Random geometry in two and three dimensions

by

Catherine C. Wolfram

B.A. Mathematics, University of Chicago, 2019

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by:	Catherine C. Wolfram Department of Mathematics April 30, 2024
Certified by:	Scott Sheffield Leighton Family Professor of Mathematics, Thesis Supervisor
Accepted by:	Davesh Maulik Professor of Mathematics Graduate Chair, Department of Mathematics

Random geometry in two and three dimensions

by

Catherine C. Wolfram

Submitted to the Department of Mathematics
on April 30, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

A central theme in random geometry is the interplay between discrete models and continuum ones that appear in scaling limits. Surprising structure and symmetry often arises in these scaling limits, leading to an interplay between combinatorics, probability, complex analysis, and geometry.

The dimer model is one of the classical lattice models of statistical mechanics and can be defined in any dimension. In the first half of this thesis, we prove a large deviation principle for dimer tilings in three dimensions. This generalizes a two-dimensional result of Cohn, Kenyon, and Propp, and is one of the first results for dimers in any dimension $d > 2$. Many ideas and constructions used to study dimers are specific to two dimensions, so our arguments start from a smaller set of tools including Hall's matching theorem, the qualitative description of the Gibbs property, and a double dimer swapping operation.

In the second half of this thesis, we study discrete, geometrically-motivated coordinates called shears on the space of circle homeomorphisms up to Möbius transformations. The Weil–Petersson Teichmüller space is a subspace of this which has been of long-term interest in geometry and string theory and has recent connections to SLE curves in probability. We introduce and study natural ℓ^2 spaces in terms of shears, and obtain sharp results comparing them to Hölder classes of circle homeomorphisms and the Weil–Petersson class. We also give a preliminary result about i.i.d. Gaussian random shears.

Thesis supervisor: Scott Sheffield
Title: Leighton Family Professor of Mathematics

Acknowledgments

First and foremost, I thank my advisor Scott Sheffield: for his generosity with his time and ideas, for his willingness to take me as a student despite the fact that I was only just starting to learn probability, and for teaching me and directing me towards lots of interesting math. I could not have asked for a better advisor, and would not be where I am now without his guidance. I would also like to thank Alexei Borodin and Curt McMullen, for serving on my thesis committee and for other helpful advice and support they have given me during my PhD.

Thank you to my collaborators Sky Cao, Nishant Chandgotia, Richard Kenyon, Dragomir Šarić, Jinwoo Sung, and Yilin Wang, who I have had the pleasure of working with over the last few years. The work presented in this thesis comes from some of these collaborations. I owe a special thank you to Yilin, who has taught me so much, and who has always been extremely kind and encouraging, especially when I was getting started doing research. Thank you to the other members of my group at MIT for interesting conversations and company: Morris Ang, Manan Bhatia, Sky Cao, Konstantinos Kavvadias, Minjae Park, Ron Nissim, and Pu Yu.

Thank you to Benson Farb, Maryanthe Malliaris, and Peter May for their advice, mentorship and encouragement when I was an undergraduate at the University of Chicago. I thank IAS and IHES for supporting longer visits during my PhD, which were energizing and productive.

For making every day fun, I thank my officemates in 2-333C at MIT: Alex, Calder, Cameron, Marisa, Matthew, and Natalia. I feel incredibly lucky to have been able to work in such a lively, friendly environment every day. Thank you also to a few of my closest friends in Cambridge (who I didn't happen to share an office with), Naomi, Noah, Arthur, Yan Sheng, and Mary.

Finally, thank you to my parents, siblings, and Colin, for supporting and believing in me.

Contents

List of Figures

Chapter 1

Introduction

Statistical mechanics uses probability to understand the large-scale behavior of systems of many small entities. The main idea is to approximate the system by a random state of a simple-to-define combinatorial model on a lattice or graph, and then study behavior in a *scaling limit*, where the objects are suitably rescaled and the lattice mesh-size goes to zero. These lattice models are simple to define combinatorially in any dimension. Classic examples include the Ising model, dimer model, or percolation.

Despite their simplicity, surprising structure and symmetry often arises in scaling limits. In two dimensions, scaling limits of statistical mechanics models at criticality are (or are strongly conjectured to be) conformally invariant. This principle has been proven in various cases and together with the discovery of *SLE curves* in the late 1990s, has lead to the launch of the field of random conformal geometry in two dimensions, and more recently to connections between Teichmüller theory and probability.

Three-dimensional statistical mechanics models have long been of interest in physics, but so far much less general theory is understood. This is starting to become an active area, and there are a number of new interesting results, though relationships between them are not well understood even conjecturally yet.

This thesis has two main topics:

- the dimer model (a classical statistical mechanics model) in three dimensions
- discrete coordinates in Teichmüller theory

The remainder of this introduction correspondingly has two parts: about the dimer model, including some short remarks about statistical mechanics in 3D more generally, and about the connection between SLE curves and Teichmüller theory and how it motivates the work in the second half of this thesis.

After the introduction, each chapter corresponds to a research article. Since the background material needed for each chapter is in practice completely separate, both papers are left intact, though parts of their introductions are included in this section as well. Chapter 2 is about the dimer model in three dimensions and corresponds to the article [CSW23] which is joint work with Nishant Chandgotia and Scott Sheffield. Chapter 3 is about shear coordinates in Teichmüller theory, and other than Section 3.7 corresponds to the article [ŠWW22] which is joint work with Dragomir Šarić and Yilin Wang.

1.1 The dimer model

Given a bipartite graph $G = (V, E)$, a *dimer cover* or *perfect matching* is a collection of edges $\tau \subset E$ such that every vertex is contained in exactly one edge in τ. When G is a subgraph of $\mathbb{Z}^d$, a dimer cover can be viewed as a tiling by *domino* tiles that are length 2 in one direction and length 1 in all others, such that each d-dimensional cube in $\mathbb{Z}^d$ is covered exactly once.

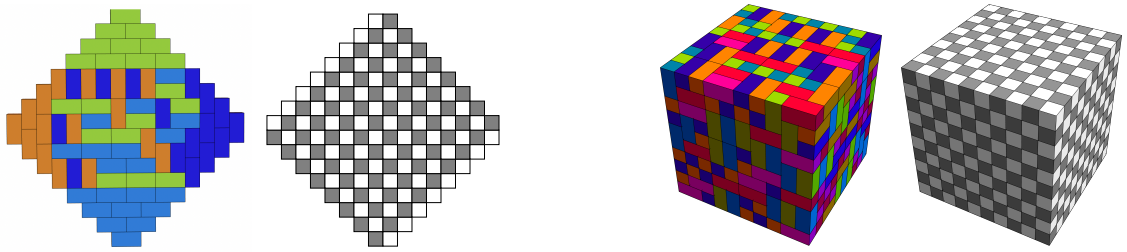


Figure 1.1: A dimer tiling of a region in $\mathbb{Z}^2$ called an *Aztec diamond* and the bipartite coloring of $\mathbb{Z}^2$ (left) and a dimer tiling of a cube and the bipartite coloring of $\mathbb{Z}^3$ (right).

For any dimension d, $\mathbb{Z}^d$ is a bipartite graph, with bipartition into *white* and *black* cubes as depicted in Figure 1.1. We view the bipartite coloring of squares in $\mathbb{Z}^d$ as completely fixed throughout. The colors of the dimer tiles in Figure 1.1 represent the cardinal direction of the tile, seen as a vector from white cube to black cube. In particular, there are $2d$ colors of tiles in a tiling of $\mathbb{Z}^d$.

There is a correspondence between a dimer tiling τ of $\mathbb{Z}^d$ and a *discrete vector field* v_τ on the oriented edges of $\mathbb{Z}^d$, where if e is oriented from white to black,

$$v_\tau(e) = \begin{cases} 1 & e \in \tau \\ 0 & e \notin \tau \end{cases} \tag{1.1}$$

If $-e$ denotes e with reversed orientation, then we define $v_\tau(-e) = -v_\tau(e)$. In this way, given a tiling τ, the colors of tiles are a graphical representation of v_τ. After subtracting a constant flow (which takes the same value on all white-to-black oriented edges, e.g. $r(e) = -1/2d$), there is a correspondence between a dimer tiling τ and a discrete vector field $f_\tau := v_\tau - r$ which is also divergence-free.

The correspondence with vector fields gives us a concrete way to compare tilings of the same region, even at different lattice scales, and we can phrase questions about random tilings τ in terms of the corresponding random divergence-free flow f_τ.

A key scaling limit question is: given a sequence of regions $R_n \subset \frac{1}{n}\mathbb{Z}^d$ approximating a region and boundary condition (R, b) in $\mathbb{R}^d$, what does the flow corresponding to a uniformly random dimer tiling of R_n look like as $n \to \infty$? The scaling limit as $n \to \infty$ of any sequence of corresponding flows f_τ is a measurable divergence-free vector field, and this question has various more precise versions:

1. Is there a *law of large numbers*, i.e. does the random flow concentrate on a particular deterministic flow as $n \to \infty$? If so, that flow is the *limit shape*.

2. Is there a *large deviation principle* (LDP), with a rate function I measuring the probability of lying close to any other limiting flow as $n \to \infty$?
3. What are the limit shapes?
4. What are the *fluctuations* around the limit shape?

These questions have been studied deeply in two dimensions, and the richness of the model is evident in the answers, which connect combinatorics, complex analysis, PDEs, probability, and algebraic geometry, see e.g. [CEP96; JPS98; CKP01; KO07; KOS06; Ken00; Ken01; Ken14; Kuc22].

The main result of Chapter 2 of this thesis is prove a large deviation principle and uniqueness of the limit shape for dimer tilings of regions in $\mathbb{R}^3$. This corresponds to answering questions 1 and 2 above. Before expanding on this more, for intuition we spend some time explaining some of the ways that tools and constructions from two dimensions do not work in three.

1.1.1 Differences between 2D and 3D

The dimer model in two dimensions has a lot of special structure which does not hold in higher dimensions. The following statements are true and special to two dimensions:

1. **Exact solvability and determinant formulas**: if G is a finite, bipartite, planar graph, then the number of dimer tilings of G can be computed as the determinant of a matrix.
2. **Local move connectedness**: we say a subgraph $G \subset \mathbb{Z}^2$ is simply connected if all cycles in G just surround one face in $\mathbb{Z}^2$. If $G \subset \mathbb{Z}^2$ is finite and simply connected, then any two dimer tilings of G are connected by a finite sequence of *flips* where the tiling is changed on a 2×2 square as in Figure 1.2.
3. **Height functions**: if $G \subset \mathbb{Z}^2$ is simply connected, there is a correspondence between dimer tilings of G and Lipschitz functions called "height functions" valued on the faces of G, up to an additive constant.

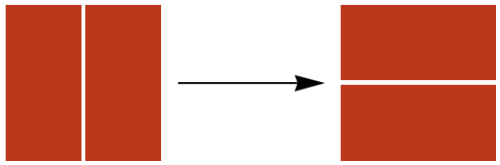


Figure 1.2: A flip or local move.

We explain why these things hold in two dimensions, and what aspects continue to hold or break in higher dimensions.

Height function from flow in 2D. Through the correspondence with divergence-free vector fields, the dimer model in general dimension can be thought of as a model of

random divergence-free 1-forms. In two dimensions only, this implies the existence of a height function. The height function correspondence is due to Thurston [Thu90], and means that dimers in two dimensions are a model of random surfaces, where the surface corresponding to a tiling τ is the graph of its height function. In two dimensions, asymptotic questions like those above are almost always studied by looking at the corresponding height function.

To recover the height function from the vector field, note that in two dimensions, a divergence-free vector field is dual (simply by rotating by 90°, i.e. $(s,t) \mapsto (-t,s)$) to a curl-free vector field. Curl-free vector fields on simply connected regions in two dimensions are always of the form ∇h for some function h. This function h is called the *scalar potential* and is the height function. In summary, there is a correspondence between a tiling τ and the height function h up to an additive constant.

In three dimensions, the nearest analog of the height function is that a divergence-free vector field f must be the curl of another vector field, i.e. $f = \nabla \times A$. In this case, A is only determined up to adding the gradient of a function. This more complicated gauge symmetry (as opposed to the gauge symmetry of adding a constant for the height function), and the fact that A is still a vector field, make it less tractable and useful for study.

Determinant formulas. The determinant formula for counting dimer tilings of planar graphs was originally discovered by Kasteleyn [Kas61] and independently by Temperley and Fisher [TF61] in the 1960s. In higher dimensions we do not know that it is *impossible* that dimer tilings are counted by a determinant—just that the methods from two dimensions do not work.

Let G be a finite bipartite graph, and let (B, W) denote its bipartition ("black" and "white"). Since every edge contains one black and one white vertex, a necessary condition for G to have a dimer tiling is $|B| = |W|$. Assuming this, let A denote the $|B| \times |W|$ adjacency matrix of G. Each entry $A_{j,k}$ corresponds to a pair of vertices $(b_j, w_k) \in B \times W$, and is 1 if $(b_j, w_k) \in E$ and 0 otherwise. In complete generality, we have the following.

Theorem 1.1.1. *For any finite bipartite graph G, the number of dimer tilings of G is the permanent of its adjacency matrix A, which is given by*

$$\operatorname{perm} A = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)}. \tag{1.2}$$

This is straightforward to see because the term for σ in the sum is 1 if $(b_1, w_{\sigma(1)}), \cdots, (b_n, w_{\sigma(n)})$ is a dimer tiling and 0 otherwise. Unfortunately, permanents are very hard to compute, so this is not so useful.

The better-known and highly computable relative of the permanent is the determinant. The determinant of a matrix M is very similar to the permanent, but has an extra sign coefficient:

$$\det M = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} M_{i,\sigma(i)}, \tag{1.3}$$

where $\operatorname{sign}(\sigma)$ is $+1$ if σ contains an even number of transpositions, and -1 otherwise.

The key insight of Kasteleyn and of Temperley Fisher is that it is possible define a *weighted version* K of the adjacency matrix A, so that $|\det K| = \operatorname{perm} A$. More precisely:

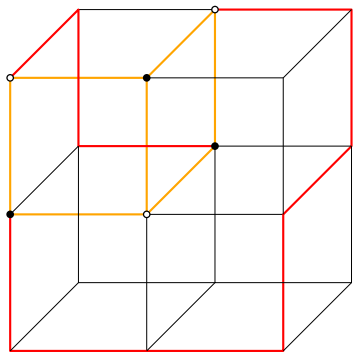


Figure 1.3: A copy of $K_{3,3}$ embedded in the $3\times3\times2$ box. Edges in red half are concatenations of edges of $\mathbb{Z}^3$, but correspond to one edge in the $K_{3,3}$.

each nonzero entry in A corresponds to an edge $e \in E$. If $A_{j,k}$ corresponds to e, then K is defined by choosing weights $\{w_e, e \in E\}$ and setting $K_{j,k} = w_e A_{j,k}$. When this works to count tilings with the determinant, these are called *Kasteleyn weights.*

Theorem 1.1.2 ([Kas61],[TF61])**.** *If G is a planar bipartite graph, there exists a choice of Kasteleyn weights so that the number of dimer tilings of G is $|\det K|$.*

If $G \subset \mathbb{Z}^2$ and simply connected, the Kasteleyn weights are quite simple. In this case, the edges of G can be split into north, south, east, and west-going edges, where the direction of an edge is its cardinal direction oriented from white to black (these categories correspond to the colors of the tiles in Figure 1.1). The weighted version K is defined so that $K_{j,k} = A_{j,k}$ if $(b_j, w_k) \in E$ and oriented north, south, or west, $K_{j,k} = -A_{j,k}$ if $(b_j, w_k) \in E$ is oriented east, and $K_{j,k} = A_{j,k} = 0$ if $(b_j, w_k) \notin E$.

To see that these weights work, given two tilings τ_1, τ_2 of G we look at the *double dimer tiling* where τ_1, τ_2 are placed on top of each other. Each vertex is incident to one tile from each tiling, so the union (τ_1, τ_2) consists of a set of edges where they agree, and a collection of finite loops $\ell_1, \ldots, \ell_m$ where the edges along the loop alternate between being in τ_1 and being in τ_2. To transform τ_2 into τ_1, we need to swap the tiles along each cycle. It suffices to show that swapping the tiles along one loop ℓ does not change the overall sign in (1.3). We say that the length of a loop ℓ is the number of edges along it from one tiling. Let γ denote the permutation which shifts the τ_2 tiles to τ_1 along a loop ℓ. This reduces to two cases to check, one where ℓ has even length (and hence $\text{sign}(\gamma) = -1$) and one where ℓ has odd length (and hence $\text{sign}(\gamma) = +1$).

Obstruction to Kasteleyn weights in $\mathbb{Z}^3$. A bipartite graph does not need to be planar for Kasteleyn weights to exist. A necessary and sufficient condition, due to C.H.C. Little [Lit75], is that the graph must not contain a $K_{3,3}$ as an embedded minor, where the vertices of the $K_{3,3}$ are vertices of the original graph, and the edges of the $K_{3,3}$ are edges or concatenations of edges of the original graph. An embedded $K_{3,3}$ is the first interpretation of a "topological obstruction" to Kasteleyn weights for subgraphs of $\mathbb{Z}^3$, as it is straightforward to see that any subgraph of $\mathbb{Z}^3$ containing a $3 \times 3 \times 2$ box contains a copy of $K_{3,3}$; see Figure 1.3.

Local moves and topological invariants. We used above that any two tilings of a graph G differ by some "cycle swaps" where we change the tiles along a finite loop. In two dimensions, if $G \subset \mathbb{Z}^2$ is a finite simply connected graph, then any two tilings differ by a finite sequence of flips as in Figure 1.2, i.e. cycle swaps length only two. See Figure 1.4 for an example sequence. Local move connectedness in 2D has an easy proof using height functions [Thu90].

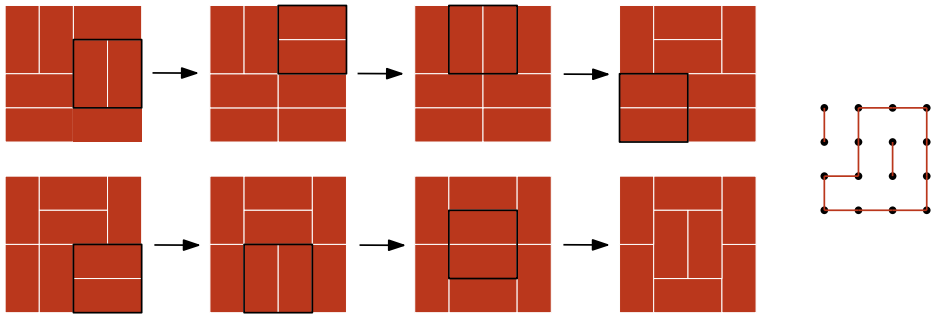


Figure 1.4: Example of a sequence of local moves.

In three dimensions, local move connectedness with flips manifestly fails. There is a simple counterexample called a *hopfion* in the $3 \times 3 \times 2$ box (term due to [FHNQ11]). There are no parallel pairs of tiles in the hopfion, so it cannot be connected under flips to other tilings of the $3 \times 3 \times 2$ box. See Figure 1.5 for a hopfion and a local move of length three called a trit [MS14a].

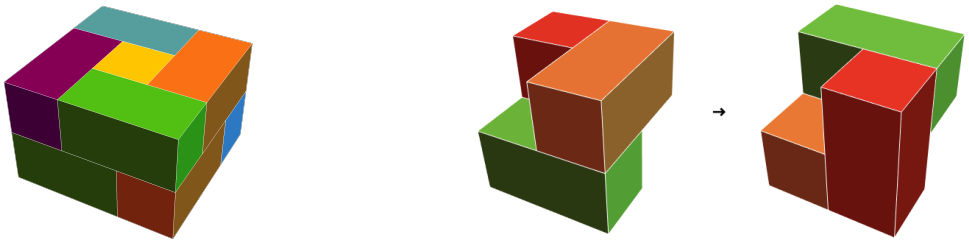


Figure 1.5: Hopfion configuration on the $3 \times 3 \times 2$ box, and the trit move.

The flip and trit together are enough to connect any two tilings of the $3 \times 3 \times 2$ box, but it is a surprisingly subtle open question whether they are enough to connect any two tilings of an $L \times M \times N$ box when $L, N, M > 2$. For an arbitrary simply connected region in $\mathbb{Z}^3$, it is a result of this thesis that there is no finite family of local moves which can connect all pairs of tilings (Corollary 2.3.7).

Local moves is one of the most active current areas of study for dimers in higher dimensions: see for example the series of works by subsets of Freire, Klivans, Milet and Saldanha [MS14a; MS14b; Mil15; MS15; FKMS22; Sal22; Sal20; KS22; Sal21], the recent work [HLT23] by Hartarsky, Lichev, and Toninelli, and physics papers by Freedman, Hastings, Nayak, Qi, and separately Bednik [FHNQ11; Bed19a; Bed19b].

The works above introduce two "topological invariants" invariant under flips (but defined in slightly different settings) to quantify the failure of local move connectedness under flips. These invariants are also intimately related to the failure of determinant formulas in three dimensions.

1. The twist, introduced in [MS14a], is defined for dimer tilings of a fixed finite region in $\mathbb{Z}^d$ and can be computed as a *linking number*. The twist is invariant under flips, and the trit move changes it by ± 1. It is valued in $\mathbb{Z}$ for tilings of $\mathbb{Z}^3$, and in $\mathbb{Z}_2$ for tilings of $\mathbb{Z}^d$ for $d > 3$ (the $d > 3$ setting was studied in [KS22]).

2. The hopf number invariant, introduced in [FHNQ11], is defined for a dimer tiling of $\mathbb{Z}^d$ which are "trivial" (i.e., all parallel tiles) outside a finite region. The hopf number is preserved by flips, and valued in $\pi_d(S^{d-1})$ for any d: in particular this is 0 when $d = 2$ (corresponding to no obstruction to connectedness under flips), $\mathbb{Z}$ when $d = 3$, and $\mathbb{Z}_2$ when $d > 2$.

3. Connection to determinant formulas [FHNQ11]: the sign of the corresponding term in a determinant computation is the hopf number mod 2. Consider $\mathbb{Z}^d$ with edge weights so that every coordinate copy of $\mathbb{Z}^2$ has its Kasteleyn weights $1, 1, -1, 1$ going around a square. If K is the weighted adjacency matrix with these weights, then $|\det K| = |A - B|$, where A is the number of tilings with even hopf number and B is the number with odd hopf number.

See Section 2.3 for a more detailed survey of works on local moves and some new results that are part of this thesis.

Another comment on topology and solvability: non-intersecting paths. In any dimension d, given a dimer tiling of $\mathbb{Z}^d$, overlaying it with a "brickwork" tiling where all the tiles point east gives a bijection between dimer tilings of $\mathbb{Z}^d$ and non-intersecting left-to-right lattice paths in $\mathbb{Z}^d$. This bijection is another way to see that counting dimer tilings has new topological complexities in three dimensions. See Figure 1.6 for an example in two dimensions.

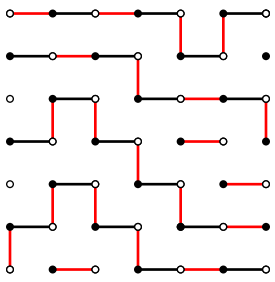


Figure 1.6: Dimer tiling (red) overlayed with brickwork tiling (black) to make non-intersecting lattice paths in two dimensions.

Non-intersecting paths in $\mathbb{Z}^2$ are fairly rigid. Since they cannot intersect, they are ordered, and given "start" and "end" points on the left and right of the square region, the connections between them are determined. The Lindström-Gessel-Viennot theorem [Lin73; GV89] gives a

way to count these non-intersecting lattice paths in $\mathbb{Z}^2$, which is therefore another equivalent way to compute the partition function for dimer tilings of $\mathbb{Z}^2$.

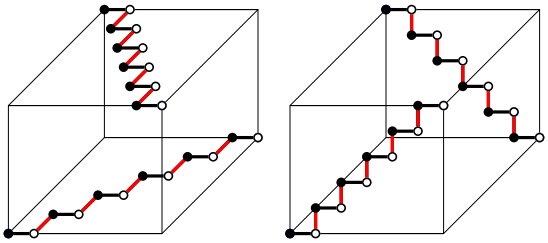


Figure 1.7: Two ways to connect the "start" and "end" points of non-intersecting paths in 3D.

Non-intersecting lattice paths in $\mathbb{Z}^3$ can exhibit much more complicated behavior and there is no known formula to count them. In three dimensions, non-intersecting paths can be braided and twisted around each other, so they are not ordered, and the "connections" between start and end point on the sides of a cube are not determined; see Figure 1.7.

1.1.2 Large deviations for the 3D dimer model

The main result of Chapter 2 is two versions of a large deviation principle for dimer tilings of regions in $\mathbb{Z}^3$, interpreted as a random divergence free vector field, with certain fixed boundary conditions. The versions of the LDP we prove differ in how we treat the boundary conditions. We further show that the large deviation rate function has a unique minimizer, implying that random tilings concentrate on a unique *limit shape* configruation in the scaling limit. See Figure 1.9 for examples of limit shapes for some simple regions which we describe more later.

These results are the three-dimensional analog of the work of Cohn, Kenyon, and Propp for two-dimensional dimers [CKP01], but the proofs use substantially different methods stemming from some of the differences mentioned in the previous section (e.g., lack of exact formulas and no height function correspondence).

A large deviation principle is a result about a sequence of probability measures $(\rho_n)_{n\geq 1}$ which quantifies the probability of rare events at an exponential scale as $n \to \infty$. More precisely, a sequence of probability measures $(\rho_n)_{n\geq 1}$ on a topological space $(X, \mathcal{B})$ is said to satisfy a large deviation principle (LDP) with rate function I and speed v_n if $I : X \to [0, \infty)$ is a lower semicontinuous function, and for all Borel measurable sets B,

$$-\inf_{x\in B^\circ} I(x) \leq \liminf_{n\to\infty} v_n^{-1} \log \rho_n(B) \leq \limsup_{n\to\infty} v_n^{-1} \log \rho_n(B) \leq -\inf_{x\in \overline{B}} I(x),$$

where $B^\circ, \overline{B}$ denote the interior and closure of B respectively. The rate function $I(x)$ is *good* if its sublevel sets $\{x : I(x) \leq a\}$ are compact. Implicit in this definition is a choice of topology $\mathcal{B}$ on X. A large deviation principle for $(\rho_n)_{n\geq 1}$ implies that random samples from ρ_n are exponentially more likely to be near the minimizers of $I(\cdot)$ as $n \to \infty$. When I is good and has a unique minimizer, this means that random samples from ρ_n *concentrate* as $n \to \infty$ in the sense that if U is any neighborhood of the unique minimizer, then as $n \to \infty$

the probability $\rho_n(X \setminus U)$ tends to zero exponentially quickly in v_n. Good references for this subject include [DZ09] and [Var16].

To formulate the large deviation principle for dimer tilings, we use the divergence-free vector field mentioned earlier:

$$f_\tau(e) = \begin{cases} +5/6 & e \in \tau \\ -1/6 & e \notin \tau. \end{cases}$$

We call this a *tiling flow*, and it will play an analogous role to the height function in two dimensions.

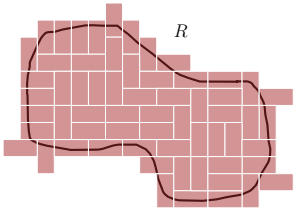


Figure 1.8: An example of a free-boundary tiling of R

Given a compact region $R \subset \mathbb{R}^3$ (which is sufficiently nice, e.g. the closure of a connected domain with piecewise smooth boundary), we define the *free-boundary tilings of R at scale n* to be tilings τ of a region $R_n \subset \frac{1}{n}\mathbb{Z}^3$ such that every point in R is contained in a tile in τ, and every tile has some intersection with R. Corresponding to one of these tilings is a *scaled* tiling flow, supported on $\frac{1}{n}\mathbb{Z}^3$ and taking values $-\frac{1}{6n^3}$ and $\frac{5}{6n^3}$. The large deviation principle we prove is about these objects as $n \to \infty$.

The natural action on dimer tilings is by even translations of $\mathbb{Z}^3$, as these preserve the underlying bipartite structure of the graph. If τ is a random tiling of $\mathbb{Z}^3$ whose law is invariant under even translations, then the *mean current per even vertex* can be defined $s = \mathbb{E}[\eta]$ where η is the vertex of $\mathbb{Z}^3$ matched to the origin by τ. Note that s lies in the set

$$\mathcal{O} = \{(s_1, s_2, s_3) \in \mathbb{R}^3 : |s_1| + |s_2| + |s_3| \leq 1\},$$

which is the convex hull of $\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\}$, and which we call the *mean-current octahedron*. This quantity measures the expected total amount of current in v_τ (or equivalently f_τ) per even vertex; see Section 2.2.2. The vertices of $\mathcal{O}$ arise for a random τ that is a.s. equal to one of the six brickwork tilings in three dimensions.

The topology we use is the *weak topology* on the space of flows obtained by interpreting each vector component of the flow as a signed measure, see Section 2.5. This topology can also be generated by the *Wasserstein distance* for flows; as such we also called this the *Wasserstein topology*. In this topology, the scaling limit of tiling flows are as expected measureable, divergence-free vector fields supported in R taking values in $\mathcal{O}$ (Theorem 2.5.19). We call these *asymptotic flows* and denote the set of them by $AF(R)$.

To state the versions of the large deviation principle, we give two ways of defining the boundary conditions in the discrete. In both cases, we fix the continuum boundary condition

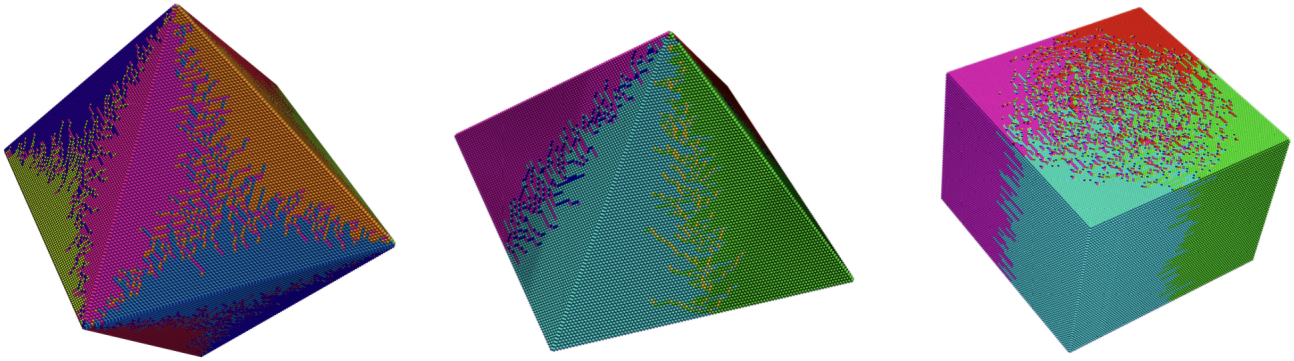


Figure 1.9: Simulations of some 3D limit shapes.

b on ∂R, which is the boundary value of some element of $AF(R)$. Like with height functions, τ_1, τ_2 are tilings of the same region $R_n \subset \mathbb{Z}^3$ if and only if their tiling flows have the same boundary values (i.e., same flow of vector field through the boundary).

- **Hard boundary (HB)**: fix a sequence of regions $R_n \subset \frac{1}{n}\mathbb{Z}^3$ with boundary values b_n approximating b and let $\overline{\rho}_n$ be uniform measure on dimer tilings of R_n.
- **Soft boundary (SB)**: choose a sequence of "thresholds" $(\theta_n)_{n\geq 0}$ with $\theta_n \to 0$ slowly enough as $n \to \infty$ and let ρ_n be uniform measure on free-boundary tilings of $R \cap \frac{1}{n}\mathbb{Z}^3$ with boundary values within θ_n of b.

The analog of the statement given in [CKP01] is the large deviation principle with hard boundary conditions as defined above. We also consider the soft boundary measures because a hard boundary large deviation principle is not true in the same generality in 3D, due to subtleties related to the fact that in 3D the function ent (which will be described below, and appears in the rate function) can be nonzero even on the boundary of $\mathcal{O}$ (see Section 2.4).

We can now give rough statements of both versions of the large deviation principle.

Theorem 1.1.3 (See Theorem 2.8.6 and Theorem 2.8.15)**.** *Suppose that $R \subset \mathbb{R}^3$ is the closure of an open set with piecewise smooth boundary and b is a nice boundary value.*

1. *(SB) Let θ_n be a sequence of thresholds with $\theta_n \to 0$ sufficiently slowly as $n \to \infty$, and let ρ_n be uniform measure on $TF_n(R)$ conditioned on the boundary values being with θ_n of b.*
2. *(HB) Assume also the the mild condition that (R, b) is "flexible" (Definition 2.8.12). Let $R_n \subset \frac{1}{n}\mathbb{Z}^3$ be a sequence of tileable regions with boundary values b_n converging to b, and let $\overline{\rho}_n$ be uniform measure on dimer tilings of R_n.*

Both $(\rho_n)_{n\geq 1}$ and $(\overline{\rho}_n)_{n\geq 1}$ satisfy a large deviation principle in the Wasserstein topology on flows with the same good rate function $I_b(\cdot)$, where

$$I_b(g) = C_b - |R|^{-1} \int_R ent(g(x))\,\mathrm{d}x \tag{1.4}$$

and ent(s) is the maximum specific entropy $h(\cdot)$ of a translation-invariant measure of mean current $s \in \mathcal{O}$ and C_b is a constant.

The large deviation principle for two dimensional dimers has a rate function of an analogous form and abstract description—i.e., it is the integral of a local entropy function $\text{ent}_2(\nabla h)$, where h is the height function and ent_2 is the specific entropy of an ergodic Gibbs measure—but the two-dimensional local entropy ent_2 turns out to also have an explicit formula [CKP01].

It is straightforward to see that $AF(R, b) \subset AF(R)$, the set of asymptotic flows with boundary condition b, is compact in the Wasserstein topology. Under the extremely mild condition that (R, b) is "semi-flexible" (this is weaker than flexible, see Definition 2.7.34), strict concavity of ent implies that $I_b(\cdot)$ has a unique maximizer in $AF(R, b)$. Combined with either large deviation principle when it holds, this implies that random dimer tilings concentrate on a unique limit shape as $n \to \infty$.

Corollary 1.1.4 (See Corollary 2.8.9 and Corollary 2.8.18)**.** *If (R, b) is semi-flexible, then I_b has a unique minimizer in $AF(R, b)$. For any $\epsilon > 0$, the probability that a uniformly random tiling flow on R at scale n (sampled from ρ_n, or sampled from $\overline{\rho}_n$ if (R, b) is flexible) differs from the rate function minimizer in $AF(R, b)$ by more than ϵ goes to 0 exponentially fast in n^3 as $n \to \infty$.*

See Figure 1.9 for some example simulations of limit shapes on regions built by stacking Aztec diamonds. All these regions satisfy the flexible condition for the hard boundary large deviation principle. These are also included in large form in the introduction to Chapter 2, and horizontal slices can be seen at https://github.com/catwolfram/3d-dimers. While the Corollary above shows that the limit shapes exist and are unique, explicit descriptions of them remains an open question.

Without exact solvability and height function tools, the methods we use to prove our main results in three dimensions are very different from those used in two dimensions. We explain a bit more about two main ideas.

Patching tilings. Dimer tilings have a *local interchangeability* property: if the flows corresponding to two dimer tilings τ_1, τ_2 of $\mathbb{Z}^3$ both approximate the same constant flow $s \in \text{Int}(\mathcal{O})$ in a certain precise sense, then we can "patch" a piece of τ_1 into τ_2 by tiling a thin annular region between them. This corresponds to saying that we can tile a thin annulus A with to agree with two different tilings τ_1, τ_2 outside its inner and outer boundaries respectively (as long as each tiling could individually be extended to cover the annulus). See Figure 1.10 for a 2D example.

This local interchangability property is completely essential to the large deviation principle and the fact that the rate function decomposes as an integral of a function depending only on local behavior. In two dimensions, via height functions, this tileability question is equivalent to a question about extending a Lipschitz function. The three-dimensional flow version cannot be phrased this way. Proving the patching theorem in three dimensions is one of the most challenging aspects of this work, and it is difficult to summarize succinctly.

The key input—applicable in any dimension—is *Hall's matching theorem* which gives a necessary and sufficient condition for the existence of a perfect matching [Hal35]. Hall's

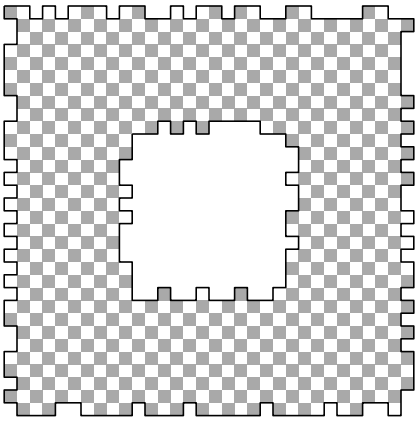


Figure 1.10: 2D example of an annular region between two tilings τ_1, τ_2.

matching theorem says that a graph $R \subset \mathbb{Z}^d$ with the same number of black and white d-dimensional cubes is tileable by d-dimensional dimers if and only if there is no *counterexample* set $U \subset R$, i.e. no set U such that

- U has more white cubes than black;
- U has only black cubes on its boundary within R, i.e.

$$\partial_0 U = \{x \in U : \exists\, y \in R \setminus U, (x, y) \text{ are connected}\}$$

  consists of only black cubes.

The *imbalance* of U is $\text{imbalance}(U) = \text{white}(U) - \text{black}(U)$, and a set U satisfying this condition on its boundary is a counterexample if and only if $\text{imbalance}(U) > 0$. A *discrete surface* is a union of two-dimensional lattice squares in $\mathbb{Z}^3$. Another key object here is the discrete surface S, which consists of the lattice squares between a cube in U and a cube in $R \setminus U$ (i.e., along $\partial_0 U$).

To prove a region is tileable with Hall's matching theorem, one has to show that no counterexamples exist. In three dimensions, if U is a counterexample in an annulus A between tilings τ_1, τ_2 that both individually extend to tilings of $\mathbb{Z}^3$, then the boundary discrete surface S is a two-dimensional surface connecting inner and outer boundaries of A. See Figure 1.11. A simple divergence theorem argument shows that $\text{imbalance}(U)$ is smaller the larger the surface area of S is. It therefore suffices to show there are no counterexamples where S has *minimal area*, subject to its constraints that it is a surface built of lattice squares, with only black cubes adjacent to it in U. Through this, the key inputs to the proof are the isoperimetric inequality and estimates using ergodic theorem, which allow us to control the existence and size of these minimal discrete surfaces. See Section 2.6 for more about these arguments.

Understanding the rate function $I_b(g)$. A key part of any large deviation principle is its rate function. For dimers, the key to understanding the rate function $I_b(g)$ is the *local entropy function* ent on $\mathcal{O}$. In particular a key step is to show that the local entropy is strictly concave on $\mathcal{O} \setminus \mathcal{E}$, where $\mathcal{E}$ is the edges of $\partial\mathcal{O}$ (ent is 0 when restricted to $\mathcal{E}$).

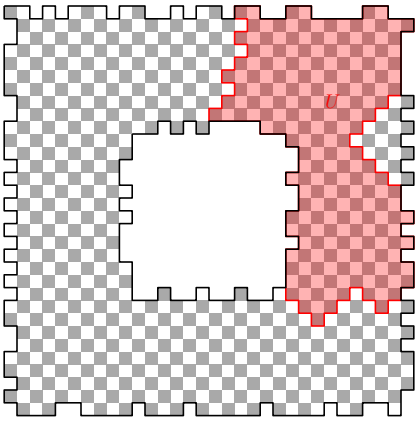


Figure 1.11: An example of a possible counterexample set $U \subset A$.

In two dimensions this is done in a very different way, since the analogous quantity ent_2 has an exact formula—derived using the Kasteleyn determinant formulas described in the previous section—which does not seem to be available in three dimensions. Instead, we have the more abstract description that $\text{ent}(s)$ is the maximum specific entropy $h(\cdot)$ of a measure of mean current s.

A pair of dimer tilings (τ_1, τ_2), when placed on top of each other, corresponds to a collection of double tiles, finite loops, and infinite paths alternating between the two tilings. Analyzing these pairs is an extremely useful way to compare tilings, or the measures they are sampled from. Instead of exact computation, the main inputs to our arguments are operations on infinite paths in pairs of dimer tilings, and the fact that entropy maximization has a qualitative description through the Gibbs property.

As a remark, we note that in two dimensions, operations on pairs of tilings leads to integrability; e.g. we used this to show that all terms in the determinant formula had the same sign, and taking one of the tilings to be fixed leads to the bijection with non-intersecting lattice paths. Perhaps the moral here is that looking at these paths is still extremely useful in higher dimensions, even though it does not seem to lead to integrability.

The connection between ent and Gibbs measures comes from a straightforward adaption of the classical variational principle of Lanford and Ruelle [LR69], from which it follows that $\text{ent}(s)$ is always realized by a Gibbs measure of mean current s (see Theorem 2.7.2). Gibbs measures are characterized by the *Gibbs property*: a measure μ on tilings is *Gibbs* if for any finite set $B \subset \mathbb{Z}^d$, μ conditioned on τ in $\mathbb{Z}^d \setminus B$ is uniform on all tilings of B that extend τ.

To prove strict concavity of ent on $\text{Int}(\mathcal{O})$, we use an operation on pairs of dimer tilings. This is a variant of the cluster swapping technique used in [She05] which we call *chain swapping*.

Given a coupling of measures $\mu = (\mu_1, \mu_2)$ on tilings, we sample a pair of tilings (τ_1, τ_2) from μ. Chain swapping constructs a pair of tilings (τ_1', τ_2') by independently "swapping" the tiles of τ_1 and τ_2 along some of the infinite paths in (τ_1, τ_2) with independent probability $1/2$ (or any probability $p \in (0, 1)$). We say that (τ_1', τ_2') is sampled from the *swapped measure* μ'. See Figure 1.12 for an example.

If μ is a coupling of measures μ_1, μ_2 on dimer tilings of mean currents s_1, s_2, then μ' is a

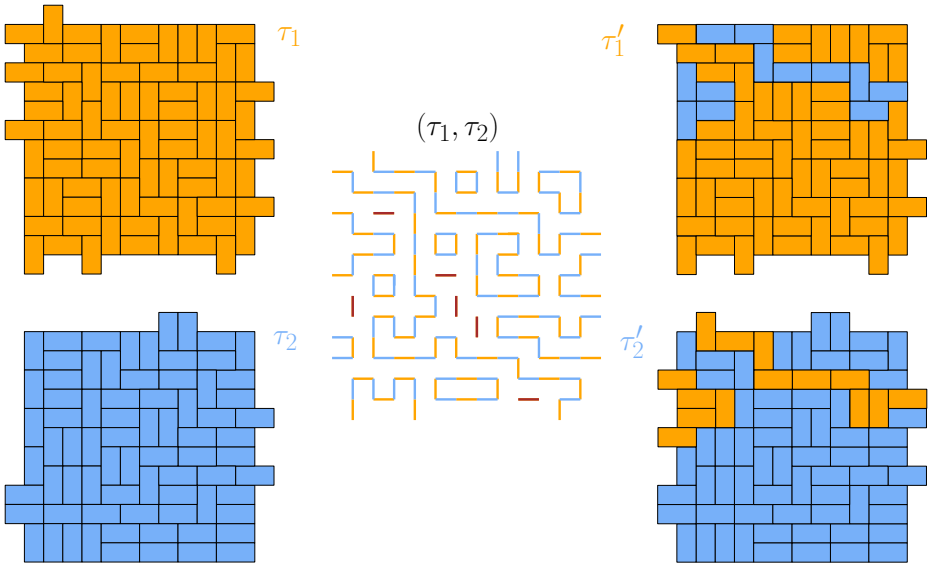


Figure 1.12: Example of chain swapping a pair of tilings.

coupling of two new measures μ_1', μ_2' on dimer tilings both of mean current $\frac{s_1+s_2}{2}$. We show that this operation preserves the total specific entropy (i.e. $h(\mu) = h(\mu')$) and ergodicity, but *breaks* the Gibbs property. More precisely, if μ_1, μ_2 are ergodic Gibbs measures of mean currents s_1, s_2 and $\frac{s_1+s_2}{2} \in \mathrm{Int}(\mathcal{O})$, then μ_1', μ_2' are ***not*** Gibbs, and hence do ***not*** have maximal entropy among measures of mean current $\frac{s_1+s_2}{2}$. The proof that the Gibbs property is broken under chain swapping requires very different techniques from those used in [She05].

Under the assumption that there exists an *ergodic* Gibbs measure μ_s of mean current s for any $s \in \mathcal{O}$, and that $\mathrm{ent}(s) = h(\mu_s)$, strict concavity would follow easily: let $\mu_1 = \mu_{s_1}, \mu_2 = \mu_{s_2}$ and apply chain swapping to get new measures μ_1', μ_2' of mean current $\frac{s_1+s_2}{2}$. Since total entropy is preserved,

$$h(\mu_1') + h(\mu_2') = h(\mu_1) + h(\mu_2) = \mathrm{ent}(s_1) + \mathrm{ent}(s_2).$$

On the other hand, since μ_1', μ_2' are not Gibbs,

$$h(\mu_1') + h(\mu_2') < 2\mathrm{ent}(\frac{s_1 + s_2}{2}),$$

which would complete the proof. A rigorous proof of the theorem is given in Section 2.7.5, and relies on casework based on ergodic decompositions as we do not know, a priori, that ergodic Gibbs measures of mean current s exist for all $s \in \mathcal{O}$. However it will then follow *from* strict concavity that this is true, and there exist ergodic Gibbs measures of all mean currents $s \in \mathcal{O}$ (Corollary 2.7.25).

1.1.3 Further questions and comparison

A lot is known about the behavior of scaling limits of dimer tilings in two dimensions, and these stories connect a lot of different areas of math. To illustrate the richness of the model—and to suggest some direction for further study in three dimensions—we say a bit more about this here.

Euler–Lagrange equations. Limit shapes can be explicitly described is via an Euler–Lagrange equation which gives the formula for minimizing the rate function subject to boundary conditions. In two dimensions, since the local entropy is a function of the height function gradient $\nabla h = (s, t)$, the Euler–Lagrange equations have the form:

$$\operatorname{div}(\nabla \operatorname{ent}_2(\nabla h)) = \frac{\partial}{\partial x}\frac{\partial \operatorname{ent}_2}{\partial s}(\nabla h) + \frac{\partial}{\partial y}\frac{\partial \operatorname{ent}_2}{\partial t}(\nabla h) = 0. \tag{1.5}$$

The explicit formula for ent_2 means that these PDEs can be written explicitly and solved [KO07]. See also [KP22b; KP24].

In three dimensions we do not know that the local entropy function ent is differentiable. Assuming it is, without the height function correspondence, the three-dimensional Euler–Lagrange equations still have a slightly different form. In fact there would be a system of Euler Lagrange equations. The aim would be to solve:

$$\max_f \int_R \operatorname{ent}(f(x,y,z))\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$$
$$\text{subject to the constraint that } \operatorname{div}(f) = 0 \text{ in } R.$$

In $\mathbb{R}^3$, $\operatorname{div}(f) = 0$ is equivalent to saying that $f = \nabla \times A$ for another vector field A called its *vector potential*. We can encode the divergence free condition in the minimization problem using A: we want to find $f = \nabla \times A$ satisfying

$$\min_{f=\nabla\times A} \int_R \operatorname{ent}(\nabla \times A)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{1.6}$$

A standard calculus of variations argument then shows that $(f_1, f_2, f_3) = f = \nabla \times A$ satisfies (1.6) if it satisfies the equations:

$$\frac{\partial}{\partial z}\frac{\partial \operatorname{ent}}{\partial f_2} - \frac{\partial}{\partial y}\frac{\partial \operatorname{ent}}{\partial f_3} = 0 \qquad \frac{\partial}{\partial x}\frac{\partial \operatorname{ent}}{\partial f_3} - \frac{\partial}{\partial z}\frac{\partial \operatorname{ent}}{\partial f_1} = 0 \qquad \frac{\partial}{\partial y}\frac{\partial \operatorname{ent}}{\partial f_1} - \frac{\partial}{\partial x}\frac{\partial \operatorname{ent}}{\partial f_2} = 0.$$

Interfaces in limit shapes. In two dimensions the interfaces between "frozen" regions (i.e., regions where only one color of dimer occurs) and "liquid" regions (where all colors appear) in limit shapes turn out to be algebraic curves. The first example of this was the *arctic circle theorem* for the Aztec diamond [JPS98], which says that the interface in this case is exactly a circle. This was later generalized to other regions in [KOS06].

The three dimensional limit shapes simulated in Figure 1.1 appear to have similar structure, but the arguments in 2D use solvability of the model and do not extend to 3D. It is still an open question even to show rigorously that any 3D limit shape has a frozen region. Assuming that frozen regions in limit shapes exist, there are interfaces that require a description: what is the law of the interfaces between two frozen regions, and what is the law of the interface between frozen and liquid regions?

The only limit shapes we can explicitly describe in 3D are extremely simple, corresponding to constant boundary conditions. The simplest one is: the limit shape for the cube with zero boundary conditions is the flow which is 0 everywhere. To see this, take the discrete regions

R_n to be just $n \times n \times n$ cube as $n \to \infty$; the zero boundary condition means no tiles are allowed to stick out of the region. Since ent : $\mathcal{O} \to \mathbb{R}$ is strictly concave, its maximum value is at 0, so the zero flow is the overall maximum-entropy flow.

Conformal invariance, Gaussian free field, and other connections. In two dimensions, scaling limits of random dimer tilings and double-dimer loops (alternating loops when two dimer tilings are stacked on each other as discussed above) exhibit conformal invariance [Ken00; Ken14]. The first of these references was one of the earliest rigorous mathematical results where a statistical mechanics model was shown to be conformally invariant.

Further, the fluctuations are shown to be the Gaussian free field [Ken01]. Boundary conditions add new subtly. The fluctuations in the liquid region are still generally believed to be the Gaussian free field in regions with boundary conditions [KO07], but with respect to a non-standard conformal structure. There are more examples, but see e.g. [BNR23; BK18; Rus18].

Even in two dimensions, convergence to the Gaussian free field is generally proved by showing that the gradient of the height function converges to the gradient of the GFF. This 1-form version of the GFF (i.e., ∇h, or its divergence-free dual) makes sense in any dimension. It is natural to conjecture that fluctuations for the three-dimensional dimer model would be related to a divergence-free 1-form version of the GFF (i.e., Gaussian white noise on edges of the graph, projected to be divergence-free), but substantial new tools and ideas would be needed to prove something like this.

It is not completely clear what analogs of conformal symmetries should be expected—and what structure they may impose—in higher dimensions. The conformal group in two dimensions is uniquely large, while conformal transformations in dimensions $d > 2$ are only Möbius transformations.

Dimers in two dimensions also enjoy connections to other central objects of random conformal geometry, e.g. the bijection between dimer tilings of Temperleyan domains and spanning trees of the same graph [KS04], the conjectural relationship between CLE_4 loops and scaling limits of level-lines of the height functions (which would follow if the convergence of height function fluctuations to the GFF were proved in a strong enough topology to imply convergence of level lines), or the bijection with Ising configurations on another planar graph (which is the original reason the dimer model was studied in the 1960s).

Other work in three dimensional statistical mechanics. There are a number of other recent works on different aspects of statistical mechanics in three dimensions. This list is undoubtedly incomplete, but we list and comment on some of these briefly here.

Quitmann and Taggi have some additional important work on the 3D dimer model, which studies the behavior of loops formed by an independently sampled pair of dimer configurations [Tag22; QT22; QT23]. Among other things, they find that when one superimposes two independent random dimer tilings on an $n \times n \times n$ torus, the union of the tilings will typically contain cycles whose length has order n^3. See also works on a generalization of rhombus tilings to higher dimensions, where there is a height function correspondence [LMN01; Lam21; WMDB02].

Another interesting group of results concern interfaces and curves. In two dimensions, loop erased random walks, percolation interfaces at criticality, and spanning trees all have

conformally-invariant scaling limits described by SLE curves. Various three-dimensional analogs have been studied, e.g. the interface in tricolor percolation on the 3D body-centered cubic lattice, which is shown to have nontrivial phase diagram depending on the color probabilities [SY14]; the scaling limit of the 3D loop-erased random walk exists and is invariant under rotations and dilations [Koz07], see also [Wil10; LS22; Shi18; Shi19]; the subsequential scaling limits of the uniform spanning tree in three dimensions exist, see e.g. [ACHS21a; ACHS21b].

Loosely speaking, three dimensional models seem to be less integrable than two dimensional ones, with fewer quantities that seem amenable to explicit computation at least with current methods. One notable exception to this seems to be certain "large N limits" of models. The most classical example of this would be mean-field expansion, others would be the $N \to \infty$ limit of $U(N)$ lattice gauge theory, where the partition function reduces to an expression in terms of Catalan numbers, or the $N \to \infty$ "multinomial dimer model" which corresponds certain measures on stacks of N dimer tilings, see [KP22a]. The branched polymer model [KW09] is another example of a model with some integrability in three dimensions.

Another extremely successful direction are results in three-dimensional statistical mechanics inspired by techniques from percolation and the random current representation, for example leading to the first rigorous proof of the continuity of the phase transition in the 3D Ising model [ADS15]. See e.g. the survey article [Dum17].

1.2 Random conformal geometry and Teichmüller theory

Here we start by explaining the connection between random conformal geometry and Teichmüller theory in two dimensions through the large deviations of Schramm-Loewner evolution (SLE) curves. This gives some of the probabilistic history which motivates the work in Chapter 3.

1.2.1 Large deviations of SLE

Given a region D and two points $a, b \in \partial D$, we say that γ is a chord in (D, a, b) if γ is a non-crossing curve contained in D from a to b. For $\kappa \geq 0$, SLE_κ is a random chord in (D, a, b), with randomness parametrized by κ. SLE_κ is conformally invariant in law, meaning that if γ has the law of an SLE_κ in (D, a, b) and $\varphi : D \to D'$ is a conformal map, then $\varphi(\gamma)$ has the law of SLE_κ defined in $(D', \varphi(a), \varphi(b))$. The connection with Teichmüller theory comes from the large deviations of SLE_κ as $\kappa \to 0^+$. We give brief and informal description here, see e.g. the survey article [Wan22] for a more detailed discussion.

Definition and history of SLE. SLE curves were discovered by O. Schramm in the late 1990s as the conjectural scaling limit of interfaces in conformally-invariant critical statistical mechanics models [Sch00]. A number of these scaling limit results have now been proved, see e.g. [LSW04; Sch07; SS09; Smi06; Smi10]. The discovery of SLE curves helped launch the field of random conformal geometry, and these curves and their relationship to random surfaces and conformal field theory is a very active area of research in probability.

The definition of SLE starts with the Loewner transform, which was introduced in the 1920s by K. Loewner [Loe23] and is completely deterministic. Suppose that γ is a parametrized simple chord in $(\mathbb{H}, 0, \infty)$. For any time $t > 0$, the slit region $\mathbb{H} \setminus \gamma[0,t]$ is still simply connected, so by the Riemann mapping theorem there exists a conformal map $g_t : \mathbb{H} \setminus \gamma[0,t] \to \mathbb{H}$. This map and the parametrization of γ can be uniquely fixed by requiring the expansion at ∞ to have the form

$$g_t(z) = z + \frac{2t}{z} + o\left(\frac{1}{z^2}\right) \qquad \text{as} \qquad z \to \infty. \tag{1.7}$$

With the conformal map and parametrization fixed, the *driving function* $W : \mathbb{R}_+ \to \mathbb{R}$ is defined to be $W_t := g_t(\gamma_t)$, the image of the tip of the $\gamma[0,t]$. See Figure 1.13.

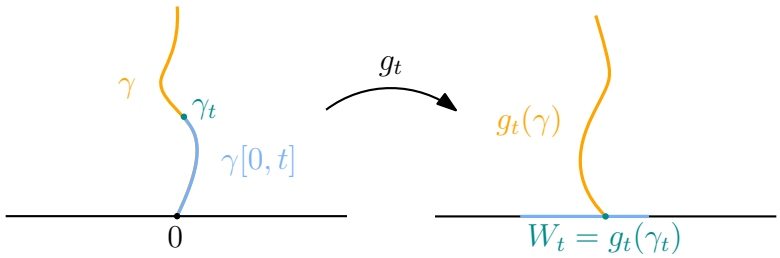


Figure 1.13: Loewner transform of a simple curve

This in fact works not just for simple curves, but for "growing hulls" $(K_t)_{t\geq 0}$ which satisfy a local growth property. Under these suitable regularity assumptions, in the more general setting the relationship between W_t and g_t can be summarized with a differential equation:

$$\partial_t g_t(z) = \frac{2}{g_t(z) - W_t}, \qquad g_0(z) = z \tag{1.8}$$

The sequence of growing hulls or curve can be recovered from the driving function W. The driving function satisfies two transformation rules, which are straightforward to check using (1.7):

- (Scaling) Fix $\lambda > 0$. If W_t is the driving function of γ, then the driving function of $\tilde{\gamma} = \lambda\gamma$ is $\tilde{W}_t = \lambda W_{\lambda^{-2}t}$.
- (Additivity) Fix $s, t > 0$. Let $\tilde{W}_t$ denote the driving function of the curve $g_s(\gamma[s, s+t]) - W_s$. Then $\tilde{W}_t = W_{s+t} - W_s$. (This corresponds to mapping out $\gamma[0,s]$, then the image of $\gamma[s, s+t]$, versus mapping out $\gamma[0, s+t]$ all at once.)

Through these transformation properties, the only driving functions that could encode a scaling limit of a conformally-invariant interface (i.e., conformally-invariant curve which cannot cross itself) are of the form $\sqrt{\kappa}B_t$, where B_t is Brownian motion.

For any $\kappa \geq 0$, chordal SLE_κ is defined to the Loewner transform of the driving function $W_t = \sqrt{\kappa}B_t$ [Sch00]. When $\kappa \in [0,4]$, SLE_κ is a simple curve. For $\kappa > 4$ it is a sequence of "growing hulls" which are traced by a curve [RS05; LSW04].

Large deviations. SLE_0 is deterministic: it is the Loewner transform of the driving function $W_t \equiv 0$. From this, it is straightforward to see that SLE_0 in $(\mathbb{H}, 0, \infty)$ is the imaginary axis $\gamma = \{\mathrm{i}t : t \geq 0\}$. By conformal invariance, SLE_0 in (D, a, b) is the hyperbolic geodesic from a to b in the unique hyperbolic metric on D.

As $\kappa \to 0^+$, SLE_κ in (D, a, b) will concentrate around the hyperbolic geodesic. Large deviations is about quantifying the exponentially small probability that SLE_κ lies close to any other chord as $\kappa \to 0^+$.

The *Dirichlet energy* of a function $W : \mathbb{R}_+ \to \mathbb{R}$ is

$$\frac{1}{2}\int_0^\infty |\dot{W}_t|^2 \, \mathrm{d}t. \tag{1.9}$$

It turns out that if $W : \mathbb{R}_+ \to \mathbb{R}$ has finite Dirichlet energy, then its Loewner transform γ is a simple chord in $(\mathbb{H}, 0, \infty)$. If W is absolutely continuous, then the *Loewner energy* of γ is defined to be the Dirichlet energy of its driving function W:

$$I_{(\mathbb{H},0,\infty)}(\gamma) = \frac{1}{2}\int_0^\infty |\dot{W}_t|^2 \, \mathrm{d}t. \tag{1.10}$$

If W is not absolutely continuous then $I_{(\mathbb{H},0,\infty)}(\gamma) = \infty$. More generally, if γ is a chord in (D, a, b) and $\varphi : D \to \mathbb{H}$ is a conformal map with $\varphi(a) = 0$ and $\varphi(b) = \infty$, then

$$I_{(D,a,b)}(\gamma) = I_{(\mathbb{H},0,\infty)}(\varphi(\gamma)). \tag{1.11}$$

This definition is independent of the conformal map φ, since φ is determined up to scaling, and scaling by a constant $\lambda > 0$ changes the driving function by $W_t \mapsto \lambda W_{\lambda^{-2}t}$, which does not change its Dirichlet energy. In various different topologies (including e.g. the Hausdorff topology [PW21]), SLE_κ satisfies a large deviation principle with the Loewner energy as its rate function:

Theorem 1.2.1 ([PW21; Wan19b])**.** SLE_κ *in* $(\mathbb{H}, 0, \infty)$ *satisfies a large deviation principle as* $\kappa \to 0^+$ *with good rate function given by the Loewner energy* $I_{\mathbb{H},0,\infty}(\gamma)$.

This result resembles and is inspired by *Schilder's theorem*, which says that scaled Brownian motion $\sqrt{\epsilon}B_t$, interpreted as random function, satisfies a large deviation principle as $\epsilon \to 0$ where the rate function is the Dirichlet energy above. The Loewner transform $W \mapsto \gamma$ is not continuous, so the proof of the large deviations of SLE does not use Schilder's theorem.

From chords to loops. By its definition, the Loewner energy is clearly Möbius invariant. On the other hand the Loewner transform $W \mapsto \gamma$ is not continuous, and it is not a priori obvious what curves with finite Loewner energy look like.

Example 1.2.2. A simple family of examples are the slanted lines $\gamma_\alpha = \{t\exp(\mathrm{i}\alpha) : t \in \mathbb{R}_+\}$ for $\alpha \in (0, \pi/2]$. When $\alpha = \pi/2$, this is just the imaginary axis, and the driving function is $W_{\pi/2}(t) \equiv 0$. For any $\alpha \in (0, \pi/2)$, the scaling property of driving functions implies that $W_\alpha(t) = C(\alpha)\sqrt{t}$, where $C(\alpha) \neq 0$. As such, $I_{(\mathbb{H},0,\infty)}(\gamma_\alpha) = \infty$ for all $\alpha \in (0, \pi/2)$.

It turns out that the Loewner energy is naturally a function of a loop instead of just a chord. Let $\hat{\mathbb{C}}$ denote the Riemann sphere, and consider a loop $\gamma : [0,1] \to \hat{\mathbb{C}}$, $\gamma(0) = \gamma(1)$. The Loewner energy of γ with respect to the root point $p = \gamma(0) = \gamma(1)$ is defined by slitting it from the root to get a chord and then taking a limit:

$$I^L(\gamma, p) := \lim_{\varepsilon \to 0} I_{\hat{\mathbb{C}} \setminus \gamma[0,\varepsilon], \gamma(\varepsilon), \gamma(1)}(\gamma[\varepsilon, 1]). \tag{1.12}$$

A chord η in $(\mathbb{H}, 0, \infty)$ can be completed to a loop in the Riemann sphere by sending $\mathbb{H} \to \mathbb{C} \setminus \mathbb{R}_+$ by the map $z \mapsto z^2$. We then complete the the chord η to a loop $\gamma := \eta \cup \mathbb{R}_+$. Applying this technique to the chords in Example 1.2.2, the vertical line (driving function $W_t \equiv 0$) corresponds to a circle, while the slanted line γ_α at angle $0 < \alpha < \pi/2$ corresponds to two semicircles meeting at corner with angle $2\alpha < \pi$.

Surprisingly, the Loewner energy of a loop turns out not to depend on the root $p \in \gamma$.

Theorem 1.2.3 ([RW19]). *$I^L(\gamma, p)$ is independent of of the root point p.*

Given this, the Loewner loop energy can be written as just $I^L(\gamma)$, without reference to a root point.

Finite Loewner energy loops and Teichmüller theory. Given $K \geq 0$, a map $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is K-quasiconformal if it takes infinitesimal circles to infinitesimal ellipses of eccentricity between $1/K$ and K. More precisely, these solve the *Beltrami equation*

$$\frac{\partial f}{\partial \bar{z}} = \mu(z) \frac{\partial f}{\partial z}, \tag{1.13}$$

for some *Beltrami coefficient* μ with L^∞ norm

$$|\mu|_\infty \leq \frac{K-1}{K+1} < 1. \tag{1.14}$$

These are a generalization of conformal maps. In this framework, conformal maps correspond to $K = 1, \mu(z) \equiv 0$. A quasicircle is the image of S^1 under a quasiconformal map. These are a subset of Jordan curves, and include for example all smooth simple loops but also some fractal curves.

The *universal Teichmüller space* $T(\mathbb{D})$ is the parameter space of quasicircles up to Möbius transformations. It is called "universal" because it contains the Teichmüller spaces of all compact Riemann surfaces.

Root invariance plus Möbius invariance imply that the Loewner loop energy I^L is a functional on $T(\mathbb{D})$. It turns out to be equal to the *universal Liouville action* $\mathbf{S}_1$ of Takhtajan and Teo [TT06], which they show is equal to the Kähler potential of Weil–Petersson Teichmüller space $T_0(\mathbb{D}) \subset T(\mathbb{D})$. In particular, $\mathbf{S}_1([\gamma]) < \infty$ if and only if γ is a Weil–Petersson quasicircle.

More precisely, let $\mathbb{D}, \mathbb{D}^*$ denote the unit disk and its complement. Assume γ is a loop not containing ∞, and let Γ, Γ^* denote the region bounded by γ containing 0 and its complement respectively. By the Riemann mapping theorem, there exist conformal maps $f : \mathbb{D} \to \Gamma$ and $g : \mathbb{D} \to \Gamma^*$ with $g(\infty) = \infty$.

Theorem 1.2.4 ([Wan19a]). *Up to a multiplicative constant, the Loewner loop energy is equal to the Liouville action of Weil-Petersson Teichmüller space:*

$$I^L(\gamma) = \frac{1}{\pi}\mathbf{S}_1(\gamma) \tag{1.15}$$

$$= \iint_{\mathbb{D}} \left|\frac{f''(z)}{f'(z)}\right|^2 \mathrm{d}A(z) + \iint_{\mathbb{D}^*} \left|\frac{g''(z)}{g'(z)}\right|^2 \mathrm{d}A(z) + 4\pi \log\frac{|f'(0)|}{|g'(\infty)|}. \tag{1.16}$$

In particular, $I^L(\gamma) < \infty$ if and only if γ is a Weil–Petersson quasicircle.

1.2.2 Weil–Petersson Teichmüller space

The Weil-Petersson Teichmüller space is a subspace of $T(\mathbb{D})$ with rich geometric structure and over twenty equivalent definitions [Bis19]. The Weil–Petersson class corresponds to the completion of the smooth curves in $T(\mathbb{D})$, under the unique maximal so-called Kähler structure (meaning compatible Riemannian metric, symplectic form, and complex structure).

This class has been studied deeply within Teichmüller theory [Cui00; Guo00; TT06; She18; ST20; STW18], and in addition to the recent link with SLE [Wan19a; Wan21; VW20a; VW20b; Wan22], it also has connections with string theory [BR87; Wit88; NV90; Pek95], hyperbolic geometry [Bis19], and Coulomb gases [Joh21; WZ22].

We explain some equivalent definitions and examples here. The variety of perspectives make this an interesting object of study, but it is often not clear how certain properties or structure manifest in different definitions.

Geometric description, examples, and remarks. Quasicircles are always Jordan curves, but include a wide range of behaviors. All smooth simple loops are quasicircles, and so are some fractals. One characterization of their regularity, due to Ahlfors, is the *bounded turning property*. Namely, γ is a quasicircle if there exists a constant $C > 0$ such that for any two points $x, y \in \gamma$ and any third point z chosen on the shorter arc of γ between x and y,

$$|x - z| + |y - z| \leq C|x - y|. \tag{1.17}$$

Weil–Petersson quasicircles have stronger regularity properties than this. One necessary (but not sufficient) condition in the same spirit as above is that they are *asymptotically smooth*. I.e., if $\gamma \subset \mathbb{C}$ is Weil–Petersson, then the shorter arc $\gamma_{x,y}$ between any two points $x, y \in \gamma$ must satisfy

$$\lim_{|x-y|\to 0} \mathrm{length}(\gamma_{x,y})/|x - y| = 1. \tag{1.18}$$

Analytically, all smooth curves are Weil–Petersson, and fractal quasicircles are never Weil–Petersson. WP quasicircles can have points where they are not differentiable, but the set of non-differentiable points must have measure zero, and the types of non-differentiable singularities allowed are very particular. One can see from the asymptotically smooth condition, for example, that any γ with a corner—i.e., a point $p \in \gamma$ where left and right derivatives exist but are not equal—is not WP (compare this to the loops corresponding to the slanted lines in Example 1.2.2: when $\alpha \neq \pi/2$, the corresponding loop has a corner and hence is not WP). Here are so more examples:

WP quasicircles	non-WP quasicircles
the circle	the square
smooth simple loops	quasicircles with corners
log log spirals	any faster spiral
C^1 piecewise geodesic	fractal quasicircles
C^1 piecewise circular	

The last two examples in the left column require definition. A curve is C^1 piecewise circular if it is C^1 and the concatenation of circular arcs. A curve γ is C^1 piecewise geodesic if it is C^1 and can be split into segments $\gamma_1, \dots, \gamma_k$ such that γ_i is a hyperbolic geodesic in $\hat{\mathbb{C}} \setminus (\gamma \setminus \gamma_i)$ for all $i = 1, \dots, k$.

C^1 piecewise geodesic curves minimize Loewner energy among curves that go through a certain set of points, and have been studied e.g. in [MRW22; BJM+23]. Further, some intriguing holographic duality between the piecewise circular and piecewise geodesic classes is described in [Wan24]. A subset of C^1 piecewise geodesic curves are also centrally studied in Chapter 3 of this thesis, since homeomorphisms with "finitely supported diamond shear coordinates" exactly correspond to C^1 piecewise gedoesic loops under the conformal welding process described below.

Remark 1.2.5. Finally, we can remark on the confusing relationship between the universal Weil–Petersson Teichmüller space $T_0(\mathbb{D})$ (which we study in this thesis) and the Weil–Petersson Teichmüller space $T_0(\Sigma)$ for a compact Riemann surface Σ. These are both contained in $T(\mathbb{D})$, but are *disjoint*.

We will see later that $T_0(\mathbb{D})$ corresponds to maps which solve a Beltrami equation where the Beltrami coefficient on the disk is are in $L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$, where $\mathrm{d}_{\mathrm{hyp}}$ is the hyperbolic metric on the disk. Similarly, $T_0(\Sigma)$ can be defined as corresponding to maps which solve a Beltrami equation where Beltrami coefficient is invariant under the action of the corresponding Fuchsian group on the disk, and is in L^2 with respect to $\mathrm{d}_{\mathrm{hyp}}$ on a fundamental domain for the action of the group. However, invariance under the group action automatically means that the integral over the whole disk as in the definition for $T_0(\mathbb{D})$ is infinite.

Conformal welding. Let $\mathbb{T}$ denote the unit circle, and let $\mathrm{Möb} \simeq \mathrm{PSU}(1,1)$ be the Möbius transformations preserving the disk. Via conformal welding, the universal Teichmüller space can be seen as a subspace of $\mathrm{Homeo}(\mathbb{T})/\mathrm{Möb}$, the space of orientation-preserving circle homeomorphisms up to Möbius transformations.

Suppose that $\gamma \subset \hat{\mathbb{C}}$ is a quasicircle, and let Γ, Γ^* denote the two components of $\hat{\mathbb{C}}$. By the Riemann mapping theorem there exist conformal maps $f : \mathbb{D} \to \Gamma$ and $g : \Gamma^* \to \mathbb{D}^*$. By the Caratheodory extension theorem, these both extend to homeomorphisms $\mathbb{T} \to \gamma$. We therefore define the *welding homeomorphism* of f and g by

$$h := g^{-1} \circ f \mid_{\mathbb{T}} . \tag{1.19}$$

See Figure 1.14. Given a circle homeomorphism h, we call a curve γ which has h as its welding a "solution to the welding problem" for h. If h is *quasisymmetric*, a solution γ to the welding problem exists and is unique up to Möbius transformations. There is no simple formula to compute the welded curve γ from the circle homeomorphism, and the general

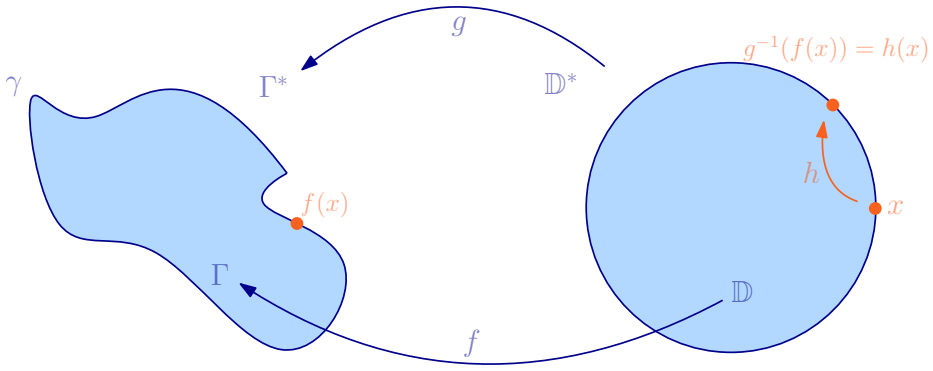


Figure 1.14: Conformal welding example.

necessary condition for a circle homeomorphism to have a unique solution to the welding problem is not known.

The universal Teichmüller space $T(\mathbb{D})$ is identified with the class of quasisymmetric circle homeomorphisms under conformal welding:

$$T(\mathbb{D}) \simeq \mathrm{QS}(\mathbb{T})/\mathrm{M\ddot{o}b}. \tag{1.20}$$

Quasisymmetric circle homeomorphisms correspond to the boundary values of quasiconformal maps preserving the disk. They can also be described directly: $h : \mathbb{T} \to \mathbb{T}$ is K-quasisymmetric if given any three points $a, b, c \in \mathbb{T}$ in cyclic order with $|a - b| = |b - c|$,

$$1/K < \frac{|h(a) - |h(b)|}{|h(b) - h(c)|} < K.$$

A map is quasisymmetric if it is K-quasisymmetric for some $K > 0$. As a subspace, the Weil–Petersson class of quasicircles is identified with a class of circle homeomorhpisms.

$$T_0(\mathbb{D}) \simeq \mathrm{WP}(\mathbb{T})/\mathrm{M\ddot{o}b}. \tag{1.21}$$

Y. Shen showed that $\mathrm{WP}(\mathbb{T})$ has a simple analytic description.

Theorem 1.2.6 ([She18]). *A circle homeomorphism $h \in \mathrm{WP}(\mathbb{T})$ if and only if h is absolutely continuous (with respect to arclength measure) and $\log h'$ belongs to the Sobolev space $H^{1/2}$. In other words,*

$$\iint_{\mathbb{T}\times\mathbb{T}} \left| \frac{\log h'(x) - \log h'(y)}{x - y} \right|^2 \mathrm{d}x\mathrm{d}y < \infty. \tag{1.22}$$

Beltrami coefficients. In this definition of Weil–Petersson, the metric, symplectic form, and L^2 structure are simple to describe. We mostly follow the presentation of [TT06], though we use $\mathbb{D}$ instead of $\mathbb{D}^*$. Let $L^\infty(\mathbb{D})_1 = \{\mu \in L^\infty(\mathbb{D}) : |\mu|_\infty < 1\}$. Given a Beltrami coefficient $\mu \in L^\infty(\mathbb{D})_1$, we can extend it to $L^\infty(\hat{\mathbb{C}})_1$ by reflection:

$$\mu(z) := \overline{\mu(1/\bar{z})}\, \frac{z^2}{\bar{z}^2} \qquad \text{for } z \in \mathbb{D}^*. \tag{1.23}$$

The reflection symmetry implies that a quasiconformal map $w : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ solving the corresponding Beltrami equation fixes $\mathbb{T} = \partial\mathbb{D}$. Choosing a normalization, there is a unique solution to this equation which fixes $1, \mathbf{i}, -1$, which we denote w_μ. Then

$$T(\mathbb{D}) = L^\infty(\mathbb{D})_1/\sim, \tag{1.24}$$

where $\mu \sim \nu$ if $w_\mu \mid_{\mathbb{T}} = w_\nu \mid_{\mathbb{T}}$, i.e. $\mu \sim \nu$ if these are both Beltrami coefficients for quasiconformal extensions of the same circle homeomorphism.

Weil–Petersson Teichmüller space is the L^2 subspace of this with respect to the hyperbolic metric $\mathrm{d}_{\mathrm{hyp}}$ on the disk $\mathbb{D}$.

Definition 1.2.7. The Weil–Petersson Teichmüller space $T_0(\mathbb{D}) \subset T(\mathbb{D})$ is the set of quasiconformal maps f (up to equivalence) such that f solves a Beltrami equation with coefficient $\mu \in L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$, i.e. for which

$$\iint_{\mathbb{D}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}z^2 < \infty. \tag{1.25}$$

This definition is slightly subtle, since it is not invariant under the equivalence relation $\mu \equiv \nu$ on Beltrami coefficients, corresponding to the fact that a circle homeomorphism can have multiple quasiconformal extensions to the disk. G. Cui showed that the Douady-Earle extension of a Weil–Petersson circle homeomorphism always has Beltrami coefficient $\mu \in L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$ [Cui00].

The infinitesimal theory of the Riemannian metric $\langle \cdot, \cdot \rangle_{\mathrm{WP}}$ and symplectic form ω on the tangent space of $\mathrm{WP}(\mathbb{T})$ can be described with the harmonic infinitesimal Beltrami coefficients $\dot{\mu}, \dot{\nu}$.

$$\langle \dot{\mu}, \dot{\nu} \rangle_{\mathrm{WP}} + \mathbf{i}\,\omega(\dot{\mu}, \dot{\nu}) = \iint_{\mathbb{D}} \frac{\dot{\mu}(z)\overline{\dot{\nu}(z)}}{(1-|z|^2)^2}\,\mathrm{d}z^2. \tag{1.26}$$

Open question. Brownian motion has many different definitions. In addition to being the scaling limit of simple random walks, one of its original definitions—due to Wiener in the 1920s—is that Brownian motion e.g. on $[0,1]$ can written as a random Fourier series, where the coefficients X_k are i.i.d. Gaussians with mean 0 and variance 1:

$$B_t = X_0 t + \sqrt{2}\sum_{k=1}^{\infty} \frac{\sin(2\pi k t)}{\pi k} X_k. \tag{1.27}$$

By analogy, in addition to the construction of SLE as a scaling limit, it would be natural to have a construction of SLE using Teichmüller theory. For example, is there a natural L^2 basis for $\mathrm{WP}(\mathbb{T})$, which when randomized constructs SLE?

While the L^2 structure for example is evident in the Beltrami coefficient description, these definitions are related in transcendental ways, e.g. through solving Beltrami equations or the welding problem, and it is often not obvious how properties manifest in different descriptions, i.e. for curves or for circle homeomorphisms.

