta.2015.07.005>

- [16] KRIEGER, A., MENZ, G. and TASSY, M. (2019). Deducing a variational principle with minimal a priori assumptions. Preprint. Available at [arXiv:1910.01042v2](https://arxiv.org/abs/1910.01042v2).
- [17] LAMMERS, P. G. and TASSY, M. (2020). Variational principle for weakly dependent random fields. *J. Stat. Phys.* **179** 846–870. [MR4102443](https://doi.org/10.1007/s10955-020-02538-8) <https://doi.org/10.1007/s10955-020-02538-8>
- [18] LIEB, E. H. (1967). Residual entropy of square ice. *Phys. Rev.* **162** 162–172.
- [19] LINDE, J., MOORE, C. and NORDAHL, M. G. (2001). An n -dimensional generalization of the rhombus tiling. In *Discrete Models: Combinatorics, Computation, and Geometry (Paris, 2001)*. *Discrete Math. Theor. Comput. Sci. Proc.*, AA 023–042. Maison Inform. Math. Discrèt. (MIMD), Paris. [MR1888761](https://doi.org/10.1007/s10955-020-02538-8)
- [20] ONSAGER, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)* **65** 117–149. [MR0010315](https://doi.org/10.1007/s10955-020-02538-8)
- [21] SHEFFIELD, S. (2005). Random surfaces. *Astérisque* **304**. [MR2251117](https://doi.org/10.1007/s10955-020-02538-8)
- [22] SUTHERLAND, B. (1967). Exact solution of a two-dimensional model for hydrogen-bonded crystals. *Phys. Rev. Lett.* **19** 103–104.
- [23] SWENDSEN, R. H. and WANG, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58** 86–88.
- [24] TEMPERLEY, H. N. V. and FISHER, M. E. (1961). Dimer problem in statistical mechanics—an exact result. *Philos. Mag. (8)* **6** 1061–1063. [MR0136398](https://doi.org/10.1007/s10955-020-02538-8)
- [25] YANG, C. N. (1952). The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev. (2)* **85** 808–816. [MR0051740](https://doi.org/10.1007/s10955-020-02538-8)
- [26] YANG, C. P. (1967). Exact solution of a model of two-dimensional ferroelectrics in an arbitrary external electric field. *Phys. Rev. Lett.* **19** 586–588.