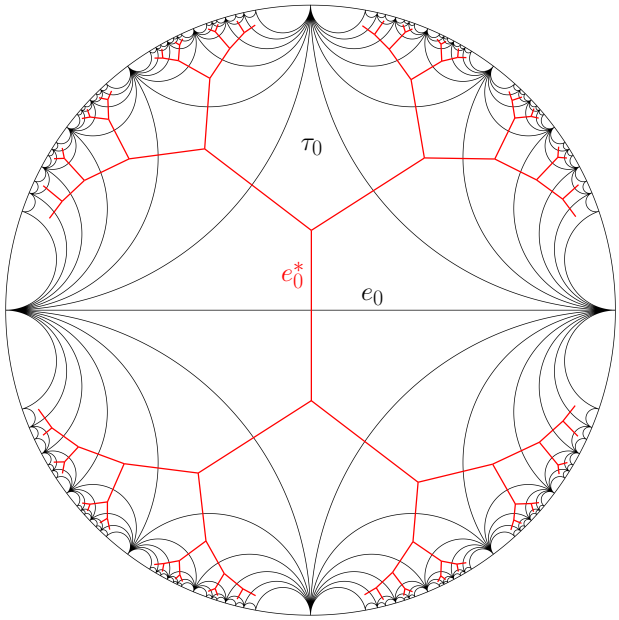


Figure 1.15: Farey tessellation of the disk (black) and its dual tree (red)

1.2.3 Results about shear coordinates

Here we describe the main results of Chapter 3, where we study ℓ^2 spaces of circle homeomorphisms in shear coordinates and compare them to the Weil–Petersson class. This is motivated in part by the question in the previous section, and the aim of finding suitable discrete coordinates for Weil–Petersson curves.

The *Farey tessellation* $\mathfrak{F}$ is a triangulation of the disk by triangles with vertices $V = \mathbb{Q}^2 \cap S^1$ and edges E. The edges of the triangles are hyperbolic geodesics, and $\mathfrak{F}$ can be generated by starting with the triangle $\tau_0 = \{1, \mathfrak{i}, -1\}$, and iteratively reflecting over edges by orientation-reversing isometries. The *generation* of a triangle in the tessellation is its number of reflections away from τ_0. See Figure 1.15 for the first few generations of $\mathfrak{F}$ (black) and the dual graph (red) which is a trivalent tree.

Any element of the universal Teichmüler space $T(\mathbb{D})$, or in fact any element h of the larger homogeneous space of orientation preserving circle homeomorphisms

$$\mathrm{Homeo}^+(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \simeq \{h \in \mathrm{Homeo}^+(\mathbb{T}) \ : \ h \text{ fixes } -1, \mathfrak{i}, 1\}$$

can be uniquely encoded by a *shear coordinate* on the edges of the Farey tessellation, namely a function $s : E \to \mathbb{R}$. More precisely, to encode a homeomorphism h fixing $1, \mathfrak{i}, -1$, we define a new tessellation $h(\mathfrak{F})$ with vertices $h(V)$ and edges $h(E) = \{(h(a), h(b)) : (a, b) \in E\}$. Given an edge $e = (a, c) \in E$, let $Q_e = (a, b, c, d)$ denote the quadrilateral in $\mathfrak{F}$ with diagonal e. The shear function $s_h : E \to \mathbb{R}$ corresponding to h is

$$s_h(e) = \log \mathrm{cr}(h(a), h(b), h(c), h(d)), \tag{1.28}$$

where cr denotes the cross ratio. Equivalently, this is the (signed) hyperbolic distance between the perpendicular bisecting geodesics from $h(b), h(d)$ to $h(e)$. See Figure 1.16. En-

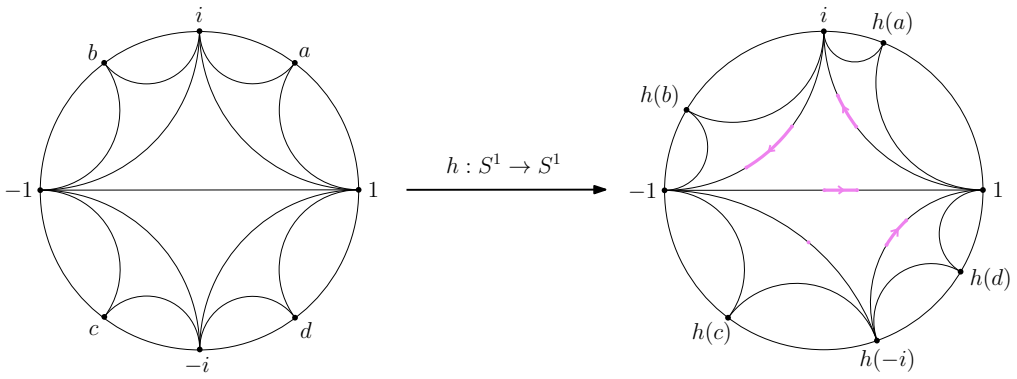


Figure 1.16: First few trianges of $\mathfrak{F}$, and the image tessellation under a circle homeomorphism h. The shears s_h are the (signed) hyperbolic lengths of the pink segments.

coding a homeomorphism by shears is an example of the more general idea of encoding a homeomorphism by an earthquake map of the disk. These come from hyperbolic geometry and were introduced by Thurston [Thu22]. In particular, earthquakes can be defined on any geodesic lamination of the disk, and shears correspond to taking the edges E of $\mathfrak{F}$ (or a subset) as this lamination.

While any homeomorphism can be encoded by a shear function, not all shear functions encode homeomorphisms. A characterization of the Weil–Petersson class $\mathrm{WP}(\mathbb{T})$ is not known, however the shear functions for $\mathrm{Homeo}(\mathbb{T})$, $\mathrm{QS}(\mathbb{T})$ have been characterized [Šar10; Šar13; Šar21]. The conditions on shear functions for certain analytic classes are often somewhat complicated. Here take an opposite perspective, and study certain ℓ^2 classes of shear functions and the analytic properties these constraints imply on the corresponding homeomorphisms.

Naïvely, the first class to consider is the set of square summable shear functions, i.e.

$$\mathscr{S} := \{s : E \to \mathbb{R} : \sum_{e \in E} s(e)^2 < \infty\}.$$

However, we find that not all $s \in \mathscr{S}$ even encode circle homeomorphisms. Conversely, we also find that there are quasisymmetric homeomorphisms which are not in $\mathscr{S}$. See Proposition 3.5.11 for simple examples to illustrate both of these facts. In summary:

$$T(\mathbb{D}) \simeq \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \not\subset \mathscr{S} \qquad \text{and} \qquad \mathscr{S} \not\subset \mathrm{Homeo}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}).$$

These observations show that a basis of shear functions each supported on a single edge is "too large" to define an ℓ^2 space of circle homeomorphisms.

Motivated by this, we investigate shear functions supported on finitely many edges. Finitely supported shear functions always induce homeomorphisms, which in particular are *piecewise Möbius* with pieces bounded by rational points in V (see Lemma 3.3.1). A homeomorphism h with finitely supported shear s_h is piecewise Möbius *and* C^1 with breakpoints in V if and only if it is Weil–Petersson, and we show that this happens if and only if it belongs to a linear subspace of all shear functions spanned by *diamond shears*. The quasicircles corresponding to these homeomorphisms are exactly the C^1 piecewise geodesic quasicircles mentioned in the previous section. See Lemma 3.3.4 and Proposition 3.3.6.

For each edge $e = (a, c) \in E$, the corresponding diamond shear basis element $\Delta_e : E \to \mathbb{R}$ has four nonzero shears. More precisely, let $e_1 = (a, b)$, $e_2 = (b, c)$, $e_3 = (c, d)$, $e_4 = (d, a)$ in E be the boundary edges in counterclockwise order of quadrilateral $Q_e = (a, b, c, d)$ consisting of the two triangles from $\mathfrak{F}$ containing $e = (a, c) \in E$. Then $\Delta_e : E \to \mathbb{R}$ is the shear function with $\Delta_e(e_1) = \Delta_e(e_3) = 1$ and $\Delta_e(e_2) = \Delta_e(e_4) = -1$. If a shear function $s : E \to \mathbb{R}$ has finite support, then there is a unique way to rewrite it as $\sum_{e \in E} \vartheta(e)\Delta_e$ (Lemma 3.3.6). We call the coefficients $\vartheta : E \to \mathbb{R}$ the *diamond shear coordinate*. Since the quad Q_e corresponds to an edge of the dual tree $\mathfrak{F}^*$, it is often more convenient to define the diamond shear coordinate on the edges of the dual tree $\mathfrak{F}^*$. As the edges of $\mathfrak{F}$ and $\mathfrak{F}^*$ are in one-to-one correspondence, this identification should not add any ambiguity. See Figure 3.1.

We can extend the definition of diamond shears to functions $s : E \to \mathbb{R}$ with infinite support, and give three descriptions of the corresponding diamond shears:

1. Combinatorial description: ϑ corresponds to an infinite sum denoted $\Psi(s)$ whenever s is in a certain subclass $\mathscr{P}$ (which can be characterized analytically in terms of differentiability of h). See Section 3.3.3.

2. Analytic description: if h is such that s_h lies in $\mathscr{P}_0 \subset \mathscr{P}$ (which implies that h is differentiable at all $v \in V$), then the diamond shear is given by $\vartheta_h(e) = \frac{1}{2} \log h'(a)h'(b) - \log \frac{h(a)-h(b)}{a-b}$. See Proposition 3.3.25.

3. Relationship with $\log \Lambda$ lengths: $\log \Lambda$-*lengths* are coordinates defined by R. Penner on decorated Teichmüller space. This is a (trivial) bundle over $T(\mathbb{D})$, with fiber $\mathbb{R}^V_{>0}$ over $h \in T(\mathbb{D})$ corresponding to choosing a horocycle at each $h(v) \in h(V)$. See [Pen93; Pen02; MP98] or the book [Pen12]. Roughly speaking, the decoration allows one to truncate and define the "renormalized hyperbolic length" of an infinite geodesic $h(e) \in h(E)$. A homeomorphism h that is differentiable on V gives a canonical way to *fix* as in [MP98] a decoration on $h(V)$, and $\vartheta_h(e)$ is equal to $-1/2$ times the renormalized length of e. In sum, if $s_h \in \mathscr{P}_0$, then for any $e = (a, b) \in E$,

$$\vartheta_h(e) = -\log \Lambda_h(e) = -\frac{1}{2} \operatorname{length}(h(e))$$

where $\operatorname{length}(h(e))$ is the signed hyperbolic length of the part of $h(e)$ between the horocycles centered at $h(a)$ and $h(b)$ chosen from the fixed decoration. See Lemma 3.3.30.

The main object of study in Chapter 3 is the set of shear functions with ℓ^2 summable diamond shear coordinates:

$$\mathscr{H} := \{s : E \to \mathbb{R} : \sum_{e \in E} \vartheta(e)^2 < \infty\}.$$

It is relatively straightforward to show that this corresponds to a class of circle homeomorphisms. Given this, we use the abuse of notation $h \in \mathscr{H}$ to mean $s_h \in \mathscr{H}$ throughout. Moreover, unlike $\mathscr{S}$, it is a subset of the universal Teichmüller space $T(\mathbb{D})$.

Proposition 1.2.8 (See Corollary 3.3.21)**.** *If $s \in \mathscr{H}$, then s induces a quasisymmetric circle homeomorphism. In other words, $\mathscr{H} \subset \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \simeq T(\mathbb{D})$.*

Our first main result is to characterize the Hölder classes of circle homeomorphisms that are contained in $\mathscr{H}$. Define for $\alpha \in (0,1]$,

$$\mathscr{C}^{1,\alpha} := \{h : \mathbb{T} \to \mathbb{T} \text{ homeomorphism} \colon \log h' \text{ is } \alpha\text{-Hölder}\}. \tag{1.29}$$

In particular, the welding homeomorphisms of $C^{1,\alpha}$ Jordan curves belong to $\mathscr{C}^{1,\alpha}$.

Theorem 1.2.9 (See Theorem 3.4.1)**.** *If $\alpha > 1/2$, then $\mathscr{C}^{1,\alpha} \subset \mathscr{H}$.*

This result is sharp as Theorem 3.1.5 will show that $\mathscr{C}^{1,1/2}$ is not in $\mathscr{H}$. The proof of this result relies on Proposition 3.1.1 and the $\ell^{2\alpha}$ summability of the lengths of the shorter arcs in $\mathbb{T}$ between a and b for each $(a,b) \in E$. The ℓ^2 summability was studied and implied by results in [Hal70; Pen02], we improve it to $\ell^{2\alpha}$ summability. See Proposition 3.4.2.

Our next main result is to compare $\mathscr{H}$ and $\mathscr{S}$ with the Weil–Petersson class and its topology.

Theorem 1.2.10. *We have $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$. Additionally if $h \in \mathrm{WP}(\mathbb{T})$, then $s_h \in \mathscr{S}$. Both inclusions are strict.*

See Theorem 3.5.5 for the first inclusion, Section 3.5.3 for why it is strict, and Theorem 3.5.12 for the last inclusion. For comparison, note that this result implies that Theorem 3.1.4 is sharp, since $\mathscr{C}^{1,1/2} \not\subseteq \mathscr{H}$ as otherwise it would also be in $\mathrm{WP}(\mathbb{T})$ (which contradicts Lemma 3.2.9). As smooth diffeomorphisms are dense in $\mathrm{WP}(\mathbb{T})$, so is $\mathscr{H}$.

The fact that the inclusion $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$ is strict basically comes from the fact that a homeomorphism $h \in \mathscr{H}$ must be differentiable at all the vertices $V = \mathbb{Q}^2 \cap \mathbb{T}$. Weil–Petersson homeomorphisms can only have a measures zero set of non-differentiable points, but of course they are not required to lie in the complement of V. As discussed in the previous section, non-differentiable points of Weil–Petersson curves cannot be simple singularities (e.g. corners), but correspond to move exotic behavior like log log spirals. These types of curves are the ones that live in-between $\mathscr{H}$ and $\mathrm{WP}(\mathbb{T})$.

A key step in the proof of $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$ is an explicit construction of a quasiconformal extension $f : \mathbb{D} \to \mathbb{D}$ of $h \in \mathscr{H}$ inspired by a construction in [KM08] by Kahn and Markovic. The construction is adapted to the cell decomposition of the Farey tessellation, and is one of the places where its discrete structure is essential. This construction crucially uses the *generalized balanced condition* satisfied by shear functions that can be written in terms of diamond shears. While characterizations of shear functions for quasisymmetric homeomorphisms are known, analogous methods for constructing their quasiconformal extensions using the shear function are not known. We further find that if $h \in \mathscr{H}$, then the Beltrami differential $\mu_f = \bar{\partial} f / \partial f$ of the extension f is in $L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$. Cui [Cui00] showed that the Douady–Earle quasiconformal extension of a Weil–Petersson homeomorphism satisfies this property, and we remark that it is notable that our construction using shears has the desired property for all $h \in \mathscr{H}$.

In fact, our construction of the quasiconformal extension for functions in $\mathscr{H}$ can be adapted to show the following stronger result that convergence in $\mathscr{H}$ endowed with its ℓ^2 topology implies convergence in the Weil–Petersson metric.

Theorem 1.2.11 (See Corollary 3.5.10). *Suppose that $h, (h_n)_{n\geq 1} \in \mathscr{H}$ with diamond shear coordinates ϑ, ϑ_n respectively. If*

$$\lim_{n\to\infty} \sum_{e\in E} (\vartheta_n(e) - \vartheta(e))^2 = 0,$$

then h_n converges to h in the Weil–Petersson metric.

We obtain immediately the following corollary.

Corollary 1.2.12. *The class of continuously differentiable and piecewise Möbius circle homeomorphisms (with break points in V) is dense in $\mathscr{H}$ and in $\mathrm{WP}(\mathbb{T})$.*

Indeed, this class is equal to the class of circle homeomorphisms with finitely supported diamond shear coordinates (Lemma 3.3.4 and Proposition 3.3.6) which is dense in $\mathscr{H}$ for the Weil–Petersson metric by the above theorem.

We also study infinitesimal shear and diamond shear coordinates on the tangent spaces of $\mathscr{H}$. We compute an explicit formula for the Weil–Petersson metric in terms of diamond shears.

Theorem 1.2.13 (See Corollary 3.6.5, Theorem 3.6.8 and Corollary 3.6.10). *Each ℓ^2-summable infinitesimal diamond shear gives rise to a $H^{3/2}$ vector field on $\mathbb{T}$. Let u_1, u_2 be the $H^{3/2}$ vector fields corresponding to the ℓ^2-summable infinitesimal diamond shears $\dot\vartheta_1, \dot\vartheta_2 \in T_h\mathscr{H} \subset T_h \mathrm{WP}(\mathbb{T})$. Then*

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = \sum_{e_1\in E}\sum_{e_2\in E} \dot\vartheta_1(e_1)\dot\vartheta_2(e_2)\, g(h(Q_{e_1}), h(e_1), h(Q_{e_2}), h(e_2)),$$

where for $Q = (a_1, a_2, a_3, a_4)$, $e = (a_1, a_3)$, $Q' = (b_1, b_2, b_3, b_4)$, $e' = (b_1, b_3)$,

$$g(Q, e, Q', e') = \frac{2}{\pi}\,\mathrm{Re} \sum_{j,k=1}^{4} \frac{(-1)^{j+k} a_j^2 \bar b_k^2 (a_{j+1} - a_{j-1})(\bar b_{k+1} - \bar b_{k-1})}{(a_{j+1} - a_j)(a_j - a_{j-1})(\bar b_{k+1} - \bar b_k)(\bar b_k - \bar b_{k-1})} \sigma(a_j, b_k),$$

and for $a, b \in \mathbb{T}$,

$$\sigma(a, b) = \sum_{p=0}^{\infty} \frac{(a\bar b)^{p+1}}{(1+p)(2+p)(3+p)}.$$

The expression of the metric tensor is relatively complicated. In contrast, the symplectic form has a very simple expression first noticed by Penner in [Pen93; Pen92]. Using the formula in [Pen93, Thm. 5.5] and the relationship between diamond shears and $\log\Lambda$-lengths that we describe in Section 3.3.5, we can rewrite the Weil–Petersson symplectic form in terms of a mixture of infinitesimal shears and diamond shears as follows.

Theorem 1.2.14 (See Theorem 3.6.11). *Let ω denote the Weil–Petersson symplectic form on $\mathrm{WP}(\mathbb{T})$ and fix $h \in \mathscr{H}$. Suppose that u_1, u_2 are the $H^{3/2}$ vector fields corresponding to the ℓ^2-summable infinitesimal diamond shears $\dot\vartheta_1, \dot\vartheta_2 \in T_h\mathscr{H} \subset T_h \mathrm{WP}(\mathbb{T})$ with infinitesimal shear coordinates $\dot s_1, \dot s_2$ respectively. Then*

$$\omega(u_1, u_2) = \sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e) = -\sum_{e\in E} \dot s_1(e)\dot\vartheta_2(e).$$

We note the resemblance of this formula with the Weil–Petersson symplectic form on the finite dimensional Teichmüller spaces $T_{g,n}$ using the Fenchel-Nielson coordinates due to Wolpert [Wol83]:

$$\omega = -\frac{1}{2}\sum_{\gamma\in\Gamma} \mathrm{d}l \wedge \mathrm{d}\tau,$$

where Γ is a maximal multicurve on a Riemann surface of finite type. Here, one may draw the analogy by interpreting $\dot{s}$ as the deformation by twisting along closed geodesics corresponding to $\mathrm{d}\tau$, and $\dot{\vartheta}$ as the deformation by changing the length of geodesics corresponding to $-\frac{1}{2}\mathrm{d}l$ by Corollary 3.1.2.

Finally, in Section 3.7, we prove a preliminary result about random diamond shear functions. More precisely, let $\Theta : E \to \mathbb{R}$ be the diamond shear function where $(\Theta_e)_{e\in E}$ are i.i.d. standard Gaussians with mean 0 and variance 1. We show:

Theorem 1.2.15 (See Theorem 3.7.1)**.** *The diamond shear function $\Theta : E \to \mathbb{R}$ induces a homeomorphism almost surely.*

Intuitively we expect this result, since $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$ suggests that the random object in $\mathscr{H}$ should be more regular than the one associated with Weil–Petersson. Naturally we conjecture that the random object associated with Weil–Petersson is the SLE welding, which is a homeomorphism. To prove Theorem 3.7.1, we write the condition for a shear function to encode a homeomorphism from [Šar10] in terms of the dual tree $\mathfrak{F}^*$, and analyze the way the diamond shears add along different types of paths in the dual tree.

Chapter 2

The 3D dimer model

This chapter corresponds to [CSW23], which is joint work with Nishant Chandgotia and Scott Sheffield.

2.1 Introduction

2.1.1 Overview

Let $G = (V, E)$ be a bipartite graph. A *dimer cover* (a.k.a. *perfect matching*) of G is a collection of edges so that every vertex is contained in exactly one edge. Throughout this paper, we will assume that G is an induced subgraph of $\mathbb{Z}^d$. We partition $\mathbb{Z}^d$ into *even* vertices (the sum of whose coordinates is even) and *odd* vertices (the sum of whose coordinates is odd). By convention, we represent an (*a priori* undirected) edge e by (a, b) where a is even and b is odd.

We can also take a dual perspective, where each vertex $a = (a_1, a_2, \ldots, a_d)$ is represented by the hypercube $[a_1 - \frac{1}{2}, a_1 + \frac{1}{2}] \times \ldots \times [a_d - \frac{1}{2}, a_d + \frac{1}{2}]$ and each matched edge is represented by a "domino" which is the union of two adjacent hypercubes. The dominoes are 2×1 (or 1×2) boxes in 2D and $2 \times 1 \times 1$ (or $1 \times 2 \times 1$ or $1 \times 1 \times 2$) boxes in 3D. From this perspective, the perfect matchings corresponding to the subgraph induced by $V \subset \mathbb{Z}^d$ correspond to *domino tilings* of the region formed by the union of the corresponding cubes. The figure below illustrates a domino tiling of a two-dimensional region called the *Aztec diamond*. On the left, the domino corresponding to (a, b) is colored one of four colors, according to the value of the unit-length vector $b - a$. On the right, squares are colored by parity.

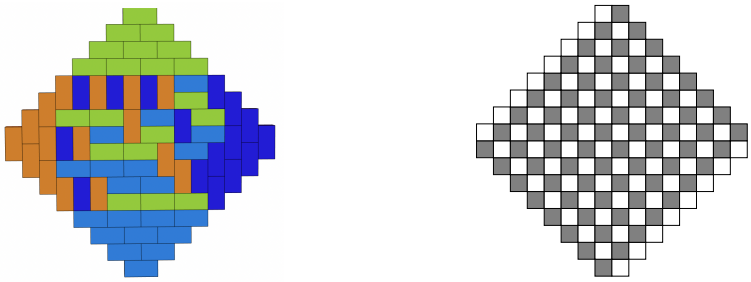


Figure 2.1: Tiling of an Aztec diamond and bipartite coloring of squares in $\mathbb{Z}^2$.

In other words, every domino in the tiling on the left contains one square that is black (in the chessboard coloring on the right) and one that is white—and the color of a domino depends on whether its white square lies north, south, east or west of its black square.

Tilings with all dominoes oriented the same way are called *brickwork tilings*. There are four brickwork orientations in dimension 2—and $2d$ brickwork orientations in dimension d.

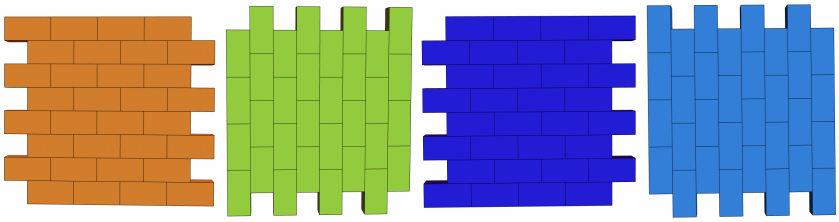


Figure 2.2: The four brickwork patterns in two dimensions.

A perfect matching τ of an induced subgraph R of $\mathbb{Z}^d$ induces an *lattice flow* v_τ that sends one unit of current from every even vertex to the odd vertex it is matched to. If we subtract a "reference flow" (which sends a current of magnitude $1/2d$ from each even vertex to each of its $2d$ odd neighbors) we obtain a *divergence-free* flow f_τ. The main problem in this paper is to understand the behavior of the *random* divergence-flow f_τ that corresponds to a τ chosen uniformly from the set of tilings of a large region, subject to certain boundary conditions.

2.1.2 Two-dimensional background

In two dimensions, the divergence-free flow on $\mathbb{Z}^2$ described above has a dual flow on the dual graph (obtained by rotating each edge 90 degrees counterclockwise about its center) that is a curl-free flow, and hence can be realized as the discrete gradient of some real-valued function defined on the vertices of the dual graph; see Section 2.2.1. This function (defined up to additive constant) is called the *height function* of the flow. Questions about the random flows associated to random perfect matchings can be equivalently formulated as questions about random height functions. For example, one can ask: when a tiling of a large region is chosen uniformly at random, what does the "typical" height function look like?

In 2000, Cohn, Kenyon and Propp studied domino tilings of a domain $R \subset \mathbb{R}^2$ like the one below, asking what happens in the limit as the mesh size tends to zero and the (appropriately

rescaled) height function on the boundary converges to a limiting function [CKP01]. Note that given any tiling τ that covers R (e.g. τ could be one of the brickwork tilings) one can form a tileable region R_n by restricting to the tiles strictly contained in R, and the choice of R_n determines how the height function changes along the boundary.

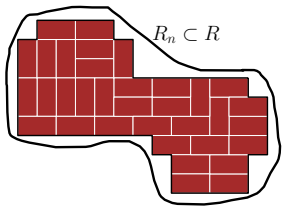


Figure 2.3: An example of a fixed boundary region $R_n \subset R$ for the LDP in two dimensions.

Cohn, Kenyon and Propp showed that as the mesh size tends to zero (and the rescaled boundary heights converge to some function on ∂R) the random height function converges in probability to the unique continuum function u that (given the boundary values) minimizes the integral

$$\int \sigma\big(\nabla u(z)\big)dz, \tag{2.1}$$

where $\sigma(s)$ is the *surface tension* function, which means that $-\sigma(s)$ is the *specific entropy* (a.k.a. *entropy per vertex*) of any ergodic Gibbs measure of slope s. More generally, they established a theory of *large deviations* by showing how exponentially unlikely the random height function would be to concentrate near any other u. Earlier work studied this problem specifically for the *Aztec diamond*, see [CEP96] and [JPS98].

The proof in [CKP01] used ingredients from the scalar height function theory (McShane-Whitney extensions, monotone couplings, stochastic domination, etc.) and the Kasteleyn determinant representation (an exact formula for the entropy function) that do not appear readily adaptable to three dimensions.

The literature on the two-dimensional dimer model is quite large and we will not attempt a detailed survey here. Introductory overviews with additional references include e.g. [Ken09] and [Gor21].

We remark that fixing the asymptotic height function boundary values on ∂R is equivalent to fixing the asymptotic rate at which current flows though ∂R in the corresponding divergence-free flow. The latter interpretation is the one that extends most naturally to higher dimensions.

2.1.3 Three-dimensional setup and simulations

The goal of this paper is to extend [CKP01] to higher dimensions, where different tools are required. For simplicity and clarity, we focus on 3D, but we expect similar arguments to work in dimensions higher than three. (We discuss possible generalizations and open problems in Section 2.9.) Before we present our main results, we provide a few illustrations. The figure below illustrates a uniformly random tiling of a $10\times10\times10$ cube, with six colors corresponding

to the six orientations. Next to it is the underlying black-and-white checkerboard grid. This figure and the others below were generated by a Monte-Carlo simulation (see Section 2.3.3) that is known to be mixing, but whose mixing rate is not known. It was run long enough that the pictures *appeared* to stabilize but we cannot quantify how close our samples are to being truly uniform. The efficient *exact sampling* algorithms that work in 2D do not have known analogs in 3D.

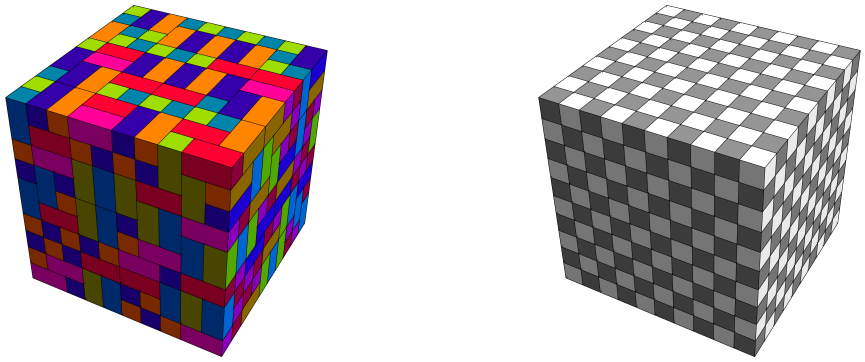


Figure 2.4: A dimer tiling of the $10 \times 10 \times 10$ cube and the bipartite coloring of the cubes in $\mathbb{Z}^3$.

The figure below represents a random tiling τ of a region R called the *Aztec pyramid* (formed by stacking Aztec diamonds of width $2, 4, 6, \ldots, 36$). Next to it is again the underlying black-and-white checkerboard coloring. Recall that (due to the reference flow) the divergence-free flow f_τ sends a $1/6$ unit of current through each square on the boundary ∂R. Such a square divides a cube inside R from a cube outside R. The flow is directed *into* R if the cube inside R is even, and *out of* R if the cube inside R is odd. Of the four triangular faces of the pyramid, two consist entirely of even cubes and the other two consist entirely of odd cubes. This means that f_τ current enters two opposite triangular faces at its maximal rate, and exits other two triangular faces at its maximal rate, while the net current through the bottom square face is zero (since on the lower boundary, the number of faces bounding even cubes in R equals the number of faces bounding odd cubes in R).

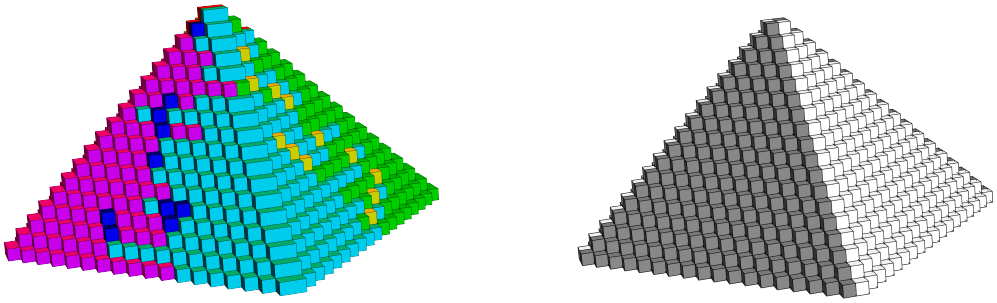


Figure 2.5: A dimer tiling of an Aztec pyramid and the bipartite coloring of the cubes in $\mathbb{Z}^3$.

Below is a larger Aztec pyramid seen from above and from underneath. One can construct a computer animation showing the horizontal cross-sections one at a time. For the three large

simulations shown here in Figures 2.6, 2.7 and 2.8, animations of the slices are available at https://github.com/catwolfram/3d-dimers. In these animations, it appears that each cross section has four "frozen" brickwork regions and a roughly circular "unfrozen" region, similar to the 2D Aztec diamond.

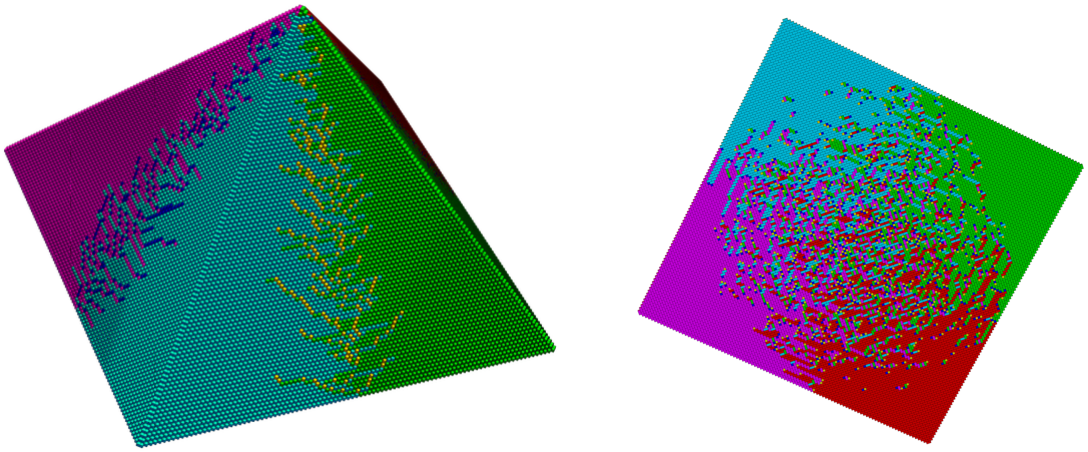


Figure 2.6: A tiling of a larger pyramid of Aztec diamonds, from the side and from below.

We now describe two larger labeled figures. Figure 2.7 illustrates a uniformly random tiling of the *Aztec octahedron* formed by gluing two Aztec pyramids along their square face. Four of the eight triangular faces of the octahedron contain only even cubes on their boundary, and the other four contain only odd cubes (and these alternate; distinct faces sharing an edge have opposite parity). In light of this, we can say that the current enters four of the faces at the maximum possible rate and exits the other four faces at the maximum possible rate. In simulations there appear to be twelve frozen regions (one for each *edge* of the octahedron) in which one of the six brickwork patterns dominates. (By contrast, tilings of the two dimensional Aztec diamond have four frozen regions, one for each *vertex* of the diamond.) Within each brickwork region, current flows at its maximum possible rate from one face (where it enters the octahedron) to an adjacent face (where it exits). Away from these brickwork regions, one sees a mix of colors, with a gradually varying density for each color. These are regions where the magnitude of the current flow is smaller, and the mean current appears to vary continuously across space.

Figure 2.8 illustrates a tiling of the *Aztec prism* (formed by stacking Aztec diamonds whose widths alternate between $2n$ and $2n+2$). Again, each slice seems to be frozen outside of a roughly circular region. The width-alternation ensures that each of the four rectangular side faces of the prism has either only even faces or only odd faces exposed. Thus, current enters two of the opposite side faces at its maximal rate, and exits the other two at its maximal rate. The net current flowing through the top and bottom faces is zero. In this figure, and in all of the examples above, the distribution of domino colors in a subset of the tiled region determines the "mean direction of current flow" in that subset. We are interested in understanding what the "typical flow" looks like in the fine mesh limit.

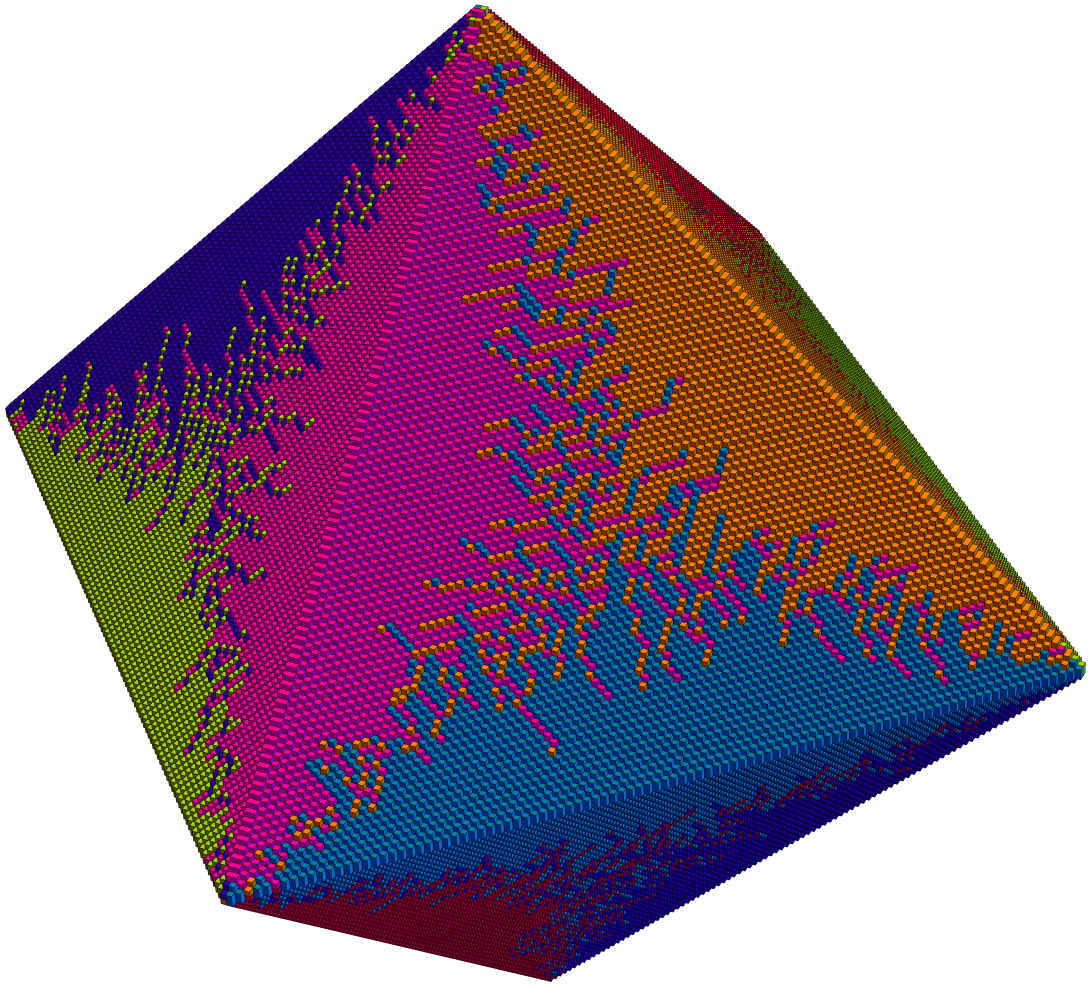


Figure 2.7: Tiling of an Aztec octahedron.

2.1.4 Main results and methods

The main results of this paper will be two versions of a *large deviation principle* (LDP) for fine-mesh limits of uniformly random dimer tilings of compact regions $R \subset \mathbb{R}^3$, with some limiting boundary condition. The versions of the LDP we prove differ in how we treat the boundary conditions.

A large deviation principle is a result about a sequence of probability measures $(\rho_n)_{n\geq 1}$ which quantifies the probability of rare events at an exponential scale as $n \to \infty$. More precisely, a sequence of probability measures $(\rho_n)_{n\geq 1}$ on a topological space $(X, \mathcal{B})$ is said to satisfy a large deviation principle (LDP) with rate function I and speed v_n if $I : X \to [0, \infty)$ is a lower semicontinuous function, and for all Borel measurable sets B,

$$-\inf_{x\in B^\circ} I(x) \leq \liminf_{n\to\infty} v_n^{-1} \log \rho_n(B) \leq \limsup_{n\to\infty} v_n^{-1} \log \rho_n(B) \leq -\inf_{x\in \overline{B}} I(x),$$

where $B^\circ, \overline{B}$ denote the interior and closure of B respectively. The rate function $I(x)$ is *good* if its sublevel sets $\{x : I(x) \leq a\}$ are compact. Implicit in this definition is a choice of

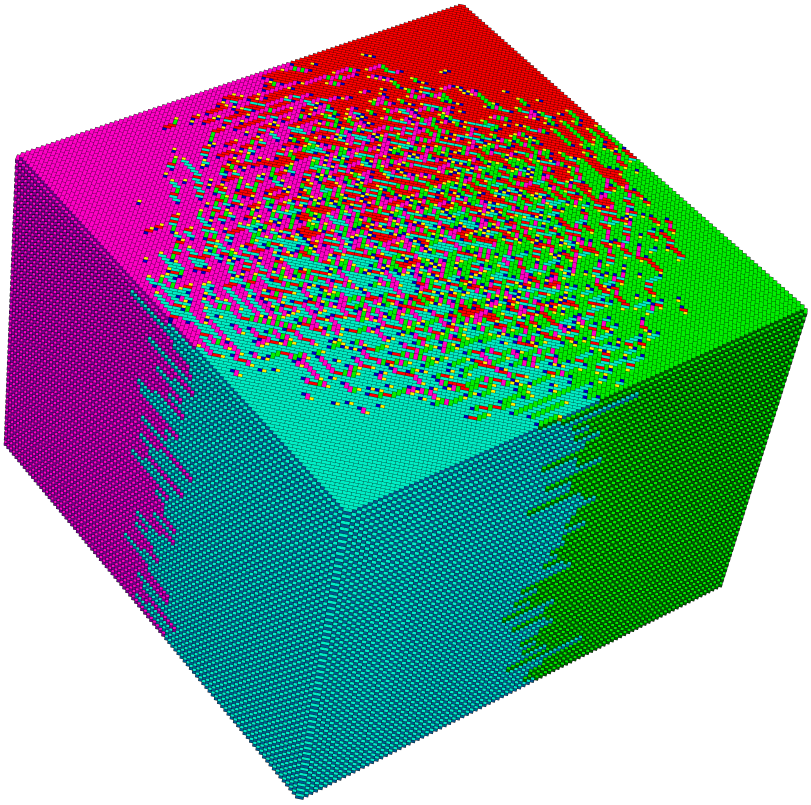


Figure 2.8: Tiling of an Aztec prism.

topology $\mathcal{B}$ on X. A large deviation principle for $(\rho_n)_{n\geq 1}$ implies that random samples from ρ_n are exponentially more likely to be near the minimizers of $I(\cdot)$ as $n \to \infty$. When I is good and has a unique minimizer, this means that random samples from ρ_n *concentrate* as $n \to \infty$ in the sense that if U is any neighborhood of the unique minimizer, then as $n \to \infty$ the probability $\rho_n(X \setminus U)$ tends to zero exponentially quickly in v_n. Good references for this subject include [DZ09] and [Var16].

To formulate the large deviation principle for dimer tilings, we associate to each tiling τ of $\mathbb{Z}^3$ a divergence-free discrete vector field. As mentioned above, we can define a flow v_τ that moves one unit of flow from the even endpoint to the odd endpoint of each e in τ. Subtracting a reference flow which sends $\frac{1}{6}$ flow along every edge from even to odd produces a divergence-free discrete vector field:

$$f_\tau(e) = \begin{cases} +5/6 & e \in \tau \\ -1/6 & e \notin \tau. \end{cases}$$

We call this a *tiling flow*, and it will play an analogous role to the height function in two dimensions. The height function in two dimensions is a scalar potential function whose gradient is the *dual* of the tiling flow (i.e., the flow on the dual lattice obtained by rotating each edge 90 degrees counterclockwise about its center). See Section 2.2.1.

Tiling flows can also be *scaled* so that they are supported on $\frac{1}{n}\mathbb{Z}^3$ instead of $\mathbb{Z}^3$. We scale them so that the total flow per edge is proportional to $\frac{1}{n^3}$, so that in the fine-mesh limit with $n \to \infty$, the total flow in a compact region in $\mathbb{R}^3$ converges to a finite value. The scaled tiling flows takes values $-\frac{1}{6n^3}$ and $\frac{5}{6n^3}$. A *scale n dimer tiling* is a dimer tiling of $\frac{1}{n}\mathbb{Z}^3$ (or a subset of it).

Fix a compact region $R \subset \mathbb{R}^3$ (which is sufficiently nice, e.g. the closure of a connected domain with piecewise smooth boundary). We define the *free-boundary tilings of R at scale n* to be tilings τ at scale n such that every point in R is contained in a tile in τ, and every tile has some intersection with R.

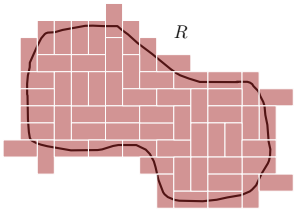


Figure 2.9: An example of a free-boundary tiling of R

We denote the corresponding *free-boundary tiling flows of R at scale n* by $TF_n(R)$. Note that $TF_n(R)$ is a finite set for all n. There is a signed *flux measure* on ∂R (supported on the points where edges of $\frac{1}{n}\mathbb{Z}^3$ cross ∂R) that encodes the net amount of flow directed into R. Since f_τ is divergence-free, the *total* flux through ∂R is zero. (See Definition 2.5.5.)

If τ is a *random* tiling of $\mathbb{Z}^3$ whose law is invariant under translations by even vectors, then one can define the *mean current per even vertex* to be $s = \mathbb{E}[\eta]$ where η is the vertex of $\mathbb{Z}^3$ matched to the origin by τ. Note that s lies in the set

$$\mathcal{O} = \{(s_1, s_2, s_3) \in \mathbb{R}^3 : |s_1| + |s_2| + |s_3| \leq 1\},$$

which is the convex hull of $\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\}$, and which we call the *mean-current octahedron*. It indicates the expected total amount of current in v_τ (or equivalently f_τ) per even vertex; see Section 2.2.2. The vertices of $\mathcal{O}$ arise for a random τ that is a.s. equal to one of the six brickwork tilings in three dimensions.

We define $AF(R)$ to be the space of measurable, divergence-free vector fields supported in R taking values in $\mathcal{O}$. The notation AF stands for *asymptotic flow* and is chosen because of the fact (formalized in Theorem 2.5.19) that these are precisely the flows that can arise as $n \to \infty$ limits of tiling flows on R, under a suitable topology.

The topology we use is the *weak topology* on the space of flows obtained by interpreting each vector component of the flow as a signed measure, see Section 2.5. This topology can also be generated by the *Wasserstein distance* for flows. Loosely speaking, two flows are considered Wasserstein close if one can be transformed into the other by moving, adding, and deleting flow with low "cost." The cost is the amount of flow moved times the distance moved, plus the amount of flow added or deleted. The large deviation principles we prove use the same topology (weak topology, which is generated by Wasserstein distance) for the fine-mesh limits of random free-boundary tiling flows of R.

Remark 2.1.1. The 2D large deviation analysis in [CKP01] is based on a topology that at first glance looks different from ours, namely the topology in which two flows are considered close if their corresponding height functions h are close in L^∞. However, it is not too hard to see that 1-Lipschitz functions on R (with fixed boundary values on ∂R) are L^∞ close if and only if their gradients (or equivalently the dual flows of their gradients) are Wasserstein close. We will not use or prove this fact here.

In three dimensions, we will also derive an LDP for random perfect matchings on finite regions approximating a continuum domain R, with boundary conditions chosen so that the flux through the boundary approximates a continuum boundary flow b, in a sense we will explain below. As in two dimensions, the rate function $I_b(\cdot)$ is a function of an asymptotic flow $g \in AF(R)$ and is (up to an additive constant) equal to

$$-\mathrm{Ent}(g) = -\frac{1}{\mathrm{Vol}(R)} \int_R \mathrm{ent}(g(x))\, \mathrm{d}x, \tag{2.2}$$

where the additive constant is $C = \max_{f \in AF(R,b)} \mathrm{Ent}(f)$. We interpret (2.2) as the three-dimensional analog of (2.1). The function $\mathrm{ent}(s)$ is the maximal *specific entropy* of a measure with mean current $s \in \mathcal{O}$. It turns out that for every $s \in \mathcal{O}$ there exists an *ergodic Gibbs measure* of mean current s that achieves this maximal entropy $\mathrm{ent}(s)$. This essentially follows from the strict concavity of ent when s is in the interior of $\mathcal{O}$, and can also be checked for $s \in \partial\mathcal{O}$. See Theorem 2.7.23 and Theorem 2.4.7.

To state our theorems, we need a way of fixing for each n a region R_n that approximates a continuum region R, such that boundary flow corresponding to R_n approximates a continuum boundary flow b on ∂R. The precise analog of the statement given in [CKP01] is not exactly true in 3D, due to subtleties related to the fact that in 3D the function ent can be nonzero even on the boundary of $\mathcal{O}$ (see Section 2.4). To briefly illustrate what can go wrong, consider a finite region S tiled with only three types of tiles: north, east and up. Then every vertex $x = (x_1, x_2, x_3)$ with $x_1 + x_2 + x_3 = c$ (with c even) must be connected to a vertex y with $y_1 + y_2 + y_3 = c + 1$, and vice versa. The vertices with coordinate sums in $\{c, c+1\}$ thus form a "slab" of points that are only connected to each other, and one can use Hall's matching theorem to show that this must be the case for *any* tiling of S. Thus we can view S as a stack of two-dimensional slabs, where the tilings within the different slabs are independent of each other. These slabs could alternate between brickwork patterns (one slab has only east-going tiles, the next has only north-going tiles, the next has only up-going tiles, etc.) but they could also all be nonzero-entropy mixtures of the three tile types. Rescalings of both types of S might approximate the same continuum (R, b), but the corresponding uniform-random-tiling measures would have very different entropy and very different large deviation behavior (see Example 2.8.17 and Section 2.4).

We will present two ways of formulating a theorem that *is* true in 3D. In the first approach we replace the *hard constraint* on the boundary behavior (where an R_n to be tiled is explicitly specified for each n) with a *soft constraint* (where all scale n tilings that cover R are allowed, provided they induce boundary flows that are "sufficiently close" to the desired limiting flow b). This "soft constraint" LDP will apply to a fairly general set of pairs (R, b). In the second approach we keep the hard constraints (i.e., the fixed R_n regions) but require (R, b) to be "flexible" in a sense that prevents the degenerate situation sketched above (where on the

discrete level there might be slab boundaries that cannot be crossed by the tiles of *any* tiling of R_n). Precisely, we say (R, b) is *flexible* if for every $v \in R$ there exists an asymptotic flow g for (R, b) such that for some neighborhood $U \ni v$ we have $\overline{g(U)} \subset \mathrm{Int}(\mathcal{O})$. Informally, this means there is no interior point near which g is *forced* to lie on $\partial\mathcal{O}$. For later purposes, we say that (R, b) is *semi-flexible* if for every $v \in R$ there exists an asymptotic flow g for (R, b) such that for some neighborhood $U \ni v$ the set $\overline{g(U)}$ is disjoint from the edges of $\mathcal{O}$. In other words, there is no interior point near which g is forced to lie on an edge of $\mathcal{O}$.

Using compactness of the space of flows, it is not hard to show that there exists g that minimizes (2.2). However it takes a bit more work to see whether such a g is unique. If ent were strictly concave everywhere, then $I_b(g_1)$ and $I_b(g_2)$ could not be minimal for distinct g_1 and g_2 (since the strict concavity would imply that $I_b\left(\frac{g_1+g_2}{2}\right)$ was even smaller). The trouble is that ent is not strictly concave on the edges of $\mathcal{O}$ (where it is constant) even though we will show that it is strictly concave everywhere else. In principle, there could still exist distinct minimizers g_1 and g_2 that (outside a set of measure zero) disagree *only* at points where they both take values on the same edge of $\mathcal{O}$. (See Problem 2.9.7.) We have not ruled out this possibility in general, but we can prove that (2.2) has a unique minimizer if (R, b) is semi-flexible. (See [Gor21, Proposition 7.10] for a 2D analog of this argument.) This in turn implies that the random flows in the corresponding LDP theorems *concentrate* near the unique minimizer g of I_b. If (R, b) is not semi-flexible then we call it *rigid*. We briefly summarize the conditions needed for each result in the following table before giving more precise statements.

(R, b)	SB LDP	I_b has unique minimizer	HB LDP
rigid	yes	not known	no
semi-flexible	yes	yes	no
flexible	yes	yes	yes

The results marked "no" in this table are provably not true in general. By taking limits of the "stack of slabs" regions discussed above, one can produce a semi-flexible (or rigid) pair (R, b) for which the hard boundary large deviation principle is false (for further discussion of this see Example 2.8.17).

Now, to introduce the soft boundary large deviation principle, we define the probability measure ρ_n to be the uniform measure on the space of flows in $TF_n(R)$ whose boundary values lie within Wasserstein distance θ_n of the desired limiting boundary flow, where $\theta_n \to 0$ as $n \to \infty$. We call θ_n the "threshold sequence" and it can be chosen arbitrarily provided that it does not tend to zero *too quickly* in a sense we explain later. A rough statement of our main theorem is the following.

Theorem 2.1.2 (See Theorem 2.8.6)**.** *Let $R \subset \mathbb{R}^3$ be a compact region which is the closure of a connected domain, with piecewise smooth boundary ∂R. Let b be a boundary value for an asymptotic flow and let $(\theta_n)_{n\geq 1}$ be a (good enough) sequence of thresholds. Let ρ_n be uniform measure on $TF_n(R)$ conditioned on the boundary values being within θ_n of b.*

Then the measures $(\rho_n)_{n\geq 1}$ satisfy a large deviation principle in the Wasserstein topology on flows with good rate function $I_b(\cdot)$ and speed $v_n = n^3\,Vol(R)$, namely for any Wasserstein-measurable set A,

$$-\inf_{g\in A^\circ} I_b(g) \leq \liminf_{n\to\infty} v_n^{-1} \log \rho_n(A) \leq \limsup_{n\to\infty} v_n^{-1} \log \rho_n(A) \leq -\inf_{g\in \overline{A}} I_b(g). \tag{2.3}$$

If g is an asymptotic flow, the rate function $I_b(g)$ is equal up to an additive constant to $-Ent(g)$. (Otherwise it is ∞.)

Remark 2.1.3. The requirements for the region R are mild—for example, we do not require that R is simply connected. In this sense, the theorem can be viewed as extending both [CKP01] (simply connected 2D) and [Kuc22] (multiply connected 2D) to three dimensions. The requirement that ∂R be piecewise smooth is probably not necessary, but if the boundary of R is allowed to be too rough, the theorem statements one can make will depend more sensitively on how the boundary conditions are handled. For example, if the boundary of R has positive volume, then tilings that cover R may have volume-order more tiles than the tilings that approximate R "from within" and the extra tiles may contribute to the limiting entropy. If the boundary of R has infinite area, then the flux through the boundary may be an infinite signed measure, which would have to be defined more carefully. (For example, we could say that two flows that vanish outside of R have "equivalent boundary values" if their difference is a flow on all of $\mathbb{R}^3$ that is divergence-free in the distributional sense, and then let b denote an equivalence class.) For simplicity, we will focus on the piecewise smooth setting in this paper.

The distinction between soft and hard boundary conditions only substantially impacts one step of the proof: the argument that there exists a tiling (in the support of ρ_n) whose flow approximates a piecewise-constant flow that in turn approximates a given $g \in AF(R)$. Theorem 2.1.2 would still apply if the boundary conditions defining the ρ_n were specified in another way (ensuring convergence to (R, b) in the limit) as long as some version of this step could be implemented. We show using the *generalized patching theorem* (Theorem 2.8.32) that under the condition that (R, b) is flexible, this step can be implemented and a hard boundary LDP holds.

We say a boundary value is a *scale n tileable* if there exists a free-boundary tiling τ of R at scale n with that boundary value. If two tilings τ_1, τ_2 have the same boundary values on ∂R, then they are tilings of the same fixed region, so fixing a sequence of scale n tileable boundary values b_n is equivalent to fixing a sequence of *regions* R_n with boundary value b_n. A rough statement of the *hard boundary* large deviation principle is as follows.

Theorem 2.1.4 (Theorem 2.8.15)**.** *Suppose that (R, b) is flexible and that $R_n \subset \frac{1}{n}\mathbb{Z}^3$ is a sequence of regions with tileable boundary values b_n on ∂R converging to b. Let $\overline{\rho}_n$ be uniform measure on dimer tilings of R_n. Then the measures $(\overline{\rho}_n)_{n\geq 1}$ satisfy a large deviation principle in the Wasserstein topology on flows with the same good rate function $I_b(\cdot)$ and speed $v_n = n^3\, Vol(R)$ as the soft boundary measures ρ_n.*

It is straightforward to show that the (R, b) pairs obtained as fine-mesh limits of the "Aztec regions" above (Figures 2.6, 2.7, 2.8) are flexible, despite the fact that *typical* tilings appear (in simulations) to have frozen brickwork regions. (See Remark 2.8.14.)

Under the condition that (R, b) is *semi-flexible* (see further discussion in Definition 2.7.34) we show that Ent has a unique maximizer (Theorem 2.7.36). This together with some basic topological results shows rigorously that *concentration* around a deterministic limit shape occurs, as we see in the simulations. This concentration holds for either the soft boundary measures ρ_n or the hard boundary measures $\overline{\rho}_n$.

Corollary 2.1.5 (See Corollary 2.8.9 and Corollary 2.8.18)**.** *Assume that (R, b) is semi-flexible. For any $\epsilon > 0$, the probability that a uniformly random tiling flow on R at scale n (either sampled from ρ_n, i.e. with boundary flow conditioned to be in a shrinking interval around b, or sampled from $\overline{\rho}_n$ if (R, b) is flexible, i.e. with boundary flow conditioned to be a fixed value b_n converging to b) differs from the entropy maximizer with boundary value b by more than ϵ goes to 0 exponentially fast in n^3 as $n \to \infty$.*

The methods in this paper are substantially different from the methods used to prove the large deviation principle for dimer tilings in two dimensions. The two-dimensional dimer model is *integrable* or *exactly solvable* in the sense that one can derive a (beautiful) explicit formula for the specific entropy function analogous to our function ent, and this explicit formula is used in the large deviations proof. The three-dimensional dimer model is not known to be integrable in this way, so we rely on "softer" arguments. We comment on a few of these below.

One of the key ingredients which does have a 2D analog in [CKP01] is the *patching argument* (Theorem 2.6.14) which essentially states that if two tilings τ_1, τ_2 satisfy a requirement that they "asymptotically have the same mean current s" for some $s \in \text{Int}(\mathcal{O})$, then we can cut out a bounded portion of τ_2 and patch it into an unbounded portion of τ_1 by tiling a thin annular region.

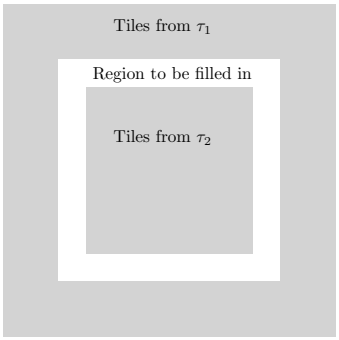


Figure 2.10: Schematic for the patching theorem.

For the hard boundary large deviation principle, we also prove a *generalized patching theorem* (Theorem 2.8.32), which says roughly that two tilings can be patched on a *general* annular region $R \setminus R'$ if they have the same boundary value b on ∂R and the inner one approximates a *flexible* flow $g \in AF(R, b)$.

Proving the patching theorems will be one of the more challenging aspects of this paper. The main input is Hall's matching theorem (proved by Hall in 1935 [Hal35]) which gives us a criterion to check if a region (e.g. the annular region between the two tilings) is tileable by dimers. It turns out that the criterion we need to check can be phrased in terms of the existence of a discretized minimal surface and leads to an interesting sequence of constructions described for boxes in Section 2.6 and generalized in Section 2.8. These arguments are more involved than the 2D patching arguments in [CKP01], which rely on height functions and Lipschitz-extension theory. It is hard to summarize the argument without giving the details, but the following is a very rough attempt (which can be skimmed on a first read).

1. For each n, define an annular region A that we want to tile (which is roughly a scale n approximation of a fixed continuum annular region, with outer boundary conditions determined by τ_1 and inner by τ_2). Use Hall's matching theorem to show that if A is non-tileable there must exist a "surface" dividing the cubes in A into two sub-regions such that 1) the cubes with faces on the surface are odd if they are in the first sub-region, even if in the second and 2) the first sub-region has more even than odd cubes overall.

2. Reduce to the case that the surface is in some sense a "minimal monochromatic surface" given its boundary, which touches both the inside and outside boundaries of A. (Here monochromatic means that all cubes on one side of the surface are odd.)

3. Use an argument involving isoperimetric bounds to show that such a surface must have at least a constant times n^2 faces when n is large.

4. Show that the even-odd imbalance in the first sub-region *cannot* be as large as it would need to be to provide a non-tileability proof. Do this by covering the first region by dominoes (from an tiling sampled from an ergodic measure in Section 2.6, or from a tiling that approximates a flow g whose existence is guaranteed by the flexible condition in Section 2.8) and use geometric considerations to show that there must be a lot of dominoes with only an odd cube in the first sub-region (including order n^2 in the middle of A) and relatively fewer dominoes with only an even cube in the first sub-region (using the ergodic theorem and the fact that both tilings approximate the same constant flow s, or in the generalized case by using Wasserstein distance bounds that apply near the boundary of A). Conclude that the first sub-region has at least as many odd cubes as even cubes, and hence does not prove non-tileability. This argument shows that there exists no surface that proves non-tileability and hence (by Hall's matching theorem) the region is tileable.

Another key step in proving the main theorems is to derive properties of the entropy function ent despite not being able to compute it explicitly on all of $\mathcal{O}$. Instead, $\text{ent}(s)$ is defined abstractly as the maximum specific entropy $h(\cdot)$ of a measure with mean current s (see Section 2.2.3). From a straightforward adaptation of the classical variational principle of Lanford and Ruelle [LR69], it follows that $\text{ent}(s)$ is always realized by a *Gibbs measure* of mean current s (see Theorem 2.7.2, see also Section 2.2 for the definition of a Gibbs measure).

To prove strict concavity of ent on $\text{Int}(\mathcal{O})$ (Theorem 2.7.22), we note that a translation invariant measure μ with mean current s and with $h(\mu) = \text{ent}(s)$ must be a Gibbs measure, and we then use a variant of the cluster swapping technique used in [She05] to compare measures of different mean currents. We call this variant *chain swapping*. It is an operation on measures on *pairs* of dimer tilings (τ_1, τ_2). From a pair of tilings (τ_1, τ_2) (sampled from μ), chain swapping constructs a pair of tilings (τ_1', τ_2') by "swapping" the tiles of τ_1 and τ_2 along some of the infinite paths in (τ_1, τ_2) with independent probability $1/2$ (or any probability $p \in (0, 1)$). We say that (τ_1', τ_2') is sampled from the *swapped measure* μ'. See Section 2.7.4 for a more detailed discussion of chain swapping, including Figure 2.29 for an example.

At a high level, chain swapping is an operation that allows us to take a coupling μ of measures μ_1, μ_2 on dimer tilings of mean currents s_1, s_2, and construct a coupling μ' of two new measures μ'_1, μ'_2 on dimer tilings both of mean current $\frac{s_1+s_2}{2}$. We show that this operation preserves the total specific entropy (i.e. $h(\mu) = h(\mu')$) and ergodicity, but *breaks* the Gibbs property. More precisely, if μ_1, μ_2 are ergodic Gibbs measures of mean currents s_1, s_2 and $\frac{s_1+s_2}{2} \in \mathrm{Int}(\mathcal{O})$, then μ'_1, μ'_2 are *not* Gibbs, and hence do *not* have maximal entropy among measures of mean current $\frac{s_1+s_2}{2}$. The proof that the Gibbs property is broken under chain swapping requires very different techniques from those used in [She05].

Under the assumption that there exists an *ergodic* Gibbs measure μ_s of mean current s for any $s \in \mathcal{O}$, and that $\mathrm{ent}(s) = h(\mu_s)$, strict concavity would follow easily: let $\mu_1 = \mu_{s_1}, \mu_2 = \mu_{s_2}$ and apply chain swapping to get new measures μ'_1, μ'_2 of mean current $\frac{s_1+s_2}{2}$. Since total entropy is preserved,

$$h(\mu'_1) + h(\mu'_2) = h(\mu_1) + h(\mu_2) = \mathrm{ent}(s_1) + \mathrm{ent}(s_2).$$

On the other hand, since μ'_1, μ'_2 are not Gibbs,

$$h(\mu'_1) + h(\mu'_2) < 2\mathrm{ent}(\frac{s_1 + s_2}{2}),$$

which would complete the proof. A rigorous proof of the theorem is given in Section 2.7.5, and relies on casework based on ergodic decompositions as we do not know, a priori, that ergodic Gibbs measures of mean current s exist for all $s \in \mathcal{O}$. However it will then follow *from* strict concavity that this is true, and there exist ergodic Gibbs measures of all mean currents $s \in \mathcal{O}$ (Corollary 2.7.25).

The above is a discussion of ent on $\mathrm{Int}(\mathcal{O})$, where no explict formula is known. We remark that ent is explicitly computable when restricted to $\partial\mathcal{O}$, since it reduces to a two-dimensional problem (see Section 2.4).

2.1.5 Three-dimensional history and pathology

The three-dimensional model is fundamentally different from the two-dimensional version in many respects. To give one example, we recall that if τ and σ are distinct perfect matchings of $\mathbb{Z}^2$ that agree on all but finitely many edges, then one can construct a sequence $\tau = \tau_0, \tau_1, \tau_2, \ldots, \tau_n = \sigma$ of perfect matchings such that for each k, the matchings τ_k and τ_{k-1} agree on all edges except those contained in a single 2×2 square — and on that square one of $\{\tau, \sigma\}$ has two parallel vertical edges and the other has two parallel horizontal edges [Thu90]. From the domino tiling point of view, we say that τ_{k-1} and τ_k differ by a *local move* that corresponds to rotating a pair of dominoes as shown below.

Figure 2.11: A local move or flip in 2D.

It turns out that the analogous statement is false in 3D. In fact, as we will explain in Section 2.3, there is *no* collection of local moves for which the analogous statement would be true in 3D. In 3D, one can construct (for any $K > 0$) a tiling τ of $\mathbb{Z}^3$ that is

1. *non-frozen* — i.e., there exists a tiling $\tau' \neq \tau$ that disagrees with τ on finitely many edges.

2. *locally frozen* to level K — i.e., there exists no tiling $\tau' \neq \tau$ that disagrees with τ' on fewer than K edges.

To understand why this is the case, recall that in two dimensions, one can superimpose an arbitrary perfect matching with a brickwork matching to obtain a collection of non-intersecting left-to-right lattice paths as follows:

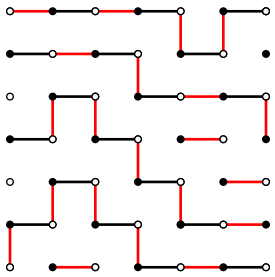


Figure 2.12: 2D non-intersecting paths.

There is an obvious bijection between non-intersecting path ensembles (as shown above) and dimer tilings (which is one way to deduce the integrability of the dimer model in two dimensions). Applying local moves corresponds to shifting these paths up and down locally. One can analogously superimpose a red three-dimensional matching with a black brickwork matching, to obtain an ensemble of left-to-right paths in three dimensions. But in this case the function that maps each "left endpoint" to the "right endpoint" on the same path may not be uniquely determined, as the following example shows. For clarity, the black and red edges that coincide with each other are not drawn—so both figures indicate a red perfect matching of $\mathbb{Z}^3$ that (when restricted to the box) consists only of right-going edges in the brickwork pattern (not shown) and a few non-right-going edges (shown together with the black right-going edges that share their endpoints).

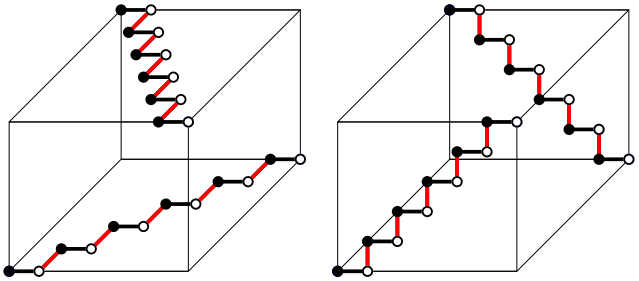


Figure 2.13: Two examples of 3D non-intersecting paths with the same endpoints.

In general, there could be many different paths, and many ways to permute the wires from the left before plugging them in on the right. In the example above, the paths are "taut" in the sense that they have no freedom to "move locally" using local moves that change only, say, three or four edges at a time (and they can be extended to taut paths on all of $\mathbb{Z}^3$). In general, 3D path ensembles are not nicely ordered from top to bottom, and do not have the same lattice structure that 2D path ensembles enjoy. They can be braided in complicated ways.

Despite this complexity, various "local move connectedness" results for 3D tilings have been obtained. See, for example, the series of works by subsets of Freire, Klivans, Milet and Saldanha [MS14a; MS14b; Mil15; MS15; FKMS22; Sal22; Sal20; KS22; Sal21], the recent work [HLT23] by Hartarsky, Lichev, and Toninelli, and physics papers by Freedman, Hastings, Nayak, Qi, and separately Bednik about topological invariants and so-called Hopfions [FHNQ11; Bed19a; Bed19b].

One of the basic observations is that even on box-shaped regions in 3D, one cannot transform any tiling to any other tiling with a sequence of *flips* (which swap two edges of a lattice square with the other two). There is a quantity associated to a tiling, called the *twist* (related to the linking number from knot theory) that is preserved by flip moves but changed by so-called *trit* moves, which involve removing three edges contained in the same $2 \times 2 \times 2$ cube and replacing them with three others, see below:

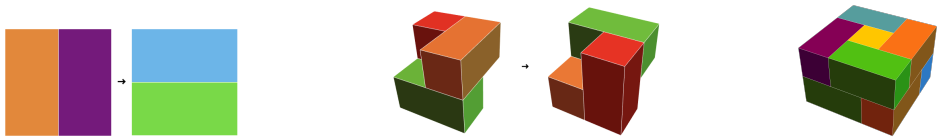


Figure 2.14: A *flip*, a *trit* and a flip-rigid configuration called a *hopfion*.

It remains open whether it is possible to connect any tiling of a rectangular box to any other using *both* the flip and trit moves shown above. It is still possible in 3D to generate random tilings of finite regions using Glauber dynamics (using an update algorithm that allows for the tiling to be modified along cycles of arbitrary length, see Section 2.3.3) but little is known about the rate of mixing (though bounds were given for another 3D tiling model in [RY00]).

Quitmann and Taggi have some additional important work on the 3D dimer model, which studies the behavior of loops formed by an independently sampled pair of dimer configurations [Tag22; QT22; QT23]. Among other things, they find that when one superimposes two independent random dimer tilings on an $n \times n \times n$ torus, the union of the tilings will typically contain cycles whose length has order n^3.

Throughout this paper our basic physical intuition is that the 3D dimer model describes a steady current through a non-isotropic medium, and we are studying how the current varies in space. But we stress that papers like the one by Freedman et al [FHNQ11] have other field theoretic phenomena in mind (*topological excitations*, *Majorana fermions*, *Abelian anyons*, etc.) and we will not attempt to explain these interpretations here, though we will briefly mention a gauge theoretic interpretation of the dimer model in Problem 2.9.14.

Let us also remark that the literature on *related* topics is quite large, including (to give just a few examples) work on large deviations for graph homomorphisms $h : \mathbb{Z}^d \to \mathbb{Z}$ [KMT20],

weakly non-planar dimer models [GMT20; GRT22], and a generalization of rhombus tilings to n dimensions [LMN01; Lam21; WMDB02].

2.1.6 Outline of the chapter

We establish notation and a few basic preliminaries in Section 2.2. We then illustrate the complexity of the 3D model with a brief discussion of the local move problem and related topics in Section 2.3.

In Section 2.4 we describe the ergodic Gibbs measures of *boundary mean current* (i.e., having mean current that lies on the boundary $\partial\mathcal{O}$, where $\mathcal{O}$ is the octahedron of possible mean currents). Not all Gibbs measures with boundary mean current have zero entropy, but we can still compute the entropy function ent on $\partial\mathcal{O}$ by reducing it to a two-dimensional problem (see Theorem 2.4.7).

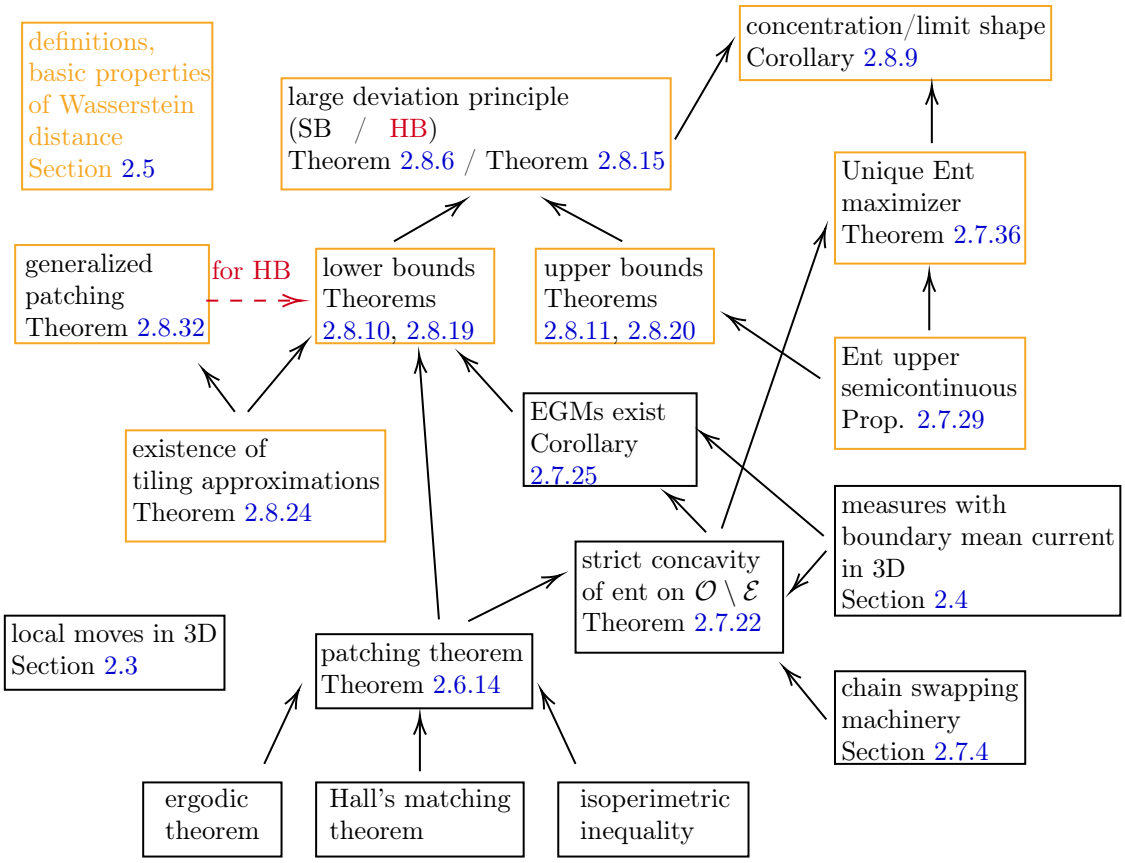
Section 2.5 is a technical section where we present some of the function-theoretic preliminaries about flows. We define scaled tiling flows, the Wasserstein metric on flows for comparing them, and asymptotic flows (which we prove are the scaling limits of tiling flows in Theorem 2.5.19). We also define boundary values for both types of flows using a *trace operator* T, and show that T is uniformly continuous as a function of the flow (Theorem 2.5.39).

In Section 2.6, we deal with the fundamental problem of how one "patches together" regions of different tilings to form one perfect matching of a large region (Theorem 2.6.14). As noted above, the key tool is *Hall's matching theorem*. We give an outline of the proof of Theorem 2.6.14 (in the "square annulus setting") in Section 2.6.3 accompanied by a sequence of two dimensional pictures. In three dimensions, Hall's matching theorem relates non-tileability to the existence of a certain type of minimal surface. The other key classical input in the proof of Theorem 2.6.14 is the isoperimetric inequality.

Section 2.7 concerns properties of the entropy functions ent and Ent, such as continuity, strict concavity, and uniqueness of maximizers. Since no exact formula for $\mathrm{ent}(s)$ is known for mean currents s in the interior of $\mathcal{O}$ this section involves interesting methods fairly different from dimension 2, in particular the chain swapping constructions in Section 2.7.4.

Section 2.8 finally ties together the ingredients of the previous sections to produce the two large deviation principles (Theorem 2.8.6 for soft boundary conditions and Theorem 2.8.15 for hard boundary conditions) which are our main results. Both of these are broken down into proving a lower bound on probabilities (Theorem 2.8.10 for soft boundary and Theorem 2.8.19 for hard boundary) and an upper bound (Theorem 2.8.11 for soft boundary and Theorem 2.8.20 for hard boundary). One of the slightly difficult parts of the paper is the explicit construction of a tiling flow approximating an asymptotic flow. This is a step in proving the lower bound which we call the "shining light" argument (Theorem 2.8.24). For the hard boundary lower bound, on top of this we also need a *generalized patching theorem* (Theorem 2.8.32) to show that any asymptotic flow can be approximated by a tiling of a fixed region. The proof of the generalized patching theorem is where we make use of the *flexible* condition on (R, b) in the hard boundary large deviation principle.

Several open problems are given in Section 2.9. See the chart below for a graphical representation of some of the dependencies and results that we highlight.



The results in orange boxes are stated using the Wasserstein metric for flows, and rely on many of its properties described in Section 2.5.

2.2 Preliminaries

As we mentioned earlier, it is sometimes convenient to represent a vertex of $\mathbb{Z}^3$ by the unit cube centered at that vertex, and to represent an edge $e = (a, b)$ of $\mathbb{Z}^3$ by the union of the two cubes centered at a and b (a domino). Both perspectives are useful for visualization, and we will use the terms *perfect matching* and *dimer tiling* somewhat interchangeably. We denote the space of dimer tilings of $\mathbb{Z}^3$ by Ω. sym]Chapter 2!Ω - the space of dimer tilings

Recall that $\mathbb{Z}^3$ is a bipartite lattice, with bipartition into even points (where the coordinate sum is even) and odd points (where the coordinate sum is odd). In a dimer tiling of $\mathbb{Z}^3$, there are six possible types of tiles corresponding to the six possible unit coordinate vectors. We denote the unit coordinate vectors by $\eta_1 = (1, 0, 0)$, $\eta_2 = (0, 1, 0)$, and $\eta_3 = (0, 0, 1)$. We denote the edge in $\mathbb{Z}^3$ connecting the origin to η_i by e_i and the edge connecting the origin to $-\eta_i$ by $-e_i$.

2.2.1 Tilings and discrete vector fields

Given a perfect matching of $\mathbb{Z}^3$, there is a natural way to associate a discrete, divergence-free vector field valued on oriented edges. We will call the flow corresponding to a tiling τ a *tiling flow*, denoted f_τ. Like height functions in two dimensions, tiling flows have well-defined

scaling limits called *asymptotic flows* (which we describe in Section 2.5). Asymptotic flows capture the broad statistics of dimer tilings in a given region. Since our main results (e.g. our large deviation principle, analogous to [CKP01]) are related to the large scale statistics of dimer tilings, they are naturally formulated in terms of tiling flows.

Let E denote the set of edges in $\mathbb{Z}^3$. A *discrete vector field* or *discrete flow* is a function from oriented edges of $\mathbb{Z}^3$ to the real numbers. Unless stated otherwise, we assume all edges are oriented from even to odd (flipping the orientation of the edge e reverses the sign of the discrete vector field on e). For a dimer tiling τ of $\mathbb{Z}^3$, we associate a discrete vector field v_τ valued on the edges $e \in E$ defined by

$$v_\tau(e) = \begin{cases} +1 & \text{if } e \in \tau, \text{ oriented even to odd} \\ 0 & \text{if } e \notin \tau \end{cases} \tag{2.4}$$

We call v_τ the *pretiling flow*. Recall that by our definition of discrete vector fields, if e is oriented odd to even, we say that $v_\tau(e) = -1$. The divergence of a discrete vector field v is given by

$$\operatorname{div} v(x) = \sum_{e \ni x} v(e) \tag{2.5}$$

where the sum is over edges e oriented away from x (e.g. if x is even, the edges in the sum are oriented from even to odd, and the opposite if x is odd). From this definition, we compute that

$$\operatorname{div} v_\tau(x) = \begin{cases} +1 & \text{if } x \text{ is even} \\ -1 & \text{if } x \text{ is odd} \end{cases}$$

Therefore v_τ itself is not divergence-free, but the divergences don't depend on τ, so we can construct a divergence-free flow corresponding to a tiling τ by subtracting a fixed reference flow r. There are lots of reasonable choices for the reference flow. For simplicity and symmetry we choose:

$$r(e) = \frac{1}{6} \qquad \text{for all edges } e \in E \text{ oriented from even to odd}$$

We can now define the *tiling flow* corresponding to a tiling τ of a region $R \subseteq \mathbb{Z}^3$.

Definition 2.2.1. Let τ be a dimer tiling of $\mathbb{Z}^3$. The divergence-free, discrete vector field corresponding to τ is $f_\tau := v_\tau - r$. We call f_τ a *tiling flow*.

$$f_\tau(e) = \begin{cases} +5/6 & \text{if } e \in \tau \\ -1/6 & \text{if } e \notin \tau \end{cases}$$

If τ is a dimer tiling of a subgraph $G \subset \mathbb{Z}^3$, we define the tiling flow by restriction.

Remark 2.2.2. In dimension 2, the analogous definition of a tiling flow f_τ also works (in this case the reference flow is $1/4$ on all edges oriented from even to odd). For every discrete flow defined on edges (whose endpoints are vertices of $\mathbb{Z}^2$) there is a dual flow on dual edges

(whose endpoints are faces of $\mathbb{Z}^2$) obtained by rotating each edge 90 degrees clockwise. If the original flow is divergence-free, then the dual flow is curl-free and is hence equal to the gradient of a function (this function is called the *height function* or *scalar potential*). It is also worth noting that there is an analog of the height function in three dimensions. Namely, since f_τ is a divergence-free flow it can be written as the curl of another flow, that is, $f_\tau = \nabla \times A$, where A is a so-called *vector potential* which is defined modulo the addition of a curl-free flow. However the set of vector potentials A is more complicated than the set of height functions (it does not have a similar lattice structure, the potentials are not uniquely defined, etc.) and is not as useful for our purposes as height functions are in two dimensions. Because of that, we do not work with the vector potential in this paper, and instead just work with the tiling flow f_τ itself.

A pair of dimer tilings $(\tau_1, \tau_2) \in \Omega \times \Omega$ is called a *double dimer tiling* or *double dimer configuration*. The double dimer model is a model of independent interest, but we mention it because it will be a tool for comparing dimer tilings. This will be used in Section 2.3, Section 2.8, and substantially in Section 2.7.

There is a natural way to associate a divergence-free discrete flow to a double dimer configuration, namely for e an edge oriented from even to odd,

$$f_{(\tau_1,\tau_2)}(e) = f_{\tau_1}(e) - f_{\tau_2}(e) = v_{\tau_1}(e) - v_{\tau_2}(e) = \begin{cases} 1 & \text{if } e \in \tau_1 \setminus \tau_2 \\ -1 & \text{if } e \in \tau_2 \setminus \tau_1 \\ 0 & \text{if } e \in \tau_1 \cap \tau_2 \text{ or if } e \notin \tau_1 \cup \tau_2. \end{cases} . \quad (2.6)$$

Unlike the tiling flow for a single tiling, the flow associated with a double dimer configuration (τ_1, τ_2) does not determine (τ_1, τ_2), since it does not specify the tiles in $\tau_1 \cap \tau_2$. However, the collection of loops formed by $\tau_1 \cup \tau_2$ (including the double tiles) and the flow $f_{(\tau_1,\tau_2)}$ together determine (τ_1, τ_2). See Section 2.7.3 for more about double dimer flows.

2.2.2 Measures on tilings and mean currents

Recall that Ω denotes the set of dimer tilings of $\mathbb{Z}^3$. The group $\mathbb{Z}^3$ acts naturally on Ω by translations, namely given $x \in \mathbb{Z}^3$ and $\tau \in \Omega$, $\tau + x$ is the tiling where $(a, b) \in \tau$ if and only if $(a + x, b + x) \in \tau + x$. There is natural topology on Ω induced by viewing it as a subset of $\{0, 1\}^E$ and giving the latter the product topology over the discrete set $\{0, 1\}$ (recall that E denotes the edges of $\mathbb{Z}^3$). This makes Ω a compact metrizable space and the translation action on it continuous.

Let $\mathbb{Z}^3_{\text{even}}$ denote the set of even vertices in $\mathbb{Z}^3$. We define $\mathcal{P}(\Omega) = \mathcal{P}$ to be the space of Borel probability measures on Ω invariant under the action of $\mathbb{Z}^3_{\text{even}}$.

To explain why we look at $\mathbb{Z}^3_{\text{even}}$-invariant measures instead of $\mathbb{Z}^3$-invariant measures, recall that $\mathbb{Z}^3$ is a bipartite lattice, with bipartition consisting of even points and odd points. In the interpretation of a dimer tiling as a flow from in Section 2.2.1, the sign of the flow on an edge oriented parallel to $(1, 0, 0)$ (for example) depends on whether the edge starts at an even point or an odd point. E.g. consider the tiling

$$\tau = \{(x, x + (1, 0, 0)) \ : \ x \in \mathbb{Z}^3 \text{ is even}\}.$$

The flow associated to τ moves current (on average) in the direction $(1,0,0)$, while the flow for $\tau + (1,0,0)$ moves current (on average) in the direction $(-1,0,0)$. We want to our measures to be invariant under an action that preserves the asymptotic direction of the flow associated to a tiling, and this is why we consider $\mathbb{Z}^3_{\text{even}}$-invariant measures instead of $\mathbb{Z}^3$-invariant measures.

The *ergodic measures* $\mathcal{P}_e$ are the extreme points of the convex set $\mathcal{P}$. A good reference for basic ergodic theory suitable for our purposes is [Kel98]. Any invariant measure $\mu \in \mathcal{P}$ can be decomposed in terms of ergodic measures, i.e. there exists a measure w_μ on $\mathcal{P}_e$ such that

$$\mu = \int_{\mathcal{P}_e} \nu \, \mathrm{d}w_\mu(\nu).$$

The measures ν in the support of w_μ are called the *ergodic components* of μ. Sampling from μ can be viewed as first sampling an ergodic component ν from w_μ and then sampling from ν.

We will also frequently make use of the so-called *uniform Gibbs measures* on Ω defined as follows: a measure $\mu \in \mathcal{P}$ is a *uniform Gibbs measure* if for any finite set $R \subset \mathbb{Z}^3$, we can say that *given* that τ contains no edges that cross the boundary of R, and *given* the tiling τ induces on $\mathbb{Z}^3 \setminus R$, the μ *conditional law* of the restriction of τ to R is the uniform measure on dimer tilings of R. We will see in the next section that the measures that maximize specific entropy are uniform Gibbs measures. Throughout the rest of the paper, we refer to uniform Gibbs measures simply as *Gibbs measures*.

A useful reference for Gibbs measures is [Geo11]. We denote the set of $\mathbb{Z}^3_{\text{even}}$-invariant Gibbs measures by $\mathcal{P}_G$ and the set of ergodic Gibbs measures (EGMs) by $\mathcal{P}_{G,e}$. A useful fact throughout is that the ergodic components of $\mathbb{Z}^3_{\text{even}}$-invariant Gibbs measures are themselves $\mathbb{Z}^3_{\text{even}}$-invariant Gibbs measures.

Proposition 2.2.3. *[Geo11, Theorem 14.15] The ergodic components of an invariant Gibbs measure are ergodic Gibbs measures almost surely.*

We remark that the analogous constructions work for *weighted Gibbs measures*. For example, one may assign weights $a_1, a_2, a_3, a_4, a_5, a_6$ to the six possible tile orientations. A $\mathbb{Z}^3_{\text{even}}$-invariant Gibbs measure μ with these weights is a measure where for any finite set R, the conditional law of μ given a tiling of $\mathbb{Z}^3 \setminus R$ is the one in which each tiling of R has probability proportional to $\prod_{i=1}^6 a_i^{N_i}$, where N_i is the number of tiles of weight a_i. We expect that our main results could be extended to weighted dimer models (and perhaps also models with weights that vary by location in a periodic way as in [She05; KOS06]) but for simplicity we focus on the unweighted case here.

A key invariant of a $\mathbb{Z}^3_{\text{even}}$-invariant measure is a quantity called the *mean current* which (as mentioned in the introduction) represents the expected current flow *per even vertex*. This is a generalization of the notion of *height function slope* from two dimensions. Recall that η_1, η_2, η_3 denote the standard basis for $\mathbb{Z}^3$, the edge connecting the origin with η_i is denoted by e_i, and the edge connecting the origin with $-\eta_i$ is denoted by $-e_i$.

Definition 2.2.4. The *mean current* of a measure $\mu \in \mathcal{P}$, denoted $s(\mu)$, is an element of $\mathbb{R}^3$ such that its i^{th}-coordinate is

$$(s(\mu))_i = \mu(e_i \in \tau) - \mu(-e_i \in \tau).$$

Note that the mean current is an affine and continuous function of the measure. The mean current is invariant under the action of $\mathbb{Z}^3_{\text{even}}$ and takes values in

$$\mathcal{O} = \{s \in \mathbb{R}^3 \ : \ |s_1| + |s_2| + |s_3| \leq 1\}$$

which we call *the mean current octahedron*.

There are a few other useful formulations of the mean current. We define the function $s_0 : \Omega \to \mathbb{R}^3$ to be the direction of the tile at the origin in τ. Then the mean current can be computed as an expected value of s_0:

$$s(\mu) = \int_\Omega s_0(\tau)\, \mathrm{d}\mu(\tau). \tag{2.7}$$

Similarly let $\Lambda_n = [-n, n]^3$, and let $\text{even}(\Lambda_n)$ denote the even points in Λ_n. We define the function

$$s_n(\tau) = \frac{1}{\text{even}(\Lambda_n)} \sum_{x \in \text{even}(\Lambda_n)} s_0(\tau + x). \tag{2.8}$$

The function $s_n(\tau)$ measures the average tile direction of τ in the box Λ_n. By $\mathbb{Z}^3_{\text{even}}$-invariance,

$$s(\mu) = \int_\Omega s_n(\tau)\, \mathrm{d}\mu(\tau). \tag{2.9}$$

We let $\mathcal{P}^s$ denote the space of $\mathbb{Z}^3_{\text{even}}$-invariant probability with mean current s. Adding the subscripts G and e will denote whether the measure is a Gibbs measure and whether it is ergodic with respect to the $\mathbb{Z}^3_{\text{even}}$ action.

2.2.3 Entropy

As is common in statistical physics models, entropy plays an important role in the large deviation principle for dimer tilings in 3D. There are a few different functions that we refer to as "entropy" (of a probability measure with finite or infinite support, of a mean current, of an asymptotic flow). Here we give some definitions and explain how these notions of entropy are related to each other. The primary reference for this section is also [Geo11].

For a probability measure ν with finite support S, its *Shannon entropy*, denoted $H(\nu)$, is

$$H(\nu) = -\sum_{\sigma \in S} \nu(\sigma) \log \nu(\sigma).$$

For a $\mathbb{Z}^3_{\text{even}}$-invariant probability measure μ with infinite support, we can define the *specific entropy* of μ as a limit of Shannon entropy per site. Given a finite region $\Lambda \subset \mathbb{Z}^3$, let $\Omega(\Lambda)$ denote the dimer tilings of Λ (i.e. tilings of $\mathbb{Z}^3$ restricted to Λ, so tiles are allowed to have one cube outside Λ). For $\sigma \in \Omega(\Lambda)$, define

$$X(\sigma) = \{\tilde{\sigma} \in \Omega : \tilde{\sigma} \mid_\Lambda = \sigma\}$$

and then

$$H_\Lambda(\mu) := -\sum_{\sigma \in \Omega(\Lambda)} \mu(X(\sigma)) \log \mu(X(\sigma)).$$

Let $\Lambda_n = [-n, n]^3$ be a sequence of growing cubes. If μ is a $\mathbb{Z}^3_{\text{even}}$-invariant probability measure on Ω, the *specific entropy* of μ, denoted $h(\mu)$, is

$$h(\mu) := \lim_{n \to \infty} |\Lambda_n|^{-1} H_{\Lambda_n}(\mu).$$

This limit exists because the terms form a subadditive sequence. In fact, one can also show that

$$h(\mu) = \inf_{\Lambda \in \mathcal{S}} |\Lambda|^{-1} H_\Lambda(\mu),$$

where $\mathcal{S}$ is the set of all possible finite regions in $\mathbb{Z}^3$. See [Geo11, Theorem 15.12]. As a function of the measure, $h(\cdot)$ is affine and upper semicontinuous [Geo11, Proposition 15.14].

The reason that Gibbs measures (introduced in the previous section) play a special role in our study is the *variational principle* which says that a measure $\mu \in \mathcal{P}$ maximizes $h(\cdot)$ if and only if μ is a Gibbs measure. This is a classical result going back to [LR69], see [Geo11, Theorem 15.39] for exposition.

The *local* or *mean-current entropy function* $\text{ent} : \mathcal{O} \to \mathbb{R}$ is defined

$$\text{ent}(s) = \max_{\mu \in \mathcal{P}^s} h(\mu).$$

This function is the main focus of Section 2.7, where we show it has a number of useful properties (continuity, concavity) and show that the maximum is always realized by an ergodic Gibbs measure of mean current s. In Theorem 2.4.7 we compute its restriction to $\partial\mathcal{O}$ by relating it to the analogous local entropy function for lozenge tilings in two dimensions.

We conclude this section with one more use of the term entropy. In Section 2.5, we will show that the "fine-mesh limits" of rescaled tiling flows are precisely the measurable vector fields we call *asymptotic flows*. Asymptotic flows are valued in $\mathcal{O}$ and supported on some compact region R. The *entropy of an asymptotic flow* g can then be defined as

$$\text{Ent}(g) = \frac{1}{\text{Vol}(R)} \int_R \text{ent}(g(x)) \, \mathrm{d}x.$$

2.3 Local moves

A number of the papers about the 3D dimer model are about local moves. Here we present some simple examples, briefly review the literature, and explain why local move connectedness fails for the torus in dimensions $d > 2$. Most of the ideas in this section are already known, but we include a few elementary observations we have not seen articulated elsewhere.

This section can be skipped on a first read, since the results are not essential for the rest of the paper. However, it is useful for understanding some of the ways that the $d = 2$ problem differs from the $d = 3$ problem (e.g., why the Kasteleyn determinant approach to computing entropy does not work in the same way) and also what makes $d = 3$ different from $d > 3$ (e.g., the integer-valued twist function is indexed by $\mathbb{Z}$ when $d = 3$ and by $\mathbb{Z}/2\mathbb{Z}$ when $d > 3$). This section will also explain how the figures in the introduction were generated.

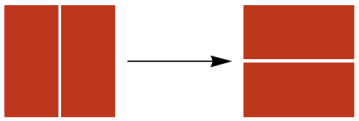


Figure 2.15: A local move or flip in two dimensions.

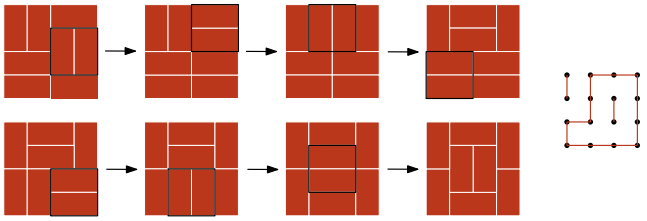


Figure 2.16: (1) an example of a sequence of local moves transforming one tiling into another and (2) the collection of cycles from overlaying the first and last tilings in this sequence.

2.3.1 Local moves in two dimensions

In two dimensions, a *local move* or *flip* is the operation of choosing a pair of parallel dimers in the tiling, and switching them out for the other pair. See Figures 2.15 and 2.16. Let R be a subgraph of $\mathbb{Z}^2$ and let $\mathcal{T}(R)$ be a graph on the set of dimer tilings of R where two tilings τ and τ' are connected by an edge if they differ by a single flip. It is shown using height functions in [Thu90] that if $R \subset \mathbb{Z}^2$ is simply connected and finite, then any two dimer tilings of R differ by a finite sequence of flips. In other words, $\mathcal{T}(R)$ is a connected graph.

Local move connectedness in the 2D dimer model means that it is possible to probe all tilings of a region using simple local updates, and this is useful for both theoretical and practical purposes. It means that uniformly random dimer tilings in 2 dimensions can be simulated using Markov chain Monte Carlo methods called *Glauber dynamics*. For the 2D dimer model, one can give an explicit polynomial bound on the mixing time of this algorithm [RT00].

2.3.2 Local moves in three dimensions

The same local moves (flips) make sense for the 3D dimer model, but local move connectedness with these manifestly fails, even for very small regions. There is a simple counterexample on the $3 \times 3 \times 2$ box which is called a *hopfion* in the physics literature (see Figure 2.17). Note that the hopfion has no parallel pairs of tiles, so it is not connected under flips to any other tiling of the $3 \times 3 \times 2$ box.

There is a series of papers by Fiere, Milet, Klivans, and Saldanha studying local move connectedness in dimension three under flips and *trits*, a new local move in three dimensions involving three tiles (see Figure 2.17). In [MS15; MS14b] they show that any two tilings of a region of the form $D \times [0, 1]$ where D is simply connected and planar are connected under flips and trits. In subsequent works [MS14a; Sal22; Mil15; FKMS22; Sal21] they introduce and study an invariant called the *twist*, related to the linking number or writhing number.

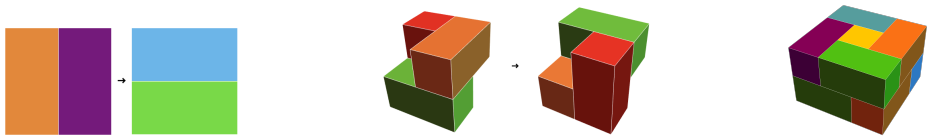


Figure 2.17: A flip, a trit, and a flip-rigid configuration called a *hopfion*.

We will present below a brief and informal overview of the various ways the twist is defined and how it is related to a linking number. More detailed exposition is found e.g. in [Sal20] or the references above.

Given two distinct smooth curves $\gamma_1, \gamma_2 : S^1 \to \mathbb{R}^3$ embedded in $\mathbb{R}^3$, one can compute their integer-valued *linking number* $L(\gamma_1, \gamma_2)$ by projecting them to a generic plane and summing the signatures of the crossings. (Recall that signature of a crossing of two oriented paths is 1 or -1 depending on whether the upper curve crosses the lower curve from right to left or left to right, when the bottom curve is viewed as being oriented from down to up.) It is a standard result that this number is independent of the plane one projects onto, see e.g. [Ada94, pages 20-21]. (The idea is to show that any one projection can be transformed into another by a sequence of Reidemeister moves, and that these moves preserve the linking number.) The linking number can also be computed with an integral formula: if r_1, r_2 are parametrizations of γ_1, γ_2, then

$$L(\gamma_1, \gamma_2) = \frac{1}{4\pi} \oint_{\gamma_1} \oint_{\gamma_2} \frac{r_1 - r_2}{|r_1 - r_2|^3} \mathrm{d}r_1 \times \mathrm{d}r_2.$$

Informally, this represents the line integral along γ_1 of the magnetic field generated by a steady current through γ_2. One can analogously compute a "linking number" of a pair of tilings in a box by summing crossings. Namely, imagine that each edge in the matching is extended $\epsilon > 0$ units in either direction. Then the crossing number is obtained by flattening these extended edges to a horizontal plane and summing the signatures of the crossings. To be more precise, we say two edges (a, b) and (c, d) constitute a crossing if their orientations are both orthogonal to the vertical (third-coordinate) direction and orthogonal to each other *and* one of the endpoints of (a, b) differs from one of the endpoints of (c, d) in the vertical coordinate and in no other coordinate. This is the same as an ordinary crossing if we assume each edge is extended ϵ units beyond its endpoints, and the sign of the crossing is defined in the usual way. We can define the linking of τ_1 and τ_2 to be the signed sum $L(\tau_1, \tau_2)$ of all crossings involving a tile in τ_1 and a tile in τ_2. This is a quadratic form, and the twist of a tiling τ is defined by $T(\tau) = \frac{1}{4}L(\tau, \tau)$. This decomposes as a sum over pairs of horizontal tiles in vertical columns. For reasonable regions (i.e., $D \times [1, N]$, where $D \subset \mathbb{Z}^2$ is simply connected and N is even), the twist is integer-valued despite the $\frac{1}{4}$ and is independent of the direction for the orthogonal projection [MS14a, Proposition 6.4]. Within a rectangular box, one can easily show that trits increment the twist and flips leave the twist unchanged (in fact this holds for any region of the form $D \times [1, N]$, [MS14a, Theorem 1]). There are also examples of tilings with twist $T(\tau) = 0$ that are not connected under flips alone ([Sal21, Figure 7]), meaning that $T(\tau) = T(\sigma)$ does not imply that τ, σ are connected under flips.

Simple questions about local move connectedness under flips and trits still remain open, for example it is not known whether all tilings of an $M \times N \times L$ box are connected under

flips and trits when $M, N, L > 2$ (see Problem 2.9.1). See [MS] for an enumeration of all tilings of the $4 \times 4 \times 4$ box, which shows explicitly that all tilings of this region are connected under flips and trits.

In dimensions $d > 3$, Klivans and Saldanha [KS22] show that the twist is valued in $\mathbb{Z}/2$. In dimension $d = 4$, even tilings of the $2 \times 2 \times 2 \times 2$ box fail to be connected under flips (see [KS22, Example 2.2]). They also show that tilings within certain larger boxes are "almost" connected under flips, i.e. they can be connected if the boxes are extended in some way.

The works of Friere, Klivans, Milet and Saldanha rely mostly on geometric and algebraic constructions to study local move problems in dimensions $d \geq 3$, but the recent work [HLT23] by Hartarsky, Lichev, and Toninelli (written concurrently with this paper) makes progress using purely combinatorial arguments. In particular it follows from their results that any tiling of a rectangular box in $\mathbb{Z}^3$ which is tileable by dimers admits at least one flip or trit [HLT23, Theorem 3], providing a partial answer to Problem 2.9.1 in Section 2.9.

In fact, [HLT23, Theorem 3] is a statement that holds for any dimension $d \geq 2$. It states that any tiling τ of a rectangular box in $\mathbb{Z}^d$ of dimensions $(n_1, \ldots, n_d)$ which is tileable by dimers contains a copy of $[0,1]^d$ such that τ restricted to this copy of $[0,1]^d$ contains at least $2^{d-2} + 1$ dimers. Specialized to the case $d = 3$, this means that there is a copy of $[0,1]^3$ which completely contains at least three dimers from τ, and the only way this can happen is if $[0,1]^3$ contains tiles which make up a flip or a trit in τ. The main idea of the proof is a clever but simple counting argument. Following the ideas in [HLT23], we present a slight modification of their proof specialized to the $d = 3$ case, with the aim of just showing the flip/trit result.

Proposition 2.3.1 ([HLT23])**.** *Let $R = [1, n_1] \times [1, n_2] \times [1, n_3] \subset \mathbb{Z}^3$ with $n_1, n_2, n_3 \geq 2$ and $n_1 n_2 n_3$ even. Any tiling τ of R admits at least one flip or trit.*

Proof. Fix a tiling τ of R. We view τ as a tiling of the torus with the same dimensions (i.e., τ is a tiling of the torus such that no dimers cross the identifications). On one hand, τ contains $n_1 n_2 n_3/2$ tiles, and each tile is contained in exactly four translates of $[0,1]^3$. On the other hand, there are $n_1 n_2 n_3$ possible choices of translates of the unit cube in the torus, so the average number of tiles per unit cube is 2.

If a unit cube contains an above-average number of tiles from τ, it contains at least three tiles. If this unit cube is in the interior of R, or is cut in half by only one face of R, then since the tiles in τ do not cross the identifications, this implies there is a flip or trit in τ as a tiling of R. The result then follows by showing that the unit cubes which are cut into four pieces along the edges (or eight pieces at the corner) by the identifications of the torus contain a below-average number of tiles from τ.

The number of such "edge unit cubes" is $(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + 1 = n_1 + n_2 + n_3 - 2$. Any dimer contained in an edge unit cube must be contained along one of the edges around R. The number of vertices in the edges around R is $4(n_1 + n_2 + n_3) - 16$ (there are 8 corners, but each one is contained in three edges), hence the maximum number of dimers contained in this region is $2n_1 + 2n_2 + 2n_3 - 8$. Given this, the average number of dimers in τ per edge unit cube is bounded by

$$\frac{2n_1 + 2n_2 + 2n_3 - 8}{n_1 + n_2 + n_3 - 2} < 2.$$

Therefore there must be a non-edge unit cube containing at least three tiles from τ, which completes the proof. □

For the hypercube $[0,1]^d \subset \mathbb{Z}^d$, Hartarsky, Lichev, and Toninelli also show that for $d \geq 3$, the connected components of the graph on dimer configurations of $[0,1]^d$ under local moves of length up to $d-1$ (here the trit is a move of length three and the flip is a move of length two) have size exponential in d [HLT23, Theorem 5], and that for $d \geq 2$, any two dimer tilings of $[0,1]^d$ are connected by a sequence of moves of length $\leq 2(d-1)$ [HLT23, Theorem 6]. For $[0,n]^d \subset \mathbb{Z}^d$, $d \geq 2$, n odd, they show that the diameter of the graph on dimer configurations of $[0,n]^d$ under local moves of length $\leq \ell$ is at least $n^{d-1}(n^2-1)/(6\ell^2)$ [HLT23, Theorem 7].

Flip connectedness has also independently been studied in the physics literature, from the perspective of looking for "topological invariants" preserved by flips. In [FHNQ11] the authors define a "Hopf number" for dimer tilings of $\mathbb{Z}^d$ valued in $\pi_d(S^{d-1})$ which is invariant under flips. The hopfion (see Figure 2.17) has Hopf number ± 1 (depending on its orientation). This construction works for any dimension $d \geq 2$. The fact that $\pi_2(S^1) = 0$ corresponds to no obstruction to connectedness under flips, and $\pi_3(S^2) = \mathbb{Z}$ corresponds to there being at least countably many connected components under flips in dimension 3. For all $d > 3$, $\pi_d(S^{d-1}) = \mathbb{Z}/2$, implying at least two connected components under flips.

In [Bed19a] it is shown in examples that the Hopf number from [FHNQ11] can be computed using discrete versions of Cherns-Simon integral formulas for the Hopf number applied to a version of the tiling flow and its vector potential. See also [Bed19b].

Remark 2.3.2. The failure of local move connectedness in three dimensions is intimately related to the failure of (at least a straightforward generalization) of Kasteleyn theory.

In two dimensions, the partition function for dimer tilings of a simply connected planar graph can be computed as the Pfaffian of an adjacency matrix of the directed graph with appropriate weights (this can also be done with a determinant when the graph is bipartite). Recall that if $M = (m_{ij})$ is an $2n \times 2n$ skew-symmetric matrix,

$$\mathrm{Pf}(M) = \frac{1}{2^n n!} \sum_{\sigma \in S_{2n}} \mathrm{sign}(\sigma) \prod_{i=1}^{n} m_{\sigma(2i-1),\sigma(2i)}.$$

There are two key observations in two dimensions. First, the weights can be chosen so that the term is ± 1 if and only if the pairing $\{\sigma(2i-1), \sigma(2i)\}_{1 \leq i \leq n}$ corresponds to a dimer tilings and otherwise it is 0. By this, it is clear that the partition function can be computed as a *permanent* (i.e., like the above without the sign terms). The second key observation, which is why this reduces to a Pfaffian computation, is that the weights can be chosen so that applying a flip does not change the sign of the term. From here, flip connectedness in two dimensions shows that the Pfaffian is counting tilings.

In three dimensions it is still possible to choose weights so that a term is ± 1 if and only if it corresponds to a dimer tiling, and all other terms are 0. Choosing certain weights such that flips do not change the sign, it is observed in [FHNQ11] that the Hopf number invariant mod 2 is equal to the sign of the term in the Pfaffian (and one can check that the trit increments this number). From this they note that if M is defined analogously to in two dimensions, then in 3D

$$\mathrm{Pf}(M) = A - B$$

where $A + B$ would be the partition function. The term A counts tilings with Hopf number $0 \mod 2$ and B counts tilings with Hopf number $1 \mod 2$.

In [KS22], the number $A - B$ is called the *defect*. The definition of the *twist invariant* discussed above is extended to dimensions $d > 3$ as the sign of the appropriate Kasteleyn determinant [KS22, Definition 3.1].

One can check by enumerating the equations for a single cube (i.e. 12 edges) that it is not possible to choose 12 nonzero weights so that the six flips (corresponding to σ with sign -1) and four trits (corresponding to σ with sign $+1$) contained in the cube all preserve the sign of the term in the Pfaffian. In fact the six flip equations plus one trit equation have no simultaneous solution with all nonzero weights.

More generally, there is a complete characterization of which graphs admit *Pfaffian weights* and thereby make it possible to compute the partition function (which is a priori a permanent) as a determinant or Pfaffian of a re-weighted matrix. It is shown that a bipartite graph G admits Pfaffian weights if and only if it does not "contain" $K_{3,3}$ [Lit75]. Here "contain" means G can be modified (by replacing a collection of disjoint paths of edges containing an even number of vertices with a single edges) to a graph H which has $K_{3,3}$ as a subgraph. One can see that in this sense $\mathbb{Z}^3$ contains $K_{3,3}$, and hence does not have Pfaffian weights. The class of graphs that have Pfaffian weights can also be described in a way so that the Pfaffian is computable by a polynomial-time algorithm [RST99].

2.3.3 Loop shift Markov chain for uniform sampling

In two dimensions, uniformly random dimer tilings of finite simply connected regions can be efficiently simulated by a Markov chain that generates random flips, see [RT00]. As we have seen, dimer tilings of topologically trivial finite regions in dimensions $d > 2$ are not connected under flips, and it is an open question even for very simple regions whether flips and trits are sufficient. Here we describe a different, non-local Markov chain method to generate uniform random dimer tilings. The algorithm works in any dimension and for regions that are not simply connected, and is how the simulations in the introduction are generated. The simple move executed at each step of our chain is to construct a "random loop" in the given dimer tiling, and "shift" the tiles along the loop. This is a well-known construction in computer science, see for instance [Bro86, Section 3]. In the physics literature, see also [HKMS03] for Monte Carlo simulations of dimers in three dimensions based on algorithms from [KM03; DK95].

Given a dimer tiling τ of a finite region $R \subset \mathbb{Z}^3$, a *loop* γ in τ is a sequence of distinct edges $e_0, e_1, \ldots e_{k-1} \in \tau$ where the odd vertex of e_i is adjacent to the even vertex of e_{i+1} for all $i \in \mathbb{Z}/k\mathbb{Z}$ for some $k \geq 2$. A *loop shift* of τ along γ is a move which replaces edges along γ by their complementary edges. Specifically the resulting tiling is

$$\tau' = (\tau \setminus \{e_0, e_1, \ldots e_{k-1}\}) \cup \{f_0, f_1, \ldots, f_{k-1}\}$$

where $\{e_0, e_1, \ldots e_{k-1}\} \cup \{f_0, f_1, \ldots, f_{k-1}\}$ form a loop in $\mathbb{Z}^3$. Since R is finite, given any two dimer tilings τ, σ of R the double dimer tiling (τ, σ) is a finite collection of double edges and loops $\gamma_1, \ldots \gamma_n$ of finite length. Loop shifting τ along γ_i for each of these transforms τ into σ. In particular, we have shown that

Proposition 2.3.3. *Let τ and σ be tilings of a finite set $R \subset \mathbb{Z}^3$. Then τ can be transformed into σ by a finite sequence of loop shifts.*

Loop shift Markov chain M. Given that any two tilings of a finite region $R \subset \mathbb{Z}^3$ differ by a finite sequence of loop shifts, we define a Markov chain M where one step proceeds as follows:

- Start with some dimer tiling τ of the region R.
- Sample an odd vertex in R uniformly at random. Start a path by following the tile from τ at this point.
- Uniformly at random choose a direction (other than the one we came from), and move in that direction for the next step.
- Repeat this (following the tile from τ, then following a uniform random choice, etc) until the path hits itself to form a loop. Call the loop γ.
- Drop any initial segment of the path which is not part of the loop γ. Then shift along γ, switching the tiles from τ for the random choices that we made along the path, and replace τ with σ which differs from τ only along γ.

By Proposition 2.3.3, M is an irreducible Markov chain and hence has a stationary distribution π. A bound on the mixing time of M is not known, see Problem 2.9.2.

Theorem 2.3.4. *The stationary distribution π of M is the uniform distribution on dimer tilings of R.*

Proof. Let P be its transition matrix. It is sufficient to prove that P is symmetric. If τ, σ are tilings such that $P(\tau, \sigma) \neq 0$, then they differ along a single loop γ.

Suppose that λ is a connected alternating-tile path in τ which consists of an initial segment α plus the loop γ. $P(\tau, \sigma)$ is a sum of the probability of paths λ of this form. We will show that the probability of generating λ in τ is the same as the probability of generating λ' in σ, where λ' has the same initial segment as λ, then traverses γ with the reverse orientation.

Let $v_1, ..., v_{2n}$ be the vertices along λ. Note that the vertices with odd index are odd, and out of these we follow a tile from τ. The vertices with even index are even, and out of these we follow a random choice. Thus the probability of generating the path λ in τ is $\prod_{k=1}^{n} \frac{1}{\deg(v_{2k})-1}$.

The sequence of vertices along λ' is the same, just in a different order. However the even vertices are still the sites where we make a random choice of direction to follow, so the probability of generating the path λ' in σ is also $\prod_{k=1}^{n} \frac{1}{\deg(v_{2k})-1}$.

Hence $P(\tau, \sigma) = P(\sigma, \tau)$. □

2.3.4 Local move connectedness on the torus and k-Gibbs measures

Here we discuss local move connectedness for dimer tilings of the torus, which is not simply connected. For any tiling τ of the d-dimensional torus $\mathbb{T}^d$, there is a standard, natural way to associate a homology class $[a(\tau)] \in H_1(\mathbb{T}^d)$. For each $i = 1, ..., d$, let P_i be any plane with normal vector η_i, the i^{th} unit coordinate vector. Let $\mathbb{T}^d_n$ denote the $n = n_1 \times n_2 \times ... \times n_d$ torus in dimension d. Without loss of generality, n_1 is even. Let τ_0 be the tiling of $\mathbb{T}^d_n$ where all tiles $t \in \tau_0$ are of the form $t = ((2x, y, z), (2x + 1, y, z))$. With slight abuse of notation, we write e to mean both an edge in $\mathbb{Z}^d$ and the unit coordinate vector parallel to e oriented even to odd, and we let $v_\tau(p) = v_\tau(e)e$ for the edge e incident to p containing a dimer. For $i = 1, \ldots, d$, we define

$$a_i(\tau) = \sum_{p \in P_i \cap \mathbb{T}^d_n} \langle v_\tau(p), \eta_i \rangle - \langle v_{\tau_0}(p), \eta_i \rangle = \sum_{p \in P_i \cap \mathbb{T}^d_n} \langle v_\tau(p), \eta_i \rangle.$$

Since $v_\tau - v_{\tau_0}$ is divergence-free, this is independent of the choice of plane P_i normal to η_i. The second equality follow from the fact that v_{τ_0} contributes 0 to the overall sum. The homology class of τ is

$$[a(\tau)] := [a_1(\tau), ..., a_d(\tau)] \in H_1(\mathbb{T}^d) \simeq \mathbb{Z}^d.$$

Note that a parallel pair of tiles contributes 0 total flow across any coordinate plane intersecting it. In particular, in any dimension $d > 1$, flips cannot change the homology class of a tiling of $\mathbb{T}^d$. However, when $d = 2$ the homology class is the only obstruction: if τ, τ' are tilings of an $n_1 \times n_2$ torus $\mathbb{T}^2_{n_1,n_2}$ and $[a(\tau)] = [a(\tau')]$, then τ, τ' are connected by a finite sequence of flips.

In dimension $d = 3$, the story is very different. In fact:

Proposition 2.3.5. *There is no finite collection of local moves that can connect all homologically equivalent dimer tilings of* $\mathbb{T}^3$.

Remark 2.3.6. The authors of [FKMS22] exhibited a tiling of the $8 \times 8 \times 4$ torus with no flips or trits, obtained by stacking horizontal brickwork patterns of different orientations. We use similar stacked brickwork patterns (but with thicker layers) in our proof of Proposition 2.3.5.

Proof. The fundamental example is the following. Let τ be a tiling of $\mathbb{T}^3_{n_1,n_2,4}$ where the first layer is an η_1 brickwork tiling, the second layer is an η_2 brickwork tiling, the third layer is a $-\eta_1$ brickwork tiling, and the fourth layer is a $-\eta_2$ brickwork tiling. By construction, $[a(\tau)] = (0, 0, 0)$. On the other hand, τ_0 also has $[a(\tau_0)] = (0, 0, 0)$, so τ and τ_0 are homologically equivalent. On the other hand, the length of the shortest alternating-tile loop in τ is $\min\{n_1, n_2, 4\}$. To see this, note that if the loop is homologically non-trivial, it must be long enough to visit at least three different horizontal layers. If it is homologically trivial, then its length must be at least $\min\{n_1, n_2, 4\}$.

More generally, for any $n = (n_1, n_2, 4n_3)$, we can construct a tiling τ of $\mathbb{T}^3_n$ which has n_3 layers of η_1 brickwork, followed by n_3 layers of η_2 brickwork, n_3 layers of $-\eta_1$ brickwork, and n_3 layers of $-\eta_2$ brickwork. Again $[a(\tau)] = (0, 0, 0)$, however the shortest contractible loop in τ has length $4n_3$ (since, again, it has to visit at least three different brickwork patterns). Therefore to connect τ, τ_0 we need loops of length at least $\min\{n_1, n_2, 4n_3\}$. These dimensions can be arbitrarily large, so this completes the proof. □

By lifting this construction from $\mathbb{T}^3$ to $\mathbb{R}^3$, we get the following corollary:

Corollary 2.3.7. *There is no finite collection of local moves which connects any two tilings of $\mathbb{Z}^3$ which differ at only finitely many places.*

Proof. Fix an integer $n > 0$. Tile all of $\mathbb{Z}^3$ with alternating brickwork layers so that there are n layers of η_1 brickwork, n layers of η_2 brickwork, n layers of $-\eta_1$ brickwork, and n layers of $-\eta_2$ brickwork. We denote the resulting tiling of $\mathbb{Z}^3$ by τ_n.

The shortest length of a cycle in τ_n is $4n$. Since there are finite cycles in τ_n, there exist tilings σ which differ from τ_n at only finitely many places. On the other hand, we need a local move of length at least $4n$ to make any change to τ_n. Since n is arbitrary this completes the proof. □

Another interesting observation can be made from the example used in the proof of Proposition 2.3.5. A measure μ is *k-Gibbs* if for any box B with side length k, it holds that conditioned on a tiling τ of $\mathbb{Z}^3 \setminus B$, μ is the uniform measure on tilings σ of B extending τ. If a measure is k-Gibbs for all k, then it is Gibbs.

In two dimensions, any two tilings of a $k \times k$ box (with the same boundary condition) are connected by some finite sequence of flips. Therefore if a measure on dimer tilings of $\mathbb{Z}^2$ is 2-Gibbs, then it is k-Gibbs for all k and hence Gibbs. The analogous statement does not hold in three dimensions.

Proposition 2.3.8. *For any integer $k \geq 2$ there exist k-Gibbs measures on Ω which are not Gibbs measures.*

Proof. Take $n = (n_1, n_2, n_3, n_4)$ and consider the tiling of $\mathbb{Z}^3$ which alternates between n_1 layers of η_1 bricks, n_2 layers of η_2 bricks, n_3 layers of $-\eta_1$ bricks, and n_4 layers of $-\eta_2$ bricks. Define a measure μ_m by averaging over translations by $\mathbb{Z}^3_{\text{even}}$ in the $m \times m \times m$ box and let μ be a subsequential limit as $m \to \infty$. The measure μ is invariant under the action of $\mathbb{Z}^3_{\text{even}}$. For $k \leq \min\{n_1, n_2, n_3, n_4\}$, μ is k-Gibbs since within any size k cube, a tiling sampled from μ is frozen for $k \leq \min\{n_1, n_2, n_3, n_4\}$. For $k > \min\{n_1, n_2, n_3, n_4\}$, μ still a.s. samples tilings which are brickwork patterns restricted to horizontal layers. However tilings of these larger boxes are not frozen, and are connected by shifting on finite loops to tilings which are not brickwork on every layer. Therefore μ is not k-Gibbs for $k > \min\{n_1, n_2, n_3, n_4\}$, hence μ is not Gibbs. □

The construction in the proof works to construct a k-Gibbs-but-not-Gibbs measure for any mean current $s = (s_1, s_2, 0)$. A more complicated construction allows us to show that there exist k-Gibbs measures which are not Gibbs and correspond to an s in the interior of $\mathcal{O}$ for which $s_1 s_2 s_3 \neq 0$. (Essentially one can arrange a periodic pattern of infinite non-intersecting taut paths like the ones shown in the introduction.) We have not found a construction that works for every $s \in \mathcal{O}$.

2.4 Measures with boundary mean current

Recall from Section 2.2.2 that $\mathbb{Z}^3_{\text{even}}$-invariant measures on dimer tilings of $\mathbb{Z}^3$ come with a parameter called the *mean current*. This definition makes sense in any dimension d. When

$d = 2$, the mean current is a 90-degree rotation of the height function *slope*, and in general it is valued in the convex polyhedron

$$\mathcal{O}_d = \{(s_1, ..., s_d) : |s_1| + ... + |s_d| \leq 1\}.$$

Recall that the mean current of a measure μ is defined in terms of tile densities (Definition 2.2.4). Given a standard basis $\eta_1, \eta_2, \ldots, \eta_d$ of $\mathbb{Z}^d$ denote by e_i the edge connecting 0 with η_i and $-e_i$ the edge connecting 0 with $-\eta_i$. The *mean current* of a measure $\mu \in \mathcal{P}(\Omega)$ is an element of $\mathbb{R}^d$ such that its i^{th}-coordinate is

$$(s(\mu))_i = \mu(e_i \in \tau) - \mu(-e_i \in \tau).$$

If $s \in \partial\mathcal{O}_d$ we say that s is a *boundary mean current*. In terms of tiles, a measure μ has boundary mean current if and only if with probability 1 it samples at most one of the two possible tile types in each coordinate direction. The purpose of this section is to describe ergodic Gibbs measures with boundary mean current in dimension three. Using this, we compute the entropy function ent$(\cdot)$ in 3D restricted to $\partial\mathcal{O} = \partial\mathcal{O}_3$ (Theorem 2.4.7).

We will see that measures with boundary mean current in 2D and 3D are qualitatively very different. While the EGMs with boundary mean current in two dimensions all have zero entropy, EGMs with boundary mean current $s \in \partial\mathcal{O}$ in three dimensions can have positive entropy when s is contained in the interior of a face of $\partial\mathcal{O}$. Further, in three dimensions for any value a between 0 and ent(s), there exists an EGM μ with specific entropy $h(\mu) = a$.

Despite these differences, in 2D and 3D the general principle is that measures with boundary mean current in dimension d correspond to sequences of measures on a $(d-1)$-dimensional lattice. This is easy to see in 2D, and we use it as a warm-up for the 3D version.

2.4.1 Review: EGMs with boundary mean current in two dimensions

Call the four possible tile directions in two dimensions (east, west) and (north, south). It is sufficient to describe measures with boundary mean current (s_1, s_2) for which $s_1, s_2 \geq 0$ and $s_1 + s_2 = 1$, i.e. measures that sample only north and east tiles. The first step is to understand what tilings containing only north and east tiles look like.

For an even point (x_1, x_2), the north tile connects it to $(x_1, x_2 + 1)$ and the east tile connects it to $(x_1 + 1, x_2)$. In other words, north and east tiles always connect points along the line $x_1 + x_2 = 2c$ to points along the line $x_1 + x_2 = 2c + 1$. Therefore a tiling consisting of only north and east tiles can be partitioned into an infinite sequence of complete dimer tilings of *strips* $T_c = \{(x_1, x_2) : x_1 + x_2 = 2c \text{ or } 2c + 1\}$.

Along each strip, there are only two complete dimer tilings: one where the tiles are all east, and one where the tiles are all north. As such, any tiling τ with only north and east tiles consists of a sequence of choices of north or east tiles along the strips. See Figure 2.18.

All tilings τ of $\mathbb{Z}^2$ containing only north and east tiles are *frozen*, meaning they contain no finite cycles. To see this, note that if τ contains a finite cycle, then local move connectedness (see Section 2.3) implies it could be broken down into cycles of length 2. However a cycle of length 2 requires a north-south or east-west pair of tiles, which is not possible if the tiling

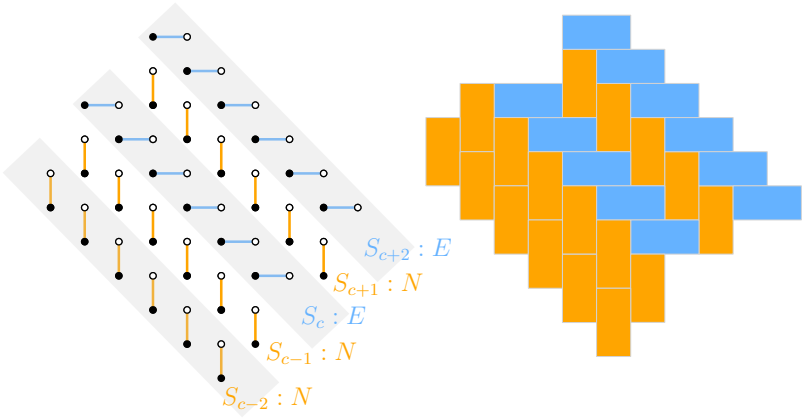


Figure 2.18: Parts of five strips drawn on the dual graph (left) and as a tiling (right)

contains only north and east tiles. Since tilings containing only north and east tiles are frozen, any measure μ which a.s. samples such tilings is automatically Gibbs. Three useful observations follow from this discussion:

- All ergodic Gibbs measures with boundary mean current in two dimensions have zero entropy. In other words, entropy is zero when restricted to $\partial\mathcal{O}_2$.
- There is a bijection between 1) Gibbs measures on dimer tilings of $\mathbb{Z}^2$ that contain only E and N tiles and 2) measures on integer-indexed $\{N, E\}$ sequences. Any sample of a process taking value E with proportion s_1 and N with proportion s_2, corresponds to a sample of a Gibbs measure on dimer tilings (obtained by placing N and E tiles on consecutive strips) of $\mathbb{Z}^2$ with mean current (s_1, s_2) and vice versa.
- There is also a bijection between 1) ergodic Gibbs measures on dimer tilings of $\mathbb{Z}^2$ that contain only E and N titles and have mean current $(p, 1-p)$ and 2) ergodic measures on integer-indexed $\{N, E\}$ sequences where the origin has probability p of being being assigned to E.

2.4.2 EGMs with boundary mean current in three dimensions

Now we will consider the three dimensional case. Let the types of tiles be (east, west), (north, south), (up, down). Without loss of generality we consider measures with boundary mean current that almost surely sample only north, east, and up tiles, i.e. mean current $s = (s_1, s_2, s_3)$ with $s_1 + s_2 + s_3 = 1$, $s_1, s_2, s_3 \geq 0$.

For an even point (x_1, x_2, x_3), an east tile connects it to $(x_1 + 1, x_2, x_3)$, a north tile connects it to $(x_1, x_2 + 1, x_3)$, and an up tile connects it to $(x_1, x_2, x_3 + 1)$. Therefore a tiling in 3D using only these three tile types corresponds to a sequence of tilings on two-dimensional *slabs*,

$$L_c = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 2c \text{ or } 2c + 1\}.$$

These slabs turn out to be a familiar two-dimensional lattice, namely the hexagonal lattice (with dimers viewed as edges) or the dual triangular lattice (with each dimer is viewed as a "lozenge" obtained as the union of two adjacent triangles), see Figure 2.19 for three things in

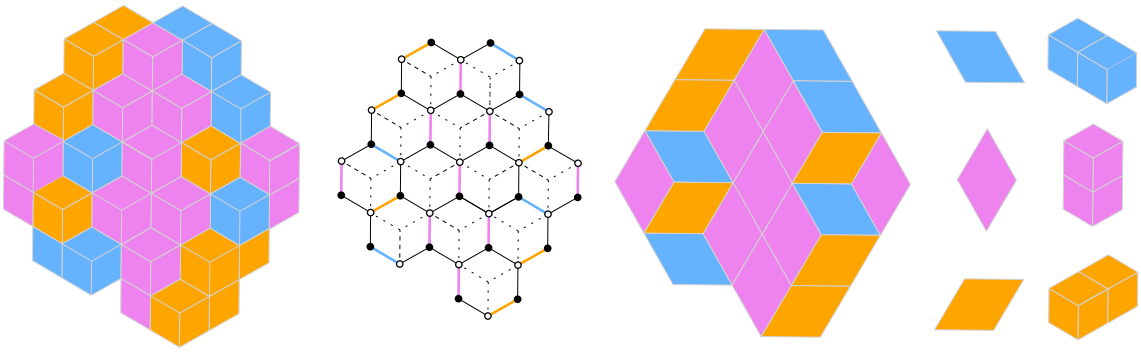


Figure 2.19: Bijections between: (1) a tiling of a slab in $\mathbb{Z}^3$, (2) perfect matching of the hexagonal lattice, and (3) a lozenge tiling.

correspondence: (1) cubes from a slab of $\mathbb{Z}^3$ visible from above the slab in a dimer tiling τ of $\mathbb{Z}^3$, (2) tiles from τ drawn on the hexagonal lattice as edges colored pink,blue and orange, (3) the same tiles drawn as lozenges obtained by taking the Voronoi cells containing these edges, and (4) a key giving the translation between lozenge tiles and 3D dimer bricks.

In the following, given a dimer tiling τ of a slab L_c, we will say that a particular tile type (north, east or up) has density s_i if the proportion of tiles of that type in $\tau \cap [-n, n]^3$ converges to s_i as $n \to \infty$. Similarly we can define the density for lozenge tilings.

Proposition 2.4.1. *For each $c \in \mathbb{Z}$, the slab L_c is a copy of the hexagonal lattice. There is a correspondence between tilings τ of $\mathbb{Z}^3$ which use only north, east and up tiles restricted to L_c and lozenge tilings. This correspondence takes a tiling of L_c with density (s_1, s_2, s_3) of the north, east and up tiles to a lozenge tiling where the density of the three lozenge tiles is also (s_1, s_2, s_3).*

Remark 2.4.2. There is a completely analogous correspondence for $s \in \partial\mathcal{O}$ when some of the components of s are negative. If the signs of s are $(\epsilon_1, \epsilon_2, \epsilon_3)$ then a tiling with boundary mean current s would restrict to a lozenge tiling on $\{(x_1, x_2, x_3) : \epsilon_1 x_1 + \epsilon_2 x_2 + \epsilon_3 x_3 = 2c \text{ or } 2c+1\}$. To simplify the presentation, some of the results in this section are stated for $s_1, s_2, s_3 \geq 0$, but the analogous statements hold for all $s \in \partial\mathcal{O}$.

Proof. Here we view the tiling as a collection of edges. Since τ uses only north, east and up tiles, the restriction $\tau_c = \tau \mid_{L_c}$ is a complete tiling of L_c. A single cube C in the $\mathbb{Z}^3$ lattice intersects four layers of the form $x_1 + x_2 + x_3 = a$. Let $\mathcal{C}_a$ be the collection of cubes in $\mathbb{Z}^3$ which intersect the layers $x_1 + x_2 + x_3 = a - 1, a, a + 1, a + 2$. By construction, $L_c \subset \mathcal{C}_{2c}$.

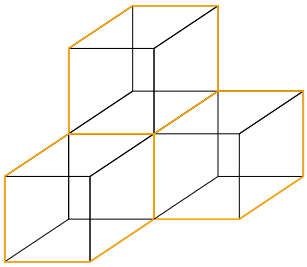


Figure 2.20: Three adjacent cubes in $\mathcal{C}_{2c}$, with intersection with L_c in orange.

For each $C \in \mathcal{C}_{2c}$, $C \cap L_c$ is a hexagon, hence the faces of L_c are hexagons. By observation we see that any two adjacent hexagons meet in an edge, any three adjacent hexagons meet at a vertex, and there are no collections of > 3 adjacent hexagons. Hence L_c is a copy of the hexagonal lattice. Finally Figure 2.19 gives the correspondence between the north, east and up tiles with the three kinds of lozenges which preserves their densities. □

Recall that $\mathcal{P}^s$ denotes the set of $\mathbb{Z}^3_{\text{even}}$-invariant probability measures on dimer tilings of $\mathbb{Z}^3$ of mean current $s = (s_1, s_2, s_3)$. We add subscripts G and e to denote Gibbs and ergodic measures respectively. Consider the group

$$\mathbb{Z}_{\text{loz}} = \mathbb{Z}^3_{\text{even}} \cap \{(x_1, x_2, x_3) \ : \ x_1 + x_2 + x_3 = 0\}.$$

Let $\mathcal{P}_{\text{loz}}$ denote the space of probability measures on dimer tilings of the slab L_0 (i.e. lozenge tilings) which are invariant under the $\mathbb{Z}_{\text{loz}}$ action. The *slope* of a measure ρ on lozenge tilings is the triple $s = (s_1, s_2, s_3)$ of expected densities of the three types of lozenges with respect to ρ. A lozenge tiling slope satisfies $s_1, s_2, s_3 \geq 0$ and $s_1 + s_2 + s_3 = 1$.

We abuse notation slightly and write $s(\mu)$ to mean the mean current or slope depending on what space μ is a measure on. To reduce notation issues, for the rest of the subsection we denote measures on dimer tilings of $\mathbb{Z}^3$ by μ or ν and measures on lozenge tilings by ρ or λ.

In this section, we use the notation τ_B to mean τ restricted to $B \subset \mathbb{Z}^3$.

Proposition 2.4.3. *Suppose μ is an ergodic Gibbs measure on dimer tilings of $\mathbb{Z}^3$ with mean current $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$, $s_1, s_2, s_3 \geq 0$. Let ρ_0 be the marginal measure of μ on the slab L_0. Then ρ_0 is a $\mathbb{Z}_{loz}$-invariant Gibbs measure with lozenge tiling slope $s(\rho_0) = s$.*

Remark 2.4.4. Since μ is $\mathbb{Z}^3_{\text{even}}$-invariant, μ is invariant under the $\mathbb{Z}$-action of translating by $(0, 0, 2c)$, which takes L_0 to L_c. Therefore ρ_c and ρ_0 are identically distributed for all $c \in \mathbb{Z}$.

Proof. Since $\mathbb{Z}_{\text{loz}} \subset \mathbb{Z}^3_{\text{even}}$, ρ_0 is $\mathbb{Z}_{\text{loz}}$-invariant. Consider a finite connected set $B \subset L_0$ with boundary ∂B in $\mathbb{Z}^3$. Suppose $\tau \in \Omega$ is a tiling in the support of μ, implying that it is a tiling using only north, east, and up tiles. Since μ is a Gibbs measure, and since there is μ-a.s. no tile in τ connecting L_0 and L_c for $c \neq 0$, we have for any tiling $\sigma \in \Omega$,

$$\mu(\sigma_B \mid \tau_{\mathbb{Z}^3 \setminus B}) = \mu(\sigma_B \mid \tau_{\partial B \cap L_0}) = \rho_0(\sigma_B \mid \tau_{\partial B \cap L_0}).$$

In the above we use the notation that for a tiling $\sigma \in \Omega$ and a set $A \subset \mathbb{Z}^3$, σ_A means σ restricted to A. Since μ is a Gibbs measure, the left hand side is uniform. Therefore ρ_0 is also a Gibbs measure.

Relating $s(\rho_0)$ to $s(\mu)$ is straightforward. Recall from Section 2.2.2 that $s_0(\tau)$ denotes the direction of the tile at the origin in τ, and that $s(\mu) = \mathbb{E}_\mu[s_0(\tau)]$. The same function s_0 can be used to compute the slope of a lozenge tiling measure, and $s(\rho_0) = \mathbb{E}_\rho[s_0(\sigma)]$ where σ is a full-plane lozenge tiling.

Let $\tau_0 = \tau_{L_0}$ be τ restricted to L_0. Since μ has boundary mean current, τ_0 is a full-plane lozenge tiling μ a.s. and $s_0(\tau) = s_0(\tau_0)$. Thus

$$s(\mu) = \mathbb{E}_\mu[s_0(\tau)] = \mathbb{E}_{\rho_0}[s_0(\tau_0)] = s(\rho_0).$$

□

To show that $h(\mu) = h(\rho_0)$ (Proposition 2.4.5), we use the fact that any Gibbs measure can be uniquely decomposed into *extreme* Gibbs measures [Geo11, Theorem 7.26]. Extreme Gibbs measures are the extreme points of the convex set of Gibbs measures (analogous to how ergodic measures are the extreme points of the convex set of invariant measures). A Gibbs measure is extreme if and only if it is tail trivial [Geo11, Theorem 7.7].

If ρ is a Gibbs measure, there is a unique weight function g_ρ on the extreme Gibbs measures which gives its *extreme Gibbs decomposition*,

$$\rho = \int \lambda \, \mathrm{d}g_\rho(\lambda).$$

This decomposition means that sampling from a Gibbs measure ρ can be thought of as a two step process: 1) sample an extreme Gibbs component λ from $\mathrm{d}g_\rho$, 2) sample a tiling τ from λ. Given a tiling τ sampled from a Gibbs measure ρ, we can a.s. recover the extreme Gibbs component λ that τ was sampled from [Geo11, Theorem 7.12]. If λ is the extreme Gibbs component that τ is sampled from, we say that τ is *generic* for λ.

Proposition 2.4.5. *Suppose μ is an EGM on Ω with mean current $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$, $s_1, s_2, s_3 \geq 0$. Let ρ_c be the marginal measure of μ on the slab L_c for $c \in \mathbb{Z}$. Sampling a tiling τ from μ induces a choice of extreme Gibbs component λ_c of ρ_c for all $c \in \mathbb{Z}$. For each $c \in \mathbb{Z}$, let $\tau_c = \tau_{L_c}$.*

1. *Conditional on the choice of extreme Gibbs component λ_c of ρ_c for each $c \in \mathbb{Z}$, the samples $(\tau_c)_{c\in\mathbb{Z}}$ are independent.*
2. *For any $c \in \mathbb{Z}$, $h(\mu) = h(\rho_c)$.*

Remark 2.4.6. Since ρ_0 and ρ_c are identically distributed, $h(\rho_c) = h(\rho_0)$ for all $c \in \mathbb{Z}$. Thus it suffices to prove (2) for $c = 0$. We also note that we could have used the ergodic decomposition instead of the extreme Gibbs decomposition to prove this theorem. The upshot of using the extreme Gibbs decomposition is that conditional on a choice of extreme Gibbs component λ_c on each slab L_c, the samples τ_c from λ_c for all $c \in \mathbb{Z}$ are independent. Conditioned on a choice of ergodic component η_c on each slab L_c, only the samples from η_c with $s(\eta_c) = (l_1^c, l_2^c, l_3^c)$ and $l_1^c, l_2^c, l_3^c > 0$ are necessarily independent [She05, Theorem 9.1.1].

Proof of Proposition 2.4.5. Since μ is Gibbs, it has an extreme Gibbs decomposition

$$\mu = \int \nu \, \mathrm{d}g_\mu(\nu).$$

Sampling a tiling τ from μ is equivalent to sampling an extreme Gibbs component ν of μ (from g_μ), and then sampling a tiling τ from ν. Since ν is tail-trivial, its marginal λ_c on L_c is also tail-trivial. For all $c \in \mathbb{Z}$, the extreme Gibbs decomposition of the marginal ρ_c can be written as

$$\rho_c = \int \lambda_c \, \mathrm{d}g_{\rho_c}(\lambda_c)$$

where g_{ρ_c} is the extreme Gibbs decomposition of ρ_c.

Let $B_n = [-n, n]^3$. Since λ_0 is extreme Gibbs it is tail trivial, so

$$\lim_{m\to\infty} \mu(\tau_{B_n \cap L_0} \mid \tau_{\mathbb{Z}^3 \setminus (B_m \cap L_0)}, \lambda_0) = \lambda_0(\tau_{B_n \cap L_0}).$$

Therefore conditioned on λ_0, $\tau_0 = \tau_{L_0}$ is independent of $\tau_{(\mathbb{Z}^3 \setminus L_0)}$. In particular, conditioned on the sequence of measures $(\lambda_c)_{c\in\mathbb{Z}}$ (equivalently, conditioned on choosing an extreme Gibbs component of μ), the tilings on each slab $(\tau_c)_{c\in\mathbb{Z}}$ are independent.

Now we relate the specific entropies. Recall from Section 2.2.3 that for a region $\Delta \subset \mathbb{Z}^3$ and an invariant measure μ on tilings of $\mathbb{Z}^3$,

$$H_\Delta(\mu) = -\sum_{\sigma \in \Omega(\Delta)} \mu(X(\sigma)) \log \mu(X(\sigma)),$$

where $\Omega(\Delta)$ is the free-boundary tilings of Δ, and $X(\sigma)$ is the collection of tilings of $\mathbb{Z}^3$ which extend σ. Taking $A_n(0) = B_n \cap L_0$, let $A_n(c) = A_n(0) + (0, 0, 2c)$, and finally let $A_{n,m} = \cup_{c=-m}^{m} A_n(c)$. It is well known that the specific entropy can be computed as

$$h(\mu) = \lim_{n\to\infty} |A_{n,n}|^{-1} H_{A_{n,n}}(\mu).$$

Instead of free-boundary tilings, we can choose $\tau \in \Omega$ and let $\Omega_\tau(\Delta) = \{\sigma \in \Omega(\Delta) : \sigma \mid_{\partial\Delta} = \tau\}$ be the tilings of Δ with boundary condition agreeing with τ. Then we define the entropy of μ given a fixed boundary condition τ:

$$H_\Delta(\mu|\tau) = -\sum_{\sigma \in \Omega_\tau(\Delta)} \mu(\sigma \mid \tau_{\mathbb{Z}^3 \setminus \Delta}) \log \mu(\sigma \mid \tau_{\mathbb{Z}^3 \setminus \Delta}).$$

Again for $A \subset \mathbb{Z}^3$, τ_A means τ restricted to A. We remark that this is not the usual definition of conditional entropy where we condition on a random variable or a sigma algebra. Instead we are fixing the value of the random variable, namely, the boundary condition of the tiling in Δ. Indeed, if τ is generic for an extreme Gibbs measure component ν of μ then we have that $H_\Delta(\mu|\tau) = H_\Delta(\nu|\tau)$. We will restrict this non-standard usage to this proof. It is standard that the specific entropy of μ can also be computed using this conditional definition as

$$h(\mu) = \lim_{n\to\infty} |A_{n,n}|^{-1} \Big(\int_\Omega H_{A_{n,n}}(\mu|\tau) \, \mathrm{d}\mu(\tau) + H_{\partial A_{n,n}}(\mu) \Big) = \lim_{n\to\infty} |A_{n,n}|^{-1} \int_\Omega H_{A_{n,n}}(\mu|\tau) \, \mathrm{d}\mu(\tau). \tag{2.10}$$

In the second equality, we use that the entropy term for $\partial A_{n,n}$ is of order n^2 so it does not contribute in the limit. Now we rewrite the argument of the limit using the extreme Gibbs decomposition.

$$\int_\Omega H_{A_{n,n}}(\mu|\tau)\,\mathrm{d}\mu(\tau) = \int\int_\Omega H_{A_{n,n}}(\nu|\tau)\,\mathrm{d}\nu(\tau)\,\mathrm{d}g_\mu(\nu).$$

Recall that $\tau_c = \tau_{L_c}$. Conditional on sampling the process $(\lambda_c)_{c\in\mathbb{Z}}$ (equivalently, conditional on sampling ν), the samples $(\tau_c)_{c\in\mathbb{Z}}$ are independent. Thus

$$H_{A_{n,n}}(\nu|\tau) = \sum_{c=-n}^{n} H_{A_n(c)}(\lambda_c|\tau_c).$$

Recall that Ω_{loz} is the set of full-plane lozenge tilings. Thus

$$\int\int_\Omega H_{A_{n,n}}(\nu|\tau)\,\mathrm{d}\nu(\tau)\,\mathrm{d}g_\mu(\nu) = \sum_{c=-n}^{n}\int\int_{\Omega_{\mathrm{loz}}} H_{A_n(c)}(\lambda_c|\tau_c)\,\mathrm{d}\lambda_c(\tau_c)\,\mathrm{d}g_{\rho_c}(\lambda_c) \tag{2.11}$$

$$= \sum_{c=-n}^{n}\int_{\Omega_{\mathrm{loz}}} H_{A_n(c)}(\rho_c|\tau_c)\,\mathrm{d}\rho_c(\tau_c). \tag{2.12}$$

Since ρ_c is equal in distribution to ρ_0, for all $c \in \mathbb{Z}$,

$$|A_n(c)|^{-1}\int_{\Omega_{\mathrm{loz}}} H_{A_n(c)}(\rho_c|\tau_c)\,\mathrm{d}\rho_c(\tau_c) = |A_n(0)|^{-1}\int H_{A_n(0)}(\rho_0|\tau_0)\,\mathrm{d}\rho_0(\tau_0).$$

At the same time,

$$\lim_{n\to\infty}|A_n(0)|^{-1}\int_{\Omega_{\mathrm{loz}}} H_{A_n(0)}(\rho_0|\tau_0)\,\mathrm{d}\rho_0(\tau_0) = h(\rho_0). \tag{2.13}$$

Therefore

$$\begin{aligned} h(\mu) &= \lim_{n\to\infty}|A_{n,n}|^{-1}\int_\Omega H_{A_{n,n}}(\mu|\tau)\,\mathrm{d}\mu(\tau) \\ &= \lim_{n\to\infty}\frac{1}{2n+1}(2n+1)|A_n(0)|^{-1}\int_{\Omega_{\mathrm{loz}}} H_{A_n(0)}(\rho_0|\tau_0)\,\mathrm{d}\rho_0(\tau_0) = h(\rho_0). \end{aligned}$$

□

Recall from Section 2.2.3 that the *mean-current entropy function* $\mathrm{ent} : \mathcal{O} \to [0,\infty)$ is defined by

$$\mathrm{ent}(s) = \max_{\mu\in\mathcal{P}^s} h(\mu).$$

This entropy function plays a central role in our work and will be studied extensively in Section 2.7.

Let $\mathcal{T}_2 = \{s_1, s_2, s_3 \geq 0 : s_1 + s_2 + s_3 = 1\}$ be the space of possible lozenge tiling slopes. The *slope entropy function* $\text{ent}_{\text{loz}} : \mathcal{T}_2 \to [0, \infty)$ for lozenge tilings is defined by

$$\text{ent}_{\text{loz}}(s) = \max_{\rho \in \mathcal{P}^s_{\text{loz}}} h(\rho).$$

It was shown in [CKP01, Theorem 9.2] that ent_{loz} has the explicit form

$$\text{ent}_{\text{loz}}(s_1, s_2, s_3) = \frac{1}{\pi}(L(\pi s_1) + L(\pi s_2) + L(\pi s_3))$$

where $L : [0, \pi] \to \mathbb{R}$ is the Lobachevsky function given by

$$L(\theta) = -\int_0^\theta \ln(2\sin(x))dx.$$

Using this two dimensional result, we can explicitly compute ent on $\partial\mathcal{O}$. Let $\mathcal{E} \subset \partial\mathcal{O}$ denote the edges of $\mathcal{O}$.

Theorem 2.4.7. *For $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$,*

$$ent(s) = ent_{loz}(|s_1|, |s_2|, |s_3|) = \frac{1}{\pi}(L(\pi|s_1|) + L(\pi|s_2|) + L(\pi|s_3|)).$$

Further, if $s \notin \mathcal{E}$, then any measure μ realizing $h(\mu) = ent(s)$ is an ergodic Gibbs measure on Ω with respect to the $\mathbb{Z}^3_{even}$ action. If $s \in \mathcal{E}$, then $ent(s) = 0$.

It is well-known that ent_{loz} is strictly concave as a function of slope on the interior of allowed slopes [CKP01, Theorem 10.1]. Thus as an immediate corollary, we get that

Corollary 2.4.8. *Let $\mathcal{F}$ be any face of $\partial\mathcal{O}$. The entropy function $ent(\cdot)$ is strictly concave on the interior of $\mathcal{F}$.*

Proof of Theorem 2.4.7. By Theorem 2.7.2, if $\mu \in \mathcal{P}^s$ satisfies $h(\mu) = \text{ent}(s)$, then μ is a Gibbs measure. While we include this result later in the paper for organizational reasons, it follows easily from the classical variational principle for Gibbs measures [LR69] (the only adaptation is that we are looking at the maximizer with a fixed mean current).

Without loss of generality assume that $s_1, s_2, s_3 \geq 0$. First suppose that $\mu \in \mathcal{P}^s$ is an EGM, and as usual let ρ_c denote its marginal on L_c. By Proposition 2.4.3 and Proposition 2.4.5,

$$s = s(\mu) = s(\rho_0) \qquad \text{and} \qquad h(\mu) = h(\rho_0).$$

Combining the results of [CKP01] and [She05],

- If s has $s_1, s_2, s_3 > 0$, then $\rho_0 \in \mathcal{P}^s_{\text{loz}}$ satisfies $h(\rho_0) = \text{ent}_{\text{loz}}(s)$ if and only if ρ_0 is the unique ergodic Gibbs measure of slope s, which we denote by λ_s.
- If s has $s_i = 0$ for some $i = 1, 2, 3$, then $h(\rho_0) = \text{ent}_{\text{loz}}(s) = 0$ for all $\rho_0 \in \mathcal{P}^s_{\text{loz}}$.

Therefore if $s_1, s_2, s_3 > 0$ and $h(\rho_0) = \mathrm{ent}_{\mathrm{loz}}(s)$, then by strict concavity of $\mathrm{ent}_{\mathrm{loz}}$ [CKP01, Theorem 10.1], ρ_c are identically distributed and equal to λ_s a.s. By [She05, Theorem 9.1.1], when $s_1, s_2, s_3 > 0$, the unique ergodic Gibbs measure λ_s is an extreme Gibbs measure, and thus $\{\rho_c\}_{c\in\mathbb{Z}}$ is i.i.d. by Proposition 2.4.5. Alternatively if $s_i = 0$ for some i, then $h(\rho_0) = 0$, and hence $h(\mu) = 0$.

If μ is not ergodic with respect to the $\mathbb{Z}^3_{\mathrm{even}}$ action, then it can be decomposed

$$\mu = \int_{\mathcal{P}_e} \nu \, \mathrm{d}w_\mu(\nu),$$

where

$$s(\mu) = \int_{\mathcal{P}_e} s(\nu) \, \mathrm{d}w_\mu(\nu).$$

Note that w_μ almost surely, $s(\nu)$ is contained in the same face of $\partial\mathcal{O}$ as $s = s(\mu)$. By the analysis above for an ergodic measure, if $s(\nu) \notin \mathcal{E}$ then $h(\nu) = \mathrm{ent}(s(\nu)) = \mathrm{ent}_{\mathrm{loz}}(s(\nu))$ if and only if ν is an EGM of mean current $s(\nu)$ with marginals on each slab i.i.d. and equal to the lozenge tiling EGM of slope $s(\nu)$ (if $s(\nu) \in \mathcal{E}$, then $s(\nu) = 0$). Since $\mathrm{ent}_{\mathrm{loz}}$ is strictly concave on the interior of allowed slopes, if s is contained in the interior of a face of $\partial\mathcal{O}$, then $h(\mu) = \mathrm{ent}(s)$ if and only if μ is an ergodic Gibbs measure of mean current s. □

As seen in the proof of Theorem 2.4.7, we get an explicit description of the entropy maximizers for $s \in \partial\mathcal{O}$. In contrast to two dimensions, the maximum entropy is positive for mean currents in the interior of faces.

Corollary 2.4.9. *Suppose $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$.*

- *If $s \in \mathcal{E}$ (i.e. $s_i = 0$ for some $i = 1, 2, 3$), then $h(\mu) = 0$ for any $\mu \in \mathcal{P}^s$.*
- *If $s_1 s_2 s_3 \neq 0$, then the entropy maximizer in $\mathcal{P}^s$ is an ergodic Gibbs measure such that for all $c \in \mathbb{Z}$, $\rho_c = \lambda_s$ a.s., where λ_s is the unique ergodic Gibbs measure on lozenge tilings with slope $(|s_1|, |s_2|, |s_3|)$. Here ρ_c is the marginal on the slab $\{(x_1, x_2, x_3 : \epsilon_1 x_1 + \epsilon_2 x_2 + \epsilon_3 x_3 = 2c \text{ or } 2c + 1\}$, where ϵ_i is the sign of s_i.*

It is also straightforward to show that there exist EGMs of a fixed boundary mean current with a range of different entropies.

Proposition 2.4.10. *Suppose $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$, $s_1, s_2, s_3 > 0$. Then for all $0 \leq \theta \leq 1$ there is an ergodic Gibbs measure μ such that $h(\mu) = \theta$ ent(s).*

Proof. Let $\rho_{\max}, \rho_1, \rho_2, \rho_3 \in \mathcal{P}_{\mathrm{loz}}$ be EGMs on lozenge tilings of slopes (s_1, s_2, s_3), $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ respectively. Now consider an i.i.d. process $(\eta_c)_{c\in\mathbb{Z}}$ with state space

$$\{\rho_{\max}, \rho_1, \rho_2, \rho_3\}$$

such that the probability of $\rho_{\max}$ is θ, and the probability of ρ_i is $(1 - \theta)s_i$ for $i = 1, 2, 3$.

Let μ be a measure on Ω given by taking a sample from $(\eta_c)_{c\in\mathbb{Z}}$, this gives a tiling of $\mathbb{Z}^3$ such that the restriction to each slab L_c is an independent sample from η_c. Clearly μ is a Gibbs measure on Ω. Since $(\eta_c)_{c\in\mathbb{Z}}$ is an i.i.d. process it is ergodic so μ is ergodic with respect to $\mathbb{Z}^3_{\mathrm{even}}$. By Proposition 2.4.3 $s(\mu) = s$ and by Proposition 2.4.5 $h(\mu) = \theta \, \mathrm{ent}(s)$. □

We now summarize the results from this section to illustrate the similarities and differences with the two dimensional case.

- In three dimensions, EGMs of the same boundary mean current s can have different specific entropy values (Proposition 2.4.10).
- Every EGM μ on dimer tilings that contains only east, north, and up tiles gives rise to a Gibbs measure on integer-indexed stationary sequences of extreme Gibbs measures $(\lambda_c)_{c\in\mathbb{Z}}$ on lozenge tilings (Proposition 2.4.5).
- If $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$ is such that $s_1, s_2, s_3 \neq 0$ then the entropy-maximizing measure with mean current s is an EGM such that $(\lambda_c)_{c\in\mathbb{Z}}$ is an i.i.d. sequence of copies of the unique EGM on lozenge tilings with slope $(|s_1|, |s_2|, |s_3|)$ (Corollary 2.4.9).

2.5 Free-boundary tilings, asymptotic flows, and Wasserstein distance

This section sets up some of the key function-theoretic preliminaries for the large deviation principle in Section 2.8.

A domain is an open subset of $\mathbb{R}^3$. Let $R \subset \mathbb{R}^3$ be a compact region which is the closure of a connected domain and has piecewise smooth boundary ∂R. We say that a grid region G is *scale* n if $G \subseteq \frac{1}{n}\mathbb{Z}^3$. If R_n is a scale n grid region, then with a slight abuse of notation we say that $R_n \supseteq R$ if the collection of $\frac{1}{n}$-width cubes centered at points in R_n contains R. If τ is a tiling of R_n, we define the restriction of τ to R, denoted τ_R, to be the collection of tiles from τ which intersect R.

Definition 2.5.1. The *free-boundary tilings of R at scale n* are

$$T_n(R) := \bigcup_{R_n \supseteq R} \{\tau_R : \tau \text{ is a tiling of } R_n\}.$$

The *free-boundary tiling flows on R at scale n* are

$$TF_n(R) := \{f_\tau : \tau \in T_n(R)\}.$$

Finally, we define the space of all free-boundary tiling flows on R to be $TF(R) := \bigcup_{n\geq 1} TF_n(R)$.

The edges in $\frac{1}{n}\mathbb{Z}^3$ have length $\frac{1}{n}$. To ensure that the total flow of a tiling flow is roughly constant in n, we need the flow per edge of $f_\tau \in TF_n(R)$ to be of order $\frac{1}{n^3}$. We can achieve that by rescaling the flow by a factor of n^3 so that it has magnitude $\frac{5}{6n^3}$ on each matched edge and $\frac{1}{6n^3}$ on each unmatched edge.

Remark 2.5.2. Note that $TF_n(R)$ may contain elements that do not arise as restrictions of tilings of all of $\frac{1}{n}\mathbb{Z}^3$ to R. That is, there may be free-boundary tilings of R that cover R but do not extend to tilings of all of $\mathbb{Z}^3$. (These might exist, for example, if R is a concave region.)

We define a metric on flows (Section 2.5.2), denoted d_W, which is an adaptation of *generalized Wasserstein distance* from signed measures to flows. Intuitively we want to consider two flows f, g "close" if we don't have to change the flow of f too much—either by moving flow over, or by adding or deleting it—to transform it into g. This is what $d_W(f, g)$ will measure. In terms of this metric, the main question of this section is: if $f_n \in TF_n(R)$ for all $n \in \mathbb{N}$, what are the possible limits of the form $\lim_{n\to\infty} f_n$?

We show (Theorem 2.5.19) that any fine-mesh limit of free-boundary tiling flows on R is an *asymptotic flow on* R, defined by:

Definition 2.5.3. Let R be a compact region which is the closure of a connected domain and has piecewise smooth boundary. We say that f is an *asymptotic flow on* R if it satisfies the following properties:

- f is a Borel-measureable vector field with support contained in R;
- f is valued in $\mathcal{O}$ (since f is measurable, this means that f is valued in $\mathcal{O}$ Lebesgue-a.e.);
- f is divergence-free in the interior of R, i.e. $\operatorname{div} f = 0$ as a distribution (so for any smooth function ϕ compactly supported in the interior of R, $\int_R \phi \operatorname{div} f := \int_R \nabla\phi \cdot f = 0$.)

We denote the set of asymptotic flows on R by $AF(R)$.

In Theorem 2.5.22 we will show that $(AF(R), d_W)$ is a compact metric space. In Sections 2.5.4 and 2.5.5, we define a boundary value operator T (*trace operator*) which takes a flow to its boundary value on ∂R. In fact we do something more general, and define the trace of a flow for any compact, piecewise smooth surface contained in R. After defining T for asymptotic flows, we define the space of asymptotic flows with boundary value b, denoted $AF(R, b)$, and show that it is compact with respect to d_W (Corollary 2.5.32).

The boundary value operator is defined in different but analogous ways for asymptotic flows (Section 2.5.4) and tiling flows (Section 2.5.5). The main essential result about T is that these definitions are compatible and that T is uniformly continuous (Theorem 2.5.39).

We remark that the main important property of Wasserstein distance in our analysis is that it metrizes weak convergence, and that it thereby formalizes the intuitive notions that the scaling limits of tiling flows are asymptotic flows, and that boundary values depend continuously on the flow. While the Wasserstein metric and other transportation metrics have a number of additional special properties, we do not use this theory here. All the properties of the Wasserstein metric that we use are described in Section 2.5.1.

2.5.1 Background on (generalized) Wasserstein distance

The original *Wasserstein distance* or *earth-movers distance* is a metric on probability measures on a fixed metric space. Suppose that (X, d) is a compact, separable metric space. The L^1 Wasserstein distance W_1 is a metric on $\mathcal{P}(X)$, the space of probability measures on X and is given by

$$W_1(\mu, \nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} d(x, y) d\gamma(x, y)$$

where $\Gamma(\mu, \nu)$ is the collection of all couplings of μ and ν. Intuitively W_1 measures the cost—i.e. how much mass and how far it has to be moved—required to transform μ into ν by redistributing the mass of μ. This metric was developed in the theory of optimal transport and has been applied in many different contexts including probability, Riemannian geometry, and image processing. For more see [Vil09].

We will define and use versions of Wasserstein distance for flows and their boundary values. In the next section, we define a mapping between flows and measures, where flow (with direction) corresponds to mass (with sign). The measures corresponding to flows do not necessarily have the same mass and can be signed. Given this, our Wasserstein distance on flows will be based on a version of *generalized Wasserstein distance*.

Let $\mathcal{M}(\mathbb{R}^d)$ denote the space of Borel regular measures on $\mathbb{R}^d$ with finite total mass. In [PR14] and [PR16], they define a generalized Wasserstein distance on $\mathcal{M}(\mathbb{R}^d)$ by introducing an L^1 cost for adding and deleting mass. It is denoted $W_1^{1,1}$ and defined as

$$W_1^{1,1}(\mu, \nu) = \inf_{\tilde{\mu}, \tilde{\nu}} |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}| + W_1(\tilde{\mu}, \tilde{\nu})$$

where the infimum is taken over $\mathcal{M}(\mathbb{R}^d)$.

In [AMS11] the L^1 Wasserstein distance was generalized to signed probability measures. This metric is denoted $\mathbb{W}_1$. If μ, ν are signed measures with Jordan decompositions $\mu = \mu_+ - \mu_-$ and $\nu = \nu_+ - \nu_-$, then

$$\mathbb{W}_1(\mu, \nu) = W_1(\mu_+ + \nu_-, \nu_+ + \mu_-).$$

In fact, note that this definition does not depend on the decomposition of the measures μ, ν.

In [PRT19], they combine these to give a definition of Wasserstein distance for signed measures of different masses. Let $\mathcal{M}^s(\mathbb{R}^d)$ denote the space of signed Radon measures on $\mathbb{R}^d$, i.e. measures μ that can be written $\mu_+ - \mu_-$ for $\mu\pm \in \mathcal{M}(\mathbb{R}^d)$. Denoted $\mathbb{W}_1^{1,1}$, the generalized Wasserstein distance on $\mathcal{M}^s(\mathbb{R}^d)$ is defined

$$\mathbb{W}_1^{1,1}(\mu, \nu) = W_1^{1,1}(\mu_+ + \nu_-, \nu_+ + \mu_-).$$

This is the definition of Wasserstein distance that we will use in this paper. We note a few important facts about $\mathbb{W}_1^{1,1}$ that we will use.

Lemma 2.5.4 (See [PRT19, Lemma 18]). *If $\mu, \nu, \rho \in \mathcal{M}^s(\mathbb{R}^d)$, then*

$$\mathbb{W}_1^{1,1}(\mu, \nu) = \mathbb{W}_1^{1,1}(\mu + \rho, \nu + \rho).$$

Proposition 2.5.5 (See [PRT19, Proposition 23]). *Let*

$$\mathcal{C}_b^{0,Lip} = \{f : \mathbb{R}^d \to \mathbb{R} \ : \ f \text{ continuous, bounded, Lipschitz}\}.$$

Then

$$\mathbb{W}_1^{1,1}(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^d} \varphi \, \mathrm{d}(\mu - \nu) \ : \ \varphi \in \mathcal{C}_b^{0,Lip}, \|\varphi\|_\infty \leq 1, \|\varphi\|_{Lip} \leq 1 \right\}.$$

From this it clearly follows that

Corollary 2.5.6. *If* $\lim_{n\to\infty} \mathbb{W}_1^{1,1}(\mu_n, \mu) = 0$*, then* μ_n *converges weakly to* μ*.*

The non-signed generalized Wasserstein distance $W_1^{1,1}$ metrizes weak convergence for tight sequences of measures in $\mathcal{M}(\mathbb{R}^d)$ [PR14, Theorem 13], as does the original L^1 Wasserstein distance for probability measures [Vil09, Theorem 6.9]. With signed measures, slightly stranger behavior can occur in general, see e.g. [PRT19, Remark 26]. However we show that the Wasserstein distance for flows defined below does metrize weak convergence, see Remark 2.5.12.

For $R \subset \mathbb{R}^d$, we define $\mathcal{M}(R)$ to be the Radon measures supported in R, and $\mathcal{M}^s(R)$ to be signed Radon measures supported in R. We let $\mathcal{M}_{\text{ac}}(R)$ (resp. $\mathcal{M}^s_{\text{ac}}(R)$) denote the Radon measures (resp. signed Radon measures) supported in R and absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$. We say that $\mathcal{M}^s_{\text{ac}}(R, a, b)$ denotes absolutely continuous signed measures with densities valued between a and b. By Lemma 2.5.4, for any $a < 0$ and $b > 0$,

$$(\mathcal{M}^s_{\text{ac}}(R, a, b), \mathbb{W}_1^{1,1}) \cong (\mathcal{M}_{\text{ac}}(R, 0, b - a), W_1^{1,1}) \tag{2.14}$$

as metric spaces. This identification has some useful consequences. From Equation (2.14) and [PR14, Proposition 15] it follows that

Proposition 2.5.7. *If* R *is compact, then* $(\mathcal{M}^s_{ac}(R, a, b), \mathbb{W}_1^{1,1})$ *is a compact metric space for* $a, b \in \mathbb{R}$*.*

2.5.2 Wasserstein distance for flows

Let R be a compact region which is the closure of a connected domain and has piecewise smooth boundary. If $f_\tau \in TF_n(R)$, then f_τ is supported in $B_{2/n}(R) = \{x : d(x, R) \leq 2/n\}$. We will define a correspondence between

1. vector fields f on $R \subset \mathbb{R}^3$ or $f_\tau \in TF_n(R)$, and
2. triples of signed measures (μ_1, μ_2, μ_3) supported in $B_{2/n}(R)$.

The idea is that the flow of the vector field in coordinate direction i corresponds to mass of the i^{th} measure, with sign coming from the direction of the flow. We define Wasserstein distance on vector fields through this correspondence:

Definition 2.5.8. The *Wasserstein distance on flows*, denoted d_W, is the sum of the generalized Wasserstein distances between the component measures. For any two vector fields f, g with corresponding triples of measures (μ_1, μ_2, μ_3) and (ν_1, ν_2, ν_3), we define

$$d_W(f, g) := \mathbb{W}_1^{1,1}(\mu_1, \nu_1) + \mathbb{W}_1^{1,1}(\mu_2, \nu_2) + \mathbb{W}_1^{1,1}(\mu_3, \nu_3).$$

To complete the definition of the metric, we need to define the correspondences between vector fields and triples of measures. There will be two definitions, one for a measurable vector field on R and one for a discrete vector field $f_\tau \in TF_n(R)$. Let $x = (x_1, x_2, x_3)$ denote a point in $\mathbb{R}^3$.

Definition 2.5.9. (Measures corresponding to a measurable vector field.) Let f be a measurable vector field supported in R. The components of $f(x) = (f_1(x), f_2(x), f_3(x))$ are measurable functions, and we define the corresponding triple of measures (μ_1, μ_2, μ_3) by

$$\mathrm{d}\mu_i(x) = f_i(x)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \qquad\qquad i = 1, 2, 3$$

where $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$ denotes Lebesgue measure on $\mathbb{R}^3$.

Definition 2.5.10. (Measures corresponding to a free-boundary tiling flow on R.) Suppose that $f = f_\tau \in TF_n(R)$ for some n. Let η_1, η_2, η_3 be the positively-oriented unit basis vectors. Orient all the edges e of $\frac{1}{n}\mathbb{Z}^3$ to be parallel to η_i, which we denote by $e \parallel \eta_i$. (Recall that changing the orientation of e changes the sign of $f(e)$.) The triple of measures (μ_1, μ_2, μ_3) corresponding to f is given by

$$\mathrm{d}\mu_i(x) = \sum_{e \parallel \eta_i} \frac{2}{n^2} f(e) \mathbb{1}_e(x) \mathrm{d}x_i \qquad\qquad i = 1, 2, 3$$

where $\mathbb{1}_e$ denotes the indicator of the edge $e \in \frac{1}{n}\mathbb{Z}^3$, and $\mathrm{d}x_i$ is 1-dimensional Lebesgue measure in the direction of η_i. Note that μ_i is supported in $B_{2/n}(R)$.

Remark 2.5.11. The scaling factor $\frac{2}{n^2}$ ensures that each edge e such that $e \parallel \eta_i$ contributes $\frac{2f(e)}{n^3}$ total mass to μ_i. The normalization is justified by looking at the extreme examples corresponding to the brickwork tilings (i.e. tilings where all tiles are the same type). Each cube in the $\frac{1}{n}\mathbb{Z}^3$ mesh can be viewed as corresponding to its lower left edge. In the brickwork pattern, exactly half of these cubes will have a dimer in the lower left edge.

Remark 2.5.12. Now that Wasserstein distance on flows is defined, we can explain why it metrizes weak convergence of the component measures. We do this by explaining how we could "shift" everything to have positive mass and use Equation (2.14). For asymptotic flows, we can just add a copy of the 3-dimensional Lebesgue measure $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$. For tiling flows, we note that we could have defined the corresponding measures to be positive by translating the mean-current octahedron $\mathcal{O}$ by $\eta_1 + \eta_2 + \eta_3$. After the translation, a scale n tiling flow measure would take values in $\{1/(3n^2), 5/(3n^2), 7/(3n^2), 11/(3n^2)\}$ instead of $\{-5/(3n^2), -1/(3n^2), 1/(3n^2), 5/(3n^2)\}$.

In terms of adding measures, translating $\mathcal{O}$ corresponds to adding a copy of 1-dimensional Lebesgue $\frac{2}{n^2}\mathrm{d}x_i$ along each edge $e \parallel \eta_i$ in $\frac{1}{n}\mathbb{Z}^3$ to the scale n tiling flow measure $\mathrm{d}\mu_i$. In the scaling limit as $n \to \infty$, this sum of 1-dimensional Lebesgue measures converges in $\mathbb{W}_1^{1,1}$ to $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$. By Equation (2.14), this implies the "translated" tiling flows measures converge (i.e. ones defined on the translated $\mathcal{O}$) to the "translated" asymptotic flow measures (i.e. ones shifted by adding $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$) in $W_1^{1,1}$ if and only if the tiling flow measures converge to the asymptotic flow measures in $\mathbb{W}_1^{1,1}$. Since $W_1^{1,1}$ metrizes weak convergence [PR14, Proposition 15], d_W metrizes weak convergence of the component measures corresponding to tiling and asymptotic flows.

Proposition 2.5.13. *The measures corresponding to tiling flows are divergence-free on the interior of R in the sense of distributions, i.e. if f is a tiling flow with corresponding measures*

(μ_1, μ_2, μ_3), then for any ϕ smooth and supported in a compact set C contained in the interior of R,

$$\int_R \frac{\partial \phi}{\partial x_1} \, \mathrm{d}\mu_1 + \int_R \frac{\partial \phi}{\partial x_2} \, \mathrm{d}\mu_2 + \int_R \frac{\partial \phi}{\partial x_3} \, \mathrm{d}\mu_3 = 0.$$

Proof. For $i = 1, 2, 3$, let $e_1^i, ..., e_{k_i}^i$ be the edges from $\frac{1}{n}\mathbb{Z}^3$ such that $e_j^i \parallel \eta_i$, is oriented parallel to η_i, and which intersect R. Let (a_j^i, b_j^i) be the endpoints of e_j^i such that $b_j^i - a_j^i = \eta_i$. By the fundamental theorem of calculus,

$$\sum_{i=1}^{3} \int_R \frac{\partial \phi}{\partial x_i} \, \mathrm{d}\mu_i = \frac{2}{n^2} \sum_{i=1}^{3} \sum_{j=1}^{k_i} (\phi(b_j^i) - \phi(a_j^i)) f(e_j^i)$$

If $v = a_j^i$ or b_j^i is not contained in the interior of R, then $\phi(v) = 0$. Therefore we can rewrite the above as a sum over vertices $v \in \frac{1}{n}\mathbb{Z}^3$ contained in the interior of R:

$$\sum_{i=1}^{3} \int_R \frac{\partial \phi}{\partial x_i} \, \mathrm{d}\mu_i = \frac{2}{n^2} \sum_{v} \phi(v) F(v),$$

where $F(v)$ is a sum (with appropriate signs) of the six $f(e)$ terms for e incident to v. We show $F(v) = 0$.

Let e_i^-, e_i^+ denote the edges incident to v and oriented parallel to η_i. Let e_i^+ be the one for which the orientation parallel to η_i coincides with the orientation even to odd. Then

$$F(v) = \sum_{i=1}^{3} f(e_i^+) - f(e_i^-).$$

But this is equal to $\sum_{\tilde{e} \ni v} f(\tilde{e})$, where the edges $\tilde{e}$ incident to v are all oriented even to odd. Therefore $F(v) = 0$ since f is divergence-free as a discrete vector field, see Equation (2.5). □

Next we prove a lemma about generalized Wasserstein distance for signed measures, in the case that both signed measures correspond to either tiling or asymptotic flows. This is an elementary result that we will use repeatedly.

Lemma 2.5.14. *Suppose that μ and ν are measures supported on a common compact set K corresponding to components of tiling or asymptotic flows. Suppose there is a partition of K into sets $\mathcal{B} = \{B_1, ..., B_M\}$ of diameter at most ϵ such that $\left|\mu(B) - \nu(B)\right| < \delta$ for all $B \in \mathcal{B}$. If one of the measures corresponds to a scale n tiling flow, then we require that $\frac{1}{n} \leq \epsilon$. Then*

$$\mathbb{W}_1^{1,1}(\mu, \nu) \leq M(10\epsilon^4 + \delta).$$

Proof. Let $\mu = \mu_+ - \mu_-$ and $\nu = \nu_+ - \nu_-$ be decompositions into positive measures and recall that

$$\mathbb{W}_1^{1,1}(\mu, \nu) = W_1^{1,1}(\mu_+ + \nu_-, \mu_- + \nu_+).$$

Let $\tilde{\mu} = \mu_+ + \nu_-$ and $\tilde{\nu} = \mu_- + \nu_+$. To get an upper bound for the distance, it suffices to give a method for redistributing and deleting mass to transform $\tilde{\mu}$ into $\tilde{\nu}$.

We proceed as follows: transform $\tilde{\mu} \mid_{B_1}$ into $\tilde{\nu} \mid_{B_1}$, the cost of this is at most $W_1^{1,1}(\tilde{\mu} \mid_{B_1}, \tilde{\nu} \mid_{B_1})$. Denote the new version of $\tilde{\mu}$ by $\tilde{\mu}'$. $\tilde{\mu}'$ will agree with $\tilde{\mu}$ on $R \setminus B_1$ and with $\tilde{\nu}$ on B_1. Next transform $\tilde{\mu}'$ into $\tilde{\nu}$ on B_2. This will cost at most $W_1^{1,1}(\tilde{\mu}' \mid_{B_2}, \tilde{\nu} \mid_{B_2}) \leq W_1^{1,1}(\tilde{\mu} \mid_{B_2}, \tilde{\nu} \mid_{B_2})$ with equality if B_2 is disjoint from B_1. Iterating this we get that

$$\mathbb{W}_1^{1,1}(\mu, \nu) \leq \sum_{j=1}^{k} W_1^{1,1}(\tilde{\mu} \mid_{B_j}, \tilde{\nu} \mid_{B_j})$$

Now we just have to compute the distance for a single B_j. First spend $\delta > 0$ to delete the difference in mass on B_j. The total mass of μ, ν on any $B \in \mathcal{B}$ is bounded by $10\epsilon^3$, and the furthest it would need to move is ϵ. Therefore

$$W_1^{1,1}(\tilde{\mu} \mid_{B_j}, \tilde{\nu} \mid_{B_j}) \leq 10\epsilon^4 + \delta.$$

Summing over j gives the result. □

An inverse version of the bound in Lemma 2.5.14 also holds, but with a constant depending on the small region B.

Lemma 2.5.15. *Suppose $B \subset R$ is a connected region with piecewise smooth boundary. If μ, ν are component measures of tiling or asymptotic flows and $\mathbb{W}_1^{1,1}(\mu, \nu) < \delta$, then there is a constant $C(B)$ depending only on B such that*

$$\mathbb{W}_1^{1,1}(\mu \mid_B, \nu \mid_B) < \delta + (C(B) + 1)\delta^{1/2}.$$

Remark 2.5.16. The constant $C(B)$ is not hard to understand and control. The term $C(B)\delta^{1/2}$ is bounded by 2 times the volume of the $\delta^{1/2}$ annulus with inner boundary ∂B.

Proof. The redistribution, addition, and deletion of mass $\mu \to \nu$ gives a redistribution $\mu \mid_B \to \nu \mid_B$, except any mass moved into or out of B now becomes an L^1 cost rather than a cost proportional to distance moved. Let $f(r)$ be the amount of flow moved distance r into or out of B by the $\mu \to \nu$ redistribution. Then

$$\mathbb{W}_1^{1,1}(\mu \mid_B, \nu \mid_B) \leq \delta + \int_0^\infty f(r)dr.$$

On the other hand,

$$\int_0^\infty r f(r)dr < \delta.$$

We split the integral we want to bound into two pieces:

$$\int_0^\infty f(r)dr = \int_0^{\delta^{1/2}} f(r)dr + \int_{\delta^{1/2}}^\infty f(r)dr.$$

Since μ, ν are measures corresponding to components of tiling or asymptotic flows, we have

$$\int_0^{\delta^{1/2}} f(r)dr \leq C(B)\delta^{1/2}$$

(this quantity is proportional to the volume of the $\delta^{1/2}$ annulus around B) and

$$\delta^{1/2} \int_{\delta^{1/2}}^{\infty} f(r)dr \leq \int_{\delta^{1/2}}^{\infty} rf(r)dr < \delta.$$

Combining these gives the desired bound. □

Lemma 2.5.17. *Let ν_n be a sequence of signed measures supported in R which converges in $\mathbb{W}_1^{1,1}$ to another measure ν. Further suppose the ν_n are absolutely continuous with respect to 3-dimensional Lebesgue measure, and their densities $g_n(x)$ take values in $[-m, M]$. Then ν is also absolutely continuous with respect to 3-dimensional Lebesgue measure.*

Proof. By Corollary 2.5.6, if $\mathbb{W}_1^{1,1}(\nu_n, \nu) \to 0$ as $n \to \infty$, then ν_n converges to ν in the weak topology.

Define an operator $Q : C^\infty(R) \to \mathbb{R}$ by integrating against ν:

$$Q(h) := \int h \, \mathrm{d}\nu.$$

Using Cauchy-Schwarz and the fact that $d\nu_n = g_n(x)\,\mathrm{d}x$ we have that

$$Q(h) = \lim_{n\to\infty} \int h\, g_n \,\mathrm{d}x \leq \limsup_{n\to\infty} \|h\|_{L^2(R)} \|g_n\|_{L^2(R)} \leq \mathrm{Vol}(R)^{1/2}(M+m)\|h\|_{L^2}$$

Therefore Q extends to an operator on $L^2(R)$. By the Riesz representation theorem this means there exists an L^2 function g such that $Q(h) = \langle h, g\rangle = \int h\, g \,\mathrm{d}x$. Therefore $\mathrm{d}\nu(x) = g(x)\mathrm{d}x$, and ν is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^3$. □

Proposition 2.5.18. *Suppose that $f, g \in AF(R)$ are continuous and satisfy $|f(x) - g(x)| < \epsilon$ for all $x \in R$. Then $d_W(f, g) < \epsilon\, vol(R)$.*

Proof. The d_W-distance from f to g is bounded by adding and subtracting mass for each of the component functions. Since pointwise they differ by at most ϵ, $d_W(f, g) \leq \epsilon\,\mathrm{vol}(R)$. □

2.5.3 Main theorems

Here we prove two of the theorems mentioned at the beginning of the section. First we show that fine-mesh limits of tiling flows are asymptotic flows.

Theorem 2.5.19. *Let $R \subset \mathbb{R}^3$ be a compact region which is the closure of a connected domain and has piecewise-smooth boundary. Let $f_n \in TF_{m_n}(R)$ be a free-boundary tiling flow on R at scale m_n with m_n going to infinity with n. Any d_W-subsequential limit of tiling flows $f_* = \lim_{k\to\infty} f_{n_k}$ is in $AF(R)$.*

Let $\mu_k = (\mu_k^1, \mu_k^2, \mu_k^3)$ be the measures corresponding to f_{n_k} and let $\mu_* = (\mu_*^1, \mu_*^2, \mu_*^3)$ be the measures corresponding to f_*. The main idea of the proof is to smoothen the measures $(\mu_k^1, \mu_k^2, \mu_k^3)$ corresponding to the tiling flow in an especially nice way, then apply Lemma 2.5.17 to say that their limits $(\mu_*^1, \mu_*^2, \mu_*^3)$ are absolutely continuous with respect to 3-dimensional Lebesgue measure. This shows that f_* is a measurable flow on R. Using our well-chosen smoothings, we will show that f_* has the other properties that an asymptotic flow must have. To make the argument easier to digest, we break down the construction of the smoothing into two lemmas.

Lemma 2.5.20. *Let $S_N = [0, N-1]^3$ (note that there are N vertices from $\mathbb{Z}^3$ on each edge of S_N). Let τ be a tiling of $\mathbb{Z}^3$ with tiling flow f_τ, and let (μ_1, μ_2, μ_3) be the corresponding measures. Then*

$$\frac{1}{Vol(S_N)}\left(\int_{S_N} \mathrm{d}\mu_1, \int_{S_N} \mathrm{d}\mu_2, \int_{S_N} \mathrm{d}\mu_3\right)$$

is valued in $(1 + O(1/N))\mathcal{O}$.

Proof. Let $E(S_N)$ denote the edges intersecting in S_N. All edges intersecting S_N in more than one point are contained in it and have length 1.

The measure μ_1 is supported on the edges parallel to η_1, and thus

$$\int_{S_N} \mathrm{d}\mu_1 = \sum_{E(S_N) \ni e \| \eta_1} 2f(e),$$

for edges e oriented parallel to η_1. The results for μ_2, μ_3 are analogous.

For each $i = 1, 2, 3$, there are $N^2(N-1)$ edges from $\mathbb{Z}^3$ contained in S_N parallel to η_i. This number is always even, so half of these edges from $(N^2(N-1)/2)$ have even-to-odd orientation parallel to η_i and half have the opposite orientation. Let α_+ be the number of even-to-odd oriented tiles in $\tau \cap E(S_N)$ and α_- be the number of odd-to-even oriented tiles. Then

$$\int_{S_N} \mathrm{d}\mu_1 = \sum_{E(S_N) \ni e \| \eta_i} 2f(e) = 2(\alpha_+ - \alpha_-).$$

Let s_i denote the fraction of tiles in the $+\eta_i$ direction in S_N minus the fraction of tiles in the $-\eta_i$ direction in S_N. Note that irrespective of the tiling we have that number of tiles in S_N is $N^2(N-1)/2 + O(N^2)$. Thus it follows that

$$s_1 = \frac{2(\alpha_+ - \alpha_-)}{N^2(N-1) + O(N^2)} = \frac{1}{N^2(N-1) + O(N^2)} \int_{S_N} \mathrm{d}\mu_1.$$

A similar equation holds for s_2, s_3. We have that $(s_1, s_2, s_3) \in \mathcal{O}$. Thus

$$\begin{aligned}\frac{1}{\mathrm{Vol}(S_N)}\left(\int_{S_N} \mathrm{d}\mu_1, \int_{S_N} \mathrm{d}\mu_2, \int_{S_N} \mathrm{d}\mu_3\right) &= \frac{N^2(N-1) + O(N^2)}{(N-1)^3}(s_1, s_2, s_3) \\ &= (1 + \frac{O(N^2)}{(N-1)^3})(s_1, s_2, s_3),\end{aligned}$$

for some constant C which is in $(1 + O(1/N))\mathcal{O}$. □

If we smoothen the measures corresponding to the tiling flow over a partition consisting of boxes of the form in Lemma 2.5.20, we can construct smoothings that satisfy a very nice list of properties.

Lemma 2.5.21. *Let (μ_1, μ_2, μ_3) be measures corresponding to a tiling flow $f \in TF_n(R)$. For any $\epsilon > 0$, there exists $\nu = (\nu_1, \nu_2, \nu_3)$ satisfying the following properties:*

1. *ν_i is supported in R for all $i = 1, 2, 3$;*
2. *ν_i has a density $g_i(x)$ with respect to Lebesgue measure on $\mathbb{R}^3$ for $i = 1, 2, 3$;*
3. *$g = (g_1, g_2, g_3)$ is valued in $(1 + O(\epsilon))\mathcal{O}$ as a distribution;*
4. *$d_W(\mu, \nu) < C(R)(\epsilon n)^{-1}$ where $C(R)$ is a constant depending only on R.*

Proof. Choose N such $N = \lfloor \frac{1}{\epsilon} \rfloor$ and a partition of cubes $\mathcal{B} = \{B_1, ..., B_M\}$ that cover R, where each B_i is an $N \times N \times N$ cube in $\frac{1}{n}\mathbb{Z}^3$ (we define the flow f to be 0 outside R). This can be done so that $M \sim n^3/N^3$. For $i = 1, 2, 3$ and all $B \in \mathcal{B}$, define

$$C_B^i := \frac{1}{\mathrm{Vol}(B)} \int_B \mathrm{d}\mu_i = \frac{(N-1)^3}{n^3} \int_B \mathrm{d}\mu_i.$$

Define the densities of ν_i by

$$g_i(x) = C_B^i \qquad \forall\, x \in B \cap R.$$

This satisfies properties 1 and 2. By Lemma 2.5.20, (g_1, g_2, g_3) is valued in $(1+O(1/N))\mathcal{O} = (1+O(\epsilon))\mathcal{O}$ as a distribution which completes property 3. Finally by Lemma 2.5.14 applied to the partition $\mathcal{B}$, we have that

$$d_W(\mu, \nu) \leq M(N/n)^4 \leq C(R)(n/N)^3(N/n)^4 \leq C(R)\frac{1}{\epsilon n}$$

where $C(R)$ is a constant depending only on R. □

We now return to the proof of the theorem.

Proof of Theorem 2.5.19. Fix $\epsilon > 0$. Recall that $\mu_k = (\mu_k^1, \mu_k^2, \mu_k^3)$ is the triple of measures corresponding to f_{n_k}. Choose $\epsilon_k = n_k^{-1/2}$ and let $\nu_k = (\nu_k^1, \nu_k^2, \nu_k^3)$ be the measures constructed in Lemma 2.5.21 for $\epsilon_k > 0$. For k large enough so that $d_W(\mu_*, \mu_k) < \epsilon$, by the triangle inequality

$$d_W(\mu_*, \nu_k) \leq d_W(\mu_*, \mu_k) + d_W(\mu_k, \nu_k) \leq \epsilon + C(R)n_k^{-1/2}.$$

Therefore the triple of measures ν_k also converges to μ_* in d_W. By Lemma 2.5.17, there are functions f_*^i such that $\mu_*^i = f_*^i(x)\mathrm{d}x$ for $i = 1, 2, 3$, so f_* is a measurable vector field. It remains for us to check the additional properties to show that f_* is an asymptotic flow.

Since the ν_k are supported in R for all k, so is f_*. By Lemma 2.5.21, the densities of ν_k are valued in $(1 + O(\epsilon_k))\mathcal{O}$, so the densities of μ_* are valued in $\mathcal{O}$ (any open neighborhood is a continuity set, so we get that the averages of (f_*^1, f_*^2, f_*^3) are valued in $(1 + O(\epsilon_k))\mathcal{O}$ for

all ϵ_k. This plus the Lebesgue differentiation theorem imply that f_* is valued in $\mathcal{O}$). On the other hand, convergence in d_W implies weak convergence of the component measures (Corollary 2.5.6). Since μ_k is divergence-free in the sense of distributions on the interior of R (Proposition 2.5.13) for all k, μ_* is also divergence-free in the sense of distributions on the interior of R. Thus $f_* \in AF(R)$. □

Theorem 2.5.22. *The metric space $(AF(R), d_W)$ is compact.*

Proof. By Proposition 2.5.7, the space of triples of measures absolutely continuous with respect to Lebesgue measure, supported in R, and valued in $\mathcal{O}$ is compact. Since $(AF(R), d_W)$ is a subspace of this, it suffices to show that it is closed. Suppose that $\mu_n = (\mu_n^1, \mu_n^2, \mu_n^3)$ is a sequence in $AF(R)$ with d_W-limit $\mu_* = (\mu_*^1, \mu_*^2, \mu_*^3)$. By Lemma 2.5.17, μ_*^i has a density $g_*^i(x)$ for each $i = 1, 2, 3$. Since convergence in $\mathbb{W}_1^{1,1}$ implies weak convergence (Corollary 2.5.6), $g_* = (g_*^1, g_*^2, g_*^3)$ is divergence-free. To show that g_* is valued in $\mathcal{O}$, note that any open ball $B \subset R$ is a continuity set, so since $\frac{1}{\mathrm{Vol}(B)}(\int_B \mathrm{d}\mu_n^1, \int_B \mathrm{d}\mu_n^2, \int_B \mathrm{d}\mu_n^3) \in \mathcal{O}$, the average of g_* over B is also valued in $\mathcal{O}$. Thus g_* is valued in $\mathcal{O}$ by the Lebesgue differentiation theorem. Thus $(AF(R), d_W)$ is compact. □

Now we will prove that any asymptotic flow can be approximated by a smooth flow which is divergence-free on a slightly smaller region. This is a standard construction which we provide for completeness as it will be used in the next subsection. Essentially all we need to do is to convolve the asymptotic flow with an appropriate smooth bump function. For this, given a region R and $\epsilon > 0$ we define

$$R_\epsilon := \{x \in R \ : \ d(x, \partial R) \geq \epsilon\}. \tag{2.15}$$

We will denote the smooth asymptotic flows on a region R by $AF^\infty(R) \subset AF(R)$. Given $f \in AF(R), g \in AF(R_\epsilon)$, we say that $d_W(f, g)$ is the distance between f and g, where g is extended to be 0 on $R \setminus R_\epsilon$.

Proposition 2.5.23. *Fix $f \in AF(R)$. For all $\epsilon > 0$ small enough, there is a smooth asymptotic flow $g \in AF^\infty(R_\epsilon)$ such that*

$$d_W(f, g) < K\sqrt{\epsilon}$$

where K is a constant depending only on R.

Proof. Consider a bump function $\phi \in C^\infty(B_\epsilon(0))$, that is, it is a non-negative smooth function such that $\phi|_{\partial B_\epsilon(0)} = 0$ and $\int_{B_\epsilon(0)} \phi(x)\,\mathrm{d}x = 1$.

Let $g = f * \phi|_{R_\epsilon}$ and suppose that ψ is a smooth test function with compact support in the interior of R_ϵ. To check that g is divergence-free in the interior of R_ϵ we look at the integral

$$\begin{aligned}\int_{R_\epsilon} (\nabla\psi \cdot g)(x)\,\mathrm{d}x &= \int_{R_\epsilon}\int_{B_\epsilon(0)} (\nabla\psi \cdot f)(x-y)\phi(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{B_\epsilon(0)}\int_{R_\epsilon} (\nabla\psi \cdot f)(x-y)\phi(y)\,\mathrm{d}x\,\mathrm{d}y = 0.\end{aligned}$$

Here the last equality uses that f is divergence-free in the interior of R. We have that $g(x) = f * \phi(x)$ is an average of elements in $\mathcal{O}$. Since $\mathcal{O}$ is convex it follows that g takes values in $\mathcal{O}$. Finally we estimate $d_W(f, g)$. The amount of mass which we might have to delete or add from $R \setminus R_\epsilon$ is bounded by $6(\mathrm{Vol}(R \setminus R_\epsilon)) \leq c\epsilon^2$ where c depends only on ∂R. Now let (μ_1, μ_2, μ_3) and (ν_1, ν_2, ν_3) be the measures corresponding to f and g respectively. Let $\delta > 0$. We have that if B is a box of side length δ contained in R_ϵ then

$$\nu_i(B) = \int_B \int_{B_\epsilon(0)} f_i(x - y)\phi(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{B_\epsilon(0)} \mu_i(B - y)\phi(y)\,\mathrm{d}y.$$

It follows that $|\nu_i(B) - \mu_i(B)|$ is less than the volume of the annular region around B of radius ϵ. Thus we have that

$$|\nu_i(B) - \mu_i(B)| < C\epsilon\delta^2$$

where C is independent of ϵ and δ. Partition R_ϵ into boxes $B_1, B_2, \ldots B_M$; where $M \sim \delta^{-3}$. By Lemma 2.5.14 we have that

$$\mathbb{W}_1^{1,1}(\mu_i, \nu_i) < M(10\delta^4 + C\epsilon\delta^2) + c\epsilon^2 \sim \delta^{-3}(10\delta^4 + C\epsilon\delta^2) + c\epsilon^2$$

Since δ is a free parameter, we can take $\delta \sim \sqrt{\epsilon}$ to complete the proof. □

2.5.4 Boundary values of asymptotic flows

In this section we define the boundary values of asymptotic flows on ∂R. In fact we do something slightly more general, and define the restriction of an asymptotic flow (or tiling flow in the next subsection) on a whole class of surfaces, namely

$$\mathbb{S}(R) = \{\text{compact piecewise smooth surfaces contained in } R\}$$

Note that R is closed, so $\partial R \in \mathbb{S}(R)$. This general set up will make things easier to prove. We also use the trace operator for other surfaces in Section 2.8.6. Recall that $AF^\infty(R) \subset AF(R)$ is the smooth asymptotic flows.

Definition 2.5.24. We define the *trace operator* on smooth asymptotic flows

$$T : AF^\infty(R) \times \mathbb{S}(R) \to \mathcal{M}^s(R).$$

by

$$T(f, S)(x) = \langle f(x), \xi_S(x)\rangle \mathrm{d}\sigma_S(x), \qquad x \in S,$$

where $\mathrm{d}\sigma_S$ denotes surface area measure on S and $\xi_S(x)$ denotes the L^2 unit normal vector to S at x.

We show that $T(\cdot, S)$ extends to a uniformly continuous map $(AF(R), d_W) \to (\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$ for all $S \in \mathbb{S}(R)$. We do this in three main steps:

- Show that $T(\cdot, S)$ is uniformly continuous on $AF^\infty(R)$ when S is a small patch.

- Extend this result to $AF(R)$ by approximation and compatibility results. Since we don't know if $AF^\infty(R)$ is dense in $AF(R)$, this requires slightly more care.
- Extend the uniform continuity result to general $S \in \mathbb{S}(R)$ by putting together the patches.

Proposition 2.5.25. *Suppose $S \in \mathbb{S}(R)$ and is such that there exists a nonzero vector v and a parameter $\theta > 0$ for which $S + tv$ is contained in R and disjoint for all $0 \le t \le \theta$. Then for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $f, g \in AF^\infty(R)$ with $d_W(f, g) < \delta$, we have that*

$$\mathbb{W}_1^{1,1}(T(f, S), T(g, S)) < \epsilon.$$

Proof. Fix two parameters $\gamma_1, \gamma_2 > 0$ which we will specify at the end of the proof. Partition S into patches $\alpha_1, \dots \alpha_M$ such that

- α_i is a smooth surface with piecewise smooth boundary for all $i = 1, ..., M$.
- α_i has diameter at most γ_1 for all i, and $M \le C\gamma_1^{-2}$ for some constant C depending on S.
- Let $\alpha_i(t) := \alpha_i + tv$ for $0 \le t \le \theta$. For all $i = 1, ..., M$, $\alpha_i(t) \cap \alpha_i(s) = \emptyset$ for $s \ne t$.

Let $\mu_f = T(f, S)$, $\mu_g = T(g, S)$ and define $\Delta > 0$ by

$$\Delta := \sup_{1 \le i \le M} \left| \mu_f(\alpha_i) - \mu_g(\alpha_i) \right|. \tag{2.16}$$

By the two-dimensional version of Lemma 2.5.14,

$$\mathbb{W}_1^{1,1}(T(f, S), T(g, S)) \le M(10\gamma_1^3 + \Delta) \le 10C\gamma_1 + C\gamma_1^{-2}\Delta. \tag{2.17}$$

Note that the power of γ_1 is 3 instead of 4 because S is two-dimensional. It remains to find a bound for Δ in terms of $d_W(f, g)$.

If $h \in AF^\infty(R)$, then h is divergence-free and hence its flux through any closed surface is zero. This implies that there exists a threshold $\theta_{\gamma_2} \in (0, 1)$ such that for all $0 \le t \le \theta_{\gamma_2}$ and $h \in AF^\infty(R)$,

$$\sup_{1 \le i \le M} \left| \int_{\alpha_i} \langle h(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i}(x) - \int_{\alpha_i(t)} \langle h(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i(t)}(x) \right| \le \gamma_2. \tag{2.18}$$

Here $\xi(x)$ is the normal vector on the surfaces $\alpha_i(t)$ with appropriate orientation and we are applying the divergence theorem to the boundary of the region $U_i = \cup_{s=0}^t \alpha_i(t)$. Since h takes values in the compact set $\mathcal{O}$, the threshold θ_{γ_2} can be taken independent of h. Applying this to $f, g \in AF^\infty(R)$ it follows that

$$\sup_{1 \le i \le M} \left| \int_{\alpha_i} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i}(x) - \int_{\alpha_i(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i(t)}(x) \right| \le 2\gamma_2$$

for all $0 \le t \le \theta_{\gamma_2}$. Observe that the first term in the inequality is $\mu_f(\alpha_i) - \mu_g(\alpha_i)$. Integrating over $t \in (0, \theta_{\gamma_2})$,

$$\left| \left| \int_0^{\theta_{\gamma_2}} \left(\int_{\alpha_i(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i(t)}(x) \right) \mathrm{d}t \right| - \theta_{\gamma_2}(\mu_f(\alpha_i) - \mu_g(\alpha_i)) \right| \le 2\theta_{\gamma_2}\gamma_2.$$

Taking the supremum over $i = 1, ..., M$ we get that

$$\Delta \le 2\gamma_2 + \frac{1}{\theta_{\gamma_2}} \sup_{1 \le i \le M} \left| \int_0^{\theta_{\gamma_2}} \left(\int_{\alpha_i(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha_i(t)}(x) \right) \mathrm{d}t \right|.$$

Suppose the supremum on the right hand side of the equation is achieved by the index i, and let $\alpha(t) := \alpha_i(t)$ to simplify notation. Then plugging this into (2.17) gives

$$\mathbb{W}_1^{1,1}(T(f,S), T(g,S)) \tag{2.19}$$

$$\le 10C\gamma_1 + 2C\gamma_1^{-2}\gamma_2 + C\gamma_1^{-2}\theta_{\gamma_2}^{-1} \left| \int_0^{\theta_{\gamma_2}} \left(\int_{\alpha(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha(t)}(x) \right) \mathrm{d}t \right|.$$

Since α_i is smooth, by an appropriate change of variables we can rewrite the integral above as an integral over $U = \cup_{t=0}^{\theta_{\gamma_2}} \alpha(t)$.

$$\begin{aligned} \int_0^{\theta_{\gamma_2}} \left(\int_{\alpha(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha(t)}(x) \right) \mathrm{d}t &= \int_U \langle f(x) - g(x), \xi(x) \rangle \varphi(x) \, \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \\ &= \sum_{k=1}^{3} \int_U (f_k(x) - g_k(x)) \xi_k(x) \varphi(x) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}x_3. \end{aligned}$$

Here $\varphi(x)$ is the factor coming from the Jacobian in the change of variables. Since $\alpha(t)$ is smooth for all $t \in [0, \theta_{\gamma_2}]$, $\varphi(x)$ and $\xi(x)$ are both smooth functions on U, so $\psi_k(x) := \xi_k(x)\varphi(x)$ is a smooth and therefore Lipschitz function on U. Let λ denote the maximum of Lipchitz constants of ψ_k, $k = 1, 2, 3$. Then by the dual definition of the Wasserstein metric (Proposition 2.5.5),

$$\int_0^{\theta_{\gamma_2}} \left(\int_{\alpha(t)} \langle f(x) - g(x), \xi(x) \rangle \, \mathrm{d}\sigma_{\alpha(t)}(x) \right) \mathrm{d}t \le \lambda \, d_W(f \mid_U, g \mid_U).$$

By Lemma 2.5.15, there is a constant $C(U)$ such that if $d_W(f,g) < \delta$ then $d_W(f \mid_U, g \mid_U) < \delta + C(U)\delta^{1/2}$. Therefore substituting this in to Equation (2.19) gives that if $d_W(f,g) < \delta$ then

$$\mathbb{W}_1^{1,1}(T(f,S), T(g,S)) \le 10C\gamma_1 + 2C\gamma_1^{-2}\gamma_2 + C\gamma_1^{-2}\theta_{\gamma_2}^{-1}\lambda(\delta + C(U)\,\delta^{1/2}).$$

Taking $\gamma_2 = \gamma_1^3$ completes the proof. □

We will now prove that perturbing S by a small amount does not change the trace very much.

Proposition 2.5.26. *Let $S \in \mathbb{S}(R)$ be a surface satisfying the conditions of Proposition 2.5.25 for some vector v and threshold $\theta > 0$. For all $\epsilon > 0$ there exists $\delta > 0$ such that for all $t < \delta$ and $f \in AF^\infty(R)$, we have that*

$$\mathbb{W}_1^{1,1}(T(f,S), T(f,S+tv)) < \epsilon.$$

Proof. This proof is much simpler than that of Proposition 2.5.25. As in that proof, we take parameters $\gamma_1, \gamma_2 > 0$ to be fixed later and choose a partition of $\alpha_1 \ldots, \alpha_M$ of S such that α_i has diameter less than γ_1 for all i, and $M \leq C\gamma_1^{-2}$ for some $C > 0$ independent of γ_1. In addition, by choosing a larger constant C if necessary we can assume that

$$\sigma_S(\alpha_i) \leq C\gamma_1^2.$$

As in (2.18), there exists a threshold $\theta_{\gamma_2} \in (0,1)$ such that for all $0 \leq t \leq \theta_{\gamma_2}$ and $f \in AF^\infty(R)$,

$$\sup_{1 \leq i \leq M} \left| T(f,S)(\alpha_i) - T(f,S+tv)(\alpha_i(t)) \right| \leq \gamma_2. \tag{2.20}$$

To get an upper bound for the distance between the traces on each patch, we give a method for redistributing, adding, and deleting mass to transform $T(f,S) \mid_{\alpha_i} = T(f,\alpha_i)$ into $T(f,\alpha_i(t))$. Note that these are both are signed measures absolutely continuous with respect to σ_S and σ_{S+tv} respectively, and both have densities bounded between -1 and 1. We can transform one to the other by (1) adding γ_2 flow, (2) moving flow distance at most $t + \gamma_1$. There is at most $4\sigma_S(\alpha_i)$ total flow from both measures. Hence

$$\mathbb{W}_1^{1,1}(T(f,S)|_{\alpha_i}, T(f,S+tv)|_{\alpha_i(t)}) \leq \gamma_2 + 4(t+\gamma_1)(\sigma_S(\alpha_i)) \leq \gamma_2 + 4C(t+\gamma_1)\gamma_1^2.$$

As in Lemma 2.5.14, by triangle inequality

$$\mathbb{W}_1^{1,1}(T(f,S), T(f,S+tv)) \quad \leq \quad M\left(\gamma_2 + 4C(t+\gamma_1)\gamma_1^2\right) \leq C\gamma_1^{-2}\gamma_2 + 4C^2(t+\gamma_1).$$

As before, by setting $\gamma_2 = \gamma_1^3$, the result follows. □

Using Proposition 2.5.25 and Proposition 2.5.26, when $S \in \mathbb{S}(R)$ satisfies the conditions of Proposition 2.5.25 for a vector v and threshold $\theta > 0$, we can extend the definition of $T(S,\cdot)$ to all of $AF(R)$ as follows. Given $t > 0$, there exists K such that for all $k \geq K$, $S(t) \subset R_{1/k}$. Here recall that

$$R_{1/k} = \{x \in R \ : \ d(x, \partial R) \geq 1/k\}.$$

By Proposition 2.5.23, for any $f \in AF(R)$ we can find a sequence $g_k \in AF^\infty(R_{1/k})$ such that $d_W(f, g_k) \to 0$ as $k \to \infty$. For $t > 0$, we define

$$T(f, S(t)) := \lim_{k \to \infty} T(g_k, S(t)) \tag{2.21}$$

where the limit is taken with respect to the metric d_W on flows supported in R. Note that if $R' \subset R$, then the projection map $AF^\infty(R) \to AF^\infty(R')$ given by $f \mapsto f \mid_{R'}$ is

continuous. By Proposition 2.5.25, T is uniformly continuous on $AF^\infty(R_{1/k})$ for any $k > 0$, so the limit in Equation (2.21) is independent of the approximating sequence $(g_k)_{k\geq 1}$ and converges to $T(f, S + tv)$ if $f \in AF^\infty(R)$. Since $(AF(R), d_W)$ is compact (Theorem 2.5.22), $T(\cdot, S(t)) : AF(R) \to \mathcal{M}^s(R)$ is a uniformly continuous map for $t > 0$.

Further, by a variant of Lemma 2.5.17 where 3-dimensional Lebesgue measure is replaced by the surface area measure $\sigma_{S(t)}$, for any $f \in AF(R)$, $T(f, S(t))$ is a signed measure absolutely continuous to $\sigma_{S(t)}$ with density bounded between -1 and 1. By Proposition 2.5.26, for any $\epsilon > 0$, there exists $\delta > 0$ so that if $t, s > 0$ and $|t - s| < \delta$ then

$$\mathbb{W}_1^{1,1}(T(f, S(s)), T(f, S(t)) < \epsilon. \tag{2.22}$$

Thus we take another limit in the d_W topology to define

$$T(f, S) := \lim_{t\to 0} T(f, S(t)).$$

Since $AF(R)$ is compact and the extension above is continuous, we get analogs of Proposition 2.5.25 and Proposition 2.5.26 for $AF(R)$.

Proposition 2.5.27. *If $S \in \mathbb{S}(R)$ is a surface satisfying the conditions of Proposition 2.5.25 for a vector v and threshold $\theta > 0$, then*

1. *Using the extension described above,*

$$T(\cdot, S) : (AF(R), d_W) \to (\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$$

   is uniformly continuous.

2. *Given $\epsilon > 0$ there exists $\delta > 0$ such that for all $t < \delta$ and $f \in AF(R)$,*

$$\mathbb{W}_1^{1,1}(T(f, S), T(f, S(t)) < \epsilon.$$

3. *For any $f \in AF(R)$, $T(f, S)$ is a signed measure absolutely continuous with respect to the surface area measure σ_S with density bounded between -1 and 1.*

We now prove a compatibility result for traces on overlapping surfaces. Using this, we can extend the continuity theorems to the trace operator for any $S \in \mathbb{S}(R)$ by cutting S in patches which satisfy the conditions of Proposition 2.5.25.

Lemma 2.5.28. *Suppose that $S, S' \in \mathbb{S}(R)$ are surfaces satisfying the conditions of Proposition 2.5.25 such that $S' \subset S$. Then for any $f \in AF(R)$, then*

$$T(f, S) \mid_{S'} = T(f, S').$$

Proof. If $f \in AF^\infty(R)$, then the result follows immediately from the form given in Definition 2.5.24. For general $f \in AF(R)$ this follows from Proposition 2.5.27. □

Finally we put together the pieces to prove uniform continuity of $T(\cdot, S)$ on $AF(R)$ for any $S \in \mathbb{S}(R)$.

Proposition 2.5.29. *For all $S \in \mathbb{S}(R)$, $T(\cdot, S)$ extends to a uniformly continuous map from $(AF(R), d_W)$ to $(\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$. Further, for all $f \in AF(R)$, $T(f, S)$ is a signed measure on S absolutely continuous to the surface measure σ_S with density bounded between -1 and 1.*

Remark 2.5.30. In particular this holds for $S = \partial R$.

Proof. Fix $S \in \mathbb{S}(R)$. We can cover S with finitely many open surfaces $S_1, ..., S_k$ which all satisfy the conditions of Proposition 2.5.25. By Lemma 2.5.28, if $S_i \cap S_j \neq \emptyset$, then for all $f \in AF(R)$,

$$T(f, S_i) \mid_{S_i \cap S_j} = T(f, S_j) \mid_{S_i \cap S_j}$$

Therefore we can define the trace operator $T(S, \cdot)$ for any $S \in \mathbb{S}(R)$ by

$$T(f, S) \mid_{S_i} = T(f, S_i).$$

By Proposition 2.5.27, $T(f, S)$ is uniformly continuous as a function of $f \in AF(R)$ and has the desired form. □

With this machinery, we can define the space of asymptotic flows with a fixed boundary value.

Definition 2.5.31. We say that $b \in \mathcal{M}^s(R)$ is a *boundary asmyptotic flow* if $b \in T(AF(R), \partial R)$. Further, we define $AF(R, b)$ to be the space of asymptotic flows on R with boundary value b, i.e. $f \in AF(R)$ such that $T(f, \partial R) = b$.

As a corollary of Proposition 2.5.29 and Theorem 2.5.22 we get the following.

Corollary 2.5.32. *The metric space $(AF(R, b), d_W)$ is compact.*

2.5.5 Boundary values of tiling flows

Next we define the trace operator on tiling flows, and show that it is compatible with the definition for asymptotic flows. Suppose $f \in TF_n(R)$ and that $S \in \mathbb{S}(R)$ is a surface which intersects the lattice $\frac{1}{n}\mathbb{Z}^3$ transversely, i.e. S does not contain any vertices of $\frac{1}{n}\mathbb{Z}^3$ (if S contains a vertex, we translate the lattice slightly so that it does not). As usual let e denote an edge from $\frac{1}{n}\mathbb{Z}^3$ oriented from even to odd, and let $\xi(x)$ denote the normal vector to S at x.

Definition 2.5.33. If $f \in TF_n(R)$ and $S \in \mathbb{S}(R)$ is a surface intersecting $\frac{1}{n}\mathbb{Z}^3$ transversely, we define

$$T(f, S) = \sum_e \frac{2 \operatorname{sign}\langle \xi(x), e \rangle}{n^2} f(e) \delta(e \cap S).$$

Note that since S is transverse to $\frac{1}{n}\mathbb{Z}^3$, if $e \cap S$ is nonempty it is a single point.

Using Definition 2.5.33 for tiling flows and Definition 2.5.24 for asymptotic flows, the final goal of this section is to show that

$$T(\cdot, S) : (AF(R) \cup TF(R), d_W) \to (\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$$

is uniformly continuous for any $S \in \mathbb{S}(R)$ (see Theorem 2.5.39). The sequence of results in this section building up to this mirrors the sequence of results in the previous section. The discrete setting makes things slightly more complicated. The main new step is that we start by proving a result for the trace on planes, and extend to more general surfaces by approximating them with planes. Throughout, we assume that any surface S we consider intersects $\frac{1}{n}\mathbb{Z}^3$ transversely. Any time it does not, the trace is defined by perturbing the lattice slightly so that it does and then using Definition 2.5.33.

Proposition 2.5.34. *Suppose $P \in \mathbb{S}(R)$ is a compact piece of a plane with normal vector ξ, and there exists a threshold $\theta > 0$ such that $P(t) = P + t\xi$ is contained in R for all $t \in [0, \theta]$. Let $f_n \in TF_n(R)$ be a sequence of tiling flows such that $d_W(f_n, f) \to 0$ as $n \to \infty$ for some $f \in AF^\infty(R)$. Then*

$$\lim_{n\to\infty} \mathbb{W}_1^{1,1}(T(f_n, P), T(f, P)) = 0.$$

Remark 2.5.35. The conditions here could be rephrased as saying that $P \in \mathbb{S}(R)$ is contained in a plane and satisfies the conditions of Proposition 2.5.25 with $v = \xi$.

Proof. As in Proposition 2.5.25, fix two parameters $\gamma_1, \gamma_2 > 0$ and partition P into patches $\alpha_1, ..., \alpha_M$ such that

- α_i is a smooth surface with piecewise smooth boundary for all $i = 1, \dots, M$;
- α_i has diameter at most γ_1 for all $i = 1, \dots, M$, and $M \leq C\gamma_1^{-2}$ for some constant C depending on P;
- Let $\alpha_i(t) := \alpha_i + t\xi$. For all $i = 1, \dots, M$, $\alpha_i(t) \cap \alpha_i(s) = \emptyset$ for $s \neq t$.

We define

$$\Delta_n = \sup_{1\leq i\leq M} \left| T(f_n, P)(\alpha_i) - T(f, P)(\alpha_i) \right|.$$

By the two-dimensional version of Lemma 2.5.14,

$$\mathbb{W}_1^{1,1}(T(f_n, P), T(f, P)) \leq 10C\gamma_1 + C\gamma_1^{-2}\Delta_n.$$

Let $U_i(s) = \cup_{t=0}^s \alpha_i(t)$ be the parallelopiped region between $\alpha_i = \alpha_i(0)$ and $\alpha_i(s)$. Given γ_2, we can find θ_{γ_2} small enough so that the number of edges from $\frac{1}{n}\mathbb{Z}^3$ hitting $\partial U_i(t) \setminus (\alpha_i(t) \cup \alpha_i)$ for any i is less than $\gamma_2 n^2 + K'n$, with constant K' depending on the length of $\partial\alpha_i$. Since the magnitude of f_n is of order $1/n^2$, for $t < \theta_{\gamma_2}$ the flow of f_n through $\partial U_i(t) \setminus (\alpha_i(t) \cup \alpha_i)$ is bounded by $O(n^{-1}) + \gamma_2$. Since any $f_n \in TF_n(R)$ is divergence-free as a discrete vector field, there exists a constant $K > 0$ so that for any f_n and $t \in (0, \theta_{\gamma_2})$,

$$\sup_{1\leq i\leq M} \left| T(f_n, \alpha_i)(\alpha_i) - T(f_n, \alpha_i(t))(\alpha_i(t)) \right| \leq \gamma_2 + K/n. \tag{2.23}$$

Possibly choosing a smaller θ_{γ_2}, the same result holds for f without the K/n in the upper bound. By the triangle inequality,

$$\sup_{1\le i\le M}\left|T(f_n,\alpha_i)(\alpha_i)-T(f,\alpha_i)(\alpha_i)-T(f_n,\alpha_i(t))(\alpha_i(t))+T(f,\alpha_i(t))(\alpha_i(t))\right|\le 2\gamma_2+K/n.$$

As in Proposition 2.5.25, integrating over $t\in(0,\theta_{\gamma_2})$ and solving for Δ_n gives

$$\Delta_n\le 2\gamma_2+K/n+\frac{1}{\theta_{\gamma_2}}\sup_{1\le i\le M}\left|\int_0^{\theta_{\gamma_2}}(T(f_n,\alpha_i(t))(\alpha_i(t))-T(f,\alpha_i(t))(\alpha_i(t)))\,\mathrm{d}t\right|.$$

Let α_i be the patch where the supermum is achieved, and let $\alpha:=\alpha_i$ to simplify notation. Then

$$\begin{aligned}\mathbb{W}_1^{1,1}(T(f_n,P),T(f,P))&\le 10C\gamma_1+2C\gamma_1^{-2}\gamma_2+C\gamma_1^{-2}Kn^{-1}\\&\quad+\frac{C\gamma_1^{-2}}{\theta_{\gamma_2}}\left|\int_0^{\theta_{\gamma_2}}T(f_n,\alpha(t))(\alpha(t))-T(f,\alpha(t))(\alpha(t))\,\mathrm{d}t\right|.\end{aligned}$$

Finally we bound the integral in the last term. Let $U=U_i(\theta_{\gamma_2})$ to simplify notation. Recall that $\alpha(t)\subset P+tv$ is a piece of a plane, and has constant unit normal vector ξ. By Definition 2.5.24, $T(f,\alpha(t))(x)=\langle f(x),\xi\rangle$. Therefore letting $\sigma_{\alpha(t)}$ denote the surface area measure on $\alpha(t)$, and applying change of variables,

$$\begin{aligned}\int_0^{\theta}T(f,\alpha(t))(\alpha(t))\,\mathrm{d}t&=\int_0^{\theta}\int_{\alpha(t)}\langle f(x),\xi\rangle\,\mathrm{d}\sigma_{\alpha(t))}(x)\mathrm{d}t=\int_U\langle f(x),\xi\rangle\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\\&=\sum_{j=1}^{3}\xi_j\mu_j(U),\end{aligned}$$

where $\xi=(\xi_1,\xi_2,\xi_3)$, $f=(f_1,f_2,f_3)$ and (μ_1,μ_2,μ_3) is the triple of measures corresponding to f. On the other hand, for the tiling flow f_n,

$$\int_0^{\theta}T(f_n,\alpha(t))(\alpha(t))\,\mathrm{d}t=\sum_{j=1}^{3}\xi_j\,\mu_j^n(U),$$

where $(\mu_1^n,\mu_2^n,\mu_3^n)$ is the triple of measures corresponding to f_n. Therefore

$$\begin{aligned}&\mathbb{W}_1^{1,1}(T(f_n,P),T(f,P))\\&\le 10C\gamma_1+2C\gamma_1^{-2}\gamma_2+C\gamma_1^{-2}Kn^{-1}+C\gamma_1^{-2}\theta_{\gamma_2}^{-1}\sum_{j=1}^{3}|\xi_j||\mu_j^n(U)-\mu_j(U)|.\end{aligned}$$

By Lemma 2.5.15, $d_W(f_n,f)\to 0$ implies that $|\mu_j^n(U)-\mu_j(U)|\to 0$ as $n\to\infty$ for $j=1,2,3$. Taking $n\to\infty$ gives

$$\limsup_{n\to\infty}\mathbb{W}_1^{1,1}(T(f_n,P),T(f,P))\le 10C\gamma_1+2C\gamma_1^{-2}\gamma_2.$$

Setting $\gamma_2=\gamma_1^3$ and taking $\gamma_1\to 0$ completes the proof. □

Next we prove a version of Proposition 2.5.26 for tiling flows and small patch surfaces as in Proposition 2.5.25.

Proposition 2.5.36. *Suppose that $S \in \mathbb{S}(R)$ satisfies the conditions of Proposition 2.5.25. For all $\epsilon > 0$ there exists $\delta > 0$ and $N > 0$ such that for all $t < \delta$, all $n \geq N$, and all $f \in TF_n(R)$,*

$$\mathbb{W}_1^{1,1}(T(f,S), T(f,S(t))) < \epsilon.$$

Proof. The proof is analogous to the proof of Proposition 2.5.26. Again we take parameters $\gamma_1, \gamma_2 > 0$ to be fixed later and a partition $\alpha_1, ..., \alpha_M$ of S into patches of diameter at most γ_1 for all i, and such that $M \leq C\gamma_1^{-2}$ for some constant C independent of γ_1. Analogous to Equation (2.23), given γ_2 we can find a threshold $\theta_{\gamma_2} > 0$ such that for all $0 \leq t \leq \theta_{\gamma_2}$ and all $f \in TF_n(R)$,

$$\sup_{1 \leq i \leq M} \left| T(f,S)(\alpha_i) - T(f,S(t))(\alpha_i(t)) \right| \leq \gamma_2 + K/n. \tag{2.24}$$

Using this, we get an upper bound for the distance by giving a method for redistributing, adding, and deleting mass to transform $T(f,S) \mid_{\alpha_i}$ into $T(f,S(t)) \mid_{\alpha_i(t)}$. Both measures are a sum of delta masses of weights $2/n^2(\pm 5/6)$ or $2/n^2(\pm 1/6)$. The number of delta masses in α_i or $\alpha_i(t)$ is bounded above by $\text{area}(\alpha_i)n^2$. Since α_i has diameter bounded by γ_1, there is a constant $A > 0$ independent of α_i such that $\text{area}(\alpha_i) \leq A\gamma_1^2$. Hence the total mass in each patch is bounded between $-2A\gamma_1^2$ and $2A\gamma_1^2$. Hence adding $\gamma_2 + K/n$ mass plus moving at most $8A\gamma_1^2$ mass distance at most $t + \gamma_1$, we get the bound

$$\mathbb{W}_1^{1,1}(T(f,S) \mid_{\alpha_i}, T(f,S(t)) \mid_{\alpha_i(t)}) \leq \gamma_2 + K/n + 8A\gamma_1^2(t + \gamma_1)$$

Summing over i we get that

$$\begin{aligned} \mathbb{W}_1^{1,1}(T(f,S), T(f,S(t))) &\leq M(\gamma_2 + K/n + 8\gamma_1^2(t + \gamma_1)) \\ &= C\gamma_1^{-2}\gamma_2 + CK\gamma_1^{-2}n^{-1} + 8A(t + \gamma_1). \end{aligned}$$

Take $\gamma_2 = \gamma_1^3$ and γ_1, t small enough so that

$$(C + 8A)\gamma_1 + 8At < \epsilon/2. \tag{2.25}$$

Then take n large enough so that

$$CK\gamma_1^{-2}n^{-1} < \epsilon/2,$$

and the result follows with $\delta = \min\{\theta_{\gamma_1^3}, \frac{1}{8A}(\epsilon/2 - (C + 8A)\gamma_1)\}$. Here the first term in the minimum comes from (2.24) and the second comes from (2.25). □

By approximation we can extend Proposition 2.5.34 to any surface $S \in S(\mathbb{R})$. For technical reasons, we first show this for S contained strictly in the interior of R. Note that we also remove the condition that the limiting flow is smooth.

Proposition 2.5.37. *Suppose that $S \in \mathbb{S}(R)$ is contained strictly in the interior of R. Let $f_n \in TF_n(R)$ be a sequence of tiling flows such that $d_W(f_n, f) \to 0$ as $n \to \infty$ for some $f \in AF(R)$. Then*

$$\lim_{n\to\infty} \mathbb{W}_1^{1,1}(T(f_n, S), T(f, S)) = 0.$$

Proof. As usual, let $\gamma_1, \gamma_2 > 0$ be small parameters to be fixed later. Since S is contained strictly in the interior of R, we can cover S by very small patch surfaces $\alpha_1, ..., \alpha_M$ so that:

- Each α_i is smooth with piecewise smooth boundary,
- The diameter of α_i is at most γ_1.
- There is a constant $C > 0$ such that $M \leq C\gamma_1^{-2}$.
- For all i, α_i satisfies the conditions of Proposition 2.5.25 for some threshold $\theta > 0$ with vector $v = \xi(q_i)$, where $\xi(q_i)$ is the normal vector to the surface at q_i for some $q_i \in \alpha_i$ with the property that the distance between q_i and any other $x \in \alpha_i$ is at most γ_1.
- Let P_i denote the tangent plane to α_i at q_i, and let $\pi_i \subset P_i$ be the patch of the plane corresponding to projecting α_i onto P_i. Potentially making $\theta > 0$ or γ_1 smaller, we can assume that π_i also satisfies the conditions of Proposition 2.5.25 for $v = \xi(q_i)$. This is where we are using the fact that S is contained in the interior of R.

As usual we denote $\alpha_i(t) = \alpha_i + t\xi(q_i)$ and $\pi_i(t) = \pi_i + t\xi(q_i)$. Note by Definition 2.5.33 and Lemma 2.5.28 that

$$T(f_n, S) \mid_{\alpha_i} = T(f_n, \alpha_i) \qquad \text{and} \qquad T(f, S) \mid_{\alpha_i} = T(f, \alpha_i) \qquad i = 1, ..., M.$$

Define

$$\Delta_n := \sup_{1\leq i\leq M} \Big| T(f_n, S)(\alpha_i) - T(f, S)(\alpha_i) \Big|.$$

By the two-dimensional version of Lemma 2.5.14,

$$\mathbb{W}_1^{1,1}(T(f_n, S), T(f, S)) \leq M(10\gamma_1^3 + \Delta_n) \leq 10C\gamma_1 + C\gamma_1^{-2}\Delta_n. \tag{2.26}$$

Since $f_n \in TF_n(R)$ is discrete divergence-free and S is compact and piecewise smooth, given γ_2 there exists $K > 0$ depending on S and a threshold θ_{γ_2} such that for all $0 \leq t \leq \theta_{\gamma_2}$,

$$\sup_{1\leq i\leq M} \Big| T(f_n, \alpha_i)(\alpha_i) - T(f_n, \alpha_i(t))(\alpha_i(t)) \Big| \leq \gamma_2 + K/n. \tag{2.27}$$

By Proposition 2.5.27, up to making θ_{γ_2} smaller, for all $0 \leq t \leq \theta_{\gamma_2}$,

$$\sup_{1\leq i\leq M} \Big| T(f, \alpha_i)(\alpha_i) - T(f, \alpha_i(t))(\alpha_i(t)) \Big| \leq \gamma_2. \tag{2.28}$$

Combining Equations (2.27) and (2.28), and as in Proposition 2.5.25 integrating over $t \in (0, \theta_{\gamma_2})$ then solving for Δ_n gives

$$\Delta_n \leq 2\gamma_2 + K/n + \frac{1}{\theta_{\gamma_2}} \sup_{1 \leq i \leq M} \left| \int_0^{\theta_{\gamma_2}} T(f_n, \alpha_i(t))(\alpha_i(t)) - T(f, \alpha_i(t))(\alpha_i(t)) \, \mathrm{d}t \right|.$$

Let i be the index where the supremum is achieved, and let $\alpha(t) := \alpha_i(t)$ and $\pi(t) := \pi_i(t)$. We now bound

$$T(f_n, \alpha(t))(\alpha(t)) - T(f, \alpha(t))(\alpha(t))$$

using four terms. Let $(g_m) \in AF^\infty(R_{\epsilon_m})$, $\epsilon_m \to 0$ as $m \to \infty$, be a sequence of smooth asymptotic flows such that $\lim_{m\to\infty} d_W(g_m, f) = 0$. Since $\alpha(t), \pi(t)$ are contained strictly in the interior of R for $0 \leq t \leq \theta$, we can assume they are all contained in R_ϵ for some $\epsilon > 0$. In particular, for m large enough $\epsilon_m < \epsilon$ and hence g_m is defined on $\pi(t)$. We have for any $0 \leq t \leq \theta_{\gamma_2}$,

$$\begin{aligned} T(f_n, \alpha(t))(\alpha(t)) - T(f, \alpha(t))(\alpha(t)) =& T(f_n, \alpha(t))(\alpha(t)) - T(f_n, \pi(t))(\pi(t)) \\ &+ T(f_n, \pi(t))(\pi(t)) - T(g_m, \pi(t))(\pi(t)) \\ &+ T(g_m, \pi(t))(\pi(t)) - T(f, \pi(t))(\pi(t)) \\ &+ T(f, \pi(t))(\pi(t)) - T(f, \alpha(t))(\alpha(t)). \end{aligned}$$

Consider the region $V(t)$ with boundary $\alpha(t) \cup \pi(t)$ plus sides to enclose it. Since $\pi(t)$ is the tangent plane to $\alpha(t)$ at $q(t) = q_i + tv$, the height of the sides needed to enclose this region is bounded by $C_2\gamma_1^2$, where since S is compact, $C_2 > 0$ is a constant depending only on S (C_2 is basically the maximum curvature at a smooth point on S). On the other hand, the length of the boundary of $\alpha(t)$ is bounded by $C_3\gamma_1$ for some constant $C_3 > 0$. Therefore since f_n is divergence-free, for some constant $K > 0$,

$$|T(f_n, \alpha(t))(\alpha(t)) - T(f_n, \pi(t))(\pi(t))| \leq C_2C_3\gamma_1^3 + K/n. \tag{2.29}$$

The analogous result holds for f (without the K/n term), controlling the fourth term above. By Proposition 2.5.27,

$$\lim_{m\to\infty} |T(g_m, \pi(t))(\pi(t)) - T(f, \pi(t))(\pi(t))| = 0.$$

As in Proposition 2.5.34, if μ_j^n denotes the component measures of f_n, ν_j^m denotes the component measures of g_m, and μ_j denotes the component measures of f, and letting $U = \cup_{t=0}^{\theta_{\gamma_2}} \pi(t)$, we have

$$\left| \int_0^{\theta_{\gamma_2}} T(f_n, \pi(t))(\pi(t)) - T(g_m, \pi(t))(\pi(t)) \, \mathrm{d}t \right|$$

$$\leq \sum_{j=1}^{3} |\xi_j(q)||\mu_j^n(U) - \nu_j^m(U)| \leq \sum_{j=1}^{3} |\xi_j(q)|(|\mu_j^n(U) - \mu_j(U)| + |\mu_j(U) - \nu_j^m(U)|).$$

Taking the limit as $m \to 0$ gives

$$\limsup_{m\to\infty} \left| \int_0^{\theta_{\gamma_2}} T(f_n, \pi(t))(\pi(t)) - T(g_m, \pi(t))(\pi(t))\, \mathrm{d}t \right| \leq \sum_{j=1}^{3} |\xi_j(q)||\mu_j^n(U) - \mu_j(U)|.$$

Therefore

$$\left| \int_0^{\theta_{\gamma_2}} T(f_n, \alpha(t))(\alpha(t)) - T(f, \alpha(t))(\alpha(t))\, \mathrm{d}t \right|$$
$$\leq 2C_2C_3\theta_{\gamma_2}\gamma_1^3 + \sum_{j=1}^{3} |\xi_j(q)||\mu_j^n(U) - \mu_j(U)| + K/n.$$

Plugging back in to Equation (2.26), we get

$$\mathbb{W}_1^{1,1}(T(f_n, S), T(f, S)) \leq 10C\gamma_1 + 2C\gamma_1^{-2}\gamma_2 + C\gamma_1^{-2}Kn^{-1} + 2CC_2C_3\gamma_1 + C\gamma_1^{-2}\theta_{\gamma_2}^{-1}Kn^{-1} +$$
$$C\gamma_1^{-2}\theta_{\gamma_2}^{-1} \sum_{j=1}^{3} |\xi_j(q)||\mu_j^n(U) - \mu_j(U)|.$$

Take $\gamma_2 = \gamma_1^3$. Then taking γ_1 small makes terms 1, 2, and 4 small. Taking n large makes terms 3 and 5 small and by Lemma 2.5.15 also makes term 6 small. □

Next we remove the condition that S is contained in the interior of R.

Proposition 2.5.38. *For any $S \in \mathbb{S}(R)$ and any sequence of tiling flows $f_n \in TF_n(R)$ such that $d_W(f_n, f) \to 0$ as $n \to \infty$ for some $f \in AF(R)$,*

$$\lim_{n\to\infty} \mathbb{W}_1^{1,1}(T(f_n, S), T(f, S)) = 0.$$

Proof. We can cover S with finitely many surfaces $S_1, ..., S_M$ which satisfy the conditions of Proposition 2.5.25 for vectors $v_1, ..., v_k$ and a threshold $\theta > 0$. We can do this so that $d = \max_{1\leq j\leq M} \operatorname{diam}(S_j)$ and there is a constant C independent of d such that $M = Cd^{-2}$. Fix $\epsilon > 0$. There exists $\delta > 0$, $N > 0$ such that for all $j = 1, ..., M$, and all $0 \leq t \leq \delta$, by Proposition 2.5.27,

$$\mathbb{W}_1^{1,1}(T(f, S_j), T(f, S_j(t))) < \epsilon$$

and by Proposition 2.5.36, for $n \geq N$ and $0 \leq t \leq \delta$,

$$\mathbb{W}_1^{1,1}(T(f_n, S_j), T(f_n, S_j(t))) < \epsilon.$$

On the other hand, by Proposition 2.5.37, for all $j = 1, .., M$ and all $t > 0$, for n large enough

$$\mathbb{W}_1^{1,1}(T(f_n, S_j(t)), T(f, S_j(t))) < \epsilon.$$

Hence by the triangle inequality, for all $j = 1, ..., M$

$$\mathbb{W}_1^{1,1}(T(f, S_j), T(f_n, S_j)) < 3\epsilon.$$

By the two-dimensional version of Lemma 2.5.14, for n large enough,

$$\mathbb{W}_1^{1,1}(T(f, S), T(f_n, S)) \leq M(10d^3 + 3\epsilon) \leq 10Cd + 3d^{-2}\epsilon.$$

Taking $d = \epsilon^{1/3}$ would complete the proof. □

Finally we can prove the main theorem about boundary values that we will refer to later in paper.

Theorem 2.5.39. *For any $S \in \mathbb{S}(R)$, the trace operator*

$$T(\cdot, S) : (AF(R) \cup TF(R), d_W) \to (\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$$

is uniformly continuous. In particular this holds for $S = \partial R$.

Proof. By Theorem 2.5.22, $(AF(R), d_W)$ is compact. Since $TF_n(R)$ is finite for each n, Theorem 2.5.19 implies that the d_W limit points of $TF(R)$ are contained in $AF(R)$. Therefore $(AF(R) \cup TF(R), d_W)$ is compact.

On the other hand, for any $S \in \mathbb{S}(R)$, Proposition 2.5.29 and Proposition 2.5.38 combine to show that $T(\cdot, S)$ is a continuous map from $(AF(R) \cup TF(R), d_W)$ to $(\mathcal{M}^s(R), \mathbb{W}_1^{1,1})$. Therefore by compactness $T(\cdot, S)$ is uniformly continuous. □

2.6 Patching

The main goal of this section is to prove a *patching theorem* (Theorem 2.6.14) which will be an essential tool throughout this paper. We show that if the flows associated to tilings τ_1, τ_2 of $\mathbb{Z}^3$ are *nearly-constant* (Definition 2.6.12 below) with value $s \in \mathrm{Int}(\mathcal{O})$ (which loosely speaking means the flows associated with τ_1, τ_2 both approximate the constant flow equal to s), then we can remove a bounded piece from τ_1, and patch it to τ_2 by tiling a thin (cubic) annulus.

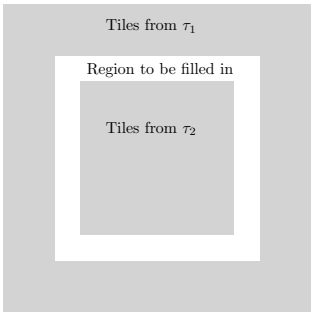


Figure 2.21: Two dimensional schematic for patching.

Equivalently, we want to show that this annular region can be tiled by dimers exactly so that it agrees with τ_1 on one boundary and with τ_2 on the other boundary. To do this, we need a condition to show that a region is tileable.

A general condition for tileability, which works in any dimension, is given by the classical *Hall's matching theorem* ([Hal35], stated here as Theorem 2.6.2), which says a region R is tileable if and only if every $U \subset R$ has certain properties; a U that does not have such properties is a "counterexample" to the condition that implies tilability, as we explain in Section 2.6.1. In two dimensions, Hall's matching theorem implies that a simple condition on height function differences along the boundary of the region is equivalent to tileabilty [Fou96]. In three dimensions, we show in Section 2.6.2 that it is sufficient to show that the

region R has no counterexample set U whose boundary is a certain type of *minimal surface*, built out of squares from the $\mathbb{Z}^3$ lattice (Corollary 2.6.9). We call surfaces built out of lattice squares *discrete surfaces*.

In Section 2.6.3 we give the statement of the patching theorem, and explain the main ideas of the proof accompanied by a series of two dimensional figures.

The main new difficulty in higher dimensions is that the counterexample sets U can have more complicated geometry. In two dimensions, the boundary of the counterexample region is a union of curves. In three dimensions it is a union of surfaces. However the fact that these surfaces can be assumed to be in some sense *minimal* gives us some control their geometry. In Section 2.6.4, we prove some straightforward adaptations of the isoperimetric inequality for discrete surfaces. In Section 2.6.5, we apply these to get useful bounds on the area growth for minimal discrete surface (Proposition 2.6.18), and show that they "spread out" (Lemma 2.6.20).

In Section 2.6.6, we prove an ergodic theorem for the flow of a tiling through a coordinate plane (Theorem 2.6.23), and note that tilings sampled from ergodic measures satisfy the conditions of the patching theorem with probability going to 1 as $n \to \infty$ (Corollary 2.6.25). We show that ergodic measures of any mean current $s \in \mathcal{O}$ exist (Lemma 2.6.21), and prove some bounds for their expected flow through discrete surfaces (Lemma 2.6.27). One of the ideas in the proof of the patching theorem is to use a tiling sampled from an ergodic measure as a "measuring stick" that we compare with the tilings we want to patch.

Equipped with the lemmas from the previous sections, in Section 2.6.7 we give the proof of the patching theorem (Theorem 2.6.14). Finally in Section 2.6.8 we give some immediate corollaries of patching for ergodic Gibbs measures (EGMs).

We use the same tools and ideas developed in this section again in Section 2.8.6 to prove a *generalized patching theorem* (Theorem 2.8.32) where the flow the tilings approximate is not required to be constant, and where the annular region is allowed to have a more general shape. The main results of this paper are two versions of a large deviation principle: one with *soft* boundary conditions (Theorem 2.8.6) and one with *hard* boundary conditions (Theorem 2.8.15). The regular patching theorem proved here (Theorem 2.6.14) is sufficient to prove the LDP with soft boundary conditions, but the generalized version (Theorem 2.8.32) is needed in the final steps to prove the version with hard boundary conditions.

2.6.1 Hall's matching theorem and non-tileability

When can a finite region $R \subset \mathbb{Z}^3$ be exactly tiled by dimers, i.e. without any tiles crossing the boundary, and with all cubes covered? This is equivalent to asking: when does a finite subgraph $G \subset \mathbb{Z}^3$ have a perfect matching? A straightforward observation is that for any bipartite graph G with bipartition (A, B), a necessary condition for G to have a perfect matching is that it is *balanced*, i.e. that $|A| = |B|$. The balanced condition is not sufficient though, see Figure 2.22. Nonetheless, it turns out there is a very general necessary and sufficient condition which characterizes whether or not a finite bipartite graph G has a perfect matching. In this section we explain these results from the graph point of view where dimers are edges and dimer tilings are perfect matchings. There are two perspectives, both of which can be useful:

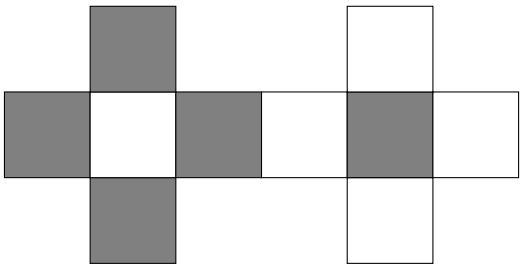


Figure 2.22: A region that is balanced but has no dimer tilings.

- the min-cut, max-flow principle
- Hall's matching theorem

We first describe the classical min-cut, max-flow principle. Let $G = (A, B)$ be the bipartition of the graph (A is "even" and B is "odd"). If G has a perfect matching τ, then there is a flow v_τ (the "pretiling flow") which sends a unit of current from each even vertex $a \in A$ to the odd vertex $b \in B$ it is paired to. Note that v_τ has a source of $+1$ at each $a \in A$ and a sink of $+1$ (or source of -1) at each $b \in B$. The existence of a perfect matching is equivalent to the existence of a flow v_τ with the desired source/sink values and a flow of 0 or 1 on each even-to-odd edge.

A *cut* is a collection of edges in G which, if deleted, separates G into two pieces G_1 and G_2. Let F_1 be the net total flow sourced in G_1 (i.e. the number of even vertices in G_1 minus the number of odd vertices), let F_2 be the net total flow sourced in G_2, and let c be the number of cut edges. If G has a perfect matching, we must have $F_1 + F_2 = 0$.

The value F_1 measures the amount of flow that would have to travel across the cut if G has a perfect matching, so we must have $F_1 \leq c$. In other words, if G has a perfect matching, then it must be the case that for *every* cut, the excess flow on either side must be less than the size of the cut. It turns out that this is a sufficient condition too, so if G does *not* have a perfect matching then there is a cut of c edges partitioning G into G_1 and G_2 such that the excess flow F_1 that needs to cross the cut is more than c. In summary:

Theorem 2.6.1 (Min-cut, max-flow principle [FF56])**.** *A finite bipartite graph G has a perfect matching (a.k.a. dimer tiling) if and only if there is no cut consisting of c edges partitioning G into two sets G_1 and G_2 such that $F_1 > c$.*

In Hall's matching theorem, we shift our perspective from the cut to the sets in the partition. Instead of looking at sets that are a mixture of even and odd vertices, we consider a set C of only even (resp. only odd) vertices, plus their neighbors

$$N(C) = \{b \in B : a \in C, (a, b) \in E\}.$$

The set $C \cup N(C)$ is analogous to either G_1 or G_2 (without loss of generality G_1) plus some of the endpoints of the edges in the cut. The excess flow is now the number of even vertices (i.e. $|C|$) minus the number of odd vertices (i.e. $|N(C)|$). Hall's matching theorem is an analog of Theorem 2.6.1 formulated in these terms:

Theorem 2.6.2 (Hall's matching theorem [Hal35])**.** *Suppose that $G = (V, E)$ is a finite bipartite graph with bipartition $G = A \cup B$. The graph G admits a perfect matching consisting of $|A|$ edges if and only if for all $C \subset A$,*

$$N(C) = \{b \in B : a \in C, (a, b) \in E\}$$

satisfies $|N(C)| \geq |C|$.

An analogous result holds with A and B switched. If G is balanced (i.e. $|A| = |B|$), then the existence of a perfect matching with $|A|$ edges is equivalent to the existence of a perfect matching of the whole graph G. If G is not balanced (i.e. $|A| \neq |B|$) then G does not have a perfect matching of the whole graph.

Note that if $C \subset A$ satisfies $|N(C)| < |C|$, then the set $U := C \cup N(C)$ has more even than odd vertices, despite having only odd vertices on its boundary within G. Therefore when G is balanced, Theorem 2.6.2 is equivalent to the following:

Theorem 2.6.3. *Suppose that $G = (V, E)$ is a finite bipartite graph with bipartition $G = A \cup B$ with $|A| = |B|$. Then G* fails to have *a perfect matching if and only if there exists a connected set $U \subset V$ such that $|U \cap A| > |U \cap B|$ but all boundary elements of U (i.e., elements of U that are adjacent to some point in $V \setminus U$) belong to B.*

We call the U from Theorem 2.6.3 a *counterexample to (the condition equivalent to) tileability* or just a *counterexample*. In our context, A and B will always be sets of even and odd vertices in $\mathbb{Z}^3$, so for us a *counterexample* to tileability for $R \subset \mathbb{Z}^3$ is any set U that has more even than odd vertices, despite having only odd vertices on its *interior boundary*, which we define to be the set of $x \in U$ that are incident to some $y \in R \setminus U$.

To show that a graph R has a dimer tiling (a.k.a. a perfect matching), we check that it is balanced, and if it is, we have to show that there are no counterexamples. We call the excess flow of a counterexample its *imbalance*, given by

$$\text{imbalance}(U) = \text{even}(U) - \text{odd}(U). \tag{2.30}$$

Note that if $U \subset R \subset \mathbb{Z}^3$ has only odd vertices on its interior boundary (within R) then $\text{imbalance}(U) > 0$ if and only if U is a counterexample for R.

2.6.2 Discrete surfaces and minimal counterexamples

As mentioned earlier, it is often intuitively useful to think about perfect matchings as tilings by $2 \times 1 \times 1$ blocks. In this picture, a counterexample set U is a collection of unit cubes, each centered at a point in $\mathbb{Z}^3$, and the edges out of it, i.e. its boundary ∂U, is a collection of squares in the translated lattice $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) + \mathbb{Z}^3$. In other words the boundary region is a surface built out of squares from the lattice.

Definition 2.6.4. A *discrete surface* in $\mathbb{Z}^3$ is a collection of squares from the $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) + \mathbb{Z}^3$ lattice.

A discrete surface in $\mathbb{Z}^3$ is *orientable* if there is a well-defined outward pointing normal vector to the surface. An orientable discrete surface S with a choice of outward pointing normal vector is called *oriented*. For a square $s \subset S$, we call the side of s that the outward normal vector points toward the *outside*. If the outward pointing normal vector to a square in an oriented surface is from even to odd, we color the outside of the square white. Otherwise we color it black.

Definition 2.6.5. An oriented discrete surface S in $\mathbb{Z}^3$ is *monochromatic* if all the outsides of all the squares in S are black (resp. all are white).

We can rewrite Equation (2.30) for the imbalance of a counterexample U in terms of the black and white surface area of ∂U. Let (A, B) be the bipartition of $\mathbb{Z}^3$ into even and odd vertices respectively.

Proposition 2.6.6. *Suppose that R is balanced but not tileable, and that $U \subset R \subset \mathbb{Z}^3$ is a counterexample to tileability. Then*

$$0 < \mathit{imbalance}(U) = \mathit{even}(U) - \mathit{odd}(U) = \frac{1}{6}\Big(\mathit{white}(\partial U) - \mathit{black}(\partial U)\Big). \tag{2.31}$$

Proof. Define a flow r on $\mathbb{Z}^3$ such that $r(e) = \frac{1}{6}$ for every dual edge (a.k.a. face) e oriented from even to odd. Then

$$\operatorname{div} r(v) = \begin{cases} -1 & v \text{ is a odd cube} \\ +1 & v \text{ is a even cube} \end{cases}$$

By the divergence theorem, with all edges $e \in \partial U$ oriented out of U,

$$\begin{aligned} \operatorname{imbalance}(U) = \operatorname{even}(U) - \operatorname{odd}(U) &= \sum_{v \in U} \operatorname{div} r(v) = \sum_{e \in \partial U} r(e) \\ &= \frac{1}{6}\Big(\operatorname{white}(\partial U) - \operatorname{black}(\partial U)\Big). \end{aligned}$$

□

By Proposition 2.6.6, if U is a counterexample then it must have more white surface area than black surface area. By the definitions in Section 2.6.1, U must have only odd cubes along its *interior boundary*, i.e. cubes $x \in U$ which are adjacent to $y \in R \setminus U$. However U also has an *exterior boundary* consisting of cubes which are adjacent to $y \in \mathbb{Z}^3 \setminus R$. Exterior boundary cubes can be even or odd.

Correspondingly, the boundary ∂U can be split into two pieces: the *exterior boundary surface* $T = \partial R \cap \partial U$ and the *interior boundary surface* $S = \partial U \setminus T$. The interior boundary surface S must be built out of only black squares, while T can be built out of a mixture of white and black squares. Given this, only squares in T contribute positively to the imbalance of U. Intuitively to increase the imbalance of U, one should *minimize* the area of S.

A surface P embedded in $\mathbb{R}^3$ is said to *locally minimize area* if given any point $p \in P$, there is a neighborhood $V \subset \mathbb{R}^3$ containing p such that $P \cap V$ has the minimal area of any surface with boundary $\partial(P \cap V)$. Surfaces that locally minimize area are called *minimal surfaces*. We will be interested in certain discrete analogs of minimal surfaces.

Definition 2.6.7. A *minimal discrete (monochromatic) surface with boundary* X is a surface S that minimizes area subject to the constraint that it is discrete, (monochromatic), and has $\partial S = X$. In particular, there is no way to "tighten the surface locally" by changing a few faces in a way that maintains the overall boundary conditions and reduces the overall area.

Proposition 2.6.8. *Let $R \subset \mathbb{Z}^3$ be a finite balanced region which is not tileable, and suppose that U is a counterexample to tileability in R. Let $T = \partial U \cap \partial R$ and let $S = \partial U \setminus T$ be the interior boundary surface. Then there exists another counterexample U' in R such that $\partial U' \cap \partial R = T$, and $S' := \partial U' \setminus T$ is a minimal monochromatic black discrete surface.*

Proof. The new set U' is defined so that $\partial U' = T \cup S'$, where S' is a minimal monochromatic discrete surface. Since S is all black, by Proposition 2.6.6,

$$6 \cdot \text{imbalance}(U) = \text{white}(\partial U) - \text{black}(\partial U) = \text{white}(T) - \text{black}(T) - \text{area}(S) \tag{2.32}$$

replacing S by S' only makes the imbalance larger, so U' is still a counterexample. □

We call counterexamples where the internal boundary surface is a minimal surface *minimal counterexamples*. We immediately get the following corollary.

Corollary 2.6.9. *A finite balanced region $R \subset \mathbb{Z}^3$ is tileable if and only if it has no minimal counterexamples.*

2.6.3 Statement of patching theorem and outline of proof

We now give the statement of the patching theorem mentioned at the beginning of the section. We also provide illustrations of the analogous constructions in 2D (because they are easier to draw) in order to explain the 3D concepts. Let $B_n = [-n, n]^3$, and for any $\delta > 0$ define

$$A_n = B_n \setminus B_{(1-\delta)n}.$$

Given two tilings $\tau_1, \tau_2 \in \Omega$, we look at the region between $\tau_1 \mid_{\mathbb{Z}^3 \setminus B_n}$ and $\tau_2 \mid_{B_{(1-\delta)n}}$. This will be the annulus A_n, with some cubes removed along its boundary We call this A_n *with boundary conditions τ_1 and τ_2*. See Figure 2.23 for a 2D example of such an A_n. In this figure, A_n is a subset of the "square annulus" bounded between the boundary of a 12×12 box and the boundary of a 32×32 box. It is obtained from the square annulus by removing some of the squares along the outer boundary and some of the squares along the inner boundary. Given any dimer tiling τ of a region containing the 32×32 box, the union of the dimers of τ that are *strictly contained* in the square annulus would also be a region of this form.

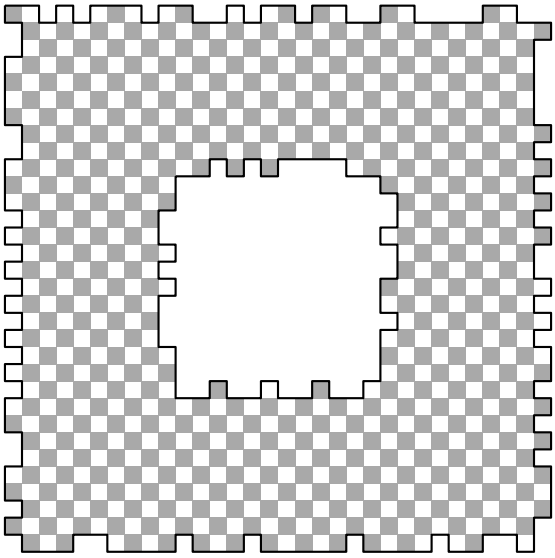


Figure 2.23: A two-dimensional example of an annular region A_n of the form with boundary conditions coming from tilings τ_1, τ_2.

The main question is: given a tiling τ_1 restricted to $\mathbb{Z}^3 \setminus B_n$ and a tiling τ_2 restricted to $B_{(1-\delta)n}$, under what conditions can we patch them together, i.e. find a tiling of A_n with inner boundary condition τ_2 and outer boundary condition τ_1? We are interested in showing that this is possible when n is large enough, when τ_1, τ_2 satisfy a consistency condition that they are *nearly constant* for the same $s \in \mathrm{Int}(\mathcal{O})$.

To specify the nearly constant condition, we give a few definitions.

Definition 2.6.10. An ϵ *patch* α on ∂B_n is an $\epsilon n \times \epsilon n$ square contained in a face of ∂B_n.

We can then measure the *flux* of a discrete flow through a patch.

Definition 2.6.11. Let S be an oriented discrete surface with outward normal vector ξ. We define the *flux* of a discrete vector field v through S by

$$\mathrm{flux}(v, S) = \sum_{e \in E(\mathbb{Z}^3), e \cap S \neq \emptyset} \mathrm{sign}\langle \xi(e \cap S), e \rangle v(e).$$

Here $E(\mathbb{Z}^3)$ denotes the edges of $\mathbb{Z}^3$ oriented from even to odd.

We now use the definition of *nearly-constant* in terms of the flux of the pre-tiling flow v_τ through patches.

Definition 2.6.12. Fix $s \in \mathcal{O}$, let $B_n = [-n, n]^3$. A tiling $\tau \in \Omega$ is *ϵ-nearly-constant with value s* if there exists $M = M(\epsilon)$ such that for all $n > M$ and all ϵ patches α on ∂B_n,

$$\mathrm{flux}(v_\tau, \alpha) = \frac{1}{2}\langle \xi_\alpha, s \rangle \mathrm{area}(\alpha) + o(\mathrm{area}(\alpha)) = \frac{1}{2}\langle \xi_\alpha, s \rangle \epsilon^2 n^2 + o(\epsilon^2 n^2),$$

where ξ_α is the outward pointing unit normal vector to α (where outward means away from B_n).

Remark 2.6.13. Any patch α is contained in a flat coordinate plane, so its area is simply $\epsilon^2 n^2$. The $\frac{1}{2}$ comes from the fact that the mean current is actually measuring the average flow per *even* vertex.

Tilings satisfying the ϵ-nearly-constant condition with value s mimic the behavior of tilings sampled from ergodic measures of mean current s. (This is made precise in Corollary 2.6.25 after Theorem 2.6.23.)

With the conditions defined, we can now state the patching theorem.

Theorem 2.6.14 (patching theorem)**.** *Fix $\delta > 0$ and a mean current $s \in Int(\mathcal{O})$. Let $B_n = [-n, n]^3$ be the cube of radius n, and let $A_n = B_n \setminus B_{(1-\delta)n}$ be the cube annulus of width δn. For $\epsilon > 0$ small enough, if $\tau_1, \tau_2 \in \Omega$ are ϵ-nearly-constant with value s, then for n large enough A_n can be tiled with outer boundary condition τ_1 and inner boundary condition τ_2.*

The main tool in the proof is Hall's matching theorem (Theorem 2.6.2). In this section we explain the main steps of the proof, guided by a series of two dimensional figures, and comment on differences between the two and three dimensional versions of this story. After this, in the remaining subsections we prove a series of lemmas (needed to control the more complicated geometric situations that can occur in three dimensions) before giving a the formal proof of Theorem 2.6.14 in Section 2.6.7.

Steps of proof

Given two tilings τ_1, τ_2 of $\mathbb{Z}^2$, we want to know whether they can be patched together. In other words, we want to know whether a region A, which is a square annulus with some squares removed along the outer boundary if they are connected by τ_1 to the outside of the annulus and some squares removed along the inner boundary if they are connected by τ_2 to the outside of the annulus, is tileable. See Figure 2.23 for an example of a region A of this form. If A is not tileable, then there exists a *counterexample set* U as in Figure 2.24.

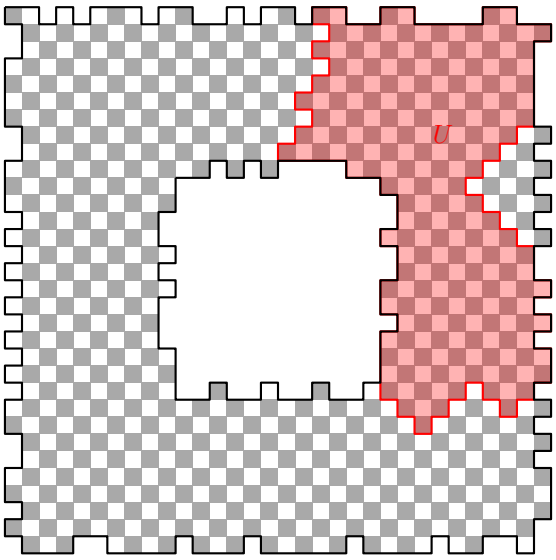


Figure 2.24: A potential counterexample set $U \subset A$.

The red set $U \subset A$ in Figure 2.24 has the property that every square on its *inner boundary* (i.e., every square of U that is incident to a square in $A \setminus U$) is black.

By Hall's matching theorem (Theorem 2.6.3), there exists a dimer tiling of A if and only if every U of this form is *not* a counterexample. In other words, every set U of this form has $\text{imbalance}(U) = \text{white}(U) - \text{black}(U) \leq 0$. We remark that the colors white and black used here stand for even and odd vertices and not for the colors we give to surfaces in 3 dimensions in the previous section. We do this because it becomes easier to illustrate the main ideas using the figures.

Given this, our strategy to show that A is tileable for n large enough proceeds by contradiction. We suppose that for all n there is a set U of the form above that is a counterexample, but then show that for n large enough, it must have $\text{white}(U) \leq \text{black}(U)$ and hence not be a counterexample. To do this, we cut up U into various smaller pieces, and bound the white minus black in each piece.

First we divide the annular A into regions as depicted in Figure 2.25. We call these the "middle region" (blue), the "thin region" (orange) and the "corner region" (green). The middle region is a centered square annulus whose size will have to be appropriately tuned. The thin region is the union of the "columns" obtained as straight-line paths of squares that go from the middle square annulus boundary to the boundary of A. The corner region is the part leftover.

We then define U' to be U intersected with the middle layer, depicted in Figure 2.26.

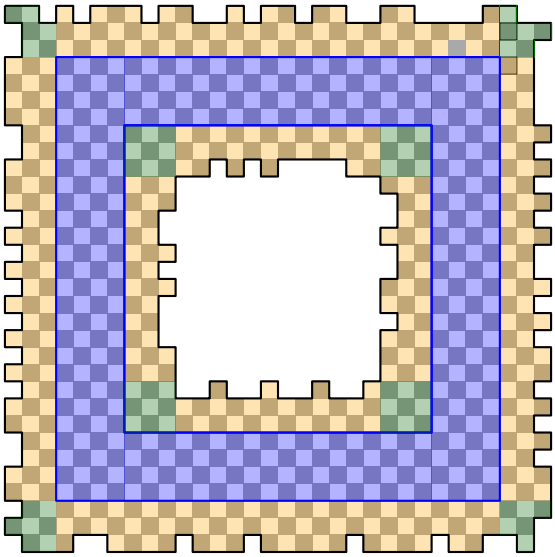


Figure 2.25: The middle (blue), thin (orange), and corner (green) regions of A.

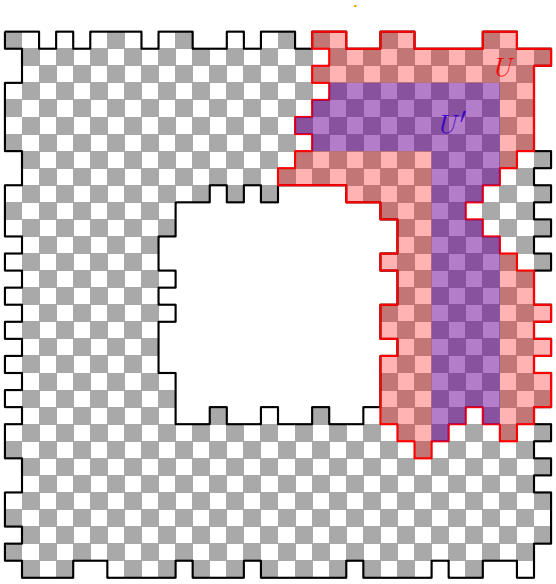


Figure 2.26: We can define U' to be the intersection of U with the middle region from Figure 2.25.

Given a tiling τ of a region containing A, we can define U_τ to be the region covered by tiles from τ intersecting U'. The set U'' shown in Figure 2.27 is the subset of U_τ that consists of the union of U' together with all of the squares covered by tiles from τ that are contained in U but in the complement of the middle region. Note that $U_\tau \setminus U''$ consists of only white squares and that U_τ is by construction evenly balanced between black and white squares. If we can show that $|U_\tau \setminus U''|$ is large then we know that U'' has many more black squares than white.

Indeed, we show that we can choose this *test tiling* τ so that $|U_\tau \setminus U''|$ is large and hence U'' has many more black than white squares. Since $U'' \subset U$, it remains to show that there cannot be enough white squares in $U \setminus U''$ for U to be a counterexample.

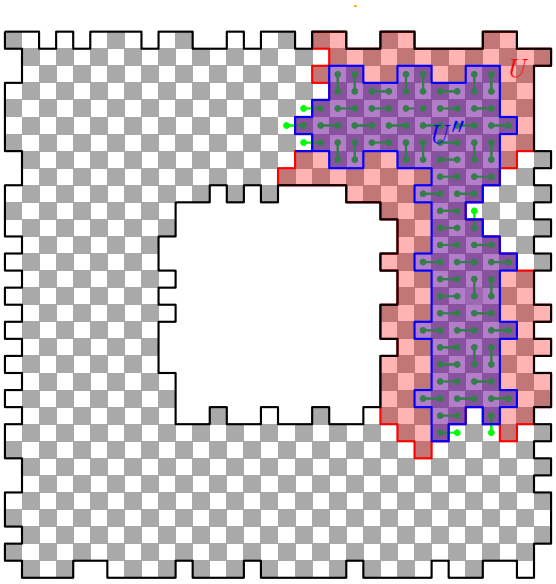


Figure 2.27: The region U'' (blue) and the tiles from τ that intersect U' (green).

In order to prove that U itself has more black than white squares, we will divide the rest of U into multiple pieces to treat separately, depicted in Figure 2.28. Here U'' is as given in Figure 2.27. The "shadow region" is $U \setminus U''$ restriced to the "thin region" from Figure 2.25. It consists of the union of the columns that can be extended all the way from ∂A to $\partial U''$. The corner region here is the intersection of U with the corner region from Figure 2.25. The "leftover pieces" are the parts of U that do not belong to one of the other three regions. Roughly speaking, we aim to show that U has more black than white by showing that (i) U'' has a lot more black than white, (ii) the shadow region can only have a little more white than black (because of the nearly-constant condition), (iii) the corner region can only have a little excess white (since it has small volume), and (iv) the leftover pieces have at least as much black as white.

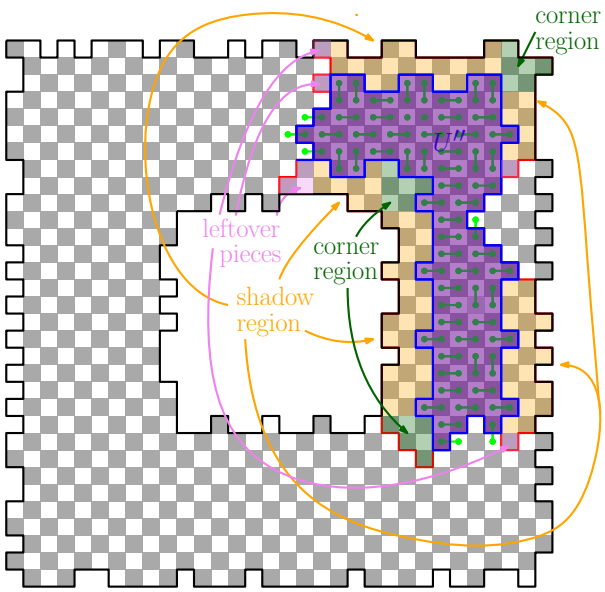


Figure 2.28: Depiction of all the regions that we divide U into: (i) U'', (ii) shadow region, (iii) corner region, and (iv) leftover pieces.

Ultimately to achieve (i), (ii), (iii), and (iv), we need to understand more about the possible shapes of counterexample sets U. The main difference between two dimensions and three dimensions is that the geometry of U can be more complicated.

In two dimensions, since A is balanced, if U is a counterexample then so is U^c (with even/odd switched). As a consequence, without loss of generality U and U^c are both connected, and (since U and U^c must both cross the annulus) the interior boundary curve S between them consists of two minimal monochromatic paths S_1, S_2 from the inner boundary to the outer boundary. There are then four points a_1, b_1 and a_2, b_2 which are the endpoints of S_1, S_2 on the inner and outer boundaries respectively. From here, one can argue as in [Fou96] to say that A is tileable as long as the boundary height differences (given by the boundary condition tilings τ_1, τ_2) at a_1 and b_1 and at a_2 and b_2 are in some sense compatible.

In three dimensions, since A is balanced, it again follows that both U and U^c are counterexamples, and that they are without loss of generality connected and have interior boundary surface S between them which is a minimal monochromatic discrete surface. However the geometry (and topology) of 2D surfaces and connected sets S in 3D can be much more complicated than that of 1D curves in 2D. For example, the set U in 3D need not be simply connected, and instead of $S \cap \partial A$ consisting of four points, it consists of a collection of curves.

To control the more complex *geometric possibilities* for counterexamples in three dimensions, we prove that the interior boundary surface S has area of order n^2 (Lemma 2.6.19), and that there is a choice for the inward blue layer where S restricted to the layer (which is some union of curves) has good properties (Lemma 2.6.20). This is the content of Section 2.6.4 and Section 2.6.5.

The other key tool is the notion of a *test tiling*. We show that we can use a tiling τ sampled from an ergodic measure of mean current $s \in \mathrm{Int}(\mathcal{O})$ to define U'' so that U'' will have order n^2 more black cubes than white cubes by showing that τ on expectation has an order n^2 number of tiles crossing S (Lemma 2.6.27). This quantity is somewhat analogous

to the height difference in two dimensions. Further, a tiling sampled from ergodic measure of mean current $s \in \text{Int}(\mathcal{O})$ is nearly constant with high probability (Corollary 2.6.25). This is the content of Section 2.6.6.

2.6.4 Discrete isoperimetric inequalities

The classical isoperimetric inequality says that the minimal area of a region D bounded by a smooth closed curve γ in $\mathbb{R}^2$ of length l is $\frac{1}{4\pi}(l^2)$. The equality case is achieved when γ is a circle and D is a disk. A lesser known fact is that this bound also holds for curves in $\mathbb{R}^3$ [Alm86]. In this section we prove a discrete version of the isoperimetric inequality.

Proposition 2.6.15. *Given any simple closed curve γ in $\mathbb{Z}^3$, there is a surface S with $\partial S = \gamma$ and*

$$\textit{area}(S) \leq \frac{1}{8}(\textit{length}(\gamma))^2.$$

Remark 2.6.16. The constant $\frac{1}{8}$ that we get from the proof is not optimal but our argument is simple and the result is sufficient for our purposes. Also note that the boundary curve γ can be replaced by a multicurve $\Gamma = \gamma_1 \cup \cdots \cup \gamma_k$ and the same result holds.

Proof. We proceed by induction. Suppose that for all simple closed curves γ in $\mathbb{Z}^3$ with $\text{length}(\gamma) = 2m \leq 2n$, there exists a surface S with $\partial S = \gamma$ and

$$\text{area}(S) \leq \frac{(m-1)m}{2}.$$

This is sufficient because it implies that

$$\text{area}(S) \leq \frac{1}{8}(\text{length}(\gamma))^2.$$

Now suppose that β is a simple closed curve of length $2n+2$, and equip β with an orientation. Choose two parallel edges $e_1 = (a_1, b_1), e_2 = (a_2, b_2)$ in β with opposite orientations. Removing e_1, e_2 and identifying a_i with b_i for $i = 1, 2$ results in a new curve β' of length $2n$. The identification means that β' is a union simple curves $\gamma_1, ..., \gamma_k$ of length $\leq 2n$ and double edges (note that each double edge contributes 2 to the length of β'). By the inductive hypothesis, there exist surfaces S_i with boundary γ_i satisfying the bound for each $i = 1, ..., k$. To get a surface with boundary β, we find a path in $\mathbb{Z}^3$ from a_1 to a_2 along $\cup_{i=1}^k S_i \cup D$, where D is the double edges in β'. Since $\text{length}(\beta') \leq 2n$, we can find a path of length $\leq n$. We add back the edges e_1, e_2, splitting the path into two parallel paths. We then add squares along the path from e_1 to e_2 to construct a surface S with boundary β satisfying

$$\text{area}(S) \leq \text{area}(\cup_{i=0}^k S_i) + n \leq \frac{n(n+1)}{2}.$$

□

This can easily be extended to the monochromatic case; the only thing that changes is the constant.

Corollary 2.6.17. *Given any collection of simple closed curves $\gamma_1, ..., \gamma_k$ in $\mathbb{Z}^3$, there is a monochromatic surface S with $\Gamma = \gamma_1 \cup \cdots \cup \gamma_k$ as its boundary and*

$$area(S) \leq \frac{5}{8} length(\Gamma)^2$$

Proof. Using Proposition 2.6.15, we can find a surface T with boundary $\gamma_1 \cup \cdots \cup \gamma_k$ satisfying $\text{area}(T) \leq \frac{1}{8}\text{length}(\Gamma)^2$. Replacing every white square in T by at most 5 black squares, we get a new surface S which is monochromatic (it is all black) satisfying the same bound with the constant $\frac{5}{8}$. □

2.6.5 Area growth of minimal monochromatic discrete surfaces

We now use the discrete isoperimetric inequalities from the previous section to show that minimal monochromatic discrete surfaces have quadratic area growth (Proposition 2.6.18). We then apply this to the cube annulus $A_n = B_n \setminus B_{(1-\delta)n}$ to get two results (Lemma 2.6.19 and Lemma 2.6.20) which will serve as lemmas for the patching theorem (Theorem 2.6.14).

Proposition 2.6.18. *Let S be a minimal monochromatic discrete surface. Let $p \in \mathbb{Z}^3 + (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ be a vertex on S and let $B_n(p) = p + [-n, n]^3$ be such that S is not contained in $B_n(p)$. Then there is a universal constant $\kappa > 0$ (i.e., independent of S and n) such that*

$$area(S) \geq \kappa n^2.$$

Proof. To do this, we show that $\text{area}(S \cap B_n(p))$ grows quadratically in n. Let $m = \lfloor \frac{n}{2} \rfloor$. For $k \leq m$, define annular regions

$$A_k = B_{m+k}(p) \setminus B_{m-k}(p)$$

Let $S_k = S \cap A_k$ be S restricted to A_k and let $\Gamma_k = \partial S_k$. By Corollary 2.6.17,

$$\text{area}(S_k) \leq \frac{5}{8}\text{length}(\Gamma_k)^2.$$

Note that $S_k \cap \partial A_k$ might be larger than Γ_k, since there might be squares from S_k contained in ∂A_k.

Any face in $S_{k+1} \setminus S_k$ corresponds to at most 4 edges along Γ_k. Therefore

$$\text{area}(S_{k+1}) - \text{area}(S_k) \geq \frac{1}{4}\text{length}(\Gamma_k) \geq \frac{1}{4}\sqrt{\frac{5}{8}}\sqrt{\text{area}(S_k)}.$$

Therefore the function $f(k) := \text{area}(S_k)$ satisfies the inequality $f(k+1) - f(k) \gtrsim \sqrt{f(k)}$. Extending f linearly to a continuous function that is differentiable from the left, this becomes $f'(k) \gtrsim \sqrt{f(k)}$, which implies that f grows at least quadratically in k. Applying this to S itself we get that $\text{area}(S) \geq \kappa n^2$, where the universal constant κ comes from the isoperimetric inequalities. □

Lemma 2.6.19. *Let $B_n = [-n, n]^3$ and suppose that $A = B_n \setminus B_{(1-\delta)n}$ for some $\delta \in (0, 1)$. Suppose that S is a minimal monochromatic discrete surface in A which connects the inner and outer boundaries of A. Then there exist constants c_1, c_2 independent of S and n such that*

$$c_1 n^2 \geq \textit{area}(S) \geq c_2 n^2$$

where $c_2 \sim \delta^2$.

Proof. For the upper bound, notice that $\partial S \subset \partial A$. From ∂A, we construct a monochromatic surface T by capping every white square on ∂A with a odd cube. There is a surface $S' \subset T$ with the same boundary as S. Since $\text{area}(T) \leq 5\,\text{area}(\partial A)$, there exists a constant c_1 such that $\text{area}(T) < c_1 n^2$. Since S is a minimal monochromatic surface,

$$c_1 n^2 \geq \text{area}(S') \geq \text{area}(S).$$

Now we prove the lower bound. Since S connects the inner and outer boundaries, there is a point $p \in S$ where we can apply Proposition 2.6.18 to $B_{(\delta/3)n}(p)$. Hence

$$\text{area}(S) \geq \kappa((\delta/3)n)^2 =: c_2 n^2$$

and c_2 is of order δ^2. □

The next application of the area growth results is loosely that minimal surfaces "spread out." If $X \subset \partial B_n$ for some n is a surface, we define the ϵ *covering area* of X to be the total area in disjoint $\epsilon n \times \epsilon n$ size squares needed to contain X. We denote this by $\text{Cov}_\epsilon(X)$.

Lemma 2.6.20 (Indenting lemma)**.** *Fix $\delta > 0$. Let $A = B_n \setminus B_{(1-\delta)n}$, and suppose that S is a minimal monochromatic discrete surface connecting the inner and outer boundaries of A. Let $\Gamma_l = \partial(S \cap \partial B_l)$. There exist constants $c, c' > 0$ independent of S and n such that for any $\epsilon > 0$, there exists $l \in ((1 - \epsilon^{1/2})n, n)$ such that*

$$\textit{Cov}_\epsilon(S \cap \partial B_l) \leq c\epsilon^{1/2} n^2$$

and

$$\textit{length}(\Gamma_l) \leq c'\epsilon^{-1/2} n.$$

The analogous statements also hold for $l \in ((1 - \delta)n, (1 - \delta + \epsilon^{1/2})n)$.

Proof. Cut the region between ∂B_n and $\partial B_{(1-\epsilon^{1/2})n}$ into $M = \lfloor \frac{1}{3\epsilon^{1/2}} \rfloor$ layers $L_1, ..., L_M$ of width $\frac{\epsilon^{1/2} n}{M} \geq 3\epsilon n$. By Lemma 2.6.19 plus the pigeonhole principle, there exists j such that

$$\text{area}(S \cap L_j) \leq \frac{c_1 n^2}{M} \leq 3c_1 \epsilon^{1/2} n^2. \tag{2.33}$$

We further subdivide L_j into three layers $L_j^{(1)}, L_j^{(2)}, L_j^{(3)}$, where $L_j^{(2)}$ is the middle one. These each have width at least ϵn. The l we find satisfying the conditions will have $\partial B_l \subset L_j^{(2)}$.

Let J be the number of $(\epsilon n)^3$ sized boxes needed to cover $S \cap L_j^{(2)}$. Then there are at least $J/9$ disjoint cubes of size $(3\epsilon n)^3$ in L_j, centered on an $(\epsilon n)^3$ cube in $L_j^{(2)}$, such that the central cube is in J (i.e., S intersects the $(\epsilon n)^3$ central cube of this $(3\epsilon n)^3$ cube). Given this, we can apply Proposition 2.6.18 to a point in each of the $J/9$ boxes with radius ϵn to get the bound

$$\text{area}(S \cap L_j) \geq \kappa \frac{J}{9}(\epsilon n)^2. \tag{2.34}$$

Combining Equations (2.33) and (2.34) and solving for J gives

$$J \leq \frac{27c_1}{\kappa}\epsilon^{-3/2}.$$

Therefore for any l such that $\partial B_l \subset L_j^{(2)}$,

$$\text{Cov}_\epsilon(S \cap \partial B_l) \leq J(\epsilon n)^2 \leq \frac{27c_1}{\kappa}\epsilon^{1/2}n^2.$$

so the first part holds with $c = 27c_1/\kappa$.

For the second part, we note that $L_j^{(2)}$ has width ϵn. Any square in $S \cap L_j^{(2)}$ contributes length at most four to the curves Γ_l for l such that $\partial B_l \subset L_j^{(2)}$, thus $\sum_{l:\partial B_l \subset L_j^{(2)}} \text{length}(\Gamma_l) \leq 4\text{area}(S \cap L_j^{(2)})$. Therefore by Equation (2.33) and the pigeonhole principle again, we can find l with $\partial B_l \subset L_j^{(2)}$ such that

$$\text{length}(\Gamma_l) \leq \frac{12c_1\epsilon^{1/2}n^2}{\epsilon n} = 12c_1\epsilon^{-1/2}n,$$

so the second part holds with $c' = 12c_1$. □

2.6.6 Tilings sampled from ergodic measures

Recall that $\mathcal{P}_e$ denotes the measures on Ω which are ergodic with respect to the action of $\mathbb{Z}^3_{\text{even}}$, and that $\{\eta_i\}_{i=1}^3$ denote the standard unit basic vectors. In this section we prove a few results for tilings sampled from ergodic measure of mean current $s \in \mathcal{O}$. In the proof of the patching theorem, we use a "test tiling" sampled from an ergodic measure which we compare with the two other tilings we want to patch. First we note that there exist ergodic measures of mean current s for all $s \in \mathcal{O}$.

Lemma 2.6.21. *For every $s \in \mathcal{O}$, there exists an ergodic measure on dimer tilings of $\mathbb{Z}^3$ of mean current s.*

Remark 2.6.22. We use methods called *chain swapping* described in Section 2.7.4 to construct ergodic measures of any edge, then face, then interior mean current from the ones for $s \in \mathcal{V}$. The only results that we use about chain swapping here (Propositions 2.7.14 and 2.7.16) are essentially computations, and do not rely on any results presented in this section.

Note that by Theorem 2.4.7, there exist ergodic measures for all mean currents $s \in \partial\mathcal{O}$, so we only need to use chain swapping to show existence for $s \in \text{Int}(\mathcal{O})$. However we choose not to rely on this here, since the chain swapping techniques allow us to show existence easily just from existence of ergodic measures at the vertices of $\partial\mathcal{O}$.

Proof of Lemma 2.6.21. Let $\mathcal{V} \subset \partial\mathcal{O}$ denote its vertices. For each $s \in \mathcal{V}$, the atomic measure which samples the corresponding brickwork pattern is an ergodic measure of mean current s.

Given any $s \in \mathrm{Int}(\mathcal{O})$ (resp. s contained in a face of $\partial\mathcal{O}$, resp. an edge of $\partial\mathcal{O}$), there exists $p \in (0,1)$ such that

$$s = (1-p)s_1 + ps_2$$

for $s_1, s_2 \in \partial\mathcal{O}$ (resp. contained in the edges of $\partial\mathcal{O}$, resp. contained in $\mathcal{V} \subset \partial\mathcal{O}$). Let μ_1 and μ_2 be ergodic measures of mean current s_1, s_2 respectively. Let μ be a measure on $\Omega \times \Omega$ which is an ergodic coupling of μ_1 and μ_2, let μ' be obtained from μ by chain swapping with swap probability p, and let μ_1' and μ_2' denote its marginals. By Proposition 2.7.16, the mean current of μ_1' is

$$s(\mu_1') = (1-p)s_1 + ps_2 = s.$$

By Proposition 2.7.14, μ' is an ergodic measure on $\Omega \times \Omega$. Therefore μ_1' is an ergodic measure of mean current s. □

Recall that the *pretiling flow* v_τ is defined for e oriented from even to odd by

$$v_\tau(e) = \begin{cases} +1 & e \in \tau \\ 0 & e \notin \tau. \end{cases}$$

Let S be an oriented discrete surface with outward normal vector ξ. Applying Definition 2.6.11 to v_τ, the *flux* of v_τ through S by

$$\mathrm{flux}(v_\tau, S) = \sum_{e \in E(\mathbb{Z}^3), e \cap S \neq \emptyset} \mathrm{sign}\langle \xi(e \cap S), e \rangle v_\tau(e).$$

As in the definition, $E(\mathbb{Z}^3)$ denotes the edges of $\mathbb{Z}^3$ oriented from even to odd. Flux of v_τ has a simple combinatorial interpretation. It counts the number of tiles in τ which cross S, with sign corresponding to the parity of the tile. If S is monochromatic black, then $\mathrm{flux}(v_\tau, S)$ is minus the number of tiles in τ which cross S.

Theorem 2.6.23. *Let P be a coordinate plane with normal vector η_i for some $i = 1, 2, 3$, and let $P_n = P \cap B_n$ (recall $B_n = [-n, n]^3$). If $\mu \in \mathcal{P}_e$ has mean current $s \in \mathcal{O}$, then*

$$\lim_{n \to \infty} \frac{1}{|P_n|} \mathit{flux}(v_\tau, P_n) = \frac{1}{2}\langle s, \eta_i \rangle$$

where this limit converges almost surely and in probability.

Remark 2.6.24. The completely equivalent statement holds with v_τ replaced by f_τ. In fact since P_n is contained in a coordinate plane $\mathrm{flux}(f_\tau, P_n) = \mathrm{flux}(v_\tau, P_n) + o(1)$. The reason for the $\frac{1}{2}$ factor is that the mean current is the average current per *even* vertex. The number of even vertices in P_n is $|P_n|/2$.

Proof. Without loss of generality let P denote the (x, y) coordinate plane, so the normal vector is $\eta_3 = (0, 0, 1)$. Recall that discrete surfaces consist of squares in $\mathbb{Z}^3 + (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, so $P = \{(x, y, 1/2) \ : \ x, y \in \mathbb{R}\}$. Let $\mathbb{Z}^2_{\text{even}} = (\mathbb{Z}^2 \times \{0\}) \cap \mathbb{Z}^3_{\text{even}}$. Since μ is invariant under the $\mathbb{Z}^2_{\text{even}}$ action, we can apply the ergodic theorem for this subaction. Let

$$T = P \cap ([-1/2, 3/2] \times [-1/2, 1/2] \times [0, 1])$$

and $\text{even}(B_n) = \mathbb{Z}^3_{\text{even}} \cap B_n$. The set T is defined so that it contains the intersection points of two adjacent edges, namely $(0, 0, 0)$ to $(0, 0, 1)$ and $(1, 0, 0)$ to $(1, 0, 1)$. Consider the function $F : \Omega \to \mathbb{R}$ given by $F(\tau) = \frac{1}{2}\text{flux}(v_\tau, T)$. We have that

$$\left| \frac{1}{|\text{even}(B_n)|} \sum_{\eta \in B_n} F(\tau + \eta) - \frac{1}{|P_n|}\text{flux}(v_\tau, P_n) \right| = o(n^2)$$

and by the ergodic theorem

$$\lim_{n\to\infty} \frac{1}{|P_n|}\text{flux}(v_\tau, P_n)$$

exists in probability and almost surely. Temporarily we call the limit $\text{flux}^\star(v_\tau, P)$. We know that this is $\mathbb{Z}^2_{\text{even}}$-invariant. By integrating the flux across T we get that

$$\int_\Omega \text{flux}^\star(v_\tau, P)\, \mathrm{d}\mu(\tau) = \int_\Omega F(\tau)\, \mathrm{d}\mu(\tau) = \frac{1}{2}\langle s, \eta_3 \rangle.$$

If we can now prove that the average $\text{flux}^\star(v_\tau, P)$ is in fact not just $\mathbb{Z}^2_{\text{even}}$-invariant but also $\mathbb{Z}^3_{\text{even}}$-invariant it will follow that it is constant almost surely and equal to $\frac{1}{2}\langle s, \eta_3 \rangle$. To show this, we use the fact that v_τ is essentially "divergence free". Indeed the tiling flow $f_\tau = v_\tau - r$ where r is a reference flow and the flux of r across T is zero. It follows that

$$|\text{flux}(f_\tau, P_n) - \text{flux}(v_\tau, P_n)| = o(n)$$

and thereby we have that

$$\lim_{n\to\infty} \frac{1}{|P_n|}\text{flux}(v_\tau, P_n) = \lim_{n\to\infty} \frac{1}{|P_n|}\text{flux}(f_\tau, P_n).$$

Since f_τ is divergence free it follows that for all $\gamma \in \mathbb{Z}^3_{\text{even}} \setminus \mathbb{Z}^2_{\text{even}}$,

$$|\text{flux}(f_\tau, P_n) - \text{flux}(f_{\tau+\gamma}, P_n)|$$

is given by the flux through the sides of parallelopiped formed by P_n and $P_n + \gamma$. Thus for a fixed $\gamma \in \mathbb{Z}^3_{\text{even}}$

$$|\text{flux}(f_\tau, P_n) - \text{flux}(f_{\tau+\gamma}, P_n)| = o(n)$$

and hence the same holds for v_τ in place of f_τ. Thus $\text{flux}^\star(v_\tau, P)$ is $\mathbb{Z}^3_{\text{even}}$-invariant, which completes the proof. □

As a straightforward corollary of Theorem 2.6.23, we see that tilings sampled from ergodic measures satisfy the ϵ-nearly-constant condition with high probability.

Corollary 2.6.25. *Fix $\epsilon > 1$. If μ is an ergodic measure of mean current s, then a tiling τ sampled from μ is ϵ-nearly-constant on B_n with value s with probability arbitrarily close to 1 for n large enough.*

The final goal of this section is to get an estimate on the expected flux of a pretiling flow across a monochromatic discrete surface (e.g. the boundary of a counterexample). To do this, we use the following combinatorial result.

Lemma 2.6.26. *Suppose that S is a monochromatic discrete surface. Let $X_1, X_2, X_3 \subset S$ be the sets of squares with normal vectors η_1, η_2, η_3 respectively.*

If S is a closed surface, then $|X_1| = |X_2| = |X_3|$. If S has boundary ∂S, then for all pairs $i \neq j$,

$$|X_i| = |X_j| + O(\mathit{length}(\partial S)).$$

Proof. If S is a closed discrete surface, then every edge of S is contained in either two or four squares from S (four can happen if the edge is an edge of non-manifold points). For an edge contained in two squares from the surface, we say those squares are *neighbors*. For an edge contained in four squares, we arbitrarily split the four into pairs, and say that the paired squares are neighbors. With this definition, every square $f \in S$ has exactly four neighbors. Since S is monochromatic, if $f \in X_1$, then two of its neighbors are in X_2 and two of its neighbors are in X_3 (and similarly for any permutation of $1, 2, 3$).

View the set of all squares as a graph, where each square corresponds to a vertex, and two vertices are connected by an edge if the corresponding squares are neighbors. The number of edges connecting X_1 to X_2 must be equal to $2|X_1|$ (since every $f \in X_1$ has two neighbors in X_2), and analogously must be equal to $2|X_2|$ (since every $f \in X_2$ has two neighbors in X_1). Therefore $|X_1| = |X_2|$. An analogous argument shows that $|X_3| = |X_1|$ and completes the proof in the closed surface case.

If S is not closed, then squares $f \in S$ which contain an edge along ∂S do not have exactly four neighbors. Therefore the result holds up to an error of length(∂S). □

Lemma 2.6.27. *Let S be a monochromatic black surface with boundary ∂S, and let Θ be the collection of odd cubes adjacent to S. For any tiling τ,*

$$|\mathit{flux}(v_\tau, S)| \leq |\Theta|.$$

If μ is an ergodic measure of mean current $s \in \mathit{Int}(\mathcal{O})$, then there is a constant $K_\mu \in (0, 1)$ independent of S such that

$$\mathbb{E}_\mu[|\mathit{flux}(v_\tau, S)|] \geq K_\mu|\Theta| + O(\mathit{length}(\partial S)).$$

Remark 2.6.28. If μ has mean current $s \in \partial\mathcal{O}$, then $K_\mu = 0$.

Remark 2.6.29. Since S is monochromatic black, $\mathrm{flux}(v_\tau, S)$ is minus the number of tiles from τ crossing S. This is why we add the absolute value.

Proof. Any tile $e \in \tau$ crossing S contains a cube from Θ. From this it follows immediately that $|\mathrm{flux}(v_\tau, S)| \leq |\Theta|$.

Let $p_1, \ldots, p_6$ be the probabilities under μ for the six types of tiles, ordered so that $s_1 = p_1 - p_2$, $s_2 = p_3 - p_4$ and $s_3 = p_5 - p_6$. Similarly let $N_1, \ldots, N_6$ be the six types

of squares on S, where the tile type parallel to the outward pointing normal vector at a square $f \in N_i$ has probability p_i. The random variable $|\text{flux}(v_\tau, S)|$ can also be written as $\sum_{f \in S} \mathbb{1}_f(\tau)$, where $\mathbb{1}_f(\tau)$ is the indicator variable which is 1 if there is a tile in τ crossing f and 0 otherwise. From this, we see that

$$\mathbb{E}_\mu[|\text{flux}(v_\tau, S)|] = \sum_{i=1}^{6} p_i N_i.$$

We minimize the right hand side to get a positive lower bound (clearly 0 is a lower bound). Let $\text{area}(S) = N = \sum_{i=1}^{6} N_i$. By Lemma 2.6.26, $N_1 + N_2$, $N_3 + N_4$ and $N_5 + N_6$ are equal to $N/3$ up to an error of $O(\text{length}(\partial S))$. Thus

$$\begin{aligned}\sum_{i=1}^{6} p_i N_i &\geq \max\{p_1 N_1 + p_2 N_2, p_3 N_3 + p_4 N_4, p_5 N_5 + p_6 N_6\} \\ &\geq \max\{\min\{p_1, p_2\}, \min\{p_3, p_4\}, \min\{p_5, p_6\}\}(N/3 + O(\text{length}(\partial S)).\end{aligned}$$

Since $s \in \text{Int}(\mathcal{O})$, at least four of $\{p_i\}_{i=1}^6$ are nonzero, including one from each pair. Noting that $N = \text{area}(S)$, and that

$$|\Theta| \leq \text{area}(S) \leq 6|\Theta|,$$

this proves the result with a constant of the form $K_\mu = p_i/3$ for some i such that $p_i \neq 0$. □

2.6.7 Proof of the patching theorem

We now give the proof of the patching theorem (Theorem 2.6.14), as described in Section 2.6.3. We refer throughout to the corresponding figures from the outline there.

Proof of Theorem 2.6.14. Since $\tau_1, \tau_2 \in \Omega$, A_n is balanced for all n. Thus by Corollary 2.6.9, if A_n is not tileable, it has a minimal counterexample U, i.e. a set with

$$\text{imbalance}(U) = \text{even}(U) - \text{odd}(U) > 0,$$

despite U having only odd cubes on its interior boundary. Let S denote the interior boundary surface of U. Since A_n is tileable with just the τ_1 boundary condition (resp. with just the τ_2 booundary condition), S must connect the inner and outer boundaries of A_n. Thus by Lemma 2.6.20, we can find $l_+ \in ((1-\epsilon^{1/2})n, n)$ and $l_- \in ((1-\delta)n, (1-\delta+\epsilon^{1/2})n)$ such that for $l = l_+$ or $l = l_-$,

$$\text{Cov}_\epsilon(S \cap \partial B_l) \leq c\epsilon^{1/2} n^2 \tag{2.35}$$

$$\text{length}(\partial(S \cap \partial B_l)) \leq c'\epsilon^{-1/2} n \tag{2.36}$$

where c, c' are constants. We define the "middle region" $A_{\text{mid}} = (B_{l_+} \setminus B_{l_-})$, see Figure 2.25. Then we let

$$U' = U \cap A_{\text{mid}}.$$

See Figure 2.26. Let μ be an ergodic measure of mean current s (this exists by Lemma 2.6.21). Let Θ be the collection of odd cubes adjacent to $S' = S \cap U'$. Note that

$$\text{area}(S') \leq |\Theta| \leq 6\text{area}(S')$$

By Lemma 2.6.19, $|\Theta| \geq \text{area}(S') \geq c_2' n^2$, where $c_2' \sim (\delta - 2\epsilon^{1/2})^2$. Thus by Lemma 2.6.27 and Equation (2.36), there is a constant $K_\mu \in (0,1)$ such that

$$\mathbb{E}_\mu[|\text{flux}(v_\tau, S')|] \geq K_\mu c_2' n^2 + O(\epsilon^{-1/2} n).$$

By Theorem 2.6.23, for any coordinate plane P, a tiling τ sampled from μ is ϵ-nearly-constant on P with value s with probability approaching 1 as n goes to ∞. Therefore we can sample a tiling τ from μ which is ϵ-nearly-constant with value s on A_{mid} for n large enough and satisfies

$$|\text{flux}(v_\tau, S')| \geq K_\mu c_2' n^2 + O(\epsilon^{-1/2} n). \tag{2.37}$$

We fix this choice of τ for the rest of the proof. Define U_τ to be the region covered by the tiles from τ which intersect U', see Figure 2.27. Since U_τ is tileable,

$$\text{imbalance}(U_\tau) = \text{even}(U_\tau) - \text{odd}(U_\tau) = 0.$$

We next define a new region U'', which is U_τ minus the even cubes adjacent to S' (see Figure 2.27 again). Note that the region $U_\tau \setminus U''$ consists of only even cubes. By Equation (2.37),

$$|U_\tau \setminus U''| \geq K_\mu c_2' n^2 + O(\epsilon^{-1/2} n).$$

Therefore

$$\text{imbalance}(U'') \leq -K_\mu c_2' n^2 + O(\epsilon^{-1/2} n).$$

It remains to show that imbalance(U) is very close to imbalance(U''), and this is where we use the ϵ-nearly-constant condition. We split $A_n \setminus A_{\text{mid}}$ into two regions. First we define the "thin region" A_{thin} to be the union of columns parallel to one of the coordinate directions which connect ∂A_n and ∂A_{mid}. See Figures 2.25 and 2.28. The complement of $A_{\text{mid}} \cup A_{\text{thin}}$ we call the "corner region" A_{corner}, and consists of a neighborhood of the edges of the outer boundary cube of ∂A_n, and the inner boundary cube of ∂A_{mid}. We note that

$$\text{area}(A_{\text{corner}} \cap \partial A_n) \leq 24\epsilon^{1/2} n^2.$$

The constant factor comes from the fact that the cube has 12 edges. Therefore

$$\text{imbalance}(U \cap A_{\text{corner}}) \leq 24\epsilon^{1/2} n^2.$$

We define $U_{\text{shadow}} = (U \setminus U'') \cap A_{\text{thin}}$. We bound the imbalance of U_{shadow} column-by-column, where each column C consists of a straight line path of single cubes from ∂A_{mid} to ∂A_n. Recall U_{shadow} is defined with boundary condition τ on ∂A_{mid} and boundary condition τ_1 or τ_2 on ∂A_n. We also note that for any column C, since S is monochromatic black, imbalance$(U \cap C) \leq +1$.

We cut ∂A_{mid} into $(\epsilon n) \times (\epsilon n)$ patches α. For each α, there is a corresponding patch β on ∂A_n matched to α by columns. For any patch $\alpha \subset \partial A_{\text{mid}}$, since τ is sampled from an ergodic measure, it is ϵ-nearly-constant with value s on ∂A_{mid} (Corollary 2.6.25). Thus we have that

$$\text{flux}(v_\tau, \alpha) = \frac{1}{2}\langle s, \xi_\alpha \rangle (\epsilon n)^2 + o(\epsilon^2 n^2). \tag{2.38}$$

Let v_* be equal to v_{τ_1} on the outer boundary of ∂A_n and v_{τ_2} on the inner boundary. For $\beta \subset \partial A_n$ the patch connected by a column to α, since τ_1, τ_2 are ϵ-nearly-constant with value s,

$$\text{flux}(v_*, \beta) = \frac{1}{2}\langle s, \xi_\alpha \rangle (\epsilon n)^2 + o(\epsilon^2 n^2). \tag{2.39}$$

(Note that $\xi_\alpha = \xi_\beta$.) For a patch $\alpha \subset \partial A_{\text{mid}}$, let $C(\alpha)$ be the union of columns incident to α. The set U_{shadow} is covered by these column sets, so it remains to control imbalance of $U_{\text{shadow}} \cap C(\alpha)$ for each patch α.

By Equation (2.35), at most $c\epsilon^{1/2} n^2$ of the area of ∂A_{mid} is in patches α which intersect U and U^c. The total imbalance contribution from these is bounded by $c\epsilon^{1/2} n^2$. If $\alpha \subset U^c$ but $C(\alpha)$ still intersects U, then all the columns in $C(\alpha)$ have at least one end on S, and the imbalance in $C(\alpha) \cap U_{\text{shadow}}$ is at most 0.

Now we look at the cases where $\alpha \subset U$. If $C(\alpha) \subset U$, then the total imbalance in $C(\alpha) \cap U$ is $o(\epsilon^2 n^2)$ by Equations (2.38) and (2.39). If $\alpha \subset U$ but $C(\alpha) \not\subset U$, then some of the columns starting from α hit S before hitting β. This means they end on an odd cube. Extending the column all the way to β would only make the imbalance larger, however then the imbalance in $C(\alpha) \cap U$ is bounded above by $o(\epsilon^2 n^2)$ by Equations (2.38) and (2.39). The number of columns is a bounded by a constant independent of n times ϵ^{-2}, hence in total the imbalance in U_{shadow} is bounded by $c\epsilon^{1/2} n^2 + o(n^2)$.

Putting everything together, we have that

$$\begin{aligned} \text{imbalance}(U) &\leq \text{imbalance}(U'') + \text{imbalance}(U_{\text{shadow}}) + \text{imbalance}(U \cap A_{\text{corner}}) \\ &\leq -K_\mu c_2' n^2 + 24\epsilon^{1/2} n^2 + c\epsilon^{1/2} n^2 + O(\epsilon^{-1/2} n) + o(n^2). \end{aligned}$$

Recall that $c_2' \sim (\delta - 2\epsilon^{1/2})^2$, so it gets larger as ϵ gets smaller. Fixing ϵ small enough as a function of the constants, for n large enough $\text{imbalance}(U) \leq 0$ and hence U is not a counterexample. Therefore by Proposition 2.6.9, for n large enough, A_n is tileable with boundary conditions τ_1, τ_2. □

2.6.8 Corollaries for ergodic Gibbs measures

Corollary 2.6.30. *If μ_1, μ_2 are EGMs of the same mean current $s \in \text{Int}(\mathcal{O})$, then $h(\mu_1) = h(\mu_2)$.*

Remark 2.6.31. The proof of the patching theorem (Theorem 2.6.14) uses that $s \notin \partial\mathcal{O}$ since this is a condition of Lemma 2.6.27. This corollary shows that $s \notin \partial\mathcal{O}$ is a necessary condition and not just an artifact of the proof, since if $s \in \partial\mathcal{O}$ then not all ergodic Gibbs measures of mean current s have the same specific entropy (see Proposition 2.4.10).

Proof. Fix $\delta > 0$, and let $B_n = [-n, n]^3$ and $A_n = B_n \setminus B_{(1-\delta)n}$. By the patching theorem (Theorem 2.6.14) with outer boundary condition sampled from μ_1 and inner boundary condition sampled from μ_2, we get that

$$h(\mu_2) \leq (1 + O(\delta))h(\mu_1).$$

Switching them, we find that

$$h(\mu_1) \leq (1 + O(\delta))h(\mu_2).$$

Therefore $h(\mu_1) = h(\mu_2)$. □

Another useful result comes from applying patching to a sequence of ϵ-nearly-constant tilings and a sample from an EGM. This relates the number of tilings of a region with fixed ϵ-nearly-constant boundary conditions to the specific entropy of an EGM. This serves as a lemma in the proof of the lower bound in the large deviation principle (Theorem 2.8.10). Recall that Ω denotes the set of dimer tilings of $\mathbb{Z}^3$.

Proposition 2.6.32. *Fix $\delta > 0$, $s \in Int(\mathcal{O})$, $\epsilon > 0$ small enough, $B_n = [-n, n]^3$, and let $A_n = B_n \setminus B_{(1-\delta)n}$. Suppose that $(\tau_n)_{n\geq 1} \subset \Omega$ is such that τ_n is ϵ-nearly-constant on ∂B_n with value s for n large enough. Let π_n be the uniform probability measure on tilings σ of B_n such that $\sigma \,|_{\partial B_n} = \tau_n$. Then for any EGM μ of mean current s and n large enough,*

$$|B_n|^{-1} H(\pi_n) \geq h(\mu)(1 + O(\delta)).$$

Remark 2.6.33. Since π_n is a uniform measure, $H(\pi_n) = \log Z_n$, where Z_n is the partition function of π_n.

Proof. We apply the patching theorem (Theorem 2.6.14) to $A_n = B_n \setminus B_{(1-\delta)n}$ with τ_n on the outer boundary and a sample from μ on the inner boundary. For n large enough, patching is possible with μ-probability $(1 - \epsilon)$ on an annulus of width δ.

For $\Lambda \subset \mathbb{Z}^3$, let $\Omega(\Lambda)$ denote the free-boundary dimer tilings of Λ. For $\sigma \in \Omega(\Lambda)$, recall that $X(\sigma)$ is the set of *extensions* of σ, i.e.

$$X(\sigma) = \{\tilde{\sigma} \in \Omega : \tilde{\sigma} \,|_\Lambda = \sigma\}.$$

Then we compute

$$\begin{aligned} H_{B_{(1-\delta)n}}(\mu) &= -\sum_{\sigma \in \Omega(B_{(1-\delta)n})} \mu(X(\sigma)) \log \mu(X(\sigma)) \\ &= -\sum_{\substack{\sigma \in \Omega(B_{(1-\delta)n}) \\ \sigma, \tau_n \text{ patchable}}} \mu(X(\sigma)) \log \mu(X(\sigma)) + O(\epsilon \log \epsilon) \end{aligned}$$

Let U_n denote the uniform probability measure on

$$\{\sigma \in \Omega(B_{(1-\delta)n}) : \sigma, \tau_n \text{ patchable}\}.$$

Since uniform measure maximizes entropy,

$$H_{B_{(1-\delta)n}}(\mu) \leq H(U_n) + O(\epsilon \log \epsilon) \leq H(\pi_n) + O(\epsilon \log \epsilon).$$

Thus for n large enough such that the patching theorem applies for τ_n and a sample from μ on $A_n = B_n \setminus B_{(1-\delta)n}$, with probability $(1-\epsilon)$ we have that

$$|B_n|^{-1} H(\pi_n) \geq (1 + O(\delta))|B_{(1-\delta)n}|^{-1} H_{B_{(1-\delta)n}}(\mu) \geq (1 + O(\delta))h(\mu).$$

□

2.7 Properties of entropy

In this section we prove results about the entropy functions ent and Ent introduced in Section 2.2.3. Recall that Ω is the set of dimer tilings of $\mathbb{Z}^3$, and that $\mathcal{P}$ denotes the space of $\mathbb{Z}^3_{\text{even}}$-invariant probability measures on Ω. Further recall that for any $s \in \mathcal{O}$, we define $\mathcal{P}^s \subset \mathcal{P}$ to be the set of measures which also have mean current s, $\mathcal{P}_e \subset \mathcal{P}$ to be the set of ergodic measures, and $\mathcal{P}_e^s$ to be the set of ergodic measures with mean current s. The mean-current entropy function $\text{ent} : \mathcal{O} \to [0, \infty)$ is defined by

$$\text{ent}(s) = \sup_{\mu \in \mathcal{P}^s} h(\mu),$$

where $h(\cdot)$ denotes specific entropy (see Section 2.2.3).

We saw in Section 2.4 that ent is equal to ent_{loz} when restricted to any face of $\partial\mathcal{O}$ (Theorem 2.4.7). In particular this implies that ent is strictly concave when restricted to any face of $\partial\mathcal{O}$ (Corollary 2.4.8). The main result of this section is that ent is strictly concave on all of $\mathcal{O} \setminus \mathcal{E}$, where $\mathcal{E}$ denotes the edges of $\partial\mathcal{O}$ (Theorem 2.7.22).

In Section 2.7.1 we show that the supremum of $\{h(\mu) : \mu \in \mathcal{P}^s\}$ is realized by a Gibbs measure for all $s \in \mathcal{O}$ (Theorem 2.7.2). It is a classical result going back to Lanford and Ruelle [LR69] that entropy maximizers in $\mathcal{P}$ are Gibbs measures. We extend this to show that the entropy maximizer in $\mathcal{P}^s$, where the mean current is fixed, is also a Gibbs measure. In Section 2.7.2 we show using elementary methods that ent is concave and continuous on $\mathcal{O}$.

The proof that ent is strict concave on $\mathcal{O} \setminus \mathcal{E}$ (Theorem 2.7.22) requires some new tools and ideas. We use a version of a technique called *cluster swapping* used in [She05, Chapter 8], which we call *chain swapping*. As set up for the proof of strict concavity in Section 2.7.5 we prove some preliminary results about flows in the double dimer model (Section 2.7.3) and introduce the chain swapping technique (Section 2.7.4). Combining chain swapping with the results for ent on $\partial\mathcal{O}$ from Section 2.4 we show that ent is strictly concave on $\mathcal{O} \setminus \mathcal{E}$.

Strict concavity has a number of important consequences. We saw in Corollary 2.6.30 that if μ_1, μ_2 are EGMs with mean current $s \in \text{Int}(\mathcal{O})$, then $h(\mu_1) = h(\mu_2)$. Combining this with strict concavity, we show that if $s \in \text{Int}(\mathcal{O})$ and $\mu \in \mathcal{P}^s$, then $h(\mu) = \text{ent}(s)$ if and only if μ is an EGM or weighted average of EGMs all of mean current s (Theorem 2.7.23), and therefore that there exists an EGM of every mean-current $s \in \mathcal{O}$ (Corollary 2.7.25).

Heuristically, the main goal of this section is to show that the mean current s captures broad statistics of dimer tilings sampled from $\mu \in \mathcal{P}_e^s$ when $s \in \mathrm{Int}(\mathcal{O})$.

Finally in Section 2.7.6, we leverage properties of ent to study the entropy functional $\mathrm{Ent} : AF(R) \to [0, \infty)$ on asymptotic flows given by

$$\mathrm{Ent}(f) = \frac{1}{\mathrm{Vol}(R)} \int_R \mathrm{ent}(f(x))\, \mathrm{d}x.$$

The rate function for the large deviation principle in Section 2.8 will be $-\mathrm{Ent}$ (up to an additive constant). Using the properties of ent from earlier in the section, we show that Ent is upper semicontinuous in the Wasserstein topology (Proposition 2.7.29) and strictly concave on the subspace of flows which never take values in the edges $\mathcal{E} \subset \partial\mathcal{O}$ (Corollary 2.7.27). After that, we adapt an argument of V. Gorin [Gor21] to show that there is a unique Ent maximizer in $AF(R, b)$ for any boundary asymptotic flow b under the mild condition that (R, b) is *semi-flexible* (Definition 2.7.34, Theorem 2.7.36).

2.7.1 Entropy maximizers of a given mean current are Gibbs measures

We first study the maximizers of specific entropy $h(\cdot)$ in $\mathcal{P}^s$ for $s \in \mathcal{O}$ fixed. It is straightforward to show that there exists a measure $\mu \in \mathcal{P}^s$ which achieves $\sup\{h(\mu) : \mu \in \mathcal{P}^s\}$ for any $s \in \mathcal{O}$.

Lemma 2.7.1. *Let $s \in \mathcal{O}$. There exists $\mu \in \mathcal{P}^s$ such that $h(\mu) = \mathit{ent}(s)$.*

Proof. The space $\mathcal{P}^s$ is compact with respect to the weak star topology. Since h is an upper semicontinuous function of the measure [Kel98, Theorem 4.2.4] it must achieve its maxima in $\mathcal{P}^s$. □

Theorem 2.7.2. *Fix $s \in \mathcal{O}$. If $\mu \in \mathcal{P}^s$ has $h(\mu) = \mathit{ent}(s)$ then μ is a Gibbs measure.*

If the mean current is not fixed, then this is a standard result originally shown by Landford and Ruelle in [LR69]. The main idea of the proof is a variational argument which says that if a measure μ is not Gibbs, then there is a "perturbation" of μ which has more entropy. We need to show that this perturbation does not change the mean current, and this is the purpose of Lemma 2.7.3. After that, our proof of Theorem 2.7.2 is inspired by the exposition in [BS94].

To show the mean current does not change, we use double dimers to compare the mean currents of the two measures. Double dimers will be a tool throughout Section 2.7. There is a natural action of the group $\mathbb{Z}^3_{\mathrm{even}}$ on the product $\Omega \times \Omega$ acting coordinate wise. Superimposing two dimer tilings τ_1 and τ_2 gives us a *double dimer configuration* (τ_1, τ_2). The union $\tau_1 \cup \tau_2$ consists of finite cycles, infinite paths and isolated double edges. Each cycle or infinite path in (τ_1, τ_2) is *oriented* in a way that agrees with the direction of the τ_1 flow (for edges in τ_1) or opposite the direction of the τ_2 flow (for edges in τ_2).

Lemma 2.7.3. *Let $\mathfrak{m}$ be a $\mathbb{Z}^3_{even}$-invariant measure on $\Omega \times \Omega$ such that for $\mathfrak{m}$ almost every (τ_1, τ_2), the union $\tau_1 \cup \tau_2$ does not contain infinite paths. Then*

$$s(\pi_1(\mathfrak{m})) = s(\pi_2(\mathfrak{m}))$$

where $\pi_i : \Omega \times \Omega \to \Omega$ *is projection onto the* i^{th} *coordinate for* $i = 1, 2$.

Remark 2.7.4. We remark that non-existence of infinite paths in a sample from a coupling like this has been used in other related but different ways in statistical mechanics, e.g. to show that two Gibbs measures are the same if there are no infinite paths in a sample from the coupling, or to compute covariances. See for example [vdBer93], [vdBS94].

Proof. Since the mean current is an affine function of measure, by the ergodic decomposition theorem it is sufficient to prove this lemma for $\mathfrak{m}$ ergodic. For the rest of the proof we assume that $\mathfrak{m}$ is an ergodic measure.

Recall that v_τ is the *pretiling flow* of τ (for the definition see Equation (2.4) in Section 2.2.1), and let (τ_1, τ_2) be a sample from $\mathfrak{m}$. By assumption all paths $\gamma \subset (\tau_1, \tau_2)$ are finite loops or double edges (which are loops with just one edge from each tiling). Let $E(\gamma)$ denote the edges along γ oriented from even to odd. (If $e \in E(\gamma)$, $-e$ is e with orientation reversed.) With a slight abuse of notation, we also view e as a vector oriented from even to odd. Since γ is a loop,

$$\sum_{e \in E(\gamma)} v_{\tau_1}(e)e = \sum_{e \in E(\gamma)} v_{\tau_2}(e)e. \tag{2.40}$$

Given $x \in \mathbb{Z}^3$ let γ_x denote the loop in (τ_1, τ_2) containing x. We denote the number of edges in a loop γ by $\text{length}(\gamma)$. For any $\epsilon > 0$, there exists k such that

$$\mathfrak{m}(\text{length}(\gamma_0) > k) < \epsilon.$$

Let $B_n := [1, n]^3$. By the mean ergodic theorem, there is n large enough such that with $\mathfrak{m}$-probability $1 - \epsilon$, the double dimer configuration (τ_1, τ_2) satisfies the following.

1. We have
$$|\{x \in B_n \ : \ \text{length}(\gamma_x) > k\}| < 2\epsilon n^3. \tag{2.41}$$

2. Let $E(B_n)$ denote the edges in B_n oriented from even to odd. For $i = 1, 2$ and any unit coordinate vector η,
$$\left| \langle s(\pi_i(\mathfrak{m})), \eta \rangle - \frac{2}{n^3} \sum_{e \in E(B_n)} \langle v_{\tau_i}(e)e, \eta \rangle \right| < \epsilon. \tag{2.42}$$

Let $C_n := \{x \in B_n \ : \ \gamma_x \subset B_n\}$ denote the $\mathfrak{m}$-random set of points on loops in (τ_1, τ_2) contained in B_n. By Equation (2.41), with $\mathfrak{m}$-probability $1 - \epsilon$,

$$|C_n| \geq (n - k)^3 - 2\epsilon n^3. \tag{2.43}$$

Clearly $\{\gamma_x : x \in C_n\}$ is a union of loops. Let $E(C_n)$ denote the edges of loops in this collection oriented from even to odd. By Equation (2.40),

$$\sum_{e \in E(C_n)} \Big(v_{\tau_1}(e)e - v_{\tau_2}(e)e \Big) = 0. \tag{2.44}$$

There are at most $\frac{n^3-|C_n|}{2}$ tiles in τ_i in $E(B_n) \setminus E(C_n)$ for $i = 1, 2$. Therefore by Equation (2.43), with $\mathfrak{m}$ probability $1 - 2\epsilon$

$$\left| \frac{2}{n^3} \sum_{e \in E(B_n)} v_{\tau_i}(e)e - \frac{2}{n^3} \sum_{e \in E(C_n)} v_{\tau_i}(e)e \right| \leq \frac{n^3 - (n-k)^3 + 2\epsilon n^3}{n^3} = 1 + 2\epsilon - \frac{(n-k)^3}{n^3}. \quad (2.45)$$

Combining Equations (2.42), (2.44), (2.45) gives that for any unit coordinate vectors η,

$$\left| \langle s(\pi_1(\mathfrak{m})), \eta \rangle - \langle s(\pi_2(\mathfrak{m})), \eta \rangle \right| < 6\epsilon + \frac{2n^3 - 2(n-k)^3}{n^3}. \quad (2.46)$$

Taking $n \to \infty$ and then $\epsilon \to 0$ completes the proof.

□

To prove Theorem 2.7.2, we mimic the perturbative argument of [BS94, Proposition 1.19], applying Lemma 2.7.3 to show that this does not change the mean current.

Proof of Theorem 2.7.2. It suffices to show that if $\mu \in \mathcal{P}^s$ is not a Gibbs measure, then there exists $\nu \in \mathcal{P}^s$ such that $h(\nu) > h(\mu)$. Under the assumption that μ is not a Gibbs measure, there exists a finite set $R \subset \mathbb{Z}^3$ and a positive measure set $\Omega' \subset \Omega$ such that for all $\tau \in \Omega'$, the conditional distribution of possible extensions of $\tau|_{\mathbb{Z}^3 \setminus R}$ to a tiling of $\mathbb{Z}^3$ is not uniform. We can assume using stationarity that R is contained in the positive quadrant. Let $n \in \mathbb{N}$ be such that $R \subset [1, n-1]^3$. Since the number of tilings of R depends only on the tiling restricted to $S := [0, n]^3 \setminus R$ there exists a tiling τ_0 in the support of μ such that the conditional distribution on the possible extensions of $\tau_0|_S$ to R is not uniform. Since entropy is maximized by the uniform measure we have that there is a $\delta > 0$ such that

$$H(\text{uniform distribution on extensions of } \tau_0|_S \text{ to } R) - H(\mu|_R \text{ conditioned on } \tau_0|_S) > \delta.$$

We now construct a modification of μ and show that it has the same mean current but more entropy. For this take a sample τ from μ. Let n be an odd integer and divide $\mathbb{Z}^3$ into translates of $B = [0, n]^3$ by $(n+1)\mathbb{Z}^3$. For each such translated box B, resample τ in B conditioned on $\tau\ |_{\partial^\circ B}$, where $\partial^\circ B = B \setminus [1, n-1]^3 \subset B$ is the inner boundary of B. This gives us a new measure ν on Ω, which is invariant with respect to the $(n+1)\mathbb{Z}^3$ subaction. By averaging ν with respect to translations by elements of $[0, n]^3 \cap \mathbb{Z}^3_{\text{even}}$ we get a $\mathbb{Z}^3_{\text{even}}$-invariant measure which we denote ν'.

Let $\mathfrak{m}$ be a measure on $\Omega \times \Omega$ which is a coupling of μ and ν', where the sample from ν' is derived by the construction above from the μ sample. If (τ_1, τ_2) is sampled from $\mathfrak{m}$, τ_1 and τ_2 differ only on the interiors of copies of B. Therefore (τ_1, τ_2) has no infinite paths $\mathfrak{m}$-a.s., so by Lemma 2.7.3,

$$s(\nu') = s(\mu).$$

On the other hand by the ergodic theorem, there exists $\epsilon > 0$ for which there is a $(n+1)\mathbb{Z}^3$-invariant set $A \subset \Omega$ with the following properties:

1. $\mu(A) > \epsilon$.

2. For all $\tau \in A$, $\tau_0|_S$ appears in translated boxes B with density greater than ϵ.

Therefore

$$h(\nu') - h(\mu) > \frac{1}{(n+1)^3}\epsilon^2\delta.$$

□

The proof of Theorem 2.7.2 also has a useful consequence for the double dimer model. Recall the maps $\pi_1, \pi_2 : \Omega \times \Omega \to \Omega$ given by $\pi_i(\tau_1, \tau_2) = \tau_i$. Let $\mathcal{P}^{s_1,s_2}$ be the space of invariant probability measures μ on $\Omega \times \Omega$ such that $\pi_i(\mu) \in \mathcal{P}^{s_i}$ for $i = 1, 2$.

Corollary 2.7.5. *Let $s_1, s_2 \in \mathcal{O}$. Then*

$$\sup_{\mu \in \mathcal{P}^{s_1,s_2}} h(\mu) = ent(s_1) + ent(s_2).$$

Further, the measures $\mu \in \mathcal{P}^{s_1,s_2}$ which maximize specific entropy on $\mathcal{P}^{s_1,s_2}$ are Gibbs measures on $\Omega \times \Omega$ and satisfy $h(\pi_i(\mu)) = ent(s_i)$ for $i = 1, 2$.

Proof. For any measure $\mu \in \mathcal{P}^{s_1,s_2}$ we have that

$$h(\mu) \leq h(\pi_1(\mu)) + h(\pi_2(\mu)) \leq \text{ent}(s_1) + \text{ent}(s_2). \tag{2.47}$$

For $i = 1, 2$, let $\nu_i \in \mathcal{P}^{s_i}$ be such that $h(\nu_i) = \text{ent}(s_i)$. The product measure $\nu = \nu_1 \times \nu_2$ has $h(\nu) = \text{ent}(s_1) + \text{ent}(s_2)$, so by Equation (2.47) ν maximizes specific entropy among measures in $\mathcal{P}^{s_1,s_2}$. Therefore if μ is a maximizer it must be in the equality case in Equation (2.47), which implies that $h(\pi_1(\mu)) = \text{ent}(s_1)$ and $h(\pi_2(\mu)) = \text{ent}(s_2)$.

Finally the proof that the entropy maximizer must be a Gibbs measure is exactly the same as the proof of Theorem 2.7.2; if a measure in $\mathcal{P}^{s_1,s_2}$ is not a Gibbs measure then we can increase its entropy by locally modifying it. □

2.7.2 Basic properties of ent

In this section we prove some straightforward properties of the mean current entropy function ent. The only tools here are basic real analysis and properties of $h(\cdot)$.

Lemma 2.7.6. *ent is a concave function on $\mathcal{O}$.*

Proof. Fix $u, v \in \mathcal{O}$ and $\alpha \in (0, 1)$. By Lemma 2.7.1 we know that the entropy function $h(\cdot)$ on $\mathcal{P}^u$ (resp. on $\mathcal{P}^v$) achieves a maximum say at μ (resp. at ν). Given this we have that $\alpha\mu + (1-\alpha)\nu \in \mathcal{P}^{\alpha u + (1-\alpha)v}$ and

$$h(\alpha\mu + (1-\alpha)\nu) = \alpha h(\mu) + (1-\alpha)h(\nu) = \alpha\, \text{ent}(u) + (1-\alpha)\text{ent}(v).$$

Thus

$$\text{ent}(\alpha u + (1-\alpha)v) \geq \alpha\, \text{ent}(u) + (1-\alpha)\, \text{ent}(v)$$

which shows that ent is concave. □

Lemma 2.7.7. *ent is an upper semi-continuous function on $\mathcal{O}$.*

Proof. Let $u_n \in \mathcal{O}$ be a sequence such that $u_n \to u$. By Lemma 2.7.1 there exist a measure μ_n maximizing the entropy function $h(\cdot)$ on $\mathcal{P}^{u_n}$. Since the mean current is a continuous function of the measure (Definition 2.2.4), we have that any subsequential limit μ of μ_n must lie in $\mathcal{P}^u$. Since $h(\cdot)$ is upper semicontinuous as a function on $\mathcal{P}$,

$$\mathrm{ent}(u) \geq h(\mu) \geq \limsup_{n\to\infty} h(\mu_n) = \limsup_{n\to\infty} \mathrm{ent}(u_n),$$

which completes the proof. □

We put these together to show that ent is continuous.

Lemma 2.7.8. *ent is a continuous function on $\mathcal{O}$.*

Proof. Fix $u \in \mathcal{O}$ and $\epsilon > 0$. Let $M := \sup_{v\in\mathcal{O}} \mathrm{ent}(v)$; M is finite since $\mathcal{O}$ is compact and ent is upper semicontinuous on $\mathcal{O}$ (Lemma 2.7.7). Let $\|\cdot\|_1$ denote the L^1 norm. Again using Lemma 2.7.7, there exists $\delta_1 > 0$ such that if $v \in \mathcal{O}$ is such that $\|v - u\|_1 < \delta_1$, then

$$\mathrm{ent}(u) - \mathrm{ent}(v) > -\epsilon.$$

Choose $L > 1+(2M/\epsilon)$ and $\delta_2 > 0$ small enough such that if $\|v-u\|_1 < \delta_2$ then $u+L(v-u) \in \mathcal{O}$ (note that this is possible even when $u \in \partial\mathcal{O}$ since $\mathcal{O}$ is convex). By Lemma 2.7.6,

$$\frac{L-1}{L}\mathrm{ent}(u) + \frac{1}{L}\mathrm{ent}(u + L(v-u)) \leq \mathrm{ent}(v).$$

Rearranging the equation we get that

$$\mathrm{ent}(u) - \mathrm{ent}(v) \leq \frac{1}{L}\mathrm{ent}(u) - \frac{1}{L}\mathrm{ent}(u + L(v-u)) \leq 2M/L \leq \epsilon.$$

Taking $\delta < \delta_1\delta_2$, if $\|v - u\|_1 < \delta$ then $|\mathrm{ent}(u) - \mathrm{ent}(v)| < \epsilon$, which proves that ent is continuous. □

2.7.3 Flows for the double dimer model

To prove that ent is strictly concave we use a double dimer model construction called *chain swapping*, which is an operation on infinite paths in a double dimer configuration related to the cluster swapping technique used in [She05]. In this section we give some of the necessary background results for the double dimer model, and in the next section we explain what chain swapping is.

Here we look at $\mathbb{Z}^3_{\text{even}}$-invariant couplings of $\mathbb{Z}^3_{\text{even}}$-invariant measures $\mu_1, \mu_2 \in \mathcal{P}_e$ and study properties of the sample (τ_1, τ_2). We distinguish between the pair of dimer tilings (τ_1, τ_2) and the union of tilings $\tau_1 \cup \tau_2$, where we forget the information of which tiling each edge $e \in \tau_1 \cup \tau_2$ belongs to. As we observed earlier, $\tau_1 \cup \tau_2$ is a set of isolated double edges, finite cycles, and infinite paths. We saw in Lemma 2.7.3 that if $\tau_1 \cup \tau_2$ consists of only finite cycles and double edges, then the marginals have the same mean current. This suggests that the infinite paths in a double dimer configuration carry a lot of information about the difference between the mean currents of the measures involved. The main results of this section are Proposition 2.7.10 and Corollary 2.7.11 which make this precise.

Recall from Section 2.2.1 that the *flow associated with a double dimer configuration* (τ_1, τ_2) is

$$f_{(\tau_1,\tau_2)} = v_{\tau_1} - v_{\tau_2},$$

where v_τ is the pretiling flow defined in Section 2.2.1, Equation (2.4). (Equivalently, $f_{(\tau_1,\tau_2)} = f_{\tau_1} - f_{\tau_2}$ where f_τ is the tiling flow, since the reference flows cancel.) Explicitly, for each edge e oriented from even to odd,

$$f_{(\tau_1,\tau_2)}(e) = \begin{cases} 1 & \text{if } e \in \tau_1 \setminus \tau_2 \\ -1 & \text{if } e \in \tau_2 \setminus \tau_1 \\ 0 & \text{if } e \in \tau_1 \cap \tau_2 \text{ or if } e \notin \tau_1 \cup \tau_2. \end{cases}$$

The vector field $f_{(\tau_1,\tau_2)}$ is divergence free, and its flow lines are the cycles and paths of the double dimer configuration $\tau_1 \cup \tau_2$. In particular, each $x \in \mathbb{Z}^3$ is in one of two cases:

1. $f_{(\tau_1,\tau_2)}$ is equal to 1 on exactly two edges e_1, e_2 incident to x, with one of the edges oriented into x and the other oriented out of x.

2. $f_{(\tau_1,\tau_2)}$ is zero on all edges e incident to x.

The set of vertices $x \in \mathbb{Z}^3$ in Case 2 is the collection of vertices covered by $\tau_1 \cap \tau_2$. In particular, we note that it is tileable by dimers.

Conversely, if a discrete vector field g satisfies these properties (i.e. all vertices are in Case 1 or Case 2, and the set of Case 2 vertices is tileable by dimers), then there exist tilings τ_1, τ_2 such that $g = f_{(\tau_1,\tau_2)}$. In fact we can explicitly construct the tilings from g. Given any tiling τ of the Case 2 vertices $\{v \in \mathbb{Z}^3 : g(e) = 0 \text{ for all } e \text{ incident to } v\}$, we define

$$\begin{aligned} \tau_1 &= \tau \cup \{e \ : \ g(e) = 1 \text{ where } e \text{ is an edge directed from an even to an odd vertex}\} \\ \tau_2 &= \tau \cup \{e \ : \ g(e) = -1 \text{ where } e \text{ is an edge directed from an even to an odd vertex}\}. \end{aligned}$$

From this we see that that the flow $f_{(\tau_1,\tau_2)}$ determines the double dimer configuration (τ_1, τ_2) up to the choice of tiling τ on the Case 2 vertices. On other other hand, the union of tilings $\tau_1 \cup \tau_2$ determines (τ_1, τ_2) on the set where $\tau_1 = \tau_2$, meaning it determines the tiling τ of the Case 2 vertices. Therefore together these are enough to recover (τ_1, τ_2). In summary, we have shown:

Proposition 2.7.9. *The pair* $(\tau_1 \cup \tau_2, f_{(\tau_1,\tau_2)})$ *uniquely determines the double dimer configuration* (τ_1, τ_2) *and vice versa.*

Shifting along flow lines. We define a $\mathbb{Z}$-action on $\Omega \times \Omega$ by translating in the direction of the double dimer flow $f_{(\tau_1,\tau_2)}$. Given $(\tau_1, \tau_2) \in \Omega \times \Omega$, let $b_1 \in \tau_1$ be the edge incident to the origin, and suppose that $b_1 = (0, a_1)$, $a_1 \in \mathbb{Z}^3$ a neighbor of the origin. Following that, let $b_2 \in \tau_2$ be the edge incident to a_1, and suppose that $b_2 = (a_1, a_1 + a_2)$, where $a_2 \in \mathbb{Z}^3$ a neighbor of the origin. These are the first two edges of a path in (τ_1, τ_2). We define $\alpha_1(\tau_1, \tau_2)$ to be the directed vector a_1 and $\alpha_2(\tau_1, \tau_2)$ to be the directed vector a_2. When the pair of tilings (τ_1, τ_2) is understood, we drop them from the notation. We then define the function $F : \Omega \times \Omega \to [-2, 2]^3$ by

$$F((\tau_1, \tau_2)) = \alpha_1 + \alpha_2.$$

We define F to be translation by two edges so that the parity of the even/odd vertices is preserved. This can be viewed as tracking the slope and speed of the flow $f_{(\tau_1,\tau_2)}$ (when there is a double edge at the origin in (τ_1,τ_2), F is 0). Finally we define a transformation $T : \Omega\times\Omega \to \Omega\times\Omega$ given by translating the double dimer tiling by $\alpha_1+\alpha_2$. If (τ_1,τ_2) has a double edge at the origin, then $T((\tau_1,\tau_2)) = (\tau_1,\tau_2)$. Otherwise, T shifts (τ_1,τ_2) along the path through the origin. The corresponding flow $T(f_{(\tau_1,\tau_2)})(e) = f_{(\tau_1,\tau_2)}(e+\alpha_1+\alpha_2)$. If μ is a $\mathbb{Z}^3_{\text{even}}$-invariant measure on $\Omega\times\Omega$, then it is also T-invariant. Thus by the ergodic theorem we have that for μ almost every $(\tau_1,\tau_2)\in\Omega\times\Omega$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} F(T^i((\tau_1,\tau_2)) =: F^\star((\tau_1,\tau_2))$$

exists. Further, $F^\star$ is invariant under T and

$$\int_{\Omega\times\Omega} F^\star((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)) = \int_{\Omega\times\Omega} F((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)).$$

By construction, $F^\star$ measures the slope or asymptotic direction of the path γ_0 through the origin in (τ_1,τ_2). We call $F^\star$ the *slope function*. If γ_0 is a double edge or finite cycle, then $F^\star$ is 0. If γ_0 is an infinite path, then it can have nonzero slope.

For any infinite path $\ell\subset(\tau_1,\tau_2)$ we can compute its slope by translating so that ℓ goes through the origin. We say that ℓ has *nonzero slope* if $F^\star((\widetilde{\tau}_1,\widetilde{\tau}_2))\neq 0$, where $(\widetilde{\tau}_1,\widetilde{\tau}_2)$ is a translation of (τ_1,τ_2) so that ℓ contains the origin (this is well-defined since $F^\star$ is T-invariant). With this we can prove the main result of this section.

Proposition 2.7.10. *Let μ be a measure on $\Omega\times\Omega$ which is a $\mathbb{Z}^3_{even}$-invariant coupling of $\mathbb{Z}^3_{even}$-invariant measures μ_1 and μ_2 on Ω. Then*

$$\int_{\Omega\times\Omega} F^\star((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)) = s(\mu_1)-s(\mu_2).$$

Proof. Since μ is T-invariant,

$$\int_{\Omega\times\Omega} F^\star((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)) = \int_{\Omega\times\Omega} (\alpha_1(\tau_1,\tau_2)+\alpha_2(\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)).$$

Since $\alpha_1(\tau_1,\tau_2)$ is the vector along the edge adjacent to 0 in τ_1 pointing away from it, it depends only on μ_1. Hence

$$\int_{\Omega\times\Omega}\alpha_1(\tau_1,\tau_2)\mathrm{d}\mu((\tau_1,\tau_2)) = \int_{\Omega}\alpha_1(\tau_1,\tau_2)\mathrm{d}\mu_1(\tau_1) = s(\mu_1).$$

The vector $\alpha_2(\tau_1,\tau_2)$ is defined similarly, but first we have to sum over the possible values of α_1.

$$\int_{\Omega\times\Omega}\alpha_2(\tau_1,\tau_2)\,\mathrm{d}\mu((\tau_1,\tau_2)) = \sum_{a_1,a_2\in\star} a_2\,\mu((0,a_1)\in\tau_1,(a_1,a_1+a_2)\in\tau_2)$$

where $\star$ is the six neighbors of the origin. By the $\mathbb{Z}^3_{\text{even}}$-invariance of μ we get that

$$\begin{aligned}\int_{\Omega\times\Omega} \alpha_2(\tau_1,\tau_2)\,\mathrm{d}\mu((\tau_1,\tau_2)) &= \sum_{a_1,a_2\in\star} a_2\,\mu((-a_1-a_2,-a_2)\in\tau_1,(-a_2,0)\in\tau_2)\\ &= \sum_{a_2\in\star} a_2\mu_2((-a_2,0)\in\tau_2) = -s(\mu_2).\end{aligned}$$

This completes the proof. □

As a corollary, we show that the mean current difference of a pair of measures (μ_1,μ_2) can be computed by looking only at the tiles on infinite paths of nonzero slope. As a consequence, note also that if $s(\mu_1)\neq s(\mu_2)$ then an invariant coupling must have order n^3 tiles along infinite paths of nonzero slope. Here recall that $s_0(\tau)$ denotes the tile direction at the origin in τ and that for a measure μ_1 on Ω, $s(\mu_1)=\mathbb{E}_\mu[s_0(\tau)]$.

Corollary 2.7.11. *Let μ be a measure on $\Omega\times\Omega$ which is a $\mathbb{Z}^3_{even}$-invariant coupling of $\mathbb{Z}^3_{even}$-invariant measures μ_1 and μ_2 on Ω. Let I_0 be the event that the origin is contained in an infinite path of nonzero slope in (τ_1,τ_2), and let*

$$s(\mu_1,I_0)-s(\mu_2,I_0)=\mathbb{E}_\mu[(s_0(\tau_1)-s_0(\tau_2))\mathbb{1}_{I_0}((\tau_1,\tau_2))].$$

Then

$$s(\mu_1)-s(\mu_2)=s(\mu_1,I_0)-s(\mu_2,I_0).$$

Proof. Note that

$$\int_{\Omega\times\Omega} F^\star((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2)) = \int_{\Omega\times\Omega} F^\star((\tau_1,\tau_2))\mathbb{1}_{\{F^\star\neq 0\}}((\tau_1,\tau_2))\,\mathrm{d}\mu((\tau_1,\tau_2))$$

By Proposition 2.7.10, the left hand side is equal to $s(\mu_1)-s(\mu_2)$. Since $I_0=\{F^\star\neq 0\}$, the right hand side is equal to $s(\mu_1,I_0)-s(\mu_2,I_0)$. □

Finally we observe that the number of infinite paths of nonzero slope in (τ_1,τ_2) that intersect two far away boxes is 0 with probability 1 as the distance between the boxes goes to ∞. This serves as a lemma for Proposition 2.7.14.

Lemma 2.7.12. *Let μ be a $\mathbb{Z}^3_{even}$-invariant probability measure on $\Omega\times\Omega$, and fix $m\in\mathbb{N}$. Given a sample (τ_1,τ_2) from μ and $x,y\in\mathbb{Z}^3_{even}$, let $L_{x,y}$ denote the number of infinite paths of nonzero slope in (τ_1,τ_2) which intersect both $x+[1,m^3]$ and $y+[1,m^3]$. Then*

$$\lim_{n\to\infty}\frac{1}{n^3}\sum_{x\in[1,n^3]}\mu(L_{0,x}=0)=1.$$

Proof. There are at most m^3 infinite paths of nonzero slope in (τ_1,τ_2) passing through $[1,m]^3$. On the other hand, for any infinite path ℓ with nonzero slope,

$$|\ell\cap[1,n]^3|=O(n).$$

Therefore each infinite path ℓ of nonzero slope intersecting $[1,m]^3$ intersects $x+[1,m]^3$ for at most $O(n)$ points $x \in [1,n]^3$. Since m^3 is a constant, this implies that the number of $x \in [1,n]^3$ such that $L_{0,x} \neq 0$ is also $O(n)$. We can rewrite

$$\frac{1}{n^3} \sum_{x \in [1,n^3]} \mu(L_{0,x} \neq 0) = \mathbb{E}_\mu \left(\frac{1}{n^3} \sum_{x \in [1,n^3]} \mathbb{1}_{(L_{0,x} \neq 0)}(\tau_1, \tau_2) \right).$$

By the dominated convergence theorem, the right hand side tends to 0 as $n \to \infty$. This completes the proof. □

2.7.4 Chain swapping

We can now introduce the main tool of this section, namely *chain swapping*, which is an operation on double dimer configurations similar to the *cluster swapping* technique used in [She05, Chapter 8].

Let $(\tau_1, \tau_2) \in \Omega \times \Omega$ be a pair of dimer tilings. Corresponding to this are the collection of (unoriented) loops $\tau_1 \cup \tau_2$ and the double dimer flow $f_{(\tau_1,\tau_2)}$. The flow $f_{(\tau_1,\tau_2)}$ determines the orientation of each loop or infinite path in $\tau_1 \cup \tau_2$.

For a fixed $p \in (0,1)$, from a random configuration (τ_1, τ_2) we define a new random pair (τ_1', τ_2') by "shifting" the tiles along each infinite path of nonzero slope $\ell \subset (\tau_1, \tau_2)$ with independent probability p. In terms of the flow $f_{(\tau_1,\tau_2)}$, shifting on the infinite path ℓ corresponds to flipping the sign of $f_{(\tau_1,\tau_2)}$ along ℓ. The new tilings τ_1', τ_2' have the following properties:

1. $\tau_1 \cup \tau_2 = \tau_1' \cup \tau_2'$, i.e. they correspond to the same collection of double edges, finite loops, and infinite paths.

2. Let $\ell_1, \ell_2, \ldots$ be the infinite paths of nonzero slope in (τ_1, τ_2). Independently for each i, either with probability $1-p$ the tiles on ℓ were not swapped, in which case

$$\tau_1' \cap \ell_i = \tau_1 \cap \ell_i, \qquad \tau_2' \cap \ell_i = \tau_2 \cap \ell_i$$

   or with probability p the tiles were swapped, in which case

$$\tau_2' \cap \ell_i = \tau_1 \cap \ell_i, \qquad \tau_1' \cap \ell_i = \tau_2 \cap \ell_i.$$

3. On the complement of the infinite paths with nonzero slope in (τ_1, τ_2), τ_1' is equal to τ_1 and τ_2' is equal to τ_2.

We call this procedure *chain swapping with probability p*. See Figure 2.29. Chain swapping transforms a measure μ on $\Omega \times \Omega$ into a new measure μ' on $\Omega \times \Omega$ which we call the *swapped measure*.

Remark 2.7.13. Note that we only swap on the infinite paths of *nonzero slope*. This is a technical point. We do not know if the "asymptotic independence" result of Lemma 2.7.12 holds for infinite paths of zero slope. However we need Lemma 2.7.12 to show that the swapped measure is still ergodic (Proposition 2.7.14).

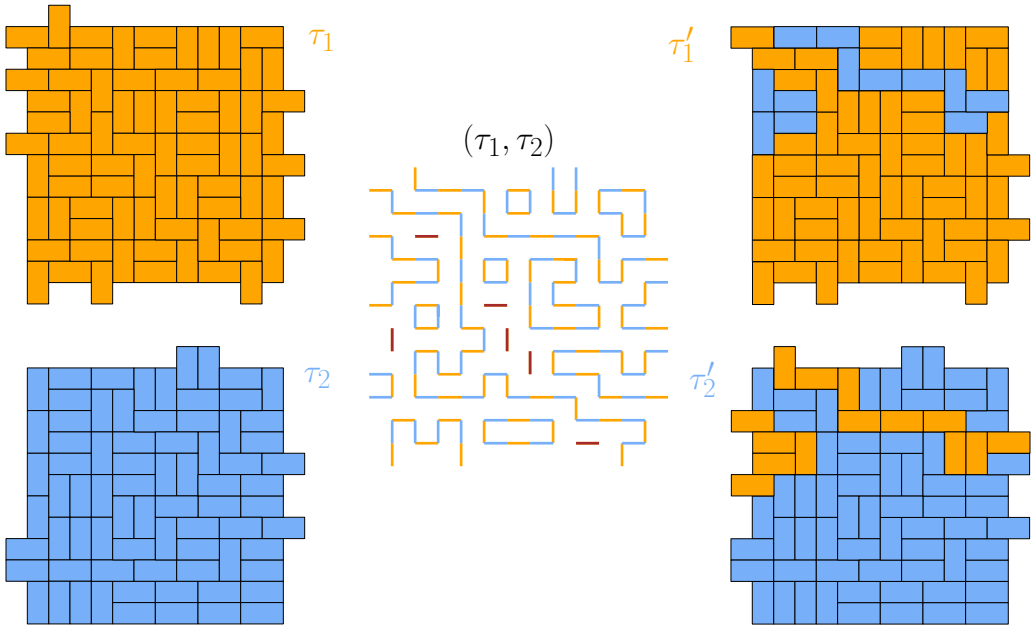


Figure 2.29: An example of tilings τ_1, τ_2, the loops in (τ_1, τ_2), and chain swapped tilings τ_1', τ_2'.

For the rest of this section, we study whether or not certain properties (ergodicity, the Gibbs property) are preserved under chain swapping, and how certain quantities (entropy, mean current) transform under chain swapping.

The first result is that chain swapping preserves ergodicity.

Proposition 2.7.14. *If μ is a ergodic measure on $\Omega \times \Omega$ with respect to the $\mathbb{Z}^3_{even}$ action and μ' is obtained from μ by chain swapping with probability p, then μ' is also ergodic.*

Proof. In this proof, we have two different parameters n (parameterizing possible translations of boxes) and m (the size of the boxes). Let $B_n = [1, n]^3$ and $\text{even}(B_n) = B_n \cap \mathbb{Z}^3_{\text{even}}$. Let $B_m + x$ denote B_m translated by $x \in \mathbb{Z}^3_{\text{even}}$. It is enough to show that for any two double dimer patterns restricted to B_m, denoted Π_1, Π_2,

$$\lim_{n\to\infty} \frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu'\Big((\tau_1', \tau_2') \mid_{B_m} = \Pi_1, (\tau_1', \tau_2') \mid_{B_m + x} = \Pi_2\Big)$$
$$= \mu'\Big((\tau_1', \tau_2') \mid_{B_m} = \Pi_1\Big)\mu'\Big((\tau_1', \tau_2') \mid_{B_m + x} = \Pi_2\Big).$$

Define the random variable L_x to be the number of infinite paths of nonzero slope in (τ_1, τ_2) sampled from μ which intersect $B_m + x$. Similarly define $L_{0,x}$ to be the number of infinite paths of nonzero slope which intersect B_m and $B_m + x$. Since the collection of tiles on infinite paths of nonzero slope is the same for μ and μ', the quantities L_0 and $L_{0,x}$ are preserved by chain swapping. Let (τ_1, τ_2) have law μ and (τ_1', τ_2') have law μ'. Finally let $\mathfrak{m}$ be the coupling of μ, μ' given by chain swapping. Then

$$\frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu'((\tau_1', \tau_2') \mid_{B_m} = \Pi_1, (\tau_1', \tau_2') \mid_{B_m + x} = \Pi_2)$$

$$= \frac{1}{\text{even}(B_n)} \sum_{\substack{k_1,k_2,k_3 \geq 0 \\ \Sigma_1,\Sigma_2 \text{ double dimer} \\ \text{tilings of } B_m}} \sum_{x \in \text{even}(B_n)} \mathfrak{m}\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, (\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2,$$
$$(\tau_1',\tau_2') \mid_{B_m} = \Pi_1, (\tau_1',\tau_2') \mid_{B_m+x} = \Pi_2,$$
$$L_0 = k_1, L_x = k_2, L_{0,x} = k_3 \Big).$$

For each infinite path of nonzero slope in (τ_1, τ_2) we have an independent probability p of reversing its direction. For any triple $l = (l_1, l_2, l_3)$ with $l_i \leq k_i$ for each i, we define the notation

$$q_{k,l} = p^{l_1+l_2-l_3}(1-p)^{k_1+k_2-k_3-l_1-l_2+l_3}.$$

This is the probability of switching (l_1, l_2, l_3) of the (k_1, k_2, k_3) paths. With this notation, for each $x \in B_n$, the x term in the sum above is equal to

$$\sum_{\substack{k_1,k_2,k_3 \geq 0 \\ \Sigma_1,\Sigma_2 \text{ double dimer tilings} \\ \text{of } B_m \text{which can swap to} \\ \Pi_1,\Pi_2 \text{with } (l_1,l_2,l_3) \text{ swaps}}} \mu\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, (\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2, L_0 = k_1, L_x = k_2, L_{0,x} = k_3 \Big) q_{k,l}. \tag{2.48}$$

For any $K > 0$,

$$\frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \Big(k_3 = K \text{ term in Equation (2.48)} \Big) \leq \frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu(L_{0,x} = K).$$

By Proposition 2.7.12, the right hand side goes to 0 as $n \to \infty$. Therefore in the limit as $n \to \infty$, it suffices to consider the terms where $k_3 = 0$ (corresponding to the set of lines hitting B_n and the set of infinite paths hitting $B_n + x$ being disjoint). Therefore

$$\frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu'\Big((\tau_1',\tau_2') \mid_{B_m} = \Pi_1, (\tau_1',\tau_2') \mid_{B_m+x} = \Pi_2 \Big)$$
$$= \sum_{\substack{k_1,k_2 \geq 0 \\ (\Sigma_1,\Sigma_2) \text{ double dimer on } B_m \text{which can swap to } \Pi_1,\Pi_2 \\ \text{with } (l_1,l_2) \text{ swaps}}}$$
$$\frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, (\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2, L_0 = k_1, L_x = k_2 \Big) r_{k,l} + o(1),$$

where $r_{k,l} = p^{l_1+l_2}(1-p)^{k_1+k_2-l_1-l_2}$ (i.e. $q_{k,l}$ when $k_3 = 0$). Since μ is ergodic, for each $\Sigma_1, \Sigma_2, k_1, k_2$,

$$\lim_{n\to\infty} \frac{1}{\text{even}(B_n)} \sum_{x \in \text{even}(B_n)} \mu\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, (\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2, L_0 = k_1, L_x = k_2 \Big) r_{k,l}$$
$$= \mu\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, L_0 = k_1 \Big) \mu\Big((\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2, L_x = k_2 \Big) r_{k,l}$$

Therefore

$$\lim_{n\to\infty} \frac{1}{\text{even}(B_n)} \sum_{x\in\text{even}(B_n)} \mu'\Big((\tau_1',\tau_2') \mid_{B_m} = \Pi_1, (\tau_1',\tau_2') \mid_{B_m+x} = \Pi_2\Big)$$

$$= \sum_{\substack{k_1,k_2\geq 0 \\ \Sigma_1,\Sigma_2 \text{ double dimer tilings of } B_m \\ \text{which can swap to } \Pi_1,\Pi_2 \\ \text{with } (l_1,l_2) \text{ swaps}}} \mu\Big((\tau_1,\tau_2) \mid_{B_m} = \Sigma_1, L_0 = k_1\Big)\mu\Big((\tau_1,\tau_2) \mid_{B_m+x} = \Sigma_2, L_x = k_2\Big) r_{k,l}$$

$$= \mu'\Big((\tau_1',\tau_2') \mid_{B_m} = \Pi_1\Big)\mu'\Big((\tau_1',\tau_2') \mid_{B_m+x} = \Pi_2\Big).$$

□

We now see how chain swapping affects the entropy and mean current of the marginal distributions.

Proposition 2.7.15. *Let μ be a measure on $\Omega\times\Omega$ which is an ergodic coupling of ergodic measures $\mu_1 \in \mathcal{P}_e^{s_1}$ and $\mu_2 \in \mathcal{P}_e^{s_2}$. If μ' is the measure obtained from μ by chain swapping with probability $p \in (0,1)$, then $h(\mu') = h(\mu)$.*

Proof. In this proof, for a stationary random field X we let $h(X)$ denote the specific entropy of the law of X. Let (τ_1,τ_2) be a sample from μ and (τ_1',τ_2') be obtained by chain swapping. By Proposition 2.7.9,

$$h(\mu') = h((\tau_1',\tau_2')) = h(\tau_1'\cup\tau_2') + h(f_{(\tau_1',\tau_2')} \mid \tau_1'\cup\tau_2').$$

Since chain swapping preserves the set of tiles, $\tau_1'\cup\tau_2' = \tau_1\cup\tau_2$, and we have automatically that $h(\tau_1'\cup\tau_2') = h(\tau_1\cup\tau_2)$. On the other hand note that

$$\begin{aligned} h(f_{(\tau_1',\tau_2')}, f_{(\tau_1,\tau_2)} \mid \tau_1'\cup\tau_2') &= h(f_{(\tau_1,\tau_2)} \mid \tau_1\cup\tau_2) + h(f_{(\tau_1',\tau_2')} \mid \tau_1\cup\tau_2,\ f_{(\tau_1,\tau_2)}) \\ &= h(f_{(\tau_1',\tau_2')} \mid \tau_1\cup\tau_2) + h(f_{(\tau_1,\tau_2)} \mid \tau_1\cup\tau_2,\ f_{(\tau_1',\tau_2')}) \end{aligned}$$

Conditioned on $\tau_1\cup\tau_2$ and $f_{(\tau_1,\tau_2)}$, the distribution of the flow $f_{(\tau_1',\tau_2')}$ is determined by independent random choices for the orientation of each infinite path of nonzero slope in $\tau_1\cup\tau_2$. Let $B_n = [1,n]^3$, and let $\ell\subset\tau_1\cup\tau_2$ be an infinite path of nonzero slope. If $\ell\cap B_n$ is nonempty, then the orientation of ℓ is determined by its direction when it intersects ∂B_n. Therefore there exists a constant $c>0$ such that

$$h(f_{(\tau_1',\tau_2')} \mid \tau_1\cup\tau_2,\ f_{(\tau_1,\tau_2)}) \leq \lim_{n\to\infty} \frac{|\partial B_n|}{|B_n|} \leq \lim_{n\to\infty} \frac{cn^2}{n^3} = 0.$$

We can analogously show that $h(f_{(\tau_1,\tau_2)} \mid \tau_1\cup\tau_2,\ f_{(\tau_1',\tau_2')}) = 0$. Therefore

$$h(f_{(\tau_1',\tau_2')} \mid \tau_1'\cup\tau_2') = h(f_{(\tau_1,\tau_2)} \mid \tau_1\cup\tau_2)$$

so

$$h(\mu') = h(\tau_1\cup\tau_2) + h(f_{(\tau_1,\tau_2)} \mid \tau_1\cup\tau_2) = h(\mu).$$

□

Proposition 2.7.16. *Let μ be a measure on $\Omega \times \Omega$ which is an ergodic coupling of ergodic measures μ_1, μ_2 with mean currents $s(\mu_1), s(\mu_2)$. If μ' is the measure obtained from μ by chain swapping with probability $p \in (0,1)$, then the marginal measures $\mu_1' = \pi_1(\mu')$ and $\mu_2' = \pi_2(\mu')$ have mean currents*

$$s(\mu_1') = (1-p)s(\mu_1) + ps(\mu_2)$$
$$s(\mu_2') = ps(\mu_1) + (1-p)s(\mu_2).$$

Proof. As in the previous section let I_0 be the event that the origin is contained on an infinite path of nonzero slope in (τ_1, τ_2). By Corollary 2.7.11,

$$s(\mu_1) - s(\mu_2) = s(\mu_1, I_0) - s(\mu_2, I_0),$$

where $s(\mu, I_0)$ is shorthand for the mean current computed as an average over only tilings where the origin is along an infinite path of nonzero slope. Since chain swapping only changes tiles that are contained on infinite paths of nonzero slope,

$$s(\mu_1') - s(\mu_1) = s(\mu_1', I_0) - s(\mu_1, I_0)$$
$$s(\mu_2') - s(\mu_2) = s(\mu_2', I_0) - s(\mu_2, I_0).$$

On the other hand, since each infinite path of nonzero slope is swapped with independent probability p,

$$s(\mu_1', I_0) = (1-p)s(\mu_1, I_0) + ps(\mu_2, I_0)$$
$$s(\mu_2', I_0) = ps(\mu_1, I_0) + (1-p)s(\mu_2, I_0).$$

Combining gives

$$s(\mu_1') - s(\mu_1) = -ps(\mu_1, I_0) + ps(\mu_2, I_0) = -ps(\mu_1) + ps(\mu_2).$$

Therefore $s(\mu_1') = (1-p)s(\mu_1) + ps(\mu_2)$. An analogous calculation gives the result for $s(\mu_2')$. □

Finally we will show that chain swapping does **not** preserve the Gibbs property. To do this, we need two technical lemmas about double dimer configurations. This result is more involved than the other chain swapping results, so for simplicity we only prove this in the $p = 1/2$ case.

Let μ be a measure on $\Omega \times \Omega$ which is an ergodic coupling of ergodic measures μ_1, μ_2 on Ω such that $s(\mu_1) \neq s(\mu_2)$. Let P be a plane with normal vector ξ such that $\langle s(\mu_1) - s(\mu_2), \xi \rangle \neq 0$. Given a sample (τ_1, τ_2) from μ, we define the random set of "last cross points" C_P by

$$C_P = \{x \in P : \text{ there is an infinite path of slope } s, \langle s, \xi \rangle \neq 0, \text{ in } (\tau_1, \tau_2) \text{ which hits } P \text{ for the } \textit{last} \text{ time at } x\}.$$

We analogously define the random set of "first cross points" A_P by

$$A_P = \{x \in P : \text{ there is an infinite path of slope } s, \langle s, \xi \rangle \neq 0, \text{ in } (\tau_1, \tau_2) \text{ which hits } P \text{ for the } \textit{first} \text{ time at } x\}.$$

Lemma 2.7.17. *With the set up above, for μ-almost every (τ_1, τ_2), both*

$$\lim_{n\to\infty} \frac{|C_P \cap [1,n]^3|}{n^2} \qquad \text{and} \qquad \lim_{n\to\infty} \frac{|A_P \cap [1,n]^3|}{n^2}$$

exist and are greater than 0.

Proof. The fact that the limits exist follows from the $\mathbb{Z}^2$ ergodic theorem applied along P.

The proofs are analogous, so we just present the proof for C_P. By Proposition 2.7.10, the μ-expected value of the slope along the component γ containing the origin in (τ_1, τ_2) is $s(\mu_1) - s(\mu_2)$. If γ is a double edge or finite cycle then the slope along γ is 0, so the set S of pairs of tilings (τ_1, τ_2) such that there is an infinite path with slope in the set $\{s : \langle s, \xi\rangle \neq 0\}$ through the origin has $\mu(S) > p$ for some $p > 0$.

Since μ is ergodic with respect to the $\mathbb{Z}^3_{\text{even}}$ action, it follows that along any $\mathbb{Z}^3_{\text{even}}$-orbit, the proportion of the orbit in S is $> p$. On the other hand, an infinite path with slope in $\{s : \langle s, \xi\rangle \neq 0\}$ only crosses P finitely many times almost surely. In particular, for any $\delta > 0$, there exists M such that

$$\mu(\ell \text{ is an infinite path passing through the origin with slope } \langle s(\ell), \xi\rangle \neq 0 \\ \text{and hits } P \text{ more than } M \text{ times}) < \delta.$$

Therefore

$$\mu\left(\lim_{n\to\infty} \frac{|C_P \cap [1,n]^3|}{n^2} > \frac{p}{M}\right) \geq 1 - \delta,$$

which completes the proof. □

The next technical lemma is about the distribution of the distance between *hit points*. Given a plane P, let $\alpha \subset (\tau_1, \tau_2)$ be an arc of a path (finite or infinite) between two points in P, such that α is disjoint from P except its endpoints $x_\alpha, y_\alpha \in P$. We define the *distance between hits* by

$$d_P(\alpha) = \text{dist}(x_\alpha, y_\alpha)$$

where dist denotes L^1 distance on P.

Lemma 2.7.18. *Let μ be an ergodic coupling of ergodic measures μ_1, μ_2 on Ω. Let $B_n = [1,n]^3$. For any $\beta > 0$, there exists M such that for all $\theta > 0$, there exists N such that if $n \geq N$, then*

$$\mu\left(\#\{\alpha \text{ arc of path hitting } P \cap B_n \ : \ d_P(\alpha) > M\} \leq \beta n^2\right) > 1 - \theta.$$

Proof. As there is some probability distribution on the distance between hit points, by $\mathbb{Z}^3_{\text{even}}$-invariance given $\epsilon > 0$ there exists M large enough such that for all $v \in P$,

$$\mu(x_\alpha = v, d_P(\alpha) > M) < \epsilon. \tag{2.49}$$

For a set of points $A \subset \mathbb{Z}^3$ let $\text{even}(A), \text{odd}(A)$ denote the subset of even, odd points respectively, and define

$$S_n^{\text{even}} = \frac{2}{n^2}\Big[\#\{\alpha \subset (\tau_1, \tau_2) \ : \ x_\alpha \in \text{even}(B_n \cap P),\, d_P(\alpha) > M\}\Big],$$
$$S_n^{\text{odd}} = \frac{2}{n^2}\Big[\#\{\alpha \subset (\tau_1, \tau_2) \ : \ x_\alpha \in \text{odd}(B_n \cap P),\, d_P(\alpha) > M\}\Big].$$

By the $\mathbb{Z}^2_{\text{even}}$ ergodic theorem applied along P, μ-almost everywhere S_n^{even} converges to a limit S^{even} as $n \to \infty$ (and similarly for S_n^{odd}). Further, we get that

$$\mu(x_\alpha = 0,\, d_P(\alpha) > M) = \int_{\Omega\times\Omega} S^{\text{even}}(\tau_1, \tau_2)\, \mathrm{d}\mu(\tau_1, \tau_2) \tag{2.50}$$

and if v is an odd point,

$$\mu(x_\alpha = v,\, d_P(\alpha) > M) = \int_{\Omega\times\Omega} S^{\text{odd}}(\tau_1, \tau_2)\, \mathrm{d}\mu(\tau_1, \tau_2). \tag{2.51}$$

Since $S_n^{\text{even}}, S^{\text{even}} \geq 0$, Equation (2.49) and Equation (2.50) (and analogously Equations (2.49), (2.51) for the odd case) combine to show that for n large enough,

$$\mu(S_n^{\text{even}} \leq 3\epsilon) \geq 1 - 2\epsilon \qquad \text{and} \qquad \mu(S_n^{\text{odd}} \leq 3\epsilon) \geq 1 - 2\epsilon.$$

Putting together the even and odd cases, for n large enough,

$$\mu\Big(\#\{\alpha \text{ arc of path hitting } P \cap B_n \ : \ d_P(\alpha) > M\} \leq 3\epsilon n^2/2\Big) \geq 1 - 2\epsilon.$$

Choosing ϵ appropriately given β, θ completes the proof. □

We can now state and prove the theorem about the effect of chain swapping (with probability $p = 1/2$) on the Gibbs property.

Theorem 2.7.19. *Let ν be a Gibbs measure on $\Omega\times\Omega$ which is an ergodic coupling of ergodic measures $\nu_1 \in \mathcal{P}_e^{s_1}$ and $\nu_2 \in \mathcal{P}_e^{s_2}$ with $s_1 \neq s_2$ and $(s_1 + s_2)/2 \in Int(\mathcal{O})$. The measure ν' obtained from ν by chain swapping with probability $p = 1/2$ is not a Gibbs measure on $\Omega\times\Omega$.*

Remark 2.7.20. The condition $(s_1 + s_2)/2 \in \text{Int}(\mathcal{O})$ is necessary in the proof so that we can use the patching theorem (Theorem 2.6.14).

Proof. Let $B_n = [1, n]^3$. Let P be a coordinate plane with normal vector denoted ξ such that $P \cap B_n$ is a face of ∂B_n (denoted F) and such that $\langle s(\nu_1) - s(\nu_2), \xi\rangle \neq 0$. By Lemma 2.7.17,

$$\nu\left(\lim_{n\to\infty} \frac{|A_P \cap B_n|}{n^2} > 0\right) = 1.$$

Recall that A_P is the collection of points $x \in P$ on infinite paths $\ell \subset (\tau_1, \tau_2)$ with $\langle s(\ell), \xi\rangle \neq 0$ such that x is the *first* time that ℓ intersects P. Given a sample (τ_1, τ_2) from ν, we look at the collection of infinite paths ℓ satisfying $\langle s(\ell), \xi\rangle \neq 0$.

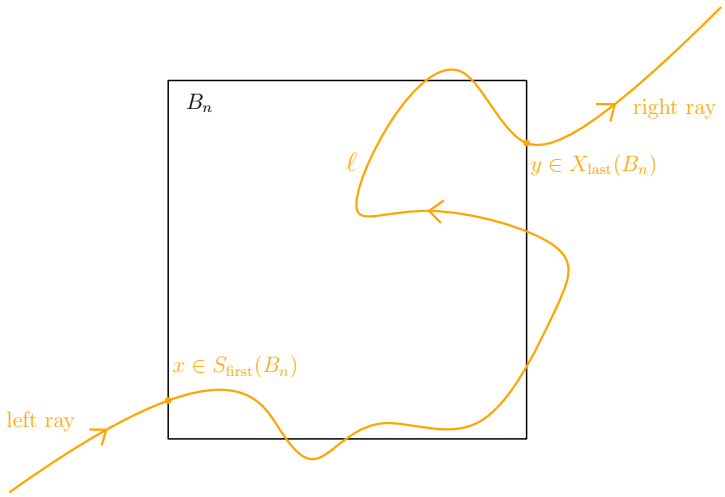


Figure 2.30: Example of an infinite path $\ell \subset (\tau_1, \tau_2)$ hitting B_n, with first entrance, last exit, and left and right rays labeled.

The part of ℓ outside B_n, $\ell \setminus B_n$, always has exactly two infinite components, a *left ray* (half-infinite path entering B_n) and a *right ray* (half-infinite path exiting B_n). We define *first entrance points* of B_n by

$$S_{\text{first}}(B_n) = \{x \in \partial B_n : \text{ there is an infinite path } \ell \subset (\tau_1, \tau_2) \text{ with } \langle s(\ell), \xi \rangle \neq 0, \\ \ell \text{ enters } B_n \text{ for the } \textit{first} \text{ time at } x\}.$$

Note that left rays hit ∂B_n at first entrance points. Similarly define *last exit points* of B_n by

$$X_{\text{last}}(B_n) = \{x \in \partial B_n : \text{ there is an infinite path } \ell \subset (\tau_1, \tau_2) \text{ with } \langle s(\ell), \xi \rangle \neq 0, \\ \ell \text{ exits } B_n \text{ for the } \textit{last} \text{ time at } x\}.$$

Right rays hit ∂B_n at last exit points. See Figure 2.30 for an illustration. We show that without loss of generality (i.e. up to translating P) there are many left rays incident to the face $F = P \cap \partial B_n$, in particular that it contains many points in $S_{\text{first}}(B_n)$. To do this, let $\tilde{B}_n$ be B_n reflected over P and notice that

$$A_P \cap B_n \subset S_{\text{first}}(B_n) \cup S_{\text{first}}(\tilde{B}_n).$$

Therefore at least one of $A_P \cap S_{\text{first}}(B_n)$ and $A_P \cap S_{\text{first}}(\tilde{B}_n)$ has size of order n^2. (It is possible for only one to have order n^2 points, for example if all paths in (τ_1, τ_2) are in the same direction.) Without loss of generality (by translating and possibly changing the choice of face F), there exists $c \in (0, 1)$ such that given $\delta > 0$, for n large enough

$$|A_P \cap S_{\text{first}}(B_n)| > cn^2 \tag{2.52}$$

with ν-probability $1 - \delta$.

Given $x \in A_P \cap S_{\text{first}}(B_n)$, there exists a unique infinite path $\ell \subset (\tau_1, \tau_2)$ with slope denoted $s(\ell)$ containing x. Since $x \in A_P$ this path will have $\langle s(\ell), \xi \rangle \neq 0$, so ℓ hits P finitely

many times almost surely. We define the function $D_P(\ell)$ to be the distance along P from $\ell \cap A_P$ to $\ell \cap C_P$ (note that this is different from $d_P(\cdot)$ defined in Lemma 2.7.18).

Without loss of generality assume that the origin is contained in P. Let ℓ_0 be the path through the origin in (τ_1, τ_2). Then for any $\theta > 0$ there exists M such that

$$\nu(D_P(\ell_0) > M \mid \langle s(\ell), \xi \rangle \neq 0) < \theta. \tag{2.53}$$

By $\mathbb{Z}^3_{\text{even}}$-invariance, this holds for any ℓ through an even point on P with $\langle s(\ell), \xi \rangle \neq 0$. An analogous statement to Equation (2.53) holds if we look at an odd point $v \in P$, and $\mathbb{Z}^3_{\text{even}}$-invariance again implies that it holds for any ℓ through an odd point on P with $\langle s(\ell), \xi \rangle \neq 0$. Putting these together, we have that for n large

$$\nu(\#\{\ell \ : \ \ell \cap C_P \notin B_n, \ell \cap A_P \in B_n\} \leq 4Mn + \theta n^2) > 1 - \theta.$$

Take $M = \epsilon n$, with $\epsilon > 0$ small to be specified below. Then for n large this becomes

$$\nu(\#\{\ell \ : \ \ell \cap C_P \notin B_n, \ell \cap A_P \in B_n\} \leq (4\epsilon + \theta)n^2) > 1 - \theta.$$

As any infinite path $\ell \subset (\tau_1, \tau_2)$ has well-defined slope, if $\langle s(\ell), \xi \rangle \neq 0$ then ℓ must be on opposite sides of P before A_P and after C_P. Hence by the above and Equation (2.52), for n large enough,

$$\nu(\#\{\ell \ : \ \ell \cap S_{\text{first}}(B_n) \in P, \ell \cap X_{\text{last}}(B_n) \notin P\} > (c - 4\epsilon - \theta)n^2) > 1 - \theta - \delta. \tag{2.54}$$

Therefore with ν-probability $1 - \theta - \delta$, at least $c'n^2 = (c - 4\epsilon - \theta)n^2$ infinite paths ℓ entering at $x \in A_P \cap B_n \subset F$ exit B_n at $y \notin F$.

On the other hand, since $s_1 \neq s_2$, we can apply chain swapping with $p = 1/2$ to get a new measure ν' distinct from ν. By Proposition 2.7.14, the marginals ν'_1, ν'_2 of ν are ergodic. By Proposition 2.7.16, they satisfy

$$s(\nu'_1) = s(\nu'_2) = \frac{s_1 + s_2}{2}.$$

Together this means that ν'_1, ν'_2 satisfy the conditions of the patching theorem (Theorem 2.6.14). Fixing $\epsilon \in (0, 1)$, let A_n be the cubic annulus between B_n and $(1 - \epsilon)B_n$. By Theorem 2.6.14 applied to ν'_1, ν'_2 on A_n, for (τ'_1, τ'_2) sampled from ν', for n large enough we can with ν'-probability $1 - \epsilon$ find a tiling τ such that

- $\tau \,|_{(1-\epsilon)B_n} = \tau'_2$
- $\tau \,|_{\mathbb{Z}^3 \setminus B_n} = \tau'_1$.

Let $Z_n \subset A_P \cap S_{\text{first}}(B_n)$ be the subset of points x such that the infinite path ℓ through x in (τ_1, τ_2) satisfies:

- ℓ has $C_P \cap \ell \in B_n$ (so that ℓ exits B_n through $\partial B_n \setminus P$);
- ℓ did not have its orientation reversed by the chain swapping (in other words, $\ell \subset (\tau'_1, \tau'_2) \cap (\tau_1, \tau_2)$.)

By Equation (2.54) and since chain swapping reverses the orientation of each infinite path with independent probability 1/2, given $\delta' > 0$, for n large enough, setting $c' = c - 4\epsilon - \theta$ we have

$$\nu'(|Z_n| > c'n^2/2) > \frac{1}{2} - \delta' > 0. \tag{2.55}$$

Conditional on the double dimer configuration (τ_1', τ_2') on $\mathbb{Z}^3 \setminus B_n$, if ν' is a Gibbs measure then it must assign the same probability to (τ_1', τ_2') and (τ, τ_2'). However since τ, τ_2' agree on $(1-\epsilon)B_n$, there are no infinite paths in (τ, τ_2') through $(1-\epsilon)B_n$.

Let $S'_{\text{first}}(B_n)$ and $X'_{\text{last}}(B_n)$ denote the first entrance and last exit points in (τ_1', τ_2'). We note that

$$S'_{\text{first}}(B_n) \cup X'_{\text{last}}(B_n) = S_{\text{first}}(B_n) \cup X_{\text{last}}(B_n)$$

because on infinite paths where the orientation was swapped, the first entrance and last exit points are swapped.

On the other hand, since (τ_1', τ_2') and (τ, τ_2') agree on $\mathbb{Z}^3 \setminus B_n$, they have the same first entrance and last exit points and the same left and right rays. If $x \in S'_{\text{first}}(B_n)$ and $y \in X'_{\text{last}}(B_n)$, we denote the left and right rays incident to them by $\ell_-(x)$ and $\ell_+(y)$ respectively. The tiling $(\tau, \tau_2') \mid_{B_n}$ pairs up all the left rays with right rays in a new way to make full infinite paths.

However recall that an infinite path in a double dimer configuration sampled from ν' has well-defined slope almost surely. We show that ν' is *not* Gibbs by showing that it is *not* possible to pair order n^2 of the left rays entering at $x \in Z_n$ with right rays of the same slope. For $x \in Z_n$, let $\gamma(x) \subset (\tau, \tau_2')$ denote the path that connects $\ell_-(x)$ to an exit point y. Then the infinite path in (τ, τ_2') through x is

$$\ell_-(x) \cup \gamma(x) \cup \ell_+(y).$$

The remainder of the proof is casework to show there only a small number of these infinite paths can have well-defined slope. Recall that $F = B_n \cap P$ and let F° denote the points in F which are distance $\geq \epsilon n$ from ∂F.

1. **Bounded by area:** We define the *thin region* T, which is a union of three things: i) $F \setminus F^\circ$, ii) $F \setminus F^\circ$ translated ϵn inward, iii) the part of ∂B_n between i) and ii). See Figure 2.31 for an illustration. Since

   $$\text{area}(T) \leq 12\epsilon n^2,$$

   the number of infinite paths in (τ, τ_2') which intersect T is bounded by $12\epsilon n^2$.

2. **Bounded by number of possible connecting paths:** choose $x \in Z_n$, and suppose that $\ell_-(x) \cup \gamma(x) \cup \ell_+(y)$ does not intersect T. To have well-defined slope, ℓ must still cross P some time after x, and to avoid T it must at some point cross P in $P \setminus F$. See Figure 2.32.

   Therefore the rest of the path $\gamma(x) \cup \ell_+(y)$ must use part of at least one finite cycle or infinite path in

   $$(\tau, \tau_2') \mid_{\mathbb{Z}^3 \setminus B_n} = (\tau_1', \tau_2') \mid_{\mathbb{Z}^3 \setminus B_n}$$

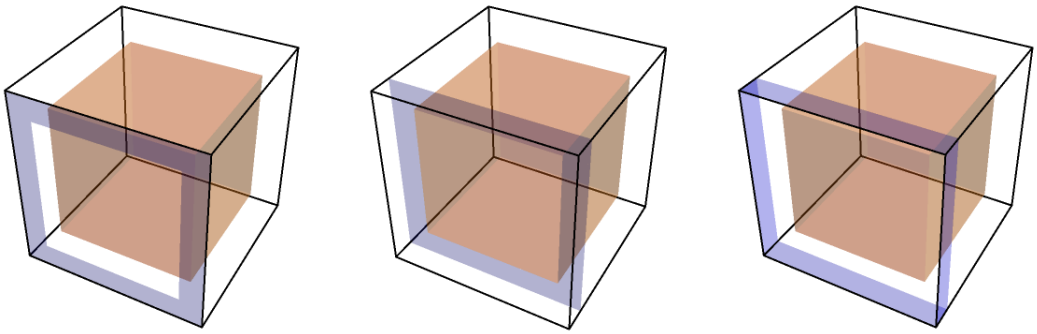


Figure 2.31: In all three pictures, the transparent cube is B_n and smaller orange cube inside it is $(1-\epsilon)B_n$. The front left face is F. In (τ, τ_2'), all tiles in $(1-\epsilon)B_n$ are double edges, so infinite paths can't enter the orange cube. The region T, corresponding to Case 1, is the union of the three blue regions.

to connect a point in F° to $P \setminus F$. This path will be an arc on P in the P-half-space on the opposite side of P from B_n. Chain swapping only changes the directions of paths, so the collection of arcs and their lengths are the same in (τ_1, τ_2) and (τ_1', τ_2'). Since the arcs are outside B_n, they are also the same in (τ, τ_2'). Thus by Lemma 2.7.18 applied to $M = \epsilon n$, we can find $\beta, \theta > 0$ small enough so that for n large enough,

$$\nu'\left(\#\{\alpha \text{ arc of path hitting } F \ : \ d_P(\alpha) > \epsilon n\} < \beta n^2\right) > 1 - \theta,$$

where recall that $d_P(\alpha)$ is the distance along P between the two intersection points of the arc α with P. Therefore with ν'-probability $1-\theta$, the number of $x \in Z_n$ such that the path

$$\ell_-(x) \cup \gamma(x) \cup \ell_+(y)$$

is disjoint from T but crosses $P \setminus F^\circ$ is at most βn^2.

3. **Remaining paths forced to have no well-defined slope:** if $x \in Z_n$ is not in Case 1 or Case 2, then the path $\ell := \ell_-(x) \cup \gamma(x) \cup \ell_+(y)$ does not intersect T and does not cross $P \setminus F^\circ$ at any time after going through x. This implies that $\ell_-(x)$ and $\ell_+(y)$ are contained in the same P half-space, in which case ℓ cannot have well-defined slope. See Figure 2.33.

In summary, with probability $1-\theta$, there are at most

$$(12\epsilon + \beta)n^2$$

points $x \in Z_n$ such that we can connect $\ell_-(x)$ to have well-defined slope. However by Equation (2.55), with positive ν-probability $|Z_n| > c'n^2/2$. We can take ϵ, β, θ to be arbitrarily small compared to c, and thus since ν' is $\mathbb{Z}^3_{\text{even}}$-invariant (meaning infinite paths must have well-defined slope a.s.), $\nu'(\cdot \mid (\tau_1', \tau_2') \mid_{\mathbb{Z}^3 \setminus B_n})$ cannot assign the same probability to (τ_1', τ_2') and (τ, τ_2'). Therefore ν' is not a Gibbs measure. □

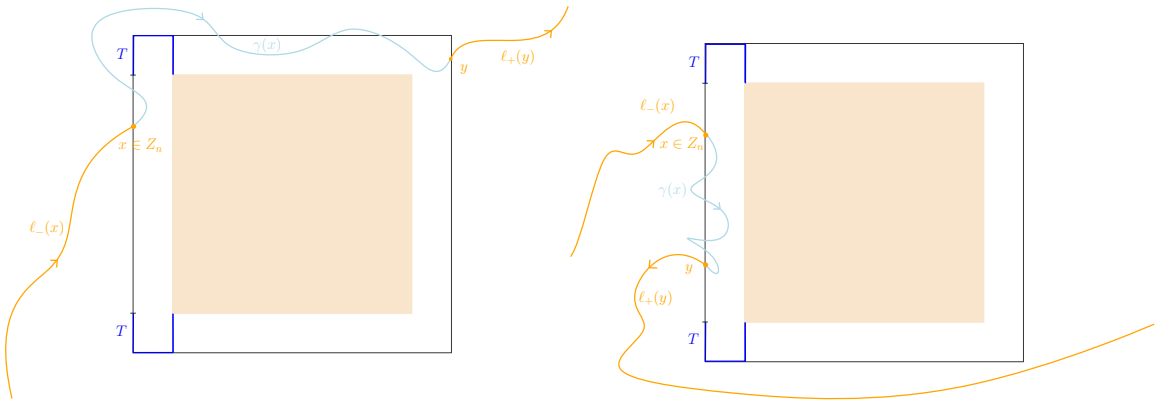


Figure 2.32: Two examples corresponding to Case 2. In this case the infinite path does not intersect T, so this can happen either if the final exit point $y \in \partial B_n \setminus F$ (left) or if the final exit point $y \in F$, but the right ray $\ell_+(y)$ crosses P again outside B_n (right).

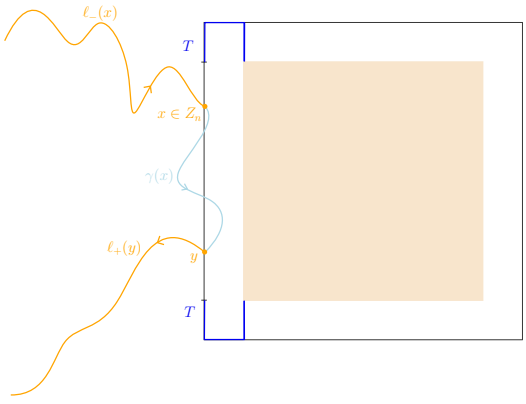


Figure 2.33: Corresponding to Case 3, if $\gamma(x) \cup \ell_+(y)$ never crosses $P \setminus F$, then the resulting infinite path cannot have well-defined slope.

Theorem 2.7.19 and Corollary 2.7.5 combine to give the following corollary.

Corollary 2.7.21. *If μ is a Gibbs measure on $\Omega\times\Omega$ which is an ergodic coupling of $\mu_1 \in \mathcal{P}_e^{s_1}$ and $\mu_2 \in \mathcal{P}_e^{s_2}$ for $s_1 \neq s_2$ and $\frac{s_1+s_2}{2} \in Int(\mathcal{O})$, then the measure μ' obtained by chain swapping with probability $p = 1/2$ does not maximize entropy in $\mathcal{P}^{\frac{s_1+s_2}{2},\frac{s_1+s_2}{2}}$.*

2.7.5 Strict concavity of ent and existence of EGMs of every mean current

With the chain swapping machinery developed in Section 2.7.4, we can now prove one of the main results of this section, namely that that ent is *strictly* concave on $\mathcal{O}\setminus\mathcal{E}$ (recall that $\mathcal{E}$ denotes the edges of $\mathcal{O}$). We already showed that ent is concave on $\mathcal{O}$ in Lemma 2.7.6. We also already showed in Section 2.4 that ent restricted to the interior of any face of $\partial\mathcal{O}$ is strictly concave (Corollary 2.4.8) by relating ent restricted to a face of $\partial\mathcal{O}$ to ent_{loz}, the slope entropy function for two-dimensional lozenge tilings (Theorem 2.4.7).

Theorem 2.7.22. *The entropy function ent is strictly concave on $\mathcal{O}\setminus\mathcal{E}$.*

Proof. By Lemmas 2.7.6 and 2.7.8, ent is concave and continuous on $\mathcal{O}$. To show strict concavity on $\mathcal{O}\setminus\mathcal{E}$, it suffices to show that if $s_1, s_2 \in \mathcal{O}$ and $(s_1+s_2)/2 \in \mathcal{O}\setminus\mathcal{E}$, then

$$\text{ent}((s_1+s_2)/2) > (\text{ent}(s_1)+\text{ent}(s_2))/2.$$

If $(s_1+s_2)/2$ is contained in the interior of a face $\mathcal{F}\subset\partial\mathcal{O}$, then we are done by Corollary 2.4.8. The remaining case is that $(s_1+s_2)/2 \in \text{Int}(\mathcal{O})$. In this case let μ_1 and μ_2 be entropy maximizers in $\mathcal{P}^{s_1}$ and $\mathcal{P}^{s_2}$ respectively (these exist by Lemma 2.7.1) and let μ be the independent coupling of μ_1 and μ_2. Then

$$\text{ent}(s_1)+\text{ent}(s_2) = h(\mu).$$

Consider the ergodic decomposition

$$\mu = \int \nu \, \mathrm{d}w_\mu(\nu)$$

where w_μ is a probability measure on the space of ergodic couplings of ergodic Gibbs measures (see Proposition 2.2.3, which says that the ergodic components of a Gibbs measure are Gibbs a.s.). Let μ' be the measure obtained by applying chain swapping with probability $p = 1/2$ to μ. By Proposition 2.7.14,

$$\mu' = \int \nu' \, \mathrm{d}w_\mu(\nu)$$

where ν' is obtained from ν by chain swapping also with $p = 1/2$. By Proposition 2.7.15, $h(\nu) = h(\nu')$. Since $h(\cdot)$ is an affine function, we get that

$$h(\mu) = \int h(\nu)\,\mathrm{d}w_\mu(\nu) = \int h(\nu')\,\mathrm{d}w_\mu(\nu) = h(\mu').$$

By Proposition 2.7.16, $s(\pi_1(\nu')) = s(\pi_2(\nu')) = (s(\nu_1) + s(\nu_2))/2$. Since $s(\cdot)$ is an affine function,

$$s(\pi_1(\mu')) = s(\pi_2(\mu')) = (s_1 + s_2)/2.$$

Let $(\mathfrak{s}_1, \mathfrak{s}_2)$ denote the random pair of mean currents for a double dimer configuration sampled from μ. To complete the proof, we proceed with cases based on w_μ. Consider the sets

$$A = \{\nu : \frac{s(\pi_1(\nu)) + s(\pi_2(\nu))}{2} \in \mathrm{Int}(\mathcal{O})\}, \qquad B = \{\nu : s(\pi_1(\nu)) \neq s(\pi_2(\nu))\}.$$

If $\nu \in A \cap B$ is an ergodic coupling of ergodic measures, then ν satisfies the conditions of Theorem 2.7.19. Since $\mathbb{E}[\mathfrak{s}_1 - \mathfrak{s}_2] = s_1 - s_2 \neq 0$, $w_\mu(B) > 0$. If $w_\mu(A) > 0$, then since μ is an independent coupling, we can argue in a few cases that $w_\mu(A \cap B) > 0$:

- If $\mathfrak{s}_1, \mathfrak{s}_2$ are both atomic, then $w_\mu(A \cap B) > 0$. For the next cases we assume without loss of generality that $\mathfrak{s}_2$ is not atomic.
- If $\{\mathfrak{s}_1 \in \mathrm{Int}(\mathcal{O})\}$ has positive probability, then given any value of $\mathfrak{s}_1$ in $\mathrm{Int}(\mathcal{O})$, $\mathfrak{s}_2$ has positive probability to be different from it. Since $\mathfrak{s}_1 \in \mathrm{Int}(\mathcal{O})$, the average is in $\mathrm{Int}(\mathcal{O})$.
- If $\mathfrak{s}_1$ has positive probability to be contained in $\partial\mathcal{O}$, then given any value of $\mathfrak{s}_1$ in $\partial\mathcal{O}$, $w_\mu(A) > 0$ implies that $\mathfrak{s}_2$ has positive probability to not be contained in the same face as $\mathfrak{s}_1$ (since on A, their average must be in $\mathrm{Int}(\mathcal{O})$). On the other hand if $\mathfrak{s}_2$ is not contained in the same face of $\partial\mathcal{O}$ as $\mathfrak{s}_1$, then it must be different from $\mathfrak{s}_1$.

Applying Theorem 2.7.19 shows that ν' is not a Gibbs measure for $\nu \in A \cap B$. Since the ergodic components of Gibbs measures are Gibbs a.s., if $w_\mu(A \cap B) > 0$ then μ' is not a Gibbs measure. By Corollary 2.7.5, μ' is not an entropy maximizer in $\mathcal{P}^{\frac{s_1+s_2}{2}, \frac{s_1+s_2}{2}}$, and hence

$$\mathrm{ent}(s_1) + \mathrm{ent}(s_2) = h(\mu) = h(\mu') < 2\mathrm{ent}((s_1 + s_2)/2).$$

This completes the proof if $w_\mu(A) > 0$.

However it can happen that $w_\mu(A) = 0$ (for example, if $\mathfrak{s}_1$ is supported at a corner vertex $v \in \partial\mathcal{O}$, and $\mathfrak{s}_2$ is supported on a square on $\partial\mathcal{O}$ around v). Since μ is an independent coupling, $w_\mu(A) = 0$ implies that $\mathfrak{s}_1, \mathfrak{s}_2$ are supported in $\partial\mathcal{O}$. There are two remaining cases.

First suppose there is a face $\mathcal{F} \subset \partial\mathcal{O}$ such that

$$C = \{\nu : \frac{s(\pi_1(\nu)) + s(\pi_2(\nu))}{2} \in \mathrm{Int}(\mathcal{F})\}$$

has $w_\mu(C) > 0$. Since μ is an independent coupling, $w_\mu(C \cap B) > 0$ by arguments analogous to those above for A, B. Let μ' be obtained from μ by chain swapping. By Proposition 2.7.14,

$$\mu' = \int \nu' \, \mathrm{d}w_\mu(\nu)$$

where ν' is obtained from ν by chain swapping. Let ν_1, ν_2 denote the marginals of ν and let ν'_1, ν'_2 denote the marginals of ν'. If $\nu \in B$, then $s(\nu_1) \neq s(\nu_2)$ and ν' is distinct from ν. By Proposition 2.7.15 and Proposition 2.7.16,

$$h(\nu') = h(\nu), \qquad s(\nu'_1) = s(\nu'_2) = \frac{s(\nu_1) + s(\nu_2)}{2}.$$

By Theorem 2.4.7, ent $|_{\mathcal{F}} =$ ent$_{\text{loz}}$. Since ent$_{\text{loz}}$ is strictly concave on Int$(\mathcal{F})$, we have that for each $\nu \in C \cap B$,

$$h(\nu') = h(\nu'_1) + h(\nu'_2) < 2\,\text{ent}((s(\nu_1) + s(\nu_2))/2)$$

Since $w_\mu(C \cap B) > 0$, Lemma 2.7.6 and the affine property of h implies that

$$\text{ent}(s_1) + \text{ent}(s_2) = h(\mu) = h(\mu') = \int h(\nu)\,\mathrm{d}w_\mu(\nu) < 2\,\text{ent}((s_1 + s_2)/2).$$

This completes the proof in the case that $w_\mu(C) > 0$.

Finally if $w_\mu(A) = 0$ and $w_\mu(C) = 0$ for all faces of $\partial\mathcal{O}$, then $\mathfrak{s}_1, \mathfrak{s}_2$ must be supported in $\mathcal{E}$ (for example, $\mathfrak{s}_1$ could be supported at one vertex $v \in \partial\mathcal{O}$, and $\mathfrak{s}_2$ could be supported on the four edges of $\partial\mathcal{O}$ incident to v). Since ent $|_{\mathcal{E}} \equiv 0$, this implies that $h(\mu_1) + h(\mu_2) = h(\mu) = 0$ and hence that $h(\mu_1) = h(\mu_2) = 0$.

However by Lemma 2.7.6 and Theorem 2.4.7, ent $|_{\mathcal{O}\setminus\mathcal{E}} > 0$. Therefore if $s_1, s_2 \in \mathcal{O} \setminus \mathcal{E}$, then μ_1, μ_2 cannot be entropy maximizers in $\mathcal{P}^{s_1}$, $\mathcal{P}^{s_2}$. This completes the proof. □

With this we can strengthen Theorem 2.7.2.

Theorem 2.7.23. *For every $s \in$ Int$(\mathcal{O})$, a measure $\mu \in \mathcal{P}^s$ satisfies $h(\mu) =$ ent(s) if and only if μ is a convex combination of ergodic Gibbs measures of mean current s. In particular, if $\nu \in \mathcal{P}^s$ is an ergodic Gibbs measure, then $h(\nu) =$ ent(s).*

Remark 2.7.24. In contrast, an EGM of mean current $s \in \partial\mathcal{O}$ can have any specific entropy between 0 and ent(s), see Proposition 2.4.10. (Note however that for $s \in \mathcal{E}$, ent$(s) = 0$.) All EGMs of mean current $s \in$ Int$(\mathcal{O})$ have the same specific entropy by Corollary 2.6.30.

Proof. Suppose $\mu \in \mathcal{P}^s$ maximizes entropy (μ exists by Lemma 2.7.1). By Theorem 2.7.2, μ is a Gibbs measure. Consider its ergodic decomposition

$$\mu = \int \nu\,\mathrm{d}w_\mu(\nu).$$

Since ergodic components of Gibbs measures are Gibbs a.s., w_μ is a probability measure on ergodic Gibbs measures. Since $h(\cdot)$ is an affine function, it follows that

$$\text{ent}(s) = h(\mu) = \int h(\nu)\,\mathrm{d}w_\mu(\nu) \leq \int \text{ent}(s(\nu))\,\mathrm{d}w_\mu(\nu).$$

Since $s \in$ Int$(\mathcal{O})$, by Theorem 2.7.22 if $s(\nu)$ is not constant then the middle inequality below is strict:

$$\text{ent}(s) \leq \int \text{ent}(s(\nu))\,\mathrm{d}w_\mu(\nu) < \text{ent}\left(\int s(\nu)\,\mathrm{d}w_\mu(\nu)\right) = \text{ent}(s).$$

Therefore all ergodic components ν of μ must have $s(\nu) = s$, i.e., the support of w_μ is contained in the set of ergodic Gibbs measures of mean current s.

By Corollary 2.6.30, if ν_1, ν_2 are EGMs of the same mean current $s \in \mathrm{Int}(\mathcal{O})$, then $h(\nu_1) = h(\nu_2)$. Therefore if $\nu \in \mathcal{P}^s$ is an ergodic Gibbs measure, $\mathrm{ent}(s) = h(\nu)$.

□

From Theorem 2.7.23 and Lemma 2.7.1 for interior mean currents and the results of Section 2.4 for boundary ones, there exist ergodic Gibbs measures of all mean currents.

Corollary 2.7.25. *For all $s \in \mathcal{O}$, there exists an ergodic Gibbs measure of mean current s.*

Remark 2.7.26. In two dimensions, there exists a *unique* ergodic Gibbs measure of every interior slope. Uniqueness of EGMs for interior mean currents is open problem, see Problem 2.9.3 and the related Problem 2.9.4.

2.7.6 Properties of Ent

Recall that $AF(R)$ denotes the space of asymptotic flows on R, and $AF(R, b)$ denotes the asymptotic flows on R with boundary value b. Both are equipped with the Wasserstein metric d_W (see Section 2.5). Here we use the properties of the mean-current entropy function ent to prove things about Ent, the entropy functional on asymptotic flows given by

$$\mathrm{Ent}(f) = \frac{1}{\mathrm{Vol}(R)} \int_R \mathrm{ent}(f(x))\,\mathrm{d}x.$$

As Corollaries of Lemma 2.7.6, Theorem 2.7.22 and Lemma 2.7.8 respectively we get that

Corollary 2.7.27. *The entropy functional Ent is concave on $AF(R)$. Further, Ent is strictly concave when restricted to the space of asymptotic flows which are valued in $\mathcal{O} \setminus \mathcal{E}$.*

Corollary 2.7.28. *If $f_n \to f$ almost everywhere in R, then $Ent(f) = \lim_{n\to\infty} Ent(f_n)$.*

From this, we show

Proposition 2.7.29. *The functional $Ent : AF(R) \to [0, \infty)$ is upper semicontinuous in the Wasserstein topology induced by d_W.*

Proof. Let $(f_n)_{n\geq 1}$ be a sequence of flows in $AF(R)$ such that $d_W(f_n, f) \to 0$ as $n \to \infty$ for some $f \in AF(R)$. For any $g \in AF(R)$, we can define its approximation g_ϵ given by

$$g_\epsilon(x) := \frac{1}{\mathrm{Vol}B_\epsilon(x)} \int_{B_\epsilon(x)} g(y)\,\mathrm{d}y.$$

Here we say that $g(y) = 0$ if $y \notin R$. While g_ϵ is not an asymptotic flow because it is not divergence-free, it is still valued in $\mathcal{O}$ and thus $\mathrm{Ent}(g_\epsilon) := \frac{1}{\mathrm{Vol}(R)} \int_R \mathrm{ent}(g_\epsilon(x))\,\mathrm{d}x$ still makes sense. By the Lebesgue differentiation theorem, g_ϵ converges to g almost everywhere as $\epsilon \to 0$. By Corollary 2.7.28,

$$\lim_{\epsilon\to 0} \mathrm{Ent}(g_\epsilon) = \mathrm{Ent}(g).$$

By Lemma 2.7.6, for any $x \in R$,

$$\operatorname{ent}(g_\epsilon(x)) = \operatorname{ent}\left(\frac{3}{4\pi\epsilon^3}\int_{B_\epsilon(x)} g(y)\,\mathrm{d}y\right) \geq \frac{3}{4\pi\epsilon^3}\int_{B_\epsilon(x)} \operatorname{ent}(g(y))\,\mathrm{d}y.$$

Therefore there is a constant C (proportional to $\mathrm{Area}(\partial R)/\mathrm{Vol}(R)$ and independent of ϵ) such that

$$\operatorname{Ent}(g_\epsilon) + C\epsilon \geq \operatorname{Ent}(g).$$

Since $d_W(f_n, f) \to 0$ as $n \to \infty$, by Corollary 2.5.6, $f_{n,\epsilon}$ converges pointwise to f_ϵ. By Corollary 2.7.28,

$$\limsup_{n\to\infty} \operatorname{Ent}(f_n) \leq \limsup_{n\to\infty} \operatorname{Ent}(f_{n,\epsilon}) + C\epsilon = \operatorname{Ent}(f_\epsilon) + C\epsilon.$$

Taking ϵ to zero, we get that

$$\limsup_{n\to\infty} \operatorname{Ent}(f_n) \leq \operatorname{Ent}(f),$$

hence Ent is upper semicontinuous. □

Remark 2.7.30. It is not difficult to see that Ent is not continuous. Indeed consider the flows $f_n \in AF([0,1]^3)$ given by

$$f_n(x_1,x_2,x_3) = \begin{cases} \eta_2 & \text{if } x_1 \in (\frac{2k}{2n}, \frac{2k+1}{2n}) \text{ for some } 0 \leq k \leq n-1 \\ -\eta_2 & \text{if } x_1 \in (\frac{2k+1}{2n}, \frac{2k+2}{2n}) \text{ for some } 0 \leq k \leq n-1. \end{cases}$$

Then f_n converges to the constant zero vector field but $\operatorname{Ent}(f_n) = 0$ while $\operatorname{Ent}(0) > 0$.

Our main goal is to show that there exists a unique Ent maximizer in $AF(R,b)$ under some mild conditions on the pair (R,b). Standard analytic arguments are enough to show existence and a weak form of uniqueness. Let $\mathfrak{e}_1, ..., \mathfrak{e}_8$ denote the eight closed edges of $\mathcal{O}$ which make up $\mathcal{E}$.

Proposition 2.7.31. *There exists $f \in AF(R,b)$ such that $Ent(f) = \sup_{g\in AF(R,b)} Ent(g)$. Further, given $f_1, f_2 \in AF(R,b)$, define*

$$A = \{x \in R : f_1(x) \neq f_2(x)\}, \qquad B = \bigcup_{i=1}^{8} \{x \in R : f_1(x), f_2(x) \in \mathfrak{e}_i\}.$$

If f_1, f_2 are both Ent maximizers, then $A \subseteq B$.

Remark 2.7.32. The problem is that ent is only strictly concave on $\mathcal{O} \setminus \mathcal{E}$, not all of $\mathcal{O}$. The same problem arises in two dimensions, and is addressed in [Gor21] and [DS10].

Proof. Since $(AF(R,b), d_W)$ is compact (Theorem 2.5.22) and Ent is upper semicontinuous (Proposition 2.7.29), the existence of the maximizer follows.

To prove weak uniqueness, recall that $\operatorname{ent}(s) = 0$ if and only if $s \in \mathcal{E}$. If f_1, f_2 are distinct maximizers then A has positive measure. If $A \cap (R \setminus B)$ has positive measure, then by strict convexity of Ent on flows valued in $\mathcal{O} \setminus \mathcal{E}$ (Corollary 2.7.27),

$$\operatorname{Ent}\left(\frac{f_1+f_2}{2}\right) > \operatorname{Ent}(f_1) + \operatorname{Ent}(f_2),$$

which would contradict the claim that f_1, f_2 are maximizers. Therefore $A \subseteq B$. □

We adapt an argument of V. Gorin in [Gor21, Proposition 7.10] to prove uniqueness under the mild condition that the pair (R, b) is *semi-flexible* as defined in Definition 2.7.34 below. We call this *semi-flexible* since it is a weaker condition than *flexible*, which will be defined at the beginning of Section 2.8.

Definition 2.7.33. Fix a boundary asymptotic flow b on R. A point $x \in R$ with boundary condition b is *frozen* if for all open sets $U \ni x$ and all entropy maximizers $f \in AF(R, b)$, there are points $y \in U$ such that $f(y) \in \mathcal{E}$. A point $x \in R$ with boundary condition b is *always frozen* if for all open sets $U \ni x$ and *all* $g \in AF(R, b)$, there are points $y \in U$ such that $f(y) \in \mathcal{E}$.

Definition 2.7.34. The pair (R, b) is *semi-flexible* if there are no always frozen points in $\mathrm{Int}(R)$. I.e., (R, b) is semi-flexible if for all $x \in \mathrm{Int}(R)$, there exists an extension $g \in AF(R, b)$ and an open set $U \ni x$ such that $g(U) \subset \mathcal{O} \setminus \mathcal{E}$. If (R, b) is not semi-flexible, we say (R, b) is *rigid*.

Remark 2.7.35. The weak uniqueness statement in Proposition 2.7.31 can be rephrased as saying that entropy maximizers are unique on the complement of the frozen points. In particular the task that remains is to show that a region (i.e. the set of frozen points) cannot both be frozen and have multiple tilings.

Theorem 2.7.36. *If (R, b) is semi-flexible, then there is a unique Ent maximizer in $AF(R, b)$.*

Remark 2.7.37. We do not know of an example of a three-dimensional region $R \subset \mathbb{R}^3$ with boundary value b such that (R, b) is rigid, so we do not know of an example in our context where the maximizer is not unique. However see Problem 2.9.7, which includes a two-dimensional, non-planar example where the maximizer is not unique.

To prove Theorem 2.7.36, we show that an equivalent definition of (R, b) semi-flexible is that b has an extension f_0 valued in $\mathcal{O} \setminus \mathcal{E}$ on $\mathrm{Int}(R)$ (Lemma 2.7.40). After that, the key step is to show that if a maximizer takes values in $\mathcal{E}$, we can perturb it by f_0 to get a flow which does not take edge values and has more entropy (Lemma 2.7.41). In particular we have the corollary that even if uniqueness fails for (R, b), it holds if b is replaced by (say) $.999b$.

Corollary 2.7.38. *Given any boundary asymptotic flow b on R and any $\delta \in (0, 1)$ there is a unique entropy maximizer in $AF(R, \delta b)$.*

Remark 2.7.39. It is also not hard to see directly that $(R, \delta b)$ is semi-flexible, and in fact flexible, see Definition 2.8.12 and Remark 2.8.14.

Lemma 2.7.40. *The pair (R, b) is semi-flexible if and only if there exists $f_0 \in AF(R, b)$ such that f_0 is valued in $\mathcal{O} \setminus \mathcal{E}$.*

Proof. The reverse implication is clear, since for any $x \in \mathrm{Int}(R)$, taking U small enough so that $U \subset \mathrm{Int}(R)$, f_0 is an extension such that $f_0(y) \notin \mathcal{E}$ for all $y \in U$.

If (R, b) is semi-flexible, for all $x \in \mathrm{Int}(R)$ there exists an open set $U_x \ni x$ and $f_x \in AF(R, b)$ such that $f_x(y) \in \mathcal{O} \setminus \mathcal{E}$ for all $y \in U_x$. If $U'_x \subset U_x$ is a smaller open set, then clearly the same property holds for U'_x.

Let $\{V_i\}_{i\in\mathbb{N}}$ be the collection of open balls centered at rational points in R with rational radii. For any pair (x, U_x) we can find V_i such that $x \in V_i$ and $V_i \subset U_x$. Therefore for each $i \in \mathbb{N}$, there exists $g_i \in AF(R, b)$ such that g_i is valued in $\mathcal{O} \setminus \mathcal{E}$ on V_i. Hence the flow

$$f_0 := \sum_{i=1}^{\infty} \frac{1}{2^i} g_i$$

is valued in $\mathcal{O} \setminus \mathcal{E}$ everywhere in $\mathrm{Int}(R)$ as desired. □

We follow the same strategy as in [Gor21, Proposition 7.10] to prove Theorem 2.7.36 using the nowhere-edge-valued extension f_0. The key step is:

Lemma 2.7.41. *Suppose that (R, b) is semi-flexible, and let $\mathcal{V} \subset \mathcal{E}$ denote the vertices of $\partial\mathcal{O}$. If $f \in AF(R, b)$ maximizes Ent, then up to a set of measure zero f does not take values in $\mathcal{E} \setminus \mathcal{V}$.*

Proof. Suppose for contradiction that f is an Ent maximizer in $AF(R, b)$ which takes values in $\mathcal{E} \setminus \mathcal{V}$ on a set A of positive measure, and that f_0 is an extension of the form guaranteed by Lemma 2.7.40. We will contradict the claim that f is a maximizer by showing that perturbing f by f_0 increases Ent.

By Theorem 2.4.7, if $s = (s_1, s_2, s_3)$ is contained in a face of $\partial\mathcal{O}$ then $\mathrm{ent}(s)$ is equal to the entropy function for two dimensional lozenge tilings, namely

$$\mathrm{ent}(s) = \mathrm{ent}_{\mathrm{loz}}(|s_1|, |s_2|, |s_3|) = \frac{1}{\pi}\Big(L(\pi|s_1|) + L(\pi|s_2|) + L(\pi|s_3|)\Big),$$

where $L(\theta) = \int_0^\theta \log(2\sin t)\,\mathrm{d}t$ ([CKP01], see Section 2.4). As in the proof in two dimensions [Gor21, Proposition 7.10], note from this formula that if $s \in \mathcal{E} \setminus \mathcal{V}$ and t is contained in a face of $\partial\mathcal{O}$ adjacent to the edge containing s, then for $\epsilon > 0$ small enough

$$\mathrm{ent}(\epsilon t + (1-\epsilon)s) > c\epsilon \log(1/\epsilon)$$

for some constant $c > 0$ depending on s, t. More generally, $t \in \mathrm{Int}(\mathcal{O})$ can be written as a weighted average of the six brickwork patterns. Simplifying a bit, this means that t can be written as a weighted average of t_1, t_2 in the faces adjacent to the edge containing s (this takes into account four brickwork patterns), and t_3 in the edge diagonally opposite the edge containing s (this takes into account the remaining two).

$$t = \alpha t_1 + \beta t_2 + \gamma t_3, \qquad \alpha + \beta + \gamma = 1.$$

By strict concavity of ent on $\mathcal{O} \setminus \mathcal{E}$ (Theorem 2.7.22),

$$\mathrm{ent}(\epsilon t + (1-\epsilon)s) > \alpha\,\mathrm{ent}(\epsilon t_1 + (1-\epsilon)s) + \beta\,\mathrm{ent}(\epsilon t_2 + (1-\epsilon)s) + \gamma\,\mathrm{ent}(\epsilon t_3 + (1-\epsilon)s). \quad (2.56)$$

We can use the two-dimensional result directly to bound the first two terms from below. For the third term, we note that $\epsilon t_3 + (1-\epsilon)s \in \mathrm{Int}(\mathcal{O})$, and for ϵ small enough this whole quantity can be written as an average of mean currents on the faces adjacent to the edge containing s. Using strict concavity of ent on $\mathcal{O} \setminus \mathcal{E}$ we can again apply the lower bound

from the two-dimensional result. In summary, for $\epsilon > 0$ small enough, there is a constant $c > 0$ depending on s, t so that

$$\text{ent}(\epsilon t + (1-\epsilon)s) > c\epsilon \log(1/\epsilon). \tag{2.57}$$

We now consider the perturbation

$$(1-\epsilon)f + \epsilon f_0 \in AF(R, b).$$

Let $M = \sup_{s \in \mathcal{O}} \text{ent}(s)$ (this is finite because ent is continuous). For all $x \in R \setminus A$, since ent is concave on all of $\mathcal{O}$ (Lemma 2.7.6) and non-negative we have

$$\text{ent}((1-\epsilon)\, f(x) + \epsilon\, f_0(x)) - \text{ent}(f(x)) \geq \epsilon\, (\text{ent}(f_0(x)) - \text{ent}(f(x)) \geq -M\epsilon.$$

Therefore

$$\int_{R \setminus A} \text{ent}((1-\epsilon)f(x) + \epsilon f_0(x))\, \mathrm{d}x - \int_{R \setminus A} \text{ent}(f(x))\, \mathrm{d}x \geq -M\epsilon \text{Vol}(R \setminus A).$$

On the other hand by Equation (2.57), for ϵ small enough there exists $A' \subset A$ of positive measure and a fixed constant $c > 0$ such that for all $x \in A'$,

$$\text{ent}((1-\epsilon)f(x) + \epsilon f_0(x)) - \text{ent}(f(x)) = \text{ent}((1-\epsilon)f(x) + \epsilon f_0(x)) > c\epsilon \log(1/\epsilon).$$

Therefore

$$\text{Ent}((1-\epsilon)f + \epsilon f_0) - \text{Ent}(f) \geq \frac{-M\epsilon \text{Vol}(R \setminus A) + \text{Vol}(A') c\epsilon \log(1/\epsilon)}{\text{Vol}(R)}.$$

For $\epsilon > 0$ small enough this implies $\text{Ent}((1-\epsilon)f + \epsilon f_0) > \text{Ent}(f)$ and contradicts the claim that f is an entropy maximizer. □

Proof of Theorem 2.7.36. Suppose that f_1, f_2 are maximizers of Ent in $AF(R, b)$. By Lemma 2.7.41, they cannot take values in $\mathcal{E} \setminus \mathcal{V}$. By Proposition 2.7.31, they can only differ on frozen points, so

$$\{x \in R : f_1(x) \neq f_2(x)\} \subseteq \{x \in R : f_1(x), f_2(x) \in \mathcal{V}\}.$$

On the other hand $\frac{1}{2}(f_1 + f_2)$ is also a maximizer. If there is a point where f_1, f_2 take different values in $\mathcal{V}$, then $\frac{1}{2}(f_1 + f_2)$ would take an edge value contradicting Lemma 2.7.41. Therefore $f_1 = f_2$. □

2.8 Large deviation principles

Here we put together the results of the previous sections to prove the main results of this paper, namely two versions of a *large deviation principle (LDP)* for fine-mesh limits of random dimer tilings of regions $R \subset \mathbb{R}^3$ with some fixed limiting boundary value b, in the topology induced by the Wasserstein metric on flows d_W introduced in Section 2.5. Section

2.1.4 also includes a discussion of our results and a brief description of what a large deviation principle is in general. For more background information, see e.g. [DZ09] or [Var16]. Here we give a slightly more detailed informal description of the main theorems and an outline of the section before getting to formal theorem statements in Sections 2.8.1 and 2.8.2. We use results here from throughout the paper, but a lot of the notation in this section was originally introduced in Section 2.5.

For the large deviation principles, we only work with the boundary flows b which are (i) *boundary asymptotic flows* meaning that b has an extension g to R which is an asymptotic flow (Definition 2.5.31) and (ii) *extendable outside* meaning there exists $\epsilon > 0$ such that b extends to a divergence-free measurable vector field valued in $\mathcal{O}$ on an ϵ neighborhood of R (Definition 2.8.1). Analogous extendability conditions are also required in the large deviation principle for dimer tilings in 2D [CKP01]; see Remark 2.8.2.

In both versions of the LDP we prove, we look at measures supported on dimer tilings of finite regions in $\frac{1}{n}\mathbb{Z}^3$ that cover R (we call these *free-boundary tilings of R at scale n*, see Definition 2.5.1). We can require that the boundary values of these flows converge as $n \to \infty$ to the fixed boundary value b with either a *soft constraint* or a *hard constraint* on the tilings.

The large deviation principle for dimer tilings in two dimensions [CKP01] uses a hard constraint. In three dimensions, new subtleties arise from the fact that ent can be nonzero on $\partial\mathcal{O}$, and the analogous hard boundary large deviation principle is not true in full generality (see the discussion in Section 2.1.4 or Example 2.8.17). Instead we prove two versions of an LDP, one with soft boundary constraint and one with hard boundary constraint that holds under an additional condition.

A *soft constraint* means that we choose a sequence of good "thresholds" $(\theta_n)_{n\geq 1}$ with $\theta_n \to 0$ as $n \to \infty$, and look at uniform measures ρ_n on free-boundary tilings of R at scale n with boundary values within θ_n of b in the Wasserstein metric $\mathbb{W}_1^{1,1}$ that we use to compare boundary values. The *soft boundary large deviation principle* (SB LDP) says that ρ_n satisfy an LDP, as long as θ_n goes to 0 slowly enough. This is stated precisely in Theorem 2.8.6.

A *hard constraint* means that we choose a sequence of fixed boundary value b_n in the discrete such that b_n converges to b in $\mathbb{W}_1^{1,1}$. We say a boundary value is a *scale n tileable* if there exists a free-boundary tiling τ of R at scale n with that boundary value. If two tilings τ_1, τ_2 have the same boundary values on ∂R, then they are tilings of the same fixed region, so fixing a sequence of scale n tileable boundary values b_n is equivalent to fixing a sequence of *regions* R_n with boundary value b_n. We define $\overline{\rho}_n$ to be uniform measure on free boundary tilings of R at scale n with boundary value b_n, or equivalently as uniform measure on tilings of the fixed region R_n. We show that the measures $(\overline{\rho}_n)_{n\geq 1}$ satisfy an LDP under two conditions: (i) the region R_n is tileable for all n (equivalently, b_n is scale n tileable boundary value) and (ii) the region and boundary value pair (R, b) is *flexible* meaning that for every $x \in \mathrm{Int}(R)$, there exists g extending b and an open set $U \ni x$ such that $\overline{g(U)} \subset \mathrm{Int}(\mathcal{O})$ or equivalently, there exists $f_0 \in AF(R, b)$ such that for every compact set $D \subset \mathrm{Int}(R)$, $\overline{f_0(D)} \subset \mathrm{Int}(\mathcal{O})$ (see Definition 2.8.12 and Lemma 2.8.13). We call this the *hard boundary large deviation principle* (HB LDP), and it is stated precisely in Theorem 2.8.15.

The condition (R, b) *flexible* is strictly stronger than (R, b) *semi-flexible*, which says that for every point $x \in \mathrm{Int}(R)$, there is an extension g and an open set $U \ni x$ such that $g(U) \subset \mathcal{O} \setminus \mathcal{E}$, or equivalently that b has an extension f_0 which is valued in $\mathcal{O} \setminus \mathcal{E}$ (see

Definition 2.7.34 and Lemma 2.7.40). Recall that if (R, b) is not semi-flexible we call it *rigid*.

If (R, b) is *semi-flexible*, then $\text{Ent}(\cdot)$ has a unique maximizer in $AF(R, b)$ (Theorem 2.7.36). Whenever this holds, as a corollary of either LDP we show that "random dimer tilings" of R with boundary values converging to b *concentrate* in the fine-mesh limit on the unique deterministic limiting flow which maximizes $\text{Ent}(\cdot)$ in $AF(R, b)$. This result holds for "random dimer tiling" defined by sampling from any sequence of measures (i.e. ρ_n or $\overline{\rho}_n$) for which an LDP holds, see Corollaries 2.8.9 and 2.8.18.

We summarize the conditions needed for each of the theorems in the following table. Note that in all cases we have the basic assumptions that $R \subset \mathbb{R}^3$ is a compact region which is the closure of a connected domain, ∂R is piecewise smooth, and b is a boundary asymptotic flow which is extendable outside.

(R, b)	SB LDP	Unique Ent maximizer in $AF(R, b)$	HB LDP
rigid	yes	not known in general	no
semi-flexible	yes	yes	no
flexible	yes	yes	yes

We remark that the "no" entries in this table are statements that are provably not true. In particular, there exists (R, b) semi-flexible for which the hard boundary LDP is false; see Example 2.8.17 or the discussion in Section 2.1.4. See Problem 2.9.7 for discussion of the "not known" entry.

In Section 2.8.1, we give the precise definitions, conditions, and statement for the soft boundary LDP (Theorem 2.8.6), and in Section 2.8.2, we do the same for the hard boundary LDP (Theorem 2.8.15), and explain why the hard boundary LDP can be false for (R, b) just semi-flexible (Example 2.8.17). In both cases, we prove *concentration* when (R, b) is semi-flexible and the LDP holds (so in hard boundary case, (R, b) must be flexible) as a corollary (Corollaries 2.8.9 and 2.8.18) and show that proving the LDP is equivalent to proving corresponding upper and lower bounds statements (Theorems 2.8.10 and 2.8.11 for the soft boundary LDP and Theorems 2.8.19 and 2.8.20 for the hard boundary LDP). The rest of the section is dedicated to proving the upper and lower bounds.

The proofs of the lower bounds are somewhat involved. In Section 2.8.3 we show that if b is extendable outside, then any $g \in AF(R, b)$ can be approximated by a piecewise-constant asymptotic flow on a region slightly larger than R (Proposition 2.8.21). This is where we use the extendable outside condition. Building on this, in Section 2.8.4 we show that any asymptotic flow can be approximated by the tiling flow of a free-boundary tiling (Theorem 2.8.24). Combined with the patching theorem (Theorem 2.6.14), this is all we need to prove the *soft boundary* lower bound (Theorem 2.8.10), so we prove this in Section 2.8.5. In Section 2.8.6 we state and prove a more powerful *generalized patching* theorem (Theorem 2.8.32) and use this to prove the hard boundary lower bound (Theorem 2.8.19).

In Section 2.8.7 we prove both upper bounds (Theorem 2.8.11 and 2.8.20). To do this, we prove the soft boundary upper bound (Theorem 2.8.11) and note that this implies the hard boundary upper bound (Theorem 2.8.20).

2.8.1 Statement and set up: soft boundary LDP

Let $R \subset \mathbb{R}^3$ be a compact region which is the closure of a connected domain, with ∂R piecewise smooth. Recall from Section 2.5 that for each n, $TF_n(R)$ is the set of all scale n free boundary tiling flows on R. The fine-mesh limits of these with respect to the Wasserstein metric on flows (Theorem 2.5.19) are the asymptotic flows $AF(R)$. The space of asymptotic flows with fixed boundary value b is denoted by $AF(R, b)$. For any compact, piecewise smooth surface $S \subset R$, $T(\cdot, S)$ denotes the *trace operator* which takes an asymptotic or tiling flow to its boundary value on S (see Sections 2.5.4, 2.5.5). Recall (Definition 2.5.31) that b is a *boundary asymptotic flow on* R if there exists $g \in AF(R)$ such that $T(g, \partial R) = b$. We restrict our attention to boundary asymptotic flows b which are also *extendable outside*.

Definition 2.8.1. A boundary asymptotic flow b on R is *extendable outside* if there exists $\epsilon > 0$ such that b extends to a divergence-free measurable vector field on an ϵ neighborhood of R.

Remark 2.8.2. The assumption that the boundary asymptotic flow is extendable outside is inherent in [CKP01]. The Lipschitz condition in [CKP01, Theorem 1.1] implies that there is extension of the flow in $\mathbb{R}^2$. Such a strong hypothesis is not necessary. However it is easy to build boundary asymptotic flows which are not extendable outside, and some of our current techniques do not work in such cases. Let $R = [-1, 1]^2 \setminus [0, 1]^2$ and consider the flow $f \in \mathrm{AF}(R)$ given by

$$f(x) = \begin{cases} (3/4, 0) & \text{if } x \in [-1, 0] \times [0, 1] \\ (0, 3/4) & \text{if } x \in [0, 1] \times [-1, 0] \\ (0, 0) & \text{if } x \in [-1, 0] \times [-1, 0]. \end{cases}$$

Any extension of such a flow close to the origin will have to be valued outside $\mathcal{O}_2$ by the divergence-free condition. We need b to be extendable outside in our arguments to construct a piecewise-constant approximation $\widetilde{g}$ of any flow $g \in AF(R, b)$, where $\widetilde{g}$ is supported on a set $\widetilde{R} \supset R$ (Proposition 2.8.21). This is an intermediate step in showing that any $g \in AF(R, b)$ can be approximated by a free-boundary tiling $\tau \in T_n(R)$ for n large enough (Theorem 2.8.24). If R is convex, then b is automatically extendable and thus we don't need to add a condition.

The version of the LDP we present in this section has *soft* boundary conditions in the discrete. The sequence of probability measures $(\rho_n)_{n \geq 1}$ which we show satisfy an LDP are uniform probability measures on tiling flows at scale n with boundary values conditioned to be in a sequence of neighborhoods around b which shrink as $n \to \infty$.

Recall that the metric on boundary values of flows is $\mathbb{W}_1^{1,1}$. To define ρ_n, we first define the following sets.

Definition 2.8.3. Let b be a boundary asymptotic flow and fix a threshold $\theta > 0$. We denote the set of scale n tiling flows on R with boundary values within θ of b by

$$TF_n(R, b, \theta) := \{f_\tau \in TF_n(R) : \mathbb{W}_1^{1,1}(T(f_\tau, \partial R), b) < \theta\}.$$

Note that if θ is too small, $TF_n(R,b,\theta)$ might be empty. However, it will follow from Theorem 2.8.24 that given a fixed θ, if n is large enough then $TF_n(R,b,\theta)$ is nonempty (Corollary 2.8.25).

We say a sequence of thresholds $(\theta_n)_{n\geq 1}$ is **admissible** if $\theta_n \to 0$ as $n \to \infty$, but sufficiently slowly so that $TF_n(R,b,\theta_n)$ is nonempty for all n. When the threshold sequence θ_n is understood, we define

$$TF(R,b) := \cup_{n\geq 1} TF_n(R,b,\theta_n).$$

We define a sequence of probability measures ρ_n using an admissible sequence of thresholds.

Definition 2.8.4. For all $n \geq 1$, ρ_n is the uniform probability measure on $TF_n(R,b,\theta_n)$. Further, we define μ_n to be the counting measure on $TF_n(R,b,\theta_n)$ and Z_n to be its partition function, so that $\rho_n = \frac{1}{Z_n}\mu_n$.

Remark 2.8.5. If Unif_n denotes the uniform probability measure on $TF_n(R)$, then ρ_n is the conditional distribution

$$\rho_n(\cdot) = \mathrm{Unif}_n(\cdot \mid D_{b,\theta_n})$$

where D_{b,θ_n} is the event that the boundary value of a flow is within θ_n of b.

Theorem 2.8.6 (Soft boundary large deviation principle)**.** *Let $R \subset \mathbb{R}^3$ be a compact region which the closure of a connected domain, with piecewise smooth boundary ∂R. Let b be a boundary asymptotic flow which is extendable outside.*

There exists a sequence of admissible thresholds $(\theta_n)_{n\geq 1}$ such that the uniform probability measures $(\rho_n)_{n\geq 1}$ on $TF_n(R,b,\theta_n)$ satisfy a large deviation principle in the topology induced by d_W with good rate function $I_b(\cdot)$ and speed $v_n = n^3\,Vol(R)$. Namely for any d_W-Borel measurable set A,

$$-\inf_{g\in A^\circ} I_b(g) \leq \liminf_{n\to\infty} v_n^{-1}\log\rho_n(A) \leq \limsup_{n\to\infty} v_n^{-1}\log\rho_n(A) \leq -\inf_{g\in\overline{A}} I_b(g) \tag{2.58}$$

Further, the rate function $I_b(g) = C_b - Ent(g)$ if g is an asymptotic flow, where $C_b = \max_{f\in AF(R,b)} Ent(f)$. If g is not an asymptotic flow then $I_b(g) = \infty$.

Remark 2.8.7. The existence of a sequence of thresholds for which the theorem holds follows from Theorem 2.8.10. The only requirement is that $(\theta_n)_{n\geq 0}$ goes to 0 sufficiently slowly.

Remark 2.8.8. The weaker, analogous theorem with free boundary values in the limit would also hold, i.e. there is a large deviation principle for the sequence of uniform measures $(\mathrm{Unif}_n)_{n\geq 1}$ on $TF_n(R)$ from Remark 2.8.5. The rate function in this case is also of the form $C - \mathrm{Ent}(\cdot)$, with $C = \max_{f\in AF(R)} \mathrm{Ent}(f)$.

Under the additional condition that the pair (R,b) is semi-flexible (see Definition 2.7.34), Ent has a unique maximizer in $AF(R,b)$ (Theorem 2.7.36). In this case, Theorem 2.8.6 implies a concentration or weak law of large numbers result for fine-mesh limits of ρ_n-random tiling flows.

Corollary 2.8.9. *Fix $\epsilon > 0$. Assume that (R, b) is semi-flexible so that Ent has a unique maximizer in $AF(R, b)$ which we denote by $f_{\max}$. Define the event*

$$A_\epsilon = \{f : d_W(f, f_{\max}) > \epsilon\}.$$

Then

$$\rho_n(A_\epsilon) \leq C^{-n^3}$$

where $C > 1$ is a constant depending only on b and R. In other words, for any $\epsilon > 0$, the probability that a tiling flow at scale n sampled from ρ_n (i.e., with boundary value conditioned to be in a shrinking interval around b) differs from the entropy maximizer by more than ϵ goes to 0 exponentially fast as $n \to \infty$ with rate n^3.

Proof. Cover $AF(R, b)$ by open neighborhoods B_g around each $g \in AF(R, b)$ so that if $g \neq f_{\max}$ then $\text{Ent}(h) < \text{Ent}(f_{\max})$ for all $h \in \overline{B}_g$, and $B_{f_{\max}}$ is the ϵ-neighborhood of $f_{\max}$. Since $AF(R, b)$ is compact, this has a finite subcover $B_1, ..., B_k$, where B_i is a neighborhood of g_i. Without loss of generality, $B_1 = B_{f_{\max}}$. By Theorem 2.8.6, for n large enough,

$$\rho_n(A_\epsilon) \leq \sum_{i=2}^{k} \rho_n(B_i) \leq \sum_{i=2}^{k} \exp(v_n(\text{Ent}(f_i) - \text{Ent}(f_{\max}))),$$

where f_i is the entropy-maximizer in $\overline{B}_i$. Since $\text{Ent}(f_i) - \text{Ent}(f_{\max}) < 0$ for all $i \neq 1$, this completes the proof. □

Recall that $\mu_n = Z_n \rho_n$ is counting measure on $TF_n(R, b, \theta_n)$. We define notation for Wasserstein open balls, namely

$$A_\delta(g) = \{h : d_W(h, g) < \delta\}.$$

By [Var16, Lemma 2.3], the large deviation principle for $(\rho_n)_{n \geq 1}$ (Theorem 2.8.6) is implied by local upper and lower bound statements (Theorem 2.8.10 and 2.8.11), plus a property called *exponential tightness*, namely that for any $\alpha < \infty$, there exists a compact set K_α such that, for any closed set C disjoint from K_α,

$$\limsup_{n \to \infty} v_n^{-1} \log \rho_n(C) \leq -\alpha. \tag{2.59}$$

By Corollary 2.5.32, $(AF(R, b), d_W)$ is compact. The space $TF(R, b)$ is countable, and by Theorems 2.5.19 and Theorem 2.5.39, the limit points of $TF(R, b)$ are contained in $AF(R, b)$. Therefore $(AF(R, b) \cup TF(R, b), d_W)$ is compact, from which exponential tightness follows. To prove the soft boundary large deviation principle (Theorem 2.8.6), it remains to prove the following upper and lower bound theorems.

Theorem 2.8.10 (Soft boundary lower bound)**.** *For any $g \in AF(R, b)$,*

$$\lim_{\delta \to 0} \liminf_{n \to \infty} v_n^{-1} \log \mu_n(A_\delta(g)) \geq \mathit{Ent}(g).$$

Theorem 2.8.11 (Soft boundary upper bound)**.** *For any $g \in AF(R, b)$,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} v_n^{-1} \log \mu_n(A_\delta(g)) \leq \mathit{Ent}(g).$$

2.8.2 Statement and set up: hard boundary LDP

This section parallels Section 2.8.1, but the LDP we prove is for measures $(\overline{\rho}_n)_{n\geq 1}$ defined with a *hard boundary constraint* in the discrete, instead of the soft constraint used to define the measures $(\rho_n)_{n\geq 1}$ in Section 2.8.1.

Again let $R \subset \mathbb{R}^3$ be a compact region which is the closure of a connected domain with ∂R piecewise smooth, and assume that b is a boundary asymptotic flow which is extendable outside. Unlike the soft boundary LDP, we add the condition that the pair (R, b) is *flexible*.

Definition 2.8.12. A pair (R, b) is *flexible* if for all $x \in \mathrm{Int}(R)$, there exists $g \in AF(R, b)$ and an open set $U \ni x$ such that $\overline{g(U)} \subset \mathrm{Int}(\mathcal{O})$.

By completely analogous arguments to the proof of Lemma 2.7.40, we have the following equivalent definition of (R, b) flexible.

Lemma 2.8.13. *A pair (R, b) is flexible if and only if there exists $f_0 \in AF(R, b)$ such that for every compact set $D \subset Int(R)$, $\overline{f_0(D)} \subset Int(\mathcal{O})$.*

Remark 2.8.14. It is not hard to see directly that the flexible definition given in Definition 2.8.12 is satisfied for the 3D regions in the introduction built out of aztec diamonds. On each 2D aztec diamond $R_a = R \cap \{z = a\}$ and each point $x \in R_a$, consider a rectangle inscribed in R_a containing x and with edges parallel to the coordinate axes. Then the flow which is 0 inside the rectangle and linear parallel to the adjacent edge of the rectangle in the four triangles is a 2D asymptotic flow with the right boundary conditions. Averaging these flows for different rectangles gives a flow valued in $\mathrm{Int}(\mathcal{O})$ on R_a, and combining them gives a flow f valued in $\mathrm{Int}(\mathcal{O})$ everywhere in $\mathrm{Int}(R)$ (in fact, f will be valued in the middle slice of $\mathrm{Int}(\mathcal{O})$ where the third coordinate is zero).

A boundary value b_n on ∂R is a *scale n tileable* if there exist a scale n free boundary tiling τ of R such that $T(f_\tau, \partial R) = b_n$. A region $R_n \subset \frac{1}{n}\mathbb{Z}^3$ is a *scale n region with boundary value b_n* if all tilings of R_n have boundary value b_n on ∂R. When b_n is a tileable boundary value, R_n is tileable. Note that implicit in this definition is that all tilings of R_n are scale n free-boundary tilings of R.

Given a sequence of regions R_n with scale n tileable boundary values b_n on ∂R, we define $\overline{\rho}_n$ to be uniform measure on tilings of R_n (equivalently, uniform measure on free-boundary tilings of R with boundary value exactly b_n). If (R, b) is flexible and the boundary values b_n converge to b in $\mathbb{W}_1^{1,1}$ as $n \to \infty$, we prove that $(\overline{\rho}_n)_{n\geq 1}$ satisfy an LDP.

Theorem 2.8.15 (Hard boundary large deviation principle)**.** *Let $R \subset \mathbb{R}^3$ be a compact region which the closure of a connected domain, with piecewise smooth boundary ∂R. Let b be a boundary asymptotic flow which is extendable outside, and assume that (R, b) is flexible.*

Let $R_n \subset \frac{1}{n}\mathbb{Z}^3$ be a sequence of scale n regions with tileable boundary values b_n converging to b in $\mathbb{W}_1^{1,1}$. Define $\overline{\rho}_n$ to be the uniform probability measure on tilings of R_n.

The measures $(\overline{\rho}_n)_{n\geq 1}$ satisfy a large deviation principle in the topology induced by d_W with good rate function $I_b(\cdot)$ and speed $v_n = n^3\,Vol(R)$. Namely for any d_W-Borel measurable set A,

$$-\inf_{g\in A^\circ} I_b(g) \leq \liminf_{n\to\infty} v_n^{-1} \log \overline{\rho}_n(A) \leq \limsup_{n\to\infty} v_n^{-1} \log \overline{\rho}_n(A) \leq -\inf_{g\in \overline{A}} I_b(g) \tag{2.60}$$

Further, the rate function $I_b(g) = C_b - Ent(g)$ *if* g *is an asymptotic flow, where* $C_b = \max_{f \in AF(R,b)} Ent(f)$. *If* g *is not an asymptotic flow then* $I_b(g) = \infty$.

Remark 2.8.16. The large deviation principle in [CKP01] has hard boundary conditions, where the regions R_n approximate R from *within*, i.e. $R_n \subset R$. We instead assume that $R_n \supset R$, and that our regions approximate R from outside.

The *flexible* condition on (R, b) is needed for the *generalized patching theorem* (Theorem 2.8.32). We do not know the exact condition on (R, b) needed for Theorem 2.8.15 to hold, however there do exist regions (R, b) which are just semi-flexible but for which the hard boundary LDP fails.

Example 2.8.17. As discussed in the introduction, there exists semi-flexible region and boundary condition pairs (R, b) with $R \subset \mathbb{R}^3$ for which the hard boundary large deviation principle is false. The region R is a "tilted tube," and the boundary value b takes values in a face of $\partial\mathcal{O}$. This construction is related to the measures with boundary mean current discussed in Section 2.4. Recall the definition of the *slabs*

$$L_c = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 2c \text{ or } 2c + 1\}.$$

Each slab is the union of two planes. Any tiling sampled from a measure with mean current $(s_1, s_2, s_3) \in \partial\mathcal{O}$ with $s_1, s_2, s_3 \geq 0$ breaks into a sequence of complete dimer tilings of the slabs (Proposition 2.4.1).

Let $B_n = [-n, n]^3$. Let $A_n(0) = L_0 \cap B_n$, let $A_n(c) = L_c \cap [B_n + (0, 0, 2c)]$, and finally let $A_{n,n} = \cup_{i=-n}^{n} A_n(i)$. The region R is then defined so that $\frac{1}{n} A_{n,n}$ is *a* sequence of discrete regions approximating it. We choose the boundary value b to be a constant mean current $s = (s_1, s_2, s_3) \in \partial\mathcal{O}$ with $s_1, s_2, s_3 > 0$. Note that the constant asymptotic flow $g(x) = s \in AF(R, b)$, and by Theorem 2.4.7 has $\text{Ent}(g) = \text{ent}_{\text{loz}}(s) > 0$. Here are two options we could choose for the sequence of discrete regions:

- We define one sequence of regions R_n^1 where for each c such that L_c intersects $A_{n,n}$, $R_n^1 \cap L_c = S_c$ is a region such that a lozenge tiling of S_c has slope s along ∂S_c.
- We define another sequence of regions R_n^2 by alternating between frozen brickwork lozenge tilings. Choose a sequence of ratios (s_1^n, s_2^n, s_3^n) converging to (s_1, s_2, s_3) as $n \to \infty$. We partition the group of indices c such that L_c intersects $A_{n,n}$, into three groups with sizes proportional to s_1^n, s_2^n, s_3^n. For $i = 1, 2, 3$, for each c in the i^{th} group, we define $R_n^2 \cap L_c$ to be the region tileable by the η_i lozenge brickwork tiling.

By results for 2D lozenge tilings and the relationships established in Section 2.4, the hard boundary LDP would hold for the sequence R_n^1. On the other hand, R_n^2 is frozen for all n, so the number of free boundary tilings of R_n^2 is 1 for all n. While this unique tiling does approximate the constant flow $g(x) = s$, the corresponding lower bound for the LDP does not hold.

More generally, one might conjecture that the hard boundary LDP fails when there exist regions in $\text{Int}(R)$ where the Ent-maximizing flow is valued in the faces of $\partial\mathcal{O}$. However we do not know if this is a necessary or sufficient condition for the hard boundary LDP to fail, or if there exists a region $R \subset \mathbb{R}^3$ where the Ent-maximizing flow takes face values in the interior of R but not all of R. See Problem 2.9.8.

The analogous concentration or weak law of large numbers result that held for the soft boundary measures ρ_n (Corollary 2.8.9) also holds for the hard boundary measures $\overline{\rho}_n$. Note that flexible implies semi-flexible, so the maximizer $f_{\max} \in AF(R,b)$ is unique by Theorem 2.7.36.

Corollary 2.8.18. *Assume that (R,b) is flexible and $\overline{\rho}_n$ are as in Theorem 2.8.15. Let $f_{\max}$ denote the unique maximizer of Ent in $AF(R,b)$. Define the event*

$$A_\epsilon = \{f : d_W(f, f_{\max}) > \epsilon\}.$$

Then

$$\overline{\rho}_n(A_\epsilon) \leq C^{-n^3}$$

where $C > 1$ is a constant depending only on R and b. In other words, for any $\epsilon > 0$, the probability that a tiling flow at scale n sampled from $\overline{\rho}_n$ (i.e., a tiling of R_n) differs from the entropy maximizer by more than ϵ goes to 0 exponentially fast as $n \to \infty$ with rate n^3.

Proof. Analogous to the proof of Corollary 2.8.9 with ρ_n replaced by $\overline{\rho}_n$ and Theorem 2.8.6 replaced by Theorem 2.8.15. □

Like Theorem 2.8.6, by [Var16, Lemma 2.3], to prove Theorem 2.8.15 it suffices to show that the measures $(\overline{\rho}_n)_{n\geq 1}$ satisfy local upper and lower bound statements (Theorems 2.8.19 and 2.8.20) plus the *exponential tightness property* stated for ρ_n in Equation (2.59), which follows by analogous straightforward arguments for $\overline{\rho}_n$.

We let $\overline{Z}_n$ denote the partition function of $\overline{\rho}_n$ and $\overline{\mu}_n = \overline{Z}_n \overline{\rho}_n$ the corresponding counting measures.

Theorem 2.8.19 (Hard boundary lower bound)**.** *For any $g \in AF(R,b)$,*

$$\lim_{\delta \to 0} \liminf_{n \to \infty} v_n^{-1} \log \overline{\mu}_n(A_\delta(g)) \geq Ent(g).$$

Theorem 2.8.20 (Hard boundary upper bound)**.** *For any $g \in AF(R,b)$,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} v_n^{-1} \log \overline{\mu}_n(A_\delta(g)) \leq Ent(g).$$

2.8.3 Piecewise constant approximation

The goal of this section is to show that if b is extendable outside, then any $g \in AF(R,b)$ is well-approximated in the Wasserstein metric on flows by an asymptotic flow $\widetilde{g}$ which is piecewise-constant, taking constant values on a mesh $\mathcal{X}$ of small tetrahedra covering R (see Remark 2.8.23 for why tetrahedra).

Proposition 2.8.21. *Fix $\epsilon > 0$, and suppose that b is a boundary asymptotic flow which is extendable outside. For any $g \in AF(R,b)$, there exists $\delta > 0$ and a δ-mesh of tetrahedra $\mathcal{X}$ covering R with the following properties. Let $\widetilde{R} = \cup_{X \in \mathcal{X}} X$. There exists a flow $\widetilde{g}$ satisfying:*

- *$\widetilde{g} \in AF(\widetilde{R})$;*

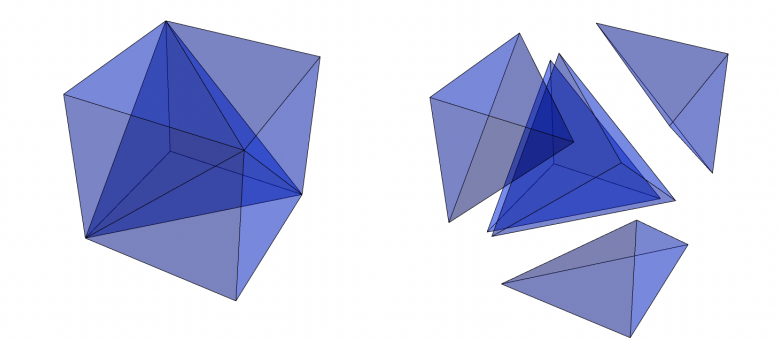


Figure 2.34: A cube cut into one regular tetrahedron and four right-angled tetrahedra. The second picture shows the same tetrahedra moved apart.

- $d_W(g, \widetilde{g}) < \epsilon$;
- *For each* $X \in \mathcal{X}$, $\widetilde{g}\,|_X = \widetilde{g}_X$ *is constant;*
- $\widetilde{g}$ *is valued strictly in* $Int(\mathcal{O})$;
- $\widetilde{g}$ *takes only rational values.*

Remark 2.8.22. We need the condition that b is extendable outside so that we can take $\widetilde{R}$ to contain R. If b is not extendable outside, a similar construction works, but the resulting piecewise-constant flow will be an asymptotic flow on a region R' slightly smaller than R instead.

Remark 2.8.23. **Tetrahedral mesh.** The fact that the mesh in this construction is built out of tetrahedra is necessary to ensure that $\widetilde{g}$ is divergence-free (needed for $\widetilde{g}$ to be an asymptotic flow). This is because a divergence-free flow on a polyhedron with F faces is determined by its flow through $F-1$ of them. Since we have 3 free parameters to specify $\widetilde{g}$ on one polyhedron, we need $F - 1 \leq 3$. The only polyhedra that satisfy this are tetrahedra.

However, regular tetrahedra alone do not tile 3-space[1], so we cannot take all elements of the mesh to be identical. Instead, 3-space can be tiled by regular tetrahedra and right-angled tetrahedra. To see this, note that cubes tile 3-space, and a cube can be cut into four right-angled tetrahedra and one regular tetrahedron (see Figure 2.34). The faces of the regular tetrahedron have normal vectors of the form $(\pm 1, \pm 1, \pm 1)$, while the right-angled tetrahedra have four coordinate plane faces and one face with normal vector $(\pm 1, \pm 1, \pm 1)$. For technical reasons (see the proof of Theorem 2.8.24, in particular Lemma 2.8.29), our arguments are simplified by assuming that the faces of the tetrahedra are always contained in one of these two types of planes. In the proof of Proposition 2.8.21 we will also use this to say that the possible normal vectors to the tetrahedra can be assumed to form a finite set. We assume throughout that our tetrahedral mesh is built out of regular and right-angled tetrahedra.

[1]Over 2,000 years ago, Aristotle (mistakenly) claimed in *De Caelo*, Book III Part 8 [AriBC] that regular tetrahedra do tile 3-space. It took around 1,000 years for the mistake to be fixed, see [LZ12] for a detailed account of the story.

Proof of Proposition 2.8.21. Since b is extendable outside, there exist $\alpha_0 > 0$ such that g can be extended to $g' \in AF(R^{\alpha_0})$, where R^{α_0} is

$$R^{\alpha_0} = \{x \in \mathbb{R}^3 : d(x, R) \leq \alpha_0\}.$$

Given this, for any $0 < \alpha < \alpha_0$, we can approximate g by a *continuous* asymptotic flow $g_\alpha \in AF(R^{\alpha_0-\alpha})$:

$$g_\alpha(x) := \frac{1}{|B_\alpha(x)|} \int_{B_\alpha(x)} g'(y)\, \mathrm{d}y, \qquad x \in R^{\alpha_0-\alpha}.$$

As $\alpha \to 0$, $d_W(g, g_\alpha \mid_R) \to 0$. We construct a piecewise-constant, divergence-free approximation u of g_α, then modify it to construct $\widetilde{g}$ also satisfying the last two conditions.

For any fixed $\delta < \alpha_0 - \alpha$ (to be specified more precisely later), we take a δ-tetrahedral mesh $\mathcal{X}$ built from regular and right-angled tetrahedra (see Remark 2.8.23) such that $X \cap R \neq \emptyset$ for all $X \in \mathcal{X}$. Let $\widetilde{R} = \cup_{X \in \mathcal{X}} X$ and note that $R \subset \widetilde{R} \subset R^{\alpha_0-\alpha}$.

Consider one tetrahedron $X \in \mathcal{X}$. Let $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ denote the faces of X and let n_1, n_2, n_3, n_4 denote their outward pointing normal vectors. Define a vector u_X by

$$u_X \cdot n_i = \frac{1}{\mathrm{area}(\zeta_i)} \int_{\zeta_i} \langle g_\alpha, n_i \rangle \, \mathrm{d}A \qquad i = 1, 2, 3.$$

Since g_α is divergence-free on X,

$$u_X \cdot n_4 = \frac{1}{\mathrm{area}(\zeta_4)} \int_{\zeta_4} \langle g_\alpha, n_4 \rangle \, \mathrm{d}A.$$

Define $u(x) := u_X$ for $x \in X$.

It remains to show that (up to multiplying by a constant $\lambda \leq 1$ but very close to 1) $\lambda u \in AF(\widetilde{R})$ and bound $d_W(\lambda u, g_\alpha \mid_{\widetilde{R}})$.

Since g_α is continuous and $R^{\alpha_0-\alpha}$ is compact, g_α is uniformly continuous on $R^{\alpha_0-\alpha}$. Thus given any $\beta > 0$ there exists $\theta > 0$ such that $|x - y| < \theta$ implies $|g_\alpha(x) - g_\alpha(y)| < \beta$.

Fixing β, we now require that $\delta < \theta$ so that uniform continuity implies that for any $X \in \mathcal{X}$ and point $x \in X$, we have that $|g_\alpha(x) - \mathrm{avg}_X g_\alpha| < \beta$. The normal vectors n_1, n_2, n_3 to three faces of X are linearly independent but not necessarily orthogonal. However since all $X \in \mathcal{X}$ are of one of five forms (see Figure 2.34), there is a constant $K > 0$ independent of $X \in \mathcal{X}$ so that

$$|u_X - \mathrm{avg}_X g_\alpha| \leq K \sum_{i=1}^{3} |\mathrm{avg}_{\zeta_i}(g_\alpha \cdot n_i) - \mathrm{avg}_X(g_\alpha \cdot n_i)| < 3K\beta.$$

Therefore for all $x \in \widetilde{R}$,

$$|u(x) - g_\alpha(x)| < (3K + 1)\beta. \tag{2.61}$$

Replacing u by λu with $\lambda = 1 - (3K + 1)\beta - \beta$, the new flow $\lambda u \in AF(\widetilde{R})$ and in fact is valued in $\mathrm{Int}(\mathcal{O})$. Further by Proposition 2.5.18, there is another constant $C > 0$ such that

$$d_W(\lambda u, g_\alpha \mid_{\widetilde{R}}) < C(6K + 3)\beta. \tag{2.62}$$

By the triangle inequality,

$$d_W(\lambda u, g) \leq d_W(\lambda u, g_\alpha \mid_{\widetilde{R}}) + d_W(g_\alpha \mid_{\widetilde{R}\setminus R}, 0) + d_W(g, g_\alpha \mid_R).$$

The first term is controlled by Equation (2.62). The second is bounded by a fixed constant times δ, and the third is bounded by a fixed constant times α. Therefore taking $\alpha, \beta > 0$ small enough and correspondingly taking $\delta < \max\{\alpha_0 - \alpha, \theta\}$, $d_W(\lambda u, g)$ can be made arbitrarily small.

The flow $\lambda u \in AF(\widetilde{R})$ and is valued in $\mathrm{Int}(\mathcal{O})$. Finally we modify λu as follows to construct $\widetilde{g}$ which also takes rational values. To do this, we solve the linear constraint problem to make the values of the flow rational without breaking the divergence-free condition.

Let M be the number of tetrahedra in the mesh $\mathcal{X}$. Enumerate the faces of the tetrahedra by $a_1, ..., a_m$. Choose a unit normal vector n_i for each face. For any flow f, let $F(f) = (F_1(f), ..., F_m(f))$, where $F_i(f) = \int_{a_i} \langle f, n_i \rangle \, \mathrm{d}x$. Note that if v is a piecewise-constant flow on the mesh, then $F(v)$ determines v.

If $F(v)$ corresponds to a divergence-free piecewise-constant flow v, then it satisfies a matrix A of M linear constraints of the form

$$\pm F_{k_1}(v) \pm F_{k_2}(v) \pm F_{k_3}(v) \pm F_{k_4}(v) = 0$$

for a_{k_j}, $j \in \{1, 2, 3, 4\}$ the faces of a tetrahedron $X \in \mathcal{X}$ (the signs are determined by the normal vector orientation, the four terms should all be for flow oriented out of X). Thus $F(v)$ solves $A\,F(v) = 0$. Since A has integer entries, there is a rational basis for the space of solutions Y of $A\,Y = 0$. Any other solution can be written as a linear combination of the rational ones, so rational solutions are dense.

Thus we can find $\widetilde{g}$ such that $\widetilde{g}_X$ takes all rational values and $|\widetilde{g}_X - (1-\delta_1)u_X|$ is as small as needed. Applying Proposition 2.5.18 again completes the proof. □

2.8.4 Existence of tiling approximations

Building on the approximation result in the previous section, we now show that if b is extendable outside then any $g \in AF(R, b)$ can be approximated in Wasserstein distance by a tiling flow. More precisely:

Theorem 2.8.24. *Fix $\delta > 0$ and suppose b is a boundary asymptotic flow which is extendable outside. For any $g \in AF(R, b)$, there exists $n(\delta)$ such that if $n \geq n(\delta)$, then there is a free boundary tiling $\tau \in T_n(R)$ such that $f_\tau \in A_\delta(g)$.*

The two dimensional analog of this theorem (i.e. [CKP01, Prop. 3.2]) is the statement that any asymptotic height function can be approximated by the height function of a tiling. In particular, one can choose the maximal height function (analog of f_τ) less than the given asymptotic height function (analog of g). There is no analogous notion of "maximal" tiling flow, so our argument in three dimensions is more complicated, and relies on an explicit construction.

We call the explicit construction in the proof of Theorem 2.8.24 the "shinning light construction." The first step is to build piecewise-linear "channels." We give a method for

tiling the channels and show that we can glue them together to construct a tiling of the whole region. The channels are tubular neighborhoods of the flow lines of a tiling flow approximating a piecewise-constant flow as constructed in Proposition 2.8.21. We call it the "shining light construction" because we imagine the flow as beams of light bending through the channels.

Before proving Theorem 2.8.24, we note that the existence of an admissible sequences of thresholds $(\theta_n)_{n\geq 1}$ follows as a straightforward corollary.

Corollary 2.8.25. *For any boundary asymptotic flow b which is extendable outside and any threshold $\theta > 0$, $TF_n(R, b, \theta)$ is nonempty for n large enough. In particular, admissible sequences of thresholds $(\theta_n)_{n\geq 1}$ exist for any boundary asymptotic flow b which is extendable outside.*

Proof. Recall that $T(\cdot, \partial R) : AF(R) \to \mathcal{M}^s(R)$ is the boundary value operator and choose $g \in T^{-1}(b)$. By Theorem 2.8.24, for any $n \geq n(\delta)$ there exists a tiling $\tau \in T_n(R)$ with $d_W(f_\tau, g) < \delta$. Since T is uniformly continuous (Theorem 2.5.39), we can choose $\delta > 0$ so that $d_W(g, f_\tau) < \delta$ implies $d_W(b, T(f_\tau)) < \theta$. □

We now proceed to the explicit construction. Recall that η_i is the i^{th} positively-oriented unit coordinate vector and that e_i denotes the edge in $\mathbb{Z}^3$ connecting the origin to η_i. Similarly, $-e_i$ is the edge connecting the origin to $-\eta_i$.

Let τ_1 denote the brickwork tiling where all tiles are $-\eta_1$ bricks. To prove Theorem 2.8.24, we show that we can construct a tiling τ so that the flow corresponding to the double dimer tiling (τ, τ_1) is close to the flow $g + \eta_1$. A double dimer tiling consists of a collection of oriented infinite paths, finite loops, and double edges. See Section 2.2.1 and Section 2.7.3.

Since τ_1 consists of only $-\eta_1$ tiles, for any other tiling τ, (τ, τ_1) consists of only infinite paths and double edges (i.e. no finite loops). The double dimer flow $f_{(\tau,\tau_1)} = f_\tau - f_{\tau_1}$ is 0 whenever the tilings agree, and otherwise points in the direction of the oriented infinite path.

For $x \in \mathbb{Z}^3$, let $\tau(x)$ denote the tile at x in τ. We say that a tiling τ of $\mathbb{Z}^3$ is *periodic* if there exist even integers $r_1, r_2, r_3 > 0$ such that $\tau(x)$ is equal to its translates $\tau(x + r_1\eta_1) = \tau(x + r_2\eta_2) = \tau(x + r_3\eta_3)$ for all $x \in \mathbb{Z}^3$. For periodic tilings, we can define a notion of the *mean current of a tiling*, denoted $s(\tau)$, as the average direction of the tiles in any $r_1 \times r_2 \times r_3$ box.

We give a method for constructing a periodic tiling τ_v of $\mathbb{Z}^3$ of a fixed, rational mean current $v \in \mathcal{O}$. This construction will serve as a building block in the proof of Theorem 2.8.24.

Construction of tiling τ_v.

First we give a construction for $v = (v_1, v_2, v_3) \in \partial\mathcal{O} \cap \mathbb{Q}^3$ with $v_1 \geq 0$, then we adapt this to the general case.

Here we view dimer tiles a in a tiling as vectors directed from even to odd. When we subtract a tiling, we reverse the direction of its dimers. Since v is rational and has nonzero norm, we can find a sequence of tiles $a_1, ..., a_r \in \{\eta_1, \text{sign}(v_2)\eta_2, \text{sign}(v_3)\eta_3\}$ such that $a_1 + ... + a_r + r\eta_1$ is parallel to $v + \eta_1$.

Below by a *plane with normal vector (1,1,1)*, we mean a collection of cubes in $\mathbb{Z}^3$ with coordinates $\{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = c\}$ for some constant $c \in \mathbb{Z}$. We analogously define a plane with normal vector $(\pm 1, \pm 1, \pm 1)$ to be the modification of this with appropriate signs.

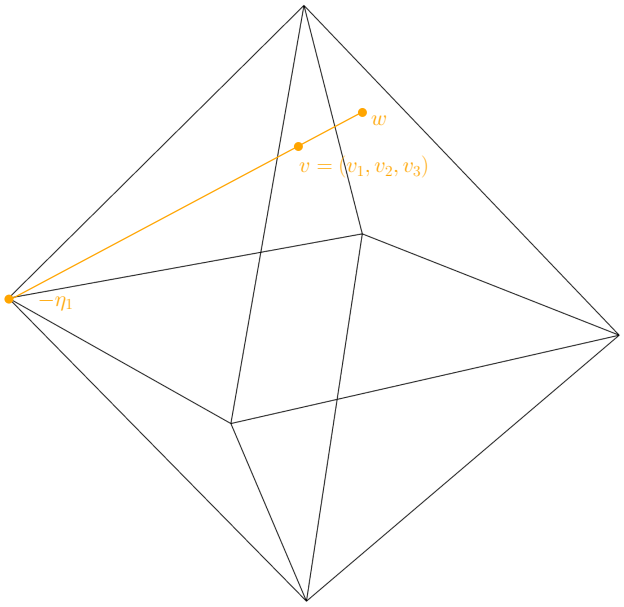


Figure 2.35: Above is an example of $v = (v_1, v_2, v_3)$ and its relationship to $w(v) = w$.

Choose a plane C_0 with normal vector $\xi = (1, \text{sign}(v_2), \text{sign}(v_3))$. Let C_k denote $C_0 + (0, 0, k)$ for all $k \in \mathbb{Z}$. Further, assume that the cubes on C_0 are even, so that edges parallel to one of $\{\eta_1, \text{sign}(v_2)\eta_2, \text{sign}(v_3)\eta_3\}$ connect cubes on C_0 to cubes on C_1. Since $v \in \partial\mathcal{O}$, any tiling with mean current v splits into perfect dimer tilings of *slabs* L_k which consist of unions of adjacent planes $L_k := C_{2k} \cup C_{2k+1}$, $k \in \mathbb{Z}$ (see Section 2.4). By Proposition 2.4.1, each slab is a copy of the 2-dimensional hexagonal lattice. There are three 3D dimer tilings of $C_{2k} \cup C_{2k+1}$ consisting of only one type of dimer, and these correspond to the three brickwork lozenge tilings using one type of lozenge. (See Figure 2.19 for a review of the correspondence between 3D dimers and lozenges.)

Restricted to each slab L_k, τ_v will be one of the three brickwork lozenge tilings. On $L_0 = C_0 \cup C_1$, τ_v will be the a_1 brickwork lozenge tiling. On $L_1 = C_2 \cup C_3$, τ_v will be the a_2 brickwork lozenge tiling. We continue this by repeating the periodic sequence $a_1, ..., a_r$ forwards and backwards in k to choose the tile type for τ_v on all other slabs $L_k = C_{2k} \cup C_{2k+1}$.

The reference tiling τ_1 consists of all $-\eta_1$ tiles, which connect C_{2k} to C_{2k-1}. Subtracting τ_1, the tiles in $-\tau_1$ connect C_{2k-1} to C_{2k}, meaning that they connected the "odd" half of L_{k-1} to the "even" half of L_k. Hence in the double dimer tiling (τ_v, τ_1), every tile is on an infinite path. Along each infinite path, (τ_v, τ_1) consists of the periodic sequence of tiles parallel to $\ldots a_1, \eta_1, a_2, \eta_1, \ldots, \eta_1, a_r, \ldots$. In particular, all infinite paths are parallel to $v + \eta_1$. This completes the construction for $v \in \partial\mathcal{O} \cap \mathbb{Q}^3$, $v_1 \geq 0$.

Now we extend the construction to any $v \in \mathcal{O} \cap \mathbb{Q}^3$, $v \neq -\eta_1$. Let p_v be the line through $-\eta_1 = (-1, 0, 0)$ and $v = (v_1, v_2, v_3)$. Let $w = w(v)$ be the intersection of p_v with $\{u = (u_1, u_2, u_3) \in \partial\mathcal{O} : u_1 \geq 0\}$. See Figure 2.35. The relationship between v, w will be sufficiently important that we record it as a definition.

Definition 2.8.26. Fix $v \in \mathcal{O}$, $v \neq -\eta_1$, and let p_v be the line through $-\eta_1$ and v. We define

$w(v)$ to be the intersection of p_v with the part of $\partial\mathcal{O}$ with non-negative first coordinate.

Note that if v is rational, $w = w(v) \in \partial\mathcal{O}$ is rational. Since the first coordinate of w is non-negative, we can construct τ_w as described above. In (τ_w, τ_1), every tile is along an infinite path. On the other hand, in (τ_1, τ_1) none of the tiles are along infinite paths. To construct (τ_v, τ_1), we interpolate between these two options by choosing an intermediate density of infinite paths.

If the line $p_v(t)$ is parameterized so that $p_v(1) = (-1, 0, 0)$ and $p_v(0) = w$, let $a \in [0, 1]$ be such that $p_v(a) = v$. Since v rational, w and a are also rational. Thus there exist periodic patterns of cubes in C_0 with density a. To construct (τ_v, τ_1), we fix a choice of periodic pattern of cubes on C_0 with density a. We delete all the infinite paths in (τ_w, τ_1) which do not go through one of the chosen cubes on C_0 and replace them with tiles parallel to $-\eta_1$. The resulting tiling is τ_v.

The tilings τ_v have a few important properties which we highlight.

Lemma 2.8.27. *Let $v \in \mathcal{O}; v \neq -\eta_1$.*

1. *τ_v has mean current v;*
2. *Let $w = w(v)$ be as in Definition 2.8.26. The infinite paths in τ_v are parallel to $w + \eta_1$.*
3. *For any rational plane P, the restriction of τ_v to P is doubly periodic, with period depending on r (the number of tiles $a_1, ..., a_r$ used to approximate $w(v)$), the choice of periodic pattern of cubes in C_0 and P.*

Remark 2.8.28. Note that τ_v is not uniquely determined by v. It depends on the sequence of tiles $a_1, ..., a_r$ used to approximate v, and on and periodic pattern of initial sites on C_0.

We now show that pieces of τ_v, τ_u can "glued" along a plane P, as long as v, u have the same flow through P. The amount of space k we need to glue is a constant depending only on the period of the tilings τ_u and τ_v.

Lemma 2.8.29. *Suppose that $u, v \in \mathcal{O} \cap \mathbb{Q}^3; u, v \neq \eta_1$ and τ_u, τ_v are tilings as in Lemma 2.8.27. Suppose that P is a coordinate plane or plane with normal vector of the form $(\pm 1, \pm 1, \pm 1)$. In both cases we denote the normal vector by n_P. Let r be such that τ_u and τ_v are periodic in P with fundamental domain an $r \times r$ parallelogram. If $v \cdot n_P = u \cdot n_P$, then there is an even integer $k > 0$ (depending on r and P) such that τ_v restricted to the left half-space of P can be connected to τ_u restricted to the right half-space of $P + k$ for some $k = O(r)$. Further, the connecting tiling τ is also periodic in P with fundamental domain an $r \times r$ parallelogram.*

Remark 2.8.30. We restrict to these two types of planes P since the the faces of tetrahedra in the mesh used to define the piecewise constant approximation (Proposition 2.8.21) are contained in one of these two types of planes; see Remark 2.8.23. The analogous result for other planes also holds, but we do not need it.

Proof. By Lemma 2.8.27, τ_u, τ_v are periodic, so there exists $r > 0$ finite and determined by τ_u, τ_v, P such that τ_u is periodic on P with fundamental domain $R_0 \subset P$, where R_0 an $r \times r$ parallelogram contained in P, and similarly τ_v is periodic on $P + k$ with fundamental

domain also an $r \times r$ parallelogram in $P + k$. Let R be the parallelopiped region parallel to n_P between one fundamental domain $R_0 \subset P$ and another $R_k \subset P + k$.

Let $R/\sim$ be R with opposite faces other than R_0 and R_k paired (i.e., $R/\sim$ is a 2-dimensional torus crossed with an interval). Given the periodicity of τ_u, τ_v, to that show the region between P and $P + k$ is tileable with $\tau_u \mid_P$ and $\tau_v \mid_{P+k}$, it suffices to show that $R/\sim$ is tileable with $\tau_u \mid_{R_0}$ and $\tau_v \mid_{R_k}$.

To show that this region is tileable we use the same techniques as in Section 2.6. In other words, first we show that $R/\sim$ with $\tau_u \mid_{R_0}$ and $\tau_v \mid_{R_k}$ is balanced, and then use Hall's matching theorem (2.6.3). The setting here is more elementary than what we consider in Section 2.6, since here the tilings defining the boundary conditions are completely periodic.

Since k is even, any perpendicular slice of $R/\sim$ is a fundamental domain for τ_u or τ_v, the condition $v \cdot n_P = u \cdot n_P$ is equivalent to

$$\sum_{e \text{ intersecting } R_0} v_{\tau_u}(e) \cdot n_P = \sum_{e \text{ intersecting } R_k} v_{\tau_v}(e) \cdot n_P. \tag{2.63}$$

Equation (2.63) is in turn equivalent to $R/\sim$ with boundary conditions $\tau_u \mid_{R_0}$ and $\tau_v \mid_{R_k}$ being balanced.

Since the region is balanced, by Hall's matching theorem (Theorem 2.6.3) it is not tileable if and only if there exists a counterexample region U which is a strict subset. Since U is a strict subset, U has a nonempty interior boundary $S \subset \partial U$. Let $T = \partial U \setminus S$. By Proposition 2.6.6,

$$\text{imbalance}(U) = \frac{1}{6}\Big(\text{white}(T) - \text{black}(T) - \text{area}(S)\Big).$$

Since $\text{area}(T) \leq 2r^2$, $\text{white}(T) \leq 2r^2$. Since the region is tileable with boundary condition from just one of the tilings, S must connect R_0 and R_k, if $k > r$, by Proposition 2.6.18,there is a universal constant κ such that $\text{area}(S) \geq \kappa k r$. Therefore by Proposition 2.6.6,

$$\text{imbalance}(U) \leq \frac{2r^2 - \kappa k r}{6}.$$

Choosing $k = cr$ for some constant $c > 2/\kappa$, U is not a counterexample which contradicts the assumption that the region is not tileable. This completes the proof. □

Using the tilings τ_v as our building blocks and their gluing properties to put them together, we now proceed to prove Theorem 2.8.24.

Proof of Theorem 2.8.24. Choose a scale $\epsilon > 0$ so that the piecewise-constant approximation $\widetilde{g}$ from Proposition 2.8.21 on an ϵ-scale tetrahedral mesh $\mathcal{X} = \{X_1, ..., X_M\}$ satisfies $d_W(g, \widetilde{g}) < \delta/2$ and hence $A_{\delta/2}(\widetilde{g}) \subset A_\delta(g)$. We assume that all $X \in \mathcal{X}$ are regular or right-angled tetrahedra so that all their faces are contained in coordinate planes or planes with normal vector $(\pm 1, \pm 1, \pm 1)$. Recall that $\widetilde{g} \in AF(\widetilde{R})$ and that $R \subset \widetilde{R}$, so any free boundary tiling of $\widetilde{R}$ can be restricted to a free boundary tiling of R. To prove the theorem, it suffices to construct $\tau \in T_n(\widetilde{R})$ with $d_W(f_\tau, \widetilde{g}) < \delta/2$.

Constructing channels. We construct *channels* $C_1, ..., C_K$ which are disjoint, partition $\widetilde{R}$ and will be nicely chosen tubular neighborhoods of a modification of the flow lines of $\widetilde{g} + \eta_1$.

For any $X_j \in \mathcal{X}$, let $v_j := \widetilde{g}\,|_{X_j}$. Recall Definition 2.8.26, which relates a vector v with $w(v)$, which is the direction of the infinite paths in a periodic tiling τ_v. For each channel C_i, the intersection $C_i \cap X_j$ will be a tube parallel to

$$w(v_j) + \eta_1.$$

Since $\widetilde{g}$ is valued strictly in $\mathrm{Int}(\mathcal{O})$, $v_j \neq -\eta_1$ for all $X_j \in \mathcal{X}$, and hence $w(v_j)$ is well-defined everywhere. As shorthand, we let $w(\widetilde{g})$ be the piecewise-constant flow equal to $w(v_j)$ on X_j. The definitions are made so that if τ_{v_j} is a periodic tiling built by the construction earlier in this section, the infinite paths in (τ_{v_j}, τ_1) move parallel to the direction of the channel on X_j. The values of $\widetilde{g}$ change on the boundaries ∂X_j of tetrahedra in the mesh. We choose the channels C_i to be thin enough as follows so that, viewing C_i as a sequence of open tubes, each end of the tube $C_i \cap X_j$ is contained in a single plane (i.e., each end is contained in a single face of ∂X_j).

Project the corners and edges of X_j onto ∂X_j along $w(v_j) + \eta_1$. Call these projections γ_j. The points $\gamma_j \subset \partial X_j$ are the ones along a flow line of $w(\widetilde{g}) + \eta_1$ that goes through an edge of X_j. The lines γ_j divide the faces of X_j into between 1 and 3 sections.

We further divide X_j by taking into account the iterated projections of the corners and edges of all the other tetrahedra in the mesh. In other words for all j, if X_k is a neighbor of X_j, then we project γ_j onto ∂X_k by orthogonal projection along $w(v_k) + \eta_1$. We iterate this for all tetrahedra until there is a projection of γ_j on ∂X_k for all $\{k, j\}$ pairs. See the left figure in Figure 2.36.

The result is that for each $j \in \{1, ..., M\}$, each triangular face of ∂X_j is partitioned into between 1 and 3^M pieces, and X_j is partitioned into tubes parallel to $w(v_j) + \eta_1$ with these pieces as their ends. Each sequence of successive tubes glued on their intersections with ∂X_j is a *channel*. The collection of channels $C_1, ..., C_K$ is pairwise disjoint and covers $\widetilde{R}$. Since $w(\widetilde{g}) + \eta_1$ has positive first coordinate everywhere, each channel connects a patch on $\partial\widetilde{R}$ to another.

Tiling a channel. Fix n large and a choice of channel C. We construct a scale n tiling of C which has only $-\eta_1$ tiles in a neighborhood of ∂C of constant-order width in n, and use this to say that we can put together the tilings of the channels together to construct one tiling of the whole region.

Let $T_1, ..., T_m$ be the sequence of tubes of the form $X_j \cap C$ in order from one intersection of C with $\partial\widetilde{R}$ to the other. Let $v_1, ..., v_m$ be the corresponding values of $\widetilde{g}$ on the tubes. Consider the $\mathbb{Z}^3$ tilings $\tau_{v_1}, ..., \tau_{v_m}$ constructed earlier in this section using the reference tiling τ_1. Recall that for each v_i, all tiles in (τ_{v_i}, τ_1) are either double tiles (which must be $-\eta_1$ tiles) or are on an infinite path, and that all infinite paths follow the same periodic sequence. Let r_j be the period of τ_{v_j}, i.e. if $\tau(x)$ denotes the tile at x in τ, then for $j = 1, ..., m$, r_j is such that translates $\tau_{v_j}(x + r_j\eta) = \tau_{v_j}(x)$ for all unit vectors η.

The main operation we will use is that for any infinite path $\ell \subset (\tau_{v_j}, \tau_1)$, we can modify τ_{v_j} by "shifting" all the tiles along ℓ, i.e. by removing all the tiles on τ_{v_j} along ℓ and replacing them with $-\eta_1$ tiles. We refer to this as *deleting* the path ℓ. The idea is to delete paths that would exit the channel before hitting $\partial\widetilde{R}$, and then to bound the number of paths that we delete.

Let $\pi_j \subset \partial X_j$ be the starting end of T_j, so that T_j is a tube connecting π_j to π_{j+1}. For each j, we start by restricting τ_{v_j} to T_j. Any infinite path $\ell \subset (\tau_{v_i}, \tau_1)$ which enters T_j in

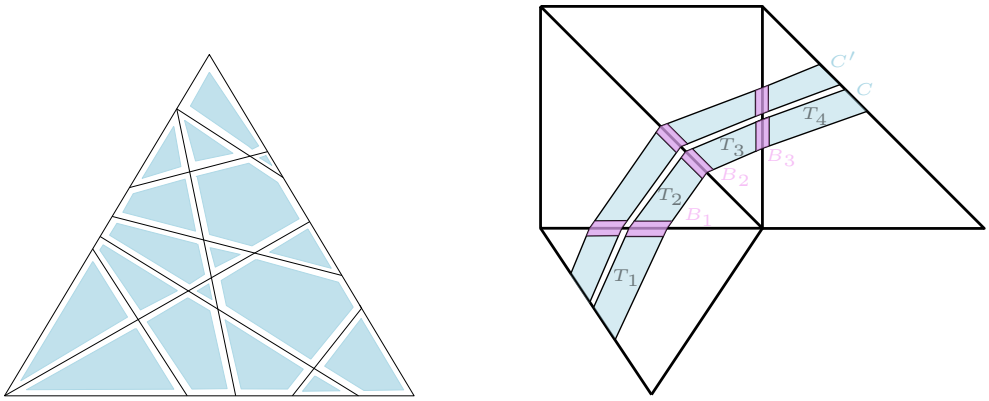


Figure 2.36: On the left is a face of one tetrahedron. The segments are the ends of channels, the smaller blue regions are places where we do not delete infinite paths. The width of the white area is $O(r/n)$. The figure on the right is a 2D schematic showing two channels C, C', with the tubes T_j and connector regions B_j labeled along C. The width of the white area between C and C' is $O(r/n)$.

π_j must exit through $\partial T_j \setminus \pi_j$, since paths in (τ_{v_j}, τ_1) always have direction with positive η_1 component.

First, we delete all infinite paths $\ell \subset (\tau_{v_j}, \tau_1)$ which do not enter T_j through π_j and exit for the first time through π_{j+1}, replacing the tiles of τ_{v_j} along these paths with $-\eta_1$ tiles. Note that this includes deleting all infinite paths which do not intersect T_j.

By Lemma 2.8.27, the infinite paths in (τ_{v_j}, τ_1) have asymptotic direction $w(v_j)$, with oscillation bounded by the length of the periodic sequence used to construct τ_{v_j}, which is $O(r_j)$. Since the direction of the tube T_j is also $w(v_j)$, any infinite path in (τ_{v_j}, τ_1) that enters T_j through π_j and exits through $T_j \setminus \pi_{j+1}$ is within $O(r_j)$ distance counted as number of edges in $\frac{1}{n}\mathbb{Z}^3$ of ∂C along T_j. Similarly, any path which enters T_j through $\partial T_j \setminus \pi_j$ would also remain within $O(r_j)$ distance in number of $\frac{1}{n}\mathbb{Z}^3$ edges of ∂C along T_j. In summary, the paths that we delete which intersect T_j are all contained in an neighborhood of $\partial C \cap T_j$ of width $O(r)$ in edge distance in $\frac{1}{n}\mathbb{Z}^3$, corresponding to a neighborhood of Euclidean width $O(r/n)$ (recall that r is a constant independent of n).

Second (to avoid issues with corners and edges of tetrahedra, and to isolate channels from each other), we delete all infinite paths which are within a Euclidean neighborhood of width $1000/n$ of $\partial T_j \setminus \{\pi_j \cup \pi_{j+1}\}$ (i.e., 1000 lattice cubes in $\frac{1}{n}\mathbb{Z}^3$). By the same logic as above, these are still contained in an $O(r/n)$-width neighborhood of ∂C. We call this tiling τ'_{v_j}.

Third, we want to glue the tiling on T_j to the tiling on T_{j+1}. To do this, we cut out a neighborhood of width $O(r) = O(r_1, ..., r_m)$ in $\frac{1}{n}\mathbb{Z}^3$ lattice cubes around π_{j+1} (the face shared by T_j and T_{j+1}) which we call the *connector region* B_j. Let α_j, α_{j+1} be the ends of the connector region (i.e. translates of π_{j+1}), see the purple region in right side figure in Figure 2.36. Let P_j be the plane containing α_j and P_{j+1} be the plane containing α_{j+1}.

Since τ_{v_j} and $\tau_{v_{j+1}}$ are periodic tilings of periods r_j, r_{j+1}, and since B_j has width $O(r)$, by Lemma 2.8.29 we can construct a tiling σ_j of $\mathbb{Z}^3$ such that it agrees with τ_{v_j} in the left half-space of P_j and $\tau_{v_{j+1}}$ on the right half-space of P_{j+1}, and fills in the region in-between in a periodic way with period $O(r)$. We can do this for all $j = 1, ..., m$.

Overlaying (σ_j, τ_1), we again get a collection of infinite paths and double tiles. First, we delete all infinite paths in σ_j which were deleted to construct τ'_{v_j} and $\tau'_{v_{j+1}}$ from τ_{v_j} and $\tau_{v_{j+1}}$. Second, we delete any infinite paths which exit C between α_j and α_{j+1} (i.e., any paths which exits C along the connector B_j). Since B_j has length $O(r)$ and since σ_j is periodic with period $O(r)$, again any infinite path which exits in B_j is contained in an $O(r)$ neighborhood of ∂C along B_j.

Finally, we can glue together the tilings $\sigma_1, ..., \sigma_m$ by going back and deleting any infinite path in (σ_j, τ_1) which connects to one which would have been deleted in (σ_i, τ_1) for all other $i \neq j$. Since the number of tubes m is constant, in the end we have a tiling τ of C where we have deleted infinite paths of (τ, τ_1) in a neighborhood of width at most constant-order in n (concretely $1000 + O(r)$, where r is constant in n) in distance measuring in edges of $\frac{1}{n}\mathbb{Z}^3$, corresponding to a neighborhood of Euclidean width of $O(r/n)$.

Since all channels C are tiled so that they have only $-\eta_1$ tiles in a neighborhood of ∂C, we can put them together. Therefore we have constructed a tiling $\tau \in T_n(\widetilde{R})$.

Bounding the final distance. To emphasize the dependence on n, let τ^n be the tiling at scale n constructed above and let τ_1^n be the $-\eta_1$ brickwork pattern at scale n. On one hand, the total flow of $\widetilde{g} + \eta_1$ over any $X_j \in \mathcal{X}$ is

$$\text{vol}(X_j)(v_j + \eta_1).$$

We claim that the double dimer flow $f_{(\tau^n, \tau_1^n)} = f_{\tau^n} - f_{\tau_1^n}$ has the same total flow, up to an $O(n^{-1})$ error. To explain the order of error, recall that for a scale n tiling flow, each $1/n^3$-volume lattice cube has flow of order $1/n^3$. The error comes from the region around the boundary of the channels where some infinite paths were deleted and replaced with $-\eta_1$ tiles. The number of lattice sites on the boundary of the channel is order n^2, and the region has width constant order in n in lattice cubes from $\frac{1}{n}\mathbb{Z}^3$, so the amount of deleted flow in this region has order $n^2/n^3 = 1/n$. Therefore there is a constant K such that

$$\left| \text{vol}(X_j)(\widetilde{g}_j + \eta_1) - \sum_{e \in \frac{1}{n}\mathbb{Z}^3,\, e \cap X_j \neq \emptyset} (f_{\tau^n} - f_{\tau_1^n})(e) \right| < Kn^{-1}.$$

There is also a constant $C = C(\widetilde{R})$ such that $M = C\epsilon^{-3}$ (recall that $M = |\mathcal{X}|$ is the number of tetrahedra in the mesh). By Lemma 2.5.14 applied to the partition $X_1, ..., X_M$ of $\widetilde{R}$,

$$d_W(f_{\tau^n} - f_{\tau_1^n}, \widetilde{g} + \eta_1) < 3M(10\epsilon^4 + Kn^{-1}) < 30C\epsilon + 3CK\epsilon^{-3}n^{-1}.$$

Taking n large enough so that $1/n < \epsilon^4$, this becomes a bound which is a constant times ϵ. A few applications of the triangle inequality and the "mass shift" property of Wasserstein distance, i.e. that $\mathbb{W}_1^{1,1}(\mu, \nu) = \mathbb{W}_1^{1,1}(\mu + \rho, \nu + \rho)$ (see Lemma 2.5.4), give that

$$d_W(f_{\tau^n}, \widetilde{g}) < d_W(f_{\tau^n} - f_{\tau_1^n}, \widetilde{g} + \eta_1) + d_W(f_{\tau_1^n}, -\eta_1).$$

Since $d_W(f_{\tau_1^n}, -\eta_1) \to 0$ as $n \to \infty$, we can make this as small as needed as $n \to \infty$. Therefore we can choose ϵ small enough and n large enough given δ so that $\tau^n \in T_n(\widetilde{R})$ has $d_W(f_{\tau^n}, \widetilde{g}) < \delta/2$. Restricting τ^n to R completes the proof. □

2.8.5 Soft boundary lower bound

With the machinery developed in the previous section we can now prove the *soft boundary lower bound*, namely Theorem 2.8.10. In particular we show that for (R, b) with b extendable outside, then for any $g \in AF(R, b)$,

$$\lim_{\delta \to 0} \liminf_{n \to \infty} v_n^{-1} \log \mu_n(A_\delta(g)) \geq \text{Ent}(g).$$

Recall that μ_n is counting measure on $TF_n(R, b, \theta_n)$, the set of free boundary tiling flows on R at scale n with boundary values within θ_n of b. The main idea of the proof is to show that from the one free boundary tiling flow $f_\tau \in A_\delta(g) \cap TF_n(R)$ constructed in previous section (Theorem 2.8.24), we can use the patching theorem (Theorem 2.6.14) to show that there are actually many tiling flows in $A_\delta(g)$.

Proof of Theorem 2.8.10. By Proposition 2.8.21, there exists $\delta_1 > 0$ such that there is a δ_1-tetrahedral mesh $\mathcal{X} = \{X_1, ..., X_M\}$, region $\widetilde{R} = \cup_{X \in \mathcal{X}} X$ containing R, and an asymptotic flow $\widetilde{g} \in AF(\widetilde{R})$ taking constant values on tetrahedra in $\mathcal{X}$ with $d_W(g, \widetilde{g}) < \delta/2$ so that

$$A_{\delta/2}(\widetilde{g}) \subset A_\delta(g).$$

Let $\widetilde{g}_i := \widetilde{g} \mid_{X_i}$. Computing directly,

$$\text{Ent}(\widetilde{g}) = \frac{1}{\text{Vol}(R)} \sum_{i=1}^{M} \text{Vol}(X_i) \text{ent}(\widetilde{g}_i).$$

On the other hand by Proposition 2.7.29,

$$\text{Ent}(\widetilde{g}) = \text{Ent}(g) + o_\delta(1).$$

Using the shining light construction from the proof of Theorem 2.8.24, for any n large enough there exists a tiling $\tau \in T_n(R)$ such that $f_\tau \in A_{\delta/2}(\widetilde{g})$ has a particular form. Let $C_1, ..., C_K$ denote the channels in the shining light construction. For each tetrahedron X and channel C that intersect, $X \cap C$ is a tube. As in the proof of Theorem 2.8.24, $\tau \mid_{X \cap C}$ is periodic at a scale independent of n, and has mean current $\widetilde{g}_X \in \text{Int}(\mathcal{O})$ on $X \cap C$ outside a neighborhood of $\partial(X \cap C)$ of width constant order in n.

We choose $\epsilon \ll \delta_1$, and partition the interior of $X \cap C$ (where τ has mean current in $\text{Int}(\mathcal{O})$) into small cubes with side length $\leq \epsilon$. For each $i = \{1, ..., M\}$, call the pieces of the partition contained in X_i

$$\{Q_1^i, ...Q_{k_i}^i\}_{i=1}^M.$$

For any (i, k) pair, $\tau \mid_{Q_k^i}$ is periodic. Recall that Q_k^i has diameter $< \epsilon$. For $\epsilon_1 \ll \epsilon$, for n large enough $\tau \mid_{\partial Q_k^i}$ is ϵ_1-nearly-constant with value $\widetilde{g}_i$ (see Definition 2.6.12), so it satisfies the conditions for the outer boundary condition in the patching theorem (Theorem 2.6.14). Fix $c \in (0, 1)$. For each (i, k) pair, we choose an EGM $\mu_{i,k}$ of mean current $\widetilde{g}_i$ (these exist by Corollary 2.7.25). Since $\widetilde{g}_i \in \text{Int}(\mathcal{O})$, a sample from $\mu_{i,k}$ satisfies the conditions of Theorem 2.6.14 with probability going to 1 as $n \to \infty$ (Corollary 2.6.25). Therefore by Theorem 2.6.14, for n large enough, with probability $(1 - c)$, τ restricted to ∂Q_k^i can be patched with

a sample σ from $\mu_{i,k}$ on an annulus of width cn. By the ergodic theorem, given any $\epsilon_2 > 0$, for n large enough we can assume that

$$d_W(f_\sigma \mid_{Q_k^i}, \widetilde{g}_i \mid_{Q_k^i}) < \epsilon_2 \tag{2.64}$$

with probability $1 - c$. Therefore with probability $1 - 2c$, Equation (2.64) holds and σ can be patched with τ.

Let $\pi_{i,k,n}$ denote uniform measure on tilings σ of Q_k^i at scale n with $\sigma \mid_{\partial Q_k^i} = \tau$ and satisfying Equation (2.64), and let $Z_n(Q_k^i)$ be its partition function. The additional constraint that Equation (2.64) is satisfied does not change the exponential order of the number of tilings, hence by Proposition 2.6.32, we get the following consequences for entropy:

$$(1 + O(c))h(\mu_{i,k}) \leq n^{-3}\mathrm{Vol}(Q_k^i)^{-1}H(\pi_{i,k,n}) = n^{-3}\mathrm{Vol}(Q_k^i)^{-1}\log Z_n(Q_k^i).$$

By Lemma 2.5.14 applied to the partition $\{Q_1^i, ..., Q_{k_i}^i\}_{i=1}^M \cup \{R \setminus \cup_{i=1}^M \cup_{k=1}^{k_i} Q_k^i\}$, if $\sigma \in T_n(R)$ is a free boundary tiling of R whose restrictions to each Q_k^i are in the support of $\pi_{i,k,n}$ then using Equation (2.64),

$$d_W(f_\tau, f_\sigma) < 3\epsilon^{-3}(10\epsilon^4 + \epsilon_2) + 3C(\partial R)\epsilon$$

where $C(\partial R)$ is a constant depending only on R. In particular, choosing $\epsilon_2 = \epsilon^4$ and taking ϵ sufficiently small, $f_\sigma \in A_{\delta/2}(\widetilde{g})$. By Theorem 2.5.39 (uniform continuity of the boundary value operator $T(\cdot, \partial R)$), we can choose ϵ small enough so that the boundary values $T(f_\sigma, \partial R)$ are within θ_n of b for all σ in the support of $\pi_{i,k,n}$. Therefore for n large enough,

$$\mu_n(A_{\delta/2}(\widetilde{g})) \geq \prod_{i=1}^M \prod_{k=1}^{k_i} Z_n(Q_k^i) \geq \prod_{i=1}^M \prod_{k=1}^{k_i} \exp\left(n^3\mathrm{Vol}(Q_k^i)h(\mu_{i,k})(1 + O(c))\right).$$

Recall that $v_n^{-1} = n^{-3}\mathrm{Vol}(R)^{-1}$. Since $\mu_{i,k}$ is an EGM of mean current $\widetilde{g}_i \in \mathrm{Int}(\mathcal{O})$, $h(\mu_{i,k}) = \mathrm{ent}(\widetilde{g}_i)$ by Theorem 2.7.23. Rearranging and taking into account the $O(\epsilon)$ proportion of each tetrahedron $X \in \mathcal{X}$ that is not included in the patched regions,

$$\begin{aligned} v_n^{-1}\log\mu_n(A_{\delta/2}(\widetilde{g})) &\geq \sum_{i=1}^M \sum_{k=1}^{k_i} \frac{\mathrm{Vol}(Q_k^i)}{\mathrm{Vol}(R)}\mathrm{ent}(\widetilde{g}_i)(1 + O(c)) \\ &\geq \sum_{i=1}^M \frac{\mathrm{Vol}(X_i)}{\mathrm{Vol}(R)}\mathrm{ent}(\widetilde{g}_i)(1 + O(c))(1 - O(\epsilon)) \\ &= \mathrm{Ent}(\widetilde{g})(1 + O(c))(1 - O(\epsilon)). \end{aligned}$$

Recall that ϵ is the size of the patched regions, and $c < \epsilon$ is the patching error. All of ϵ, c, ϵ_2 are much smaller than $\delta > 0$, and go to 0 as $\delta \to 0$. In particular for any fixed $\delta > 0$, and $\epsilon > 0$ fixed small enough given $\delta > 0$, there is a $n(\delta)$ such that for all $n > n(\delta)$,

$$\liminf_{n\to\infty} v_n^{-1}\log\mu_n(A_\delta(g)) \geq \liminf_{n\to\infty} v_n^{-1}\log\mu_n(A_{\delta/2}(\widetilde{g})) \geq \mathrm{Ent}(\widetilde{g}) + O(\epsilon) = \mathrm{Ent}(g) + o_\delta(1).$$

Taking $\delta \to 0$ completes the proof. □

2.8.6 Generalized patching and hard boundary lower bound

To prove the hard boundary lower bound (Theorem 2.8.19), we need one more tool. Recall that $\overline{\rho}_n$ is the uniform probability measure on tilings of a fixed region $R_n \subset \frac{1}{n}\mathbb{Z}^3$.

The shining light construction (Theorem 2.8.24) shows that for any $\delta > 0$ and $g \in AF(R,b)$, for n large enough there exists a free-boundary tiling $\tau \in T_n(R)$ such that $d_W(f_\tau, g) < \delta$. For hard boundary conditions, we need to know that every $g \in AF(R,b)$ can be approximated by a tiling of the *fixed* region R_n. To do this, we prove a *generalized patching theorem* (Theorem 2.8.32).

Let $B_n = [-n,n]^3$. Recall that the patching theorem (Theorem 2.6.14) says that if two tilings τ_1, τ_2 are *nearly constant* with value $s \in \mathrm{Int}(\mathcal{O})$ (Definition 2.6.12), then for any $c > 0$ there is n large enough that we can patch together $\tau_2 \mid_{B_{(1-c)n}}$ and $\tau_1 \mid_{\mathbb{Z}^3 \setminus B_n}$ by tiling the width-cn annulus between them.

In Section 2.6 where the regular patching theorem (Theorem 2.6.14) was proved, all our tools were combinatorial, and we thought of tilings τ of $\mathbb{Z}^3$ without rescaling to $\frac{1}{n}\mathbb{Z}^3$. Here we look at the tileability of more general "annular regions," where the tilings are rescaled to live in $\frac{1}{n}\mathbb{Z}^3$.

Let $R \subset \mathbb{R}^3$ be a compact set which is the closure of a connected domain and has piecewise-smooth boundary ∂R (i.e., the sort of region to which our LDP applies). For any small $c > 0$, we define

$$R^c = \{x \in R : d(x, \partial R) \geq c\}.$$

The set $R \setminus R^c$ is an annular region. On the discrete side, given a free-boundary tiling $\tau \in T_n(R)$, it restricts to $\tau' \in T_n(R^c)$ which is a free boundary tiling of R^c. We let $R_n^c \subset \frac{1}{n}\mathbb{Z}^3$ be the region covered by τ'. Given another free-boundary tiling $\sigma \in T_n(R)$, let $R_n \subset \frac{1}{n}\mathbb{Z}^3$ be the region covered by σ. The region $A = R_n \setminus R_n^c \subset \frac{1}{n}\mathbb{Z}^3$ is the type of annular region we study here.

Let (R,b) be flexible and suppose that $R_n \subset \frac{1}{n}\mathbb{Z}^3$ is a sequence of regions satisfying the conditions of the hard boundary LDP (Theorem 2.8.15), i.e. regions with tileable boundary values b_n on ∂R converging to b in $\mathbb{W}_1^{1,1}$. To prove any $g \in AF(R,b)$ can be approximated by f_τ with τ a tiling of R_n, we show that we can *patch together* suitable tilings on annuli of the form $R_n \setminus R_n^c$, with hard boundary condition on the outside.

Definition 2.8.31. We say that a flow $g \in AF(R,b)$ is *flexible* if g satisfies the condition that for any compact set $D \subset \mathrm{Int}(R)$, $\overline{g(D)} \subset \mathrm{Int}(\mathcal{O})$.

The pair (R,b) is *flexible* (see Definition 2.8.12 and Lemma 2.8.13) if and only if there exists $g \in AF(R,b)$ which is flexible.

Theorem 2.8.32 (Generalized patching theorem)**.** *Fix $c > 0$. Let (R,b) be flexible, with b a boundary asymptotic flow which is extendable outside. Let $R_n \subset \frac{1}{n}\mathbb{Z}^3$ be a sequence of regions with tileable boundary values b_n on ∂R converging to b in $\mathbb{W}_1^{1,1}$. Let σ_n be a sequence of tilings of R_n.*

Let $\tau_n \in T_n(R)$ be a sequence of tilings such that $d_W(g, f_{\tau_n}) \to 0$ as $n \to \infty$ for $g \in AF(R,b)$ flexible. Let τ_n' be τ_n restricted to a free boundary tiling of R^c, and let $R_n^c \subset \frac{1}{n}\mathbb{Z}^3$ be the cubes covered by τ_n'.

For n large enough, $R_n \setminus R_n^c$ is tileable.

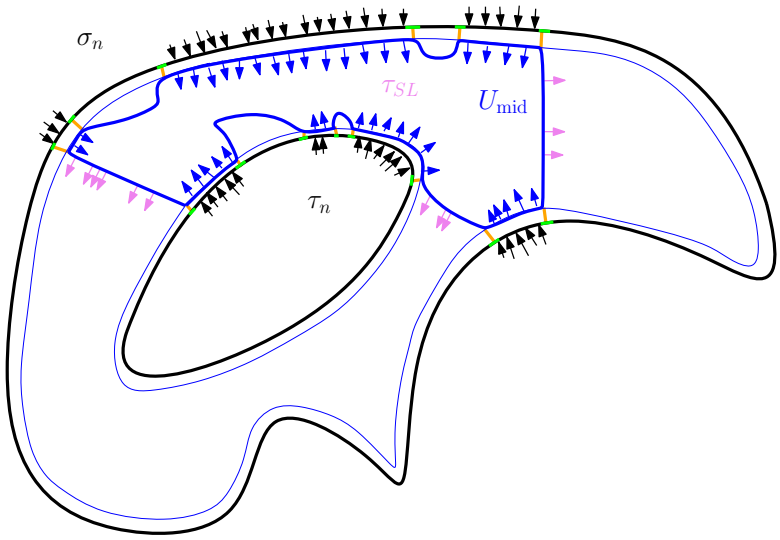


Figure 2.37: 2D schematic picture for the proof of the generalized patching theorem.

Remark 2.8.33. The *flexible* condition here is analogous to the $s \in \text{Int}(\mathcal{O})$ condition in the original patching theorem (Theorem 2.6.14). The generalized patching theorem is the reason the hard boundary LDP requires that (R, b) is flexible.

Before we prove this, we explain how it can be used to prove the hard boundary lower bound (Theorem 2.8.19). First, from the generalized patching theorem, it is straightforward to prove the fixed boundary analog of Theorem 2.8.24.

Corollary 2.8.34. *Suppose that (R, b) is flexible and b is a boundary asymptotic flow which is extendable outside. Fix a sequence of regions R_n with tileable boundary values b_n on ∂R converging to b in $\mathbb{W}_1^{1,1}$. For any $\delta > 0$ and any $g \in AF(R, b)$, there is n large enough such that there exists a tiling τ of R_n with $d_W(f_\tau, g) < \delta$.*

Proof. Since (R, b) is flexible, there exists $g_0 \in AF(R, b)$ which is flexible. For any $\epsilon > 0$, the new flow $g_\epsilon = \epsilon g_0 + (1 - \epsilon)g$ satisfies $d_W(g, g_\epsilon) < C\epsilon$ for some constant $C > 0$. Taking ϵ small enough, we can guarantee that $d_W(g, g_\epsilon) < \delta/2$.

Since g_ϵ is flexible and R_n is tileable, by Theorem 2.8.32 for n large enough there exists a tiling τ of R_n such that $d_W(g_\epsilon, f_\tau) < \delta/2$. By the triangle inequality $d_W(g, f_\tau) < \delta$, which completes the proof. □

Adding Corollary 2.8.34 as the first step, the proof of Theorem 2.8.19 (hard boundary lower bound) is the same as the proof of Theorem 2.8.10 (soft boundary lower bound).

Proof of Theorem 2.8.19. By Corollary 2.8.34, given any $\delta > 0$, for n large enough and any $g \in AF(R, b)$ we can find a tiling τ of R_n such that $d_W(f_\tau, g) < \delta$. Further, this tiling is of the form given in the shining light construction (proof of Theorem 2.8.24), other than in an annulus of width $c \in (0, 1)$ where c can be taken arbitrarily small. The remainder of the proof is identical to the proof of Theorem 2.8.10, where we use the regular patching theorem to patch in samples from ergodic Gibbs measures of appropriate mean currents. □

It remains to prove the generalized patching theorem (Theorem 2.8.32). The proof is structurally analogous to the proof of the regular patching theorem for cubes $B_n = [-n, n]^3$ (Theorem 2.6.14), and relies on a sequence of lemmas. We give an outline of the main ideas to explain where each of the lemmas is used, accompanied by the schematic picture in Figure 2.37. We then state and prove each of the lemmas, followed by a proof of Theorem 2.8.32.

Note that there are some superficial changes between the results here and their analogs in Section 2.6, since here our regions $R_n \subset \frac{1}{n}\mathbb{Z}^3$ instead of $B_n \subset \mathbb{Z}^3$. Basically this corresponds to a change in units. We introduce a few new pieces of notation to make it easier to work with the tilings τ of $\frac{1}{n}Z^3$ instead of $\mathbb{Z}^3$.

- If Q is a discrete surface built out of lattice squares in $\frac{1}{n}\mathbb{Z}^3$, we define $\text{area}_n(Q)$ to be the number of lattice squares on Q. This is n^2 times the Euclidean area of Q. If Q is a surface not built of lattice squares, we can still use $\text{area}_n(Q)$ to mean the Euclidean area of Q times n^2.

- If ℓ is a discrete curve built out of edges of squares in $\frac{1}{n}\mathbb{Z}^3$, we defined $\text{length}_n(\ell)$ to be number of lattice edges in L. This is n times the Euclidean length of ℓ.

- If τ is a tiling of $\frac{1}{n}\mathbb{Z}^3$, then the tiling flow f_τ and pretiling flow v_τ are typically *rescaled* so that for $e \in \frac{1}{n}\mathbb{Z}^3$, $v_\tau(e) = \pm 1/n^3$ or 0, and $f_\tau(e) = \pm 5/6n^3$ or $\pm 1/6n^3$. We define $\widetilde{v}_\tau$ to be *unrescaled flow* $\widetilde{v}_\tau(e) = \pm 1$ or 0 for $e \in \frac{1}{n}\mathbb{Z}^3$, and similarly $\widetilde{f}_\tau(e) = \pm 5/6$ or $\pm 1/6$.

The proof of Theorem 2.8.32 uses a mixture of combinatorial results like in Section 2.6 and more analytic results about Wasserstein distance, which are for rescaled tiling flows. These pieces of notation make it easier to go between these points of view, and to explain the purely combinatorial arguments in a way more analogous to Section 2.6.

The main combinatorial tool is again Hall's matching theorem (Theorem 2.6.3), which says that if $A = R_n \setminus R_n^c$ is not tileable, then there exists a counterexample set $U \subset A$ which proves it. The *interior boundary* S of U is without loss of generality a minimal monochromatic discrete surface with number of squares from $\frac{1}{n}\mathbb{Z}^3$ on the surface bounded above and below by a constant times n^2 (Lemma 2.8.35). The first step is to *indent* slightly and let $A_{\text{mid}} \subset A$ be a slightly smaller annulus where U is well-behaved (Lemma 2.8.36). We then define $U_{\text{mid}} = U \cap A_{\text{mid}}$.

We find a *test tiling* τ_{SL} using a shining light construction (Definition 2.8.38) and show that it satisfies a flow bound (Lemma 2.8.39). This is where we use the condition that g is flexible. Using Lemma 2.8.39, we show that there is a constant $K \in (0, 1)$ such that $\text{imbalance}(U'_{\text{mid}}) \leq -Kn^2 + O(n)$, where U'_{mid} is U_{mid} plus a few cubes from the rest of U (determined by τ_{SL}). This "slack" corresponds to flow from τ_{SL} which exits through the boundary of U_{mid} in the interior, see the pink arrows in Figure 2.37.

It remains to bound the imbalance in $U_{\text{shell}} := U \setminus U'_{\text{mid}}$. This we break into two pieces:

- Regions of U_{shell} contained in nice "cylinders" connecting ∂A to ∂A_{mid}. These are the regions where there are black and blue arrows in Figure 2.37, the sides of the cylinder are the orange regions which we call the "ribbon surface."

- The rest of U_{shell}, which we call the "leftover region."

Up to error related to the orange area in Figure 2.37 (the ribbon surface), we show that the imbalance of U_{shell} is the same as the imbalance in the cylinder regions. Finally we relate the imbalance in the cylinders to the flux of tiling flows f_{σ_n}, f_{τ_n}, and $f_{\tau_{\text{SL}}}$, which we can bound using Wasserstein convergence considerations using Lemma 2.5.15, up to an error proportional to the green area in Figure 2.37.

We now proceed to the lemmas. We first note that Lemma 2.6.19 for cubes has an analog for general regions. The only difference is that the constants c_1, c_2 change since they can depend on the regions.

Lemma 2.8.35. *Let $R_n \subset \frac{1}{n}\mathbb{Z}^3$ be a region of diameter in lattice squares at least n such that R_n are regions approximating a fixed region $R \subset \mathbb{R}^3$ with ∂R a piecewise smooth surface. Define $A = R_n \setminus R_n^c$ for some $c > 0$. Suppose that $S \subset A$ is a monochromatic minimal discrete surface in $\frac{1}{n}\mathbb{Z}^3$ with connects the inner and outer boundaries of A. Then there exist constant c_1, c_2 independent of S and n such that*

$$c_1 n^2 \geq \text{area}_n(S) \geq c_2 n^2$$

where $c_2 \sim c^2$.

Proof. This proof is completely analogous to the proof of Lemma 2.6.19.

For the upper bound, we use that S is minimal to get a bound which is a constant times the surface area of R_n. Since $R_n \subset \frac{1}{n}\mathbb{Z}^3$ are regions approximating R, and since ∂R is piecewise smooth, $\text{area}(R_n)$ is bounded by a constant times n^2, where this constant is determined by ∂R.

For the lower bound, we apply Proposition 2.6.18 to a point $p \in S$ which is distance at least $c/3$ from ∂A to get that

$$\text{area}(S) \geq \kappa((c/3)\, n)^2 = c_2 n^2,$$

where κ is a universal constant coming from the isopertimetric inequality. From this expression we see that c_2 is order c^2. □

Lemma 2.8.36 (Generalized indenting lemma)**.** *Let $A = R_n \setminus R_n^c$, and fix $\beta > 0$ small. Let S be a minimal monochromatic surface connecting inner and outer boundaries of A. There exists $\epsilon < c$ independent of S such that the following hold for n large enough:*

1. *Let $A_{a,b}$ denote the annulus between layers ∂R_n^a and ∂R_n^b. Then $\text{area}_n(S \cap A_{\epsilon,2\epsilon}) < \beta n^2$.*
2. *There exists $a \in (\epsilon, 2\epsilon)$ such that $\text{length}_n(\partial(\partial R_n^a \cap S)) \leq (\beta/\epsilon)n$.*
3. *A "ribbon surface" γ for $S \cap \partial R_n^a$ is a surface connecting ∂R_n^a to ∂R_n with boundary $\partial(S \cap \partial R_n^a)$ on ∂R_n^a. The "ribbon area" is the minimal area_n of a ribbon surface, and is bounded by $2\beta n^2$.*

The analogous bounds hold for some $a' \in (c - 2\epsilon, c - \epsilon)$.

Remark 2.8.37. For two surfaces A, B, the set $A \cap B$ is either a surface (2-dimensional), a curve (1-dimensional), or a combination of the two. In any of these cases, we take $\partial(A \cap B)$ to mean that we take union of the curve part of $A \cap B$ and the boundary of the surface part of $A \cap B$.

Proof. By Lemma 2.8.35, $c_2 n^2 \leq \text{area}_n(S) \leq c_1 n^2$. For a given $b > 0$, divide a band of the form $A_{0,b}$ in half. Each time we divide, by Lemma 2.8.35 both the halves of S have area bounded below by a fixed constant times b^2. On the other hand, one of the halves of S can have at most $1/2$ the original area. Iterating this, we can find $\epsilon > 0$ small enough so that the outer band after we split has area at most βn^2.

By the pigeonhole principle, there exists a layer an between ϵn and $2\epsilon n$ where

$$\text{length}_n(\partial(\partial R_n^a \cap S)) \leq (\beta n^2)/(\epsilon n) = (\beta/\epsilon)n.$$

Given this, we can find a ribbon surface γ (not necessarily built from lattice squares) with $\text{area}_n(\gamma) \leq (\beta/\epsilon)n(2\epsilon n) = 2\beta n^2$, and hence the ribbon area is bounded by $2\beta n^2$. □

Suppose that $g \in AF(R, b)$ is flexible. For any $\delta > 0$, we can find a piecewise-constant flow $\widetilde{g}$ with $d_W(g, \widetilde{g}) < \delta$ (Proposition 2.8.21). More precisely, $\widetilde{g}$ is piecewise-constant on a mesh of small tetrahedra $\mathcal{X}$. The region $\widetilde{R} = \cup_{X \in \mathcal{X}} X$ contains R, and $\widetilde{g} \in AF(\widetilde{R})$. In the proof of Theorem 2.8.24, to construct a tiling approximation $\tau \in T_n(R)$ of g for n large, we construct a tiling approximation $\widetilde{\tau} \in T_n(\widetilde{R})$ of $\widetilde{g}$ for n large, and restrict it to R.

The flexible condition passes from g to $\widetilde{g}$ as follows. For any compact set $D \subset \text{Int}(R)$, g flexible means that $\overline{g(D)} \subset \text{Int}(\mathcal{O})$. In particular there is a constant $k_D \in [0, 1)$ such that if $x \in D$ then

$$|g(x)|_1 \leq k_D < 1,$$

where $|\cdot|_1$ denotes the L^1 norm. When g is flexible, we can choose $\widetilde{g}$ so that for any compact set $D \subset \text{Int}(R)$, for all tetrahedra $X \in \mathcal{X}$ such that $X \subset D$, $|\widetilde{g}_X|_1 \leq k_D$. When this holds, we say that $\widetilde{g}$ is a *piecewise constant approximation of g inheriting the flexible condition.*

For any $\epsilon > 0$ and $D \subset \text{Int}(R)$ compact, for $\delta > 0$ small enough we can find $\widetilde{g}$ which is piecewise-constant on a scale δ tetrahedral mesh $\mathcal{X}$, has $d_W(g, \widetilde{g}) < \epsilon$, and inherits the flexible condition on D, i.e. $|\widetilde{g}_X| \leq k_D$ for all $X \in \mathcal{X}$ such that $X \subset D$.

In the shining light construction in the proof of Theorem 2.8.24, we cut the tetrahedra into tubes (with thin space between them), and construct the tiling approximation $\widetilde{\tau}$ of $\widetilde{g}$ by filing the tubes in X with periodic tilings of mean current approximating $\widetilde{g}_X$ on X, and fill the thin area in between with all $-\eta_1$ tiles. There is some maximum period r that we use to construct the periodic tilings on the tubes, and r is independent of n. For all n, at all the sites where $\widetilde{\tau}$ does not have mean current $\widetilde{g}$, $\widetilde{\tau}$ looks locally like the $-\eta_1$ brickwork pattern. The width of the region containing the places where $\widetilde{\tau}$ looks like the $-\eta_1$ brickwork pattern is $O(r)$ and therefore independent of n.

Definition 2.8.38 (Shining light measures)**.** Let $g \in AF(R, b)$ and let $\widetilde{g}$ be an approximation as discussed above. For each n, let $\widetilde{\tau}_n \in T_n(\widetilde{R})$ be a tiling produced by the shining light construction with $\widetilde{g}$, where the periodic tilings in the tubes have maximum period r. Fix a large constant $C = O(r)$. We define a *sequence of shining light measures* λ_n for g using $\widetilde{g}$ so that for each n, λ_n is uniform measure on tilings of the form $(\widetilde{\tau}_n + x)\,|_R \in T_n(R)$ for $x \in \mathbb{R}^3$ with $|x| \leq C$.

Using this, we prove a lemma analogous to Lemma 2.6.27 from Section 2.6. Instead of a result for ergodic measures of mean current $s \in \text{Int}(\mathcal{O})$, this lemma is for a sequence of shining light measures λ_n for a flexible flow g. This lemma is why the *flexible* condition is needed for generalized patching and hence for the hard boundary LDP.

Lemma 2.8.39 (Shining light flow bound)**.** *Let $D \subset Int(R)$, and let $g \in AF(R,b)$ be flexible. Let $\widetilde{g}$ be piecewise constant approximation of g on a tetrahedral mesh $\mathcal{X}$ such that $\widetilde{g}$ inherits the flexible condition, and such that the union of tetrahedra $\mathcal{D} \subset \mathcal{X}$ covering D is still contained in $Int(R)$.*

Let S be a monochromatic black surface in $\frac{1}{n}\mathbb{Z}^3$ with boundary ∂S and λ_n be a sequence of shining light measures as in Definition 2.8.38 for $\widetilde{g}$. Let Θ_D be the collection of odd cubes adjacent to $S \cap D$. Let $N = area_n(S \cap D)$ be the number of squares from $\frac{1}{n}\mathbb{Z}^3$ on $S \cap D$. Then there is a constant $K_D \in (0,1)$ independent of S such that for all n large enough,

$$\mathbb{E}_{\lambda_n}[|flux(\widetilde{v}_\tau, S \cap D)|] \geq K_D|\Theta_D| + O(n^{-1}N) + O(length_n(\partial(S \cap D)).$$

The constant K_D depends only on D and g. In particular, it is independent of $\widetilde{g}$, as long as $\widetilde{g}$ inherits the flexible condition and is constructed on a small enough mesh $\mathcal{X}$.

Remark 2.8.40. Recall that for τ a tiling in $\frac{1}{n}\mathbb{Z}^3$, the flow $\widetilde{v}_\tau$ is the *non-rescaled* pretiling flow.

Proof. Restricted to any tetrahedron $X \in \mathcal{X}$, λ_n samples a tiling which is periodic with mean current $\widetilde{g}_X$ up an $O(n^{-1})$ error. In particular, there are probabilities $p_1(X), ..., p_6(X)$ such that the probability of seeing a tile of type i in τ restricted to X sampled from λ_n is $p_i(X) + O(n^{-1})$ for $i = 1, ..., 6$. Let $N_1, ..., N_6$ denote the corresponding six types of squares f on S. Let $N_i(X)$ denote the number of each type of square in S restricted to X.

Recall that $\mathcal{D} \subset \mathcal{X}$ is the collection of mesh tetrahedra $X \in \mathcal{X}$ such that $X \cap D \neq \emptyset$. By assumption $\cup_{X \in \mathcal{D}} X \subset \mathrm{Int}(R)$. Since $\widetilde{g}$ inherits the flexible condition, there is a constant $k_D \in [0,1)$ such that $|\widetilde{g}_X|_1 \leq k_D$ for all $X \in \mathcal{D}$. This is related to the probabilities since

$$\widetilde{g}_X = (p_1(X) - p_2(X), p_3(X) - p_4(X), p_5(X) - p_6(X)) \tag{2.65}$$

Since S is monochromatic, the flux of $\widetilde{v}_\tau$ across S is equal to minus the number of tiles in τ which cross S. Thus the expected value of the flux (up to sign) can be split as the sum of indicator functions $\mathbb{1}_f$, where $\mathbb{1}_f(\tau)$ is 1 if f is crossed by a tile in τ and 0 otherwise. Then

$$\mathbb{E}_{\lambda_n}[|\mathrm{flux}(\widetilde{v}_\tau, S \cap D)|] \geq \sum_{X \in \mathcal{D}} \sum_{f \in S \cap X \cap D} \mathbb{E}_{\lambda_n}[\mathbb{1}_f(\tau)] = \sum_{X \in \mathcal{D}} \sum_{i=1}^{6} N_i(X)(p_i(X) + O(n^{-1})).$$

The total number of squares on the surface is $N = \sum_{X \in \mathcal{D}} \sum_{i=1}^{6} N_i(X) = \mathrm{area}_n(S \cap D)$. By Lemma 2.6.26, $N_1+N_2, N_3+N_4, N_5+N_6$ are all equal to $N/3$ up to an error of $O(\mathrm{length}_n(\partial(S \cap D))$. Let

$$p_D = \min_{X \in \mathcal{D}} \max\{\min\{p_1(X), p_2(X)\}, \min\{p_3(X), p_4(X)\}, \min\{p_5(X), p_6(X)\}\}.$$

Since $|\widetilde{g}_X|_1 \leq k_D < 1$, for all $X \in \mathcal{D}$ at least four of $p_1(X), ..., p_6(X)$ must be nonzero, including one from each pair. Combined with Equation (2.65), one can easily check from this that $p_D \geq (1 - k_D)/6 > 0$. On the other hand by the same arguments as in Lemma 2.6.27,

$$\mathbb{E}_{\lambda_n}[|\mathrm{flux}(\widetilde{v}_\tau, S \cap D)|] \geq p_D N/3 + O(n^{-1}N) + O(\mathrm{length}_n(\partial(S \cap D))).$$

Since $|\Theta_D| \leq \text{area}_n(S \cap D) = N$, this completes the proof. The constant $K_D = p_D/3 = (1 - k_D)/18$. As k_D is determined by just g and D, we note that this is independent of the choice of $\widetilde{g}$, as long as $\widetilde{g}$ inherits the flexible condition and is constructed on a small enough mesh $\mathcal{X}$. □

Equipped with these lemmas, we can now give the proof of the generalized patching theorem.

Proof of Theorem 2.8.32. If $A = R_n \setminus R_n^c$ is not tileable, then by Theorem 2.6.3 there exists a counterexample region $U \subset A$ which has only odd cubes along its interior boundary S, but has

$$\text{imbalance}(U) = \text{even}(U) - \text{odd}(U) > 0.$$

By Corollary 2.6.9, we can assume that the interior boundary $S \subset \partial U$ is a minimal monochromatic discrete surface in $\frac{1}{n}\mathbb{Z}^3$.

By Lemma 2.8.36, for any $\beta > 0$ we can find ϵ and inner and outer layers $a_+ \in (\epsilon, 2\epsilon)$ and $a_- \in (c - 2\epsilon, c - \epsilon)$ such that for $a = a_+$ or $a = a_-$,

$$\text{length}_n(\partial(S \cap \partial R_n^a)) \leq (\beta/\epsilon)n \tag{2.66}$$

and further such that there is a ribbon surface γ for $\partial(S \cap \partial R_n^a)$ such that

$$\text{area}_n(\gamma) \leq 2\beta n^2. \tag{2.67}$$

We define the middle annulus $A_{\text{mid}} = R_n^{a_+} \setminus R_n^{a_-}$. Let $U_{\text{mid}} = U \cap A_{\text{mid}}$.

We now fix a compact set $D \subset \text{Int}(R)$. We can assume that D is contained in A_{mid}. We can find a piecewise-constant approximation $\widetilde{g}$ of g on a tetrahedral mesh $\mathcal{X}$ which inherits the flexible condition. We let $\mathcal{D} \subset \mathcal{X}$ be the collection of tetrahedra X such that $X \cap D \neq \emptyset$. We can assume that the mesh scale of $\mathcal{X}$ is small enough so that the union of all $X \in \mathcal{D}$ is still contained in $\text{Int}(R)$.

By analogous pigeonhole principle arguments in the indenting lemmas, we can assume that D has width in squares from $\frac{1}{n}\mathbb{Z}^3$ of at least $cn/2$ and has $\text{length}_n(\partial(S \cap D)) = O(n)$.

Let $N = \text{area}_n(S \cap D)$. By Lemma 2.8.39, we can find a sequence of shining light measures λ_n satisfying for n large enough,

$$\mathbb{E}_{\lambda_n}[|\text{flux}(\widetilde{v}_\tau, S \cap D)|] \geq K_D|\Theta_D| + O(Nn^{-1}) + O(\text{length}_n(\partial(S \cap D)).$$

By Lemma 2.8.35, $|\Theta_D| \geq N = \text{area}_n(S \cap D) \geq (c_2/4)n^2$. Therefore

$$\mathbb{E}_{\lambda_n}[|\text{flux}(\widetilde{v}_\tau, S \cap D)|] \geq \frac{K_D c_2}{4} n^2 + O(n).$$

In particular, we can sample a tiling τ_{SL} from λ_n such that

$$|\text{flux}(\widetilde{v}_{\tau_{\text{SL}}}, S \cap D)| \geq \frac{K_D c_2}{4} n^2 + O(n). \tag{2.68}$$

Let $U_{\tau_{\text{SL}}}$ be the cubes covered by τ_{SL} restricted to U_{mid}. This is a tileable region, so

$$\text{imbalance}(U_{\tau_{\text{SL}}}) = 0.$$

Let $U'_{\text{mid}} \subset U$ be $U_{\tau_{\text{SL}}}$ minus even cubes in $A \setminus U$ with a face on S which are connected to U by τ_{SL}. By Equation (2.68),

$$\text{imbalance}(U'_{\text{mid}}) \leq -\frac{K_D c_2}{4} n^2 + O(n). \tag{2.69}$$

Let $U_{\text{shell}} = U \setminus U'_{\text{mid}}$. It remains to bound the imbalance in U_{shell}.

Consider the set $\alpha = U \cap \partial A_{\text{mid}}$. For each connected component α_i of α, we form a closed surface using a cylinder ribbon surface component γ_i of γ and corresponding patch $\alpha'_i \subset \partial A$. Let α' be the union of the α'_i components.

Let V_i be the region enclosed by α_i, γ_i, and α'_i, and let $V = \cup_i V_i$.

The regions $U_{\text{shell}} \setminus V$ are the *leftover regions*. Let W be a connected component of the leftover region. By construction, ∂W either intersects at most one of ∂R_n or ∂R_n^c. Thus the boundary condition on ∂W comes from only one tiling, either τ_n or σ_n.

Suppose it comes from σ_n, i.e. that $\partial W \cap \partial R_n \neq \emptyset$ (the version where it comes from τ_n is identical, we just make a choice for concreteness). Since σ_n can be extended to a tiling of all of R_n, we can extend σ_n to a tiling covering W. Let W_σ be the region covered by the tiles from σ_n which intersect W. Clearly $\text{imbalance}(W_\sigma) = 0$, and $W_\sigma \cap \partial R_n = W \cap \partial R_n$. If a tile in W_σ crosses ∂W, then either

- It crosses $\partial W \cap S$, in which case W_σ contains an even cube which is not contained in $W \subset U$.
- It crosses $\partial W \cap \gamma$ (recall that γ is the ribbon surface). In this case W_σ could have an odd cube which is not in W. However the number of these added cubes over all components W of $\text{U}_{\text{shell}} \setminus V$ is bounded by $\text{area}_n(\gamma) \leq 4\beta n^2$.

Therefore

$$\text{imbalance}(U_{\text{shell}} \setminus V) \leq 4\beta n^2.$$

We now show that

$$\text{imbalance}(U_{\text{shell}} \cap V) \leq \text{imbalance}(V) + 4\beta n^2.$$

The $4\beta n^2$ term again comes from the ribbon area. We use ideas analogous to those above. Let Y be a component of $V \setminus (U_{\text{shell}} \cap V)$. First note that Y intersects at most one of ∂R_n and ∂R_n^c, so its boundary condition comes from only one tiling.

Assume σ_n is the tiling which defines the boundary condition on Y, and extend it to a tiling which covers Y. Let Y_σ be the region covered by σ_n tiles which intersect Y. Clearly $\text{imbalance}(Y_\sigma) = 0$. If a tile in Y_σ crosses ∂Y, then it is in one of two cases:

- It crosses $\partial Y \cap S$. Since $Y \subset A \setminus U$, in this case Y_σ contains an odd cube which is not in Y. This makes the imbalance of Y larger.
- It crosses $\partial Y \cap \gamma$ (recall γ is the ribbon surface). In this case Y_σ could have an even cube which is not in Y. However the number of these added cubes over all components Y of $V \setminus (\text{U}_{\text{shell}} \cap V)$ is bounded by $\text{area}_n(\gamma) \leq 4\beta n^2$.

Therefore in summary,

$$\text{imbalance}(U_{\text{shell}}) \leq 8\beta n^2 + \text{imbalance}(V).$$

We now relate imbalance to flux. As the proof of Proposition 2.6.6 (where we relate black and white surface area to imbalance), given a set V, we apply the divergence theorem to the reference flow $\widetilde{r}(e) = 1/6$ for all e in $\frac{1}{n}\mathbb{Z}^3$ oriented even to odd to get that

$$\text{imbalance}(V) = \text{flux}(\widetilde{r}, \partial V) = \frac{1}{6}\Big(\text{white}(\partial V) - \text{black}(\partial V)\Big).$$

We apply this to the set $V = \cup_i V_i$. By Equation (2.67), the γ_i contribute at most $4\beta n^2$. Therefore

$$\text{imbalance}(U_{\text{shell}}) \leq \sum_i \text{flux}(\widetilde{r}, \partial V_i) \leq 12\beta n^2 + |\text{flux}(\widetilde{r}, \alpha) - \text{flux}(\widetilde{r}, \alpha')|. \tag{2.70}$$

For the second inequality, we orient α, α' to always both have inward-pointing normal vector (i.e. inward on ∂A and inward on ∂A_{mid}), meaning one has the opposite normal vector as when we compute flux for ∂V_i. This is why we get a minus sign.

The non-rescaled flow $\widetilde{f}_\tau$ is the divergence-free version of the pretiling flow $\widetilde{v}_\tau$; related by the equation $\widetilde{f}_\tau(e) = \widetilde{v}_\tau(e) - \widetilde{r}(e)$ for all edges e oriented even to odd. The rescaled version has $f_\tau = \frac{1}{n^3}\widetilde{f}_\tau$.

Since the boundary condition on α is given by τ_{SL} and the boundary conditions on α' are given by τ_n on the inner boundary and σ_n on the outer boundary, none of the tiles from the corresponding tilings cross α, α' and hence

$$\text{flux}(\widetilde{v}_{\tau_{\text{SL}}}, \alpha) = \text{flux}(\widetilde{v}_*, \alpha') = 0,$$

where $\widetilde{v}_*$ is equal to $\widetilde{v}_{\tau_n}$ on the inner boundary of ∂A and is equal to $\widetilde{v}_{\sigma_n}$ on the outer boundary of ∂A. Therefore

$$\text{imbalance}(U_{\text{shell}}) \leq 12\beta n^2 + |\text{flux}(\widetilde{f}_{\tau_{\text{SL}}}, \alpha) - \text{flux}(\widetilde{f}_*, \alpha')|, \tag{2.71}$$

where $\widetilde{f}_* = \widetilde{f}_{\tau_n}$ on the inner boundary of ∂A and $\widetilde{f}_* = \widetilde{f}_{\sigma_n}$ on the outer boundary of ∂A.

It remains to bound these flux differences, and this is where we use information about the boundary conditions. First note that for any surface X and any tiling τ of $\frac{1}{n}\mathbb{Z}^3$, the flux of the non-rescaled $\widetilde{f}_\tau$ and the rescaled f_τ are related by:

$$\text{flux}(\widetilde{f}_\tau, X) = n^2 \text{flux}(f_\tau, X). \tag{2.72}$$

We have that the *rescaled versions* of the tiling flows f_{τ_n} and $f_{\tau_{\text{SL}}}$ (rescaled, so without the tildes) converge as $n \to \infty$ to the $g \in AF(R, b)$ given in the theorem statement, that is,

$$\lim_{n\to\infty} d_W(f_{\tau_n}, g) = \lim_{n\to\infty} d_W(f_{\tau_{\text{SL}}}, g) = 0.$$

Recall that $T(\cdot, X)$ denotes the trace operator which takes a flow to its restriction to a surface X. By Theorem 2.5.39, for X fixed and any $f_1, f_2 \in AF(R)$, given any $\delta > 0$ there exists δ_1 such that if $d_W(f_1, f_2) < \delta_1$ then $\mathbb{W}_1^{1,1}(T(f_1, X), T(f_2, X)) < \delta$. Recall also that $T(g, \partial R) = b$, and that we are given that $T(f_{\sigma_n}, \partial R) = b_n$ converges to b in $\mathbb{W}_1^{1,1}$. Given these facts, we can choose n large enough to guarantee the following:

- For the outer boundary ∂R,

$$\mathbb{W}_1^{1,1}(T(f_{\sigma_n}, \partial R), b) < \delta \tag{2.73}$$

$$\mathbb{W}_1^{1,1}(T(f_{\tau_{\mathrm{SL}}}, \partial R), b) < \delta. \tag{2.74}$$

- For the inner boundary ∂R^c,

$$\mathbb{W}_1^{1,1}(T(f_{\tau_n}, \partial R^c), T(g, \partial R^c)) < \delta \tag{2.75}$$

$$\mathbb{W}_1^{1,1}(T(f_{\tau_{\mathrm{SL}}}, \partial R^c), T(g, \partial R^c)) < \delta. \tag{2.76}$$

- Finally, let $\partial R_{\mathrm{mid}} = \partial(R^{a+} \setminus R^{a-})$ be the piecewise smooth surface approximated by $\partial A_{\mathrm{mid}} = \partial(R_n^{a+} \setminus R_n^{a-})$. Then

$$\mathbb{W}_1^{1,1}(T(f_{\tau_{\mathrm{SL}}}, \partial R_{\mathrm{mid}}), T(g, \partial R_{\mathrm{mid}})) < \delta. \tag{2.77}$$

Recall also that boundary value flows correspond to measures, and note that $\mathrm{flux}(f, X) = T(f, X)(X)$ is the total mass of the measure $T(f, X)$ on X. In particular, for a surface X and tiling τ of $\frac{1}{n}\mathbb{Z}^3$ and $B \subset X$,

$$\mathrm{flux}(f_\tau, B) = T(f_\tau, X)(B).$$

Lemma 2.5.15 applied to measures μ, ν supported on a surface X says that if $\mathbb{W}_1^{1,1}(\mu, \nu) < \delta$, then for any $B \subset X$,

$$\mathbb{W}_1^{1,1}(\mu \mid_B, \nu \mid_B) \leq \delta + \delta^{1/2}(C(B) + 1),$$

where $\delta^{1/2}C(B)$ is bounded by 2 times the difference of the area of B and the $\delta^{1/2}$ neighborhood of B within X; equivalently, by the area of the annulus of width $\delta^{1/2}$ with inner boundary ∂B (see Remark 2.5.16).

To use this, we relate α, α' which are contained in the discrete surfaces $\partial A_{\mathrm{mid}}$ and ∂A built out of $\frac{1}{n}\mathbb{Z}^3$ lattice squares, to B, B' on the piecewise smooth surfaces $\partial R_{\mathrm{mid}}$ and $\partial R \cup \partial R^c$ respectively.

By Equation (2.66), $\mathrm{length}_n(\partial\alpha) \leq (2\beta/\epsilon)n$. Correspondingly the Euclidean length is bounded as $\mathrm{length}(\partial\alpha) \leq (2\beta/\epsilon)$. Given this, we can cover $\partial\alpha$ with a collection of cubes $\mathcal{C}$ with Euclidean side length 3ϵ, with $|\mathcal{C}| \leq 2\beta/(3\epsilon^2)$. Since the Euclidean width between ∂A and $\partial A_{\mathrm{mid}}$ is less than 2ϵ, $\mathcal{C}$ also covers $\partial\alpha' \subset \partial A$.

The Hausdorff distances between ∂A and $\partial R \cup \partial R^c$ and between $\partial A_{\mathrm{mid}}$ and $\partial R_{\mathrm{mid}}$ are both bounded by $2/n$. There are corresponding sets $B \subset \partial R_{\mathrm{mid}}$ and $B' \subset \partial R \cup \partial R^c$ which differ from α, α' respectively by Hausdorff distance at most $2/n$. Thus for n large enough, $\mathcal{C}$ also covers B, B'. Since $\partial R \cup \partial R^c$ and $\partial R_{\mathrm{mid}}$ are piecewise smooth, there is a constant C' such that the area of either surface restricted to one of the cubes in $\mathcal{C}$ is at most $C'\epsilon^2$. Since $|\mathcal{C}| \leq 2\beta/(3\epsilon^2)$, there is some constant C such that if $\delta^{1/2} \leq \epsilon$, then

$$\delta^{1/2}C(B) \leq C\beta \tag{2.78}$$

$$\delta^{1/2}C(B') \leq C\beta. \tag{2.79}$$

We have that length$(\partial\alpha) \leq 2\beta/\epsilon$, so the number of $\frac{1}{n}\mathbb{Z}^3$ lattice points along $\partial\alpha$ is bounded by a constant times n (the constant here depends on β/ϵ). Since the Hausdorff distance between α and B is bounded by $2/n$, for any tiling τ of $\frac{1}{n}\mathbb{Z}^3$, the flux of f_τ through a surface is proportional to the number of $\frac{1}{n}\mathbb{Z}^3$ lattice points on the surface times $\frac{1}{n^2}$. Thus for any tiling τ, since f_τ is divergence-free,

$$|\text{flux}(f_\tau, \alpha) - \text{flux}(f_\tau, B)| \leq O(n^{-1}). \tag{2.80}$$

By Equation (2.79), since length$(\partial B') \leq C(B')$, also have that length$(\partial B') \leq C\beta/\epsilon$. Since the Hausdorff distance between α' and B' is bounded by $2/n$, the number of $\frac{1}{n}\mathbb{Z}^3$ lattice points on a surface between them is also bounded by a constant times n (the constant here depends on β/ϵ). Thus we analogously get that for any tiling τ of $\frac{1}{n}\mathbb{Z}^3$,

$$|\text{flux}(f_\tau, \alpha') - \text{flux}(f_\tau, B')| \leq O(n^{-1}). \tag{2.81}$$

Therefore by Lemma 2.5.15, for δ such that $\delta^{1/2} < \epsilon$, Equations (2.73), (2.75) to relate f_* to g on B', plus Equation (2.79) where we determine the constant $C(B')$, and finally Equation (2.81) to relate f_* on α' to f_* on B', we get that

$$|\text{flux}(f_*, \alpha') - \text{flux}(g, B')| \leq \delta + \delta^{1/2} + C\beta + O(n^{-1}), \tag{2.82}$$

where f_* is f_{σ_n} on the outer boundary of ∂A and f_{τ_n} on the inner boundary. Similarly, using Equations (2.74), (2.76), and (2.77) to relate $f_{\tau_{\text{SL}}}$ to g, plus Equations (2.80), (2.78) for the constant $C(B)$ and to relate $f_{\tau_{\text{SL}}}$ on α, B, the test tiling τ_{SL} satisfies analogous bounds on both α and α':

$$|\text{flux}(f_{\tau_{\text{SL}}}, \alpha') - \text{flux}(g, B')| \leq \delta + \delta^{1/2} + C\beta + O(n^{-1}) \tag{2.83}$$

$$|\text{flux}(f_{\tau_{\text{SL}}}, \alpha) - \text{flux}(g, B)| \leq \delta + \delta^{1/2} + C\beta + O(n^{-1}). \tag{2.84}$$

Since g is divergence-free and takes values with norm bounded between -1 and 1, and B, B' differ from α, α' by Hausdorff distance bounded by $2/n$, Equation (2.67) implies that

$$|\text{flux}(g, B) - \text{flux}(g, B')| \leq 4\beta + O(n^{-1}). \tag{2.85}$$

Combining Equation (2.72) with the above,

$$|\text{flux}(\widetilde{f}_{\tau_{\text{SL}}}, \alpha) - \text{flux}(\widetilde{f}_*, \alpha')| \tag{2.86}$$

$$\leq \left[|\text{flux}(f_{\tau_{\text{SL}}}, \alpha) - \text{flux}(g, B)| + |\text{flux}(g, B) - \text{flux}(g, B')| + |\text{flux}(g, B') - \text{flux}(f_*, \alpha)|\right] n^2 \tag{2.87}$$

$$\leq (2\delta + 2\delta^{1/2} + 2C\beta + O(n^{-1}) + 4\beta)n^2. \tag{2.88}$$

Combining this with Equation (2.69) and Equation (2.71), we get that

$$\begin{aligned}\text{imbalance}(U) &= \text{imbalance}(U'_{\text{mid}}) + \text{imbalance}(U_{\text{shell}}) \\ &\leq -\frac{K_D c_2}{4} n^2 + (16 + 2C)\beta n^2 + 2\delta n^2 + 2\delta^{1/2} n^2 + O(n).\end{aligned}$$

The factor $K_D c_2/4$ and the constant C are fixed independent of δ, β. Implicit here is also the parameter ϵ that we indent by.

Taking ϵ small enough, we can make β as small as needed. The parameter $\delta > 0$ is related to the distance between tiling flows and their limits and is required to satisfy $\delta^{1/2} < \epsilon$, but this can be guaranteed for n large enough. Therefore for n large enough, imbalance(U) will be non-positive, and hence U is not a counterexample. This completes the proof. □

2.8.7 Upper bounds

To complete the proof of the large deviation principles, we need prove the upper bounds, namely Theorem 2.8.11 for the soft boundary LDP and Theorem 2.8.20 for the hard boundary LDP. We show that the soft boundary upper bound implies the hard boundary one, and then prove the soft boundary one.

Lemma 2.8.41. *Theorem 2.8.11 implies Theorem 2.8.20.*

Proof. Recall that μ_n is counting measure on $TF_n(R, b, \theta_n)$ for some sequence of thresholds $(\theta_n)_{n\geq 1}$ with $\theta_n \to 0$ as $n \to \infty$ sufficiently slowly.

On the other hand $\overline{\mu}_n$ is counting measure on tilings of fixed regions R_n with scale n tileable boundary value b_n such that $b_n \to b$ as $n \to \infty$.

We choose the sequence of thresholds θ_n so that there exists N such that if $n \geq N$ then $\mathbb{W}_1^{1,1}(b_n, b) < \theta_n$. In this case, for any $g \in AF(R, b)$, for n large enough,

$$\overline{\mu}_n(A_\delta(g)) \leq \mu_n(A_\delta(g)).$$

Therefore if Theorem 2.8.11 holds, then for all $g \in AF(R, b)$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} v_n^{-1} \log \overline{\mu}_n(A_\delta(g)) \leq \lim_{\delta\to 0} \limsup_{n\to\infty} v_n^{-1} \log \mu_n(A_\delta(g)) \leq \mathrm{Ent}(g).$$

This completes the proof of Theorem 2.8.20 from Theorem 2.8.11. □

It remains to prove Theorem 2.8.11, namely that for any $g \in AF(R, b)$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} v_n^{-1} \log \mu_n(A_\delta(g)) \leq \mathrm{Ent}(g).$$

The main idea is "coarse graining", i.e. that on a very small box, a uniform random tiling of R looks approximately like a random tiling sampled from a $\mathbb{Z}^3_{\mathrm{even}}$-invariant Gibbs measure of mean current s, where s is the expected mean current on the box.

Proof of Theorem 2.8.11. In this proof, we assume without loss of generality that R is contained in the unit cube $B = [0, 1]^2$ (this is just to avoid complicating the proof with an extra scaling parameter).

Let $\pi_{n,\delta}$ be the uniform probability measure on the set of tilings τ with tiling flow $f_\tau \in A_\delta(g) \cap TF_n(R)$ and satisfying $\mathbb{W}_1^{1,1}(T(f_\tau), b) < \theta_n$. The purpose of this is so that the partition function of $\pi_{n,\delta}$ is $Z_{n,\delta} = \mu_n(A_\delta(g))$.

Tile $\mathbb{R}^3$ by translated copies of B, each with a translated copy of R inside it. Let $\Lambda_n = \frac{1}{n}\mathbb{Z}^3 \cap B$. We define a $\mathbb{Z}^3_{\mathrm{even}}$-invariant measure ν_n on tilings of $\mathbb{Z}^3$ as follows (this

measure can sample tilings with some double tiles and some untiled sites). We take an independent sample from $\pi_{n,\delta}$ on each copy of R, and then average over translations by $x \in \text{even}(\Lambda_n)$. The measure ν_n samples tilings that are perfect matchings on the interior of each copy of R. All sites in each copy of $B \setminus R$ which are not covered by a tile connecting it to a copy of R are empty. Two copies of R might intersect on their boundaries (e.g. in the case $R = B$), in which case ν_n can sample double tiles. However the fraction of possible sites where ν_n samples double tiles is bounded by the fraction of sites in ∂B, namely

$$\frac{6n^2}{n^3} = \frac{6}{n}.$$

We define a subsequential limit

$$\nu := \lim_{j\to\infty} \nu_{n_j}.$$

Note that ν is a $\mathbb{Z}^3_{\text{even}}$-invariant measure on tilings of $\mathbb{Z}^3$, allowed to have untiled sites. It can written as a weighted average of a measure on dimer tilings and the empty ensemble.

Let $\nu_{n,0}$ be defined analogously to ν_n, but without averaging over translations. Let $\nu_{n,x}$ be the version where all tilings are translated by a fixed $x \in \text{even}(\Lambda_n)$. The Shannon entropy of $\pi_{n,\delta}$ is

$$n^{-3}\text{Vol}(R)^{-1} \log Z_{n,\delta} = v_n^{-1} \log \mu_n(A_\delta(g)).$$

By construction,

$$|\Lambda_n|^{-1} H_{\Lambda_n}(\nu_{n,0}) = n^{-3} \log \mu_n(A_\delta(g)).$$

For any other $x \in \text{even}(\Lambda_n)$, a sample on B contains pieces from up to 8 samples of $\pi_{n,\delta}$. Since ν_n is a uniform measure, and since there are more tilings when we are allowed more double tiles,

$$H_{\Lambda_n}(\nu_{n,x}) \geq H_{\Lambda_n}(\nu_{n,0}) \qquad \forall\, x \in \text{even}(\Lambda_n).$$

The specific entropy of ν_n can be computed using any sequence of boxes Δ_M with $|\Delta_M| \to \infty$ as $M \to \infty$. In particular, we can choose $\Delta_M = \Lambda_{Mn}$ so that

$$h(\nu_n) = \lim_{M\to\infty} |\Lambda_{Mn}|^{-1} H_{\Lambda_{Mn}}(\nu_n).$$

On each of the M^3 copies of Λ_n in Λ_{Mn}, ν_n samples an independent draw from $\pi_{n,\delta}$. Thus

$$h(\nu_n) \geq \lim_{M\to\infty} M^3 |\Lambda_{Mn}|^{-1} H_{\Lambda_n}(\nu_{n,0}) = |\Lambda_n|^{-1} H_{\Lambda_n}(\nu_{n,0}) = n^{-3} \log \mu_n(A_\delta(g)).$$

On the other hand, since h is upper-semicontinuous,

$$\limsup_{j\to\infty} h(\nu_{n_j}) \leq h(\nu).$$

Therefore

$$\limsup_{j\to\infty} v_{n_j}^{-1} \mu_{n_j}(A_\delta(g)) \leq \frac{1}{\text{Vol}(R)} h(\nu),$$

and we have reduced the problem to bounding $h(\nu)$. Define φ_n to be the ν_n-expected flow, namely

$$\varphi_n := Z_{n,\delta}^{-1} \sum_{\tau} f_\tau,$$

where the sum is over tilings τ in the support of $\pi_{n,\delta}$. We define a subsequential limit

$$\varphi := \lim_{j\to\infty} \varphi_{n_j}.$$

Up to taking additional subsequences we can assume that the subsequences for φ_n and ν_n are the same. Note that $\varphi \in AF(R)$. Since $A_\delta(g)$ is convex, Theorem 2.5.19 implies that $\varphi \in \overline{A_\delta(g)}$. Therefore

$$\frac{1}{\mathrm{Vol}(R)} \int_R \mathrm{ent}(\varphi(x))\,\mathrm{d}x = \mathrm{Ent}(\varphi) \le \sup_{h\in\overline{A_\delta(g)}} \mathrm{Ent}(h) = \mathrm{Ent}(g) + o_\delta(1).$$

The last equality uses that Ent is upper semi-continuous in the Wasserstein topology (Proposition 2.7.29). This reduces the problem to showing that $\frac{1}{\mathrm{Vol}(R)} h(\nu)$ is bounded by $\mathrm{Ent}(\varphi)$. To this end, we partition B into a collection $\mathcal{C}$ of k^3 smaller cubes of size $1/k^3$. We define a new flow α_k supported in R by, for all $C \in \mathcal{C}$ such that $C \cap R \neq \emptyset$,

$$\alpha_k(x) = \frac{1}{|C \cap R|} \int_{C\cap R} \varphi(y)\,\mathrm{d}y \qquad \forall x \in C \cap R.$$

For $x \notin R$, $\alpha_k(x) = 0$. Since ent is concave (Lemma 2.7.6), by Jensen's inequality,

$$\int_{C\cap R} \mathrm{ent}(\alpha_k(x))\mathrm{d}x \ge \int_{C\cap R} \mathrm{ent}(\varphi(x))\mathrm{d}x.$$

On the other hand, α_k converges to φ a.s. and $|\alpha_k| \le 1$, so α_k converges to φ in L^1, hence by Corollary 2.7.28,

$$\lim_{k\to\infty} \mathrm{Ent}(\alpha_k) = \mathrm{Ent}(\varphi).$$

Therefore it is sufficient to show that $\mathrm{Ent}(\alpha_k)$ is an upper bound for all k. We now define $\nu_{n,C}$ to be ν_n but averaged only over the translations $x \in \mathrm{even}(\Lambda_n \cap C)$ (equivalently, conditioned on the origin being in C). For each $C \in \mathcal{C}$, let ν_C be a subsequential limit of $\nu_{n,C}$. The measures ν_C are $\mathbb{Z}^3_{\mathrm{even}}$-invariant, and we can choose the subsequences so that

$$\nu = k^{-3} \sum_{C\in\mathcal{C}} \nu_C.$$

Therefore

$$h(\nu) = k^{-3} \sum_{C\in\mathcal{C}} h(\nu_C).$$

Note that if $C \cap R = \emptyset$, then ν_C is the empty ensemble and hence in that case $h(\nu_C) = 0$. When $C \cap R \neq \emptyset$, then ν_C splits as the sum of an empty ensemble (corresponding to selecting the origin in $C \setminus C \cap R$) and a measure on dimer tilings (corresponding to selecting the origin in $C \cap R$). In a slight abuse of notation we refer to the mean current of ν_C as the mean current of its component which is a measure on dimer tilings. To bound $h(\nu_C)$ when $C \cap R \neq \emptyset$ we compute this mean current. Recall from Section 2.2.2 that $s_0(\tau)$ is the vector of the tile at the origin in τ, and the mean current of a $\mathbb{Z}^3_{\text{even}}$-invariant measure μ can be computed as

$$s(\mu) = \int_\Omega s_0(\tau)\,\mathrm{d}\mu(\tau) = \mathbb{E}_\mu[s_0(\tau)].$$

We can also compute the mean current by looking at the expected tile direction over a set of points instead of just looking at the origin. Let $E(\Lambda_n)$ denote the edges in Λ_n oriented from even to odd. We can similarly define

$$s_{n,C}(\tau) := |\text{even}(\Lambda_n \cap C \cap R)|^{-1} \sum_{e \in E(\Lambda_n \cap C \cap R)} f_\tau(e)e.$$

Here note that we intersect with R because if the origin is chosen in $C \setminus C \cap R$ we get the empty ensemble, and hence the mean current is not defined. This is the average direction of f_τ over $\Lambda_n \cap C \cap R$, and by $\mathbb{Z}^3_{\text{even}}$-invariance $s(\mu)$ can also be computed

$$s(\mu) = |\text{even}(\Lambda_n \cap C \cap R)|^{-1} \sum_{x \in \text{even}(\Lambda_n \cap C \cap R)} \int_\Omega s_0(\tau + x)\,\mathrm{d}\mu(\tau) = \int_\Omega s_{n,C}(\tau)\,\mathrm{d}\mu(\tau).$$

Using this, we compute that

$$\mathbb{E}_{\nu_{n,C}}[s_{n,C}(\tau)] = Z_{n,\delta}^{-1} \sum_\tau \sum_{e \in E(\frac{1}{n}\mathbb{Z}^3 \cap C \cap R)} |C \cap R|^{-1} f_\tau(e)e = \text{avg}_{C \cap R}(\varphi_n).$$

Since ν_C is a subsequential limit of $\nu_{n,C}$ (up to choice of another subsequence),

$$\begin{aligned} s_C := s(\nu_C) = \int_\Omega s_0(\tau)\,\mathrm{d}\nu_C(\tau) = \lim_{j\to\infty} \int_\Omega s_0(\tau)\,\mathrm{d}\nu_{n_j,C}(\tau) &= \lim_{j\to\infty} \int_\Omega s_{n_j,C}(\tau)\,\mathrm{d}\nu_{n_j,C}(\tau) \\ &= \lim_{j\to\infty} \text{avg}_{C\cap R}(\varphi_{n_j}). \end{aligned}$$

On the other hand, since $\varphi_{n_j} \to \varphi$ in L^1,

$$s_C = \limsup_{j\to\infty} \text{avg}_C(\varphi_{n_j}) = \text{avg}_C(\varphi).$$

Finally we relate $h(\nu_C)$ to $\text{ent}(s_C)$. Recall that $\text{ent}(s) := \max_{\rho \in \mathcal{P}^s} h(\rho)$, where $\mathcal{P}^s$ is the space of $\mathbb{Z}^3_{\text{even}}$-invariant probability measures on dimer tilings of mean current s. The measure ν_C is a sum of an empty ensemble (corresponding to the origin being chosen in $C \setminus R$) which has zero entropy and a $\mathbb{Z}^3_{\text{even}}$-invariant on dimer tilings (corresponding to the origin being chosen in $C \cap R$) which has mean current s_C. Thus

$$h(\nu_C) \leq \text{ent}(s_C).$$

Therefore for all $k > 1$,

$$\limsup_{j\to\infty} v_{n_j}^{-1}\mu_{n_j}(A_\delta(g)) \le \frac{1}{\mathrm{Vol}(R)}h(\nu) = \frac{1}{\mathrm{Vol}(R)}\frac{1}{k^3}\sum_{C\in\mathcal{C}} h(\nu_C)$$
$$\le \frac{1}{\mathrm{Vol}(R)}\frac{1}{k^3}\sum_{C\in\mathcal{C}} \mathrm{Vol}(R\cap C)\mathrm{ent}(s_C) = \mathrm{Ent}(\alpha_k).$$

Taking $k \to \infty$, this shows that

$$\limsup_{n_j\to\infty} v_{n_j}^{-1}\mu_{n_j}(A_\delta(g)) \le \mathrm{Ent}(\varphi) = \mathrm{Ent}(g) + o_\delta(1).$$

Since this holds for any convergent subsequence n_j, taking $\delta \to 0$ completes the proof.

□

2.9 Open problems

We mentioned in the introduction that there is literature exploring the local move connectivity problem, considering moves such as the "flip" and "trit" illustrated below.

Figure 2.38: Flip and trit.

Both the flip and the trit amount to finding a cycle in $\mathbb{Z}^3$ (of length 4 or 6 respectively) that alternates between membership and non-membership in τ, and then swapping the members and non-members. Generally, a *cycle swap* is a swap of an alternating cycle of length k, and a *k-swap* is a cycle swap for which the cycle has length k. It is clear that any two perfect matchings of the same region can be connected by a sequence of such swaps of this form (simply by applying swaps to all of the cycles contained in the union of the two perfect matchings). But it is in general not so clear whether one can get from any matching to any other using only k-swaps for small k.

Problem 2.9.1. Is there a finite K such that for any positive j, m and n (at least one of which is even) it possible to get from any dimer configuration of an $j \times m \times n$ box to any other via sequence of k-swaps with $k \le K$? Is this possible using only flips and trits?

The examples we have presented in Section 2.3 already show the answer to both questions is no if one replaces boxes with general simply connected regions, such as those that can be tiled with alternating slabs of brickwork, each oriented a different direction. If we think in terms of the non-intersecting path interpretation from Section 2.1.5, we can see that the existence of taut patterns like the ones shown there are an obstruction to local move connectedness.

Progress was made on Problem 2.9.1 just after the first draft of this paper was released in [HLT23]. In particular their results show that any dimer tiling of a $j \times m \times n$ box (for jmn even, $j, m, n \geq 2$) admits at least one flip or trit [HLT23, Theorem 1]. See Section 2.3 for further description of their results.

Problem 2.9.2. What can be said about the convergence rate of the mixing algorithm described in Section 2.3.3? Is there a more efficient way to sample random perfect matchings of 3D regions?

Problem 2.9.3. Is there a unique ergodic Gibbs measure corresponding to each mean current in the interior of $\mathcal{O}$?

Problem 2.9.4. If ν_1 and ν_2 are ergodic Gibbs measures of the same mean current, and (τ_1, τ_2) is sampled uniformly from $(\nu_1 \otimes \nu_2)$, are there necessarily infinitely many infinite paths in the union of τ_1 and τ_2?

Problem 2.9.5. What can be said about the *typical fluctuations* of the flow associated to a uniformly random perfect matching of a simple region such as a cube or torus? Do they converge to a natural Gaussian process?

In 2D, Kenyon showed that domino tiling height functions converge in law to the Gaussian free field [Ken00]. This suggests that the discrete gradients of the height functions should converge (at least in some sense) to the gradient of the Gaussian free field. The dual of the discrete gradient (i.e., the discrete flow) should converge in some sense to the dual of the gradient of the Gaussian field—which can be shown to be equivalent to the field obtained by projecting vector-valued white noise orthogonally onto the space of divergence-free fields. It seems reasonable to conjecture that the same holds in any dimension.

Problem 2.9.6. Is it the case for $d \geq 3$ that the discrete divergence-free flows obtained from uniformly random perfect matchings (on a torus or box, say, or in the $\mathbb{Z}^3$ Gibbs measure setting) converge in the fine mesh limit to the Gaussian random generalized flow obtained by projecting vector-valued white noise onto the space of divergence-free flows?

Problem 2.9.7. Does there exist a three-dimensional region $R \subset \mathbb{R}^3$ and a boundary condition b for which the Ent maximizer for (R, b) is not unique? We have shown that such a system would have to be "rigid" in the sense defined in the introduction (i.e., there is an interior point x such that for any neighborhood U of x the set $\overline{g(U)}$ must intersect one of the edges of $\mathcal{O}$). But we have not ruled out the existence of multiple Ent maximizers.

In fact there do exist two dimensional surfaces R where the corresponding Ent maximizer is not unique. Consider the "slanted cylinder" below, where the left and right edges are glued following the numbers in the diagram. Here are two possible tilings of the slanted cylinder.

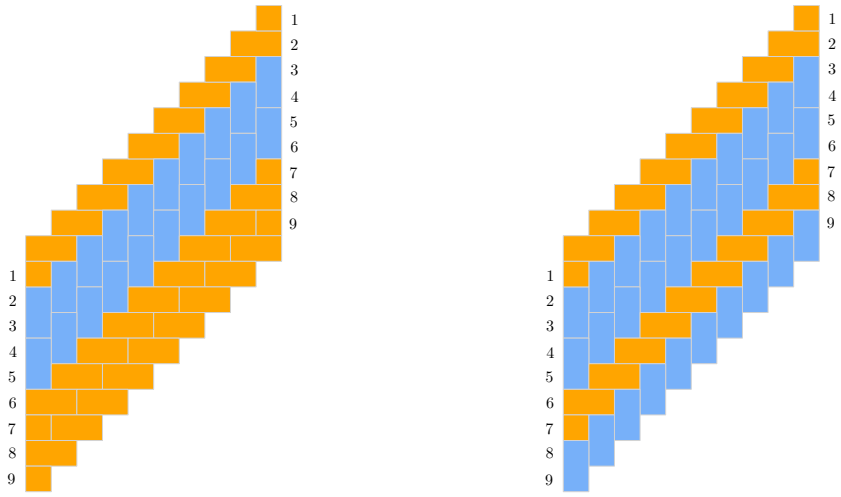


Figure 2.39: Two tilings of the slanted cylinder. The left and right edges are glued.

Any tiling of the slanted cylinder consists of a choice of north (N) or east (E) tile for each diagonal, so if the cylinder has height m then it has 2^m distinct tilings. Since there is only one choice to make on each of the diagonal "stripes" (deciding whether to color it blue or orange) the entropy per site tends to zero as the width of the cylinder tends to infinity, and the functions obtained as fine-mesh limits of these constructions are all maximizers of Ent. A

slanted cylinder can also be realized as an induced subgraph of $\mathbb{Z}^3$ as shown in Figure 2.40. Front and back sides of the surface of a cube, with five "stripes" wrapping around it, whose vertices correspond to the squares in Figure 2.39 and form a slanted cylinder embedded in $\mathbb{Z}^3$. Hall's matching theorem implies that every perfect matching of the set of vertices hit by these stripes is obtained by choosing one of the two possible perfect matchings within each stripe.

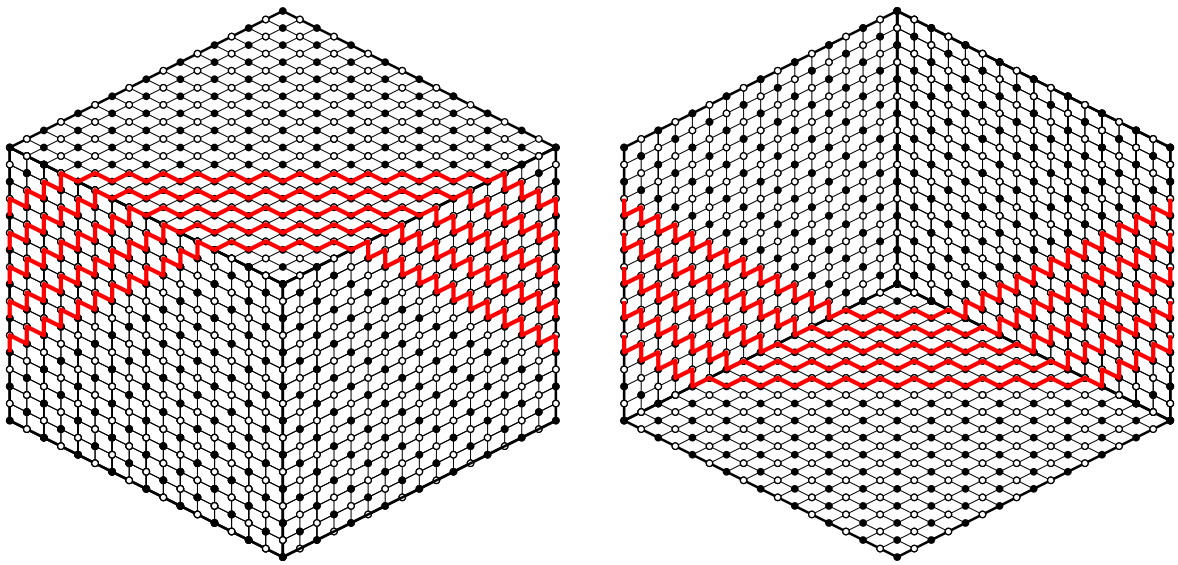


Figure 2.40: A slanted cylinder realized as an induced subgraph of $\mathbb{Z}^3$.

If we try to take a fine mesh limit of this example, we get a region with zero volume in $\mathbb{R}^3$. The question is whether this kind of phenomenon can arise for regions with non-zero volume that are ordinary subsets of $\mathbb{R}^3$ (as opposed to, say, 3D analogs of the slanted cylinder). "Thickening" the ribbon-like example above (by taking the union of multiple ribbon layers, taken on different concentric cubes) does not seem to work, as a tiling obtained that way need not be locally frozen (trit moves may be possible at the corners).

Problem 2.9.8. Is there a region $R \subset \mathbb{R}^3$ and boundary asymptotic flow b where the Ent maximizing flow takes values on a face of $\partial\mathcal{O}$ within a strict subset of the interior of R? Or within a strict subset of all of R? (Does this happen on the boundary of the Aztec octahedron?) For an example of (R, b) where the limit shape takes values in a face of $\partial\mathcal{O}$ on all of R (i.e., not only a strict subset of the region), see Example 2.8.17.

Problem 2.9.9. Given a region $R \subset \mathbb{R}^3$ and a flow b on ∂R, is there an elegant way to describe the conditions under which $AF(R, b)$ is nonempty? In other words, under what conditions does b admit an extension to R which is an asymptotic flow (measurable, divergence-free, and valued in $\mathcal{O}$)? Recall that if $R \subset \mathbb{Z}^3$ is a discrete region and b is a discrete vector field on $\partial R \subset \mathbb{Z}^3$, then Hall's matching theorem or the min cut, max flow principle say that b is extendable if and only if there is no *counterexample region* $U \subset R$ such that $S = \partial U \cap R$ is a type of discrete minimal surface, and any extension of b would be required to have too much flow across S. Is there a continuum version of Hall's matching theorem and the min cut, max flow principle that characterizes when b on ∂R can be extended to an asymptotic flow—i.e., a statement that b is extendable as long as there is no "minimal surface" cut S such that any extension of b would be required to have too much flow across S? See e.g. [Str10] for discussion of related problems.

A particularly simple case of interest is that where R is a polyhedron and the boundary value b is constant on the faces of the polyhedron.

Problem 2.9.10. Let us try to generalize Aztec prism example from the introduction. Suppose $R \subset \mathbb{R}^3$ is a prism of the form $S \times [0, 1]$ (where S is a two-dimensional region) and b is equal to 0 on the top and bottom faces of the prism. Alternatively, one may identify the top and bottom of the prism, to obtain S cross a circle. We expect that one can show from basic symmetry that the Ent minimizing flow g has zero flow in the vertical direction, that its restriction to a slice $S \times \{x\}$ does not depend on x. Understanding the behavior within this slice is then a two-dimensional flow problem. Is this behavior the same as what one would see for the corresponding two-dimensional dimer model on the slice?

Problem 2.9.11. What can be said about the interfaces between frozen regions on the boundaries of limit shapes (such as those apparent in the figures in the introduction)? How large do the fluctuations tend to be?

Problem 2.9.12. The 2D Aztec diamond has four frozen regions (one for each vertex) and the 3D Aztec octahedron appears to have twelve frozen regions (one for each edge). One might guess that in the k-dimensional analog we would see $4\binom{k}{2}$ frozen regions, one for each co-dimension-two boundary simplex. Can anything along these lines be proved, either in 3D or in higher dimensions?

Problem 2.9.13. In two dimensions, the large deviation theory [CKP01] can be generalized to many other types of random height function models [She05], even though for most of these

models we cannot compute ent explicitly. For example, instead of having height differences constrained to $\{3/4, -1/4\}$ as in the 2D dimer model, they could be constrained to some other set, like $\{-1, 1\}$ or $\{-1, 0, 1\}$. That raises a natural question for us. To what other discrete divergence-free flow models in 3D (or in higher dimensions) can the results of this paper be extended? For example, what if instead of restricting the even-to-odd flows to lie in $\{5/6, -1/6\}$ we restrict them to $\{-1, 1\}$ or to some other set? Would the max-flow-min-cut theory available in these settings allow us to complete the steps that relied on Hall's matching theorem in this paper? Could the "chain swapping" arguments used in this paper be adapted to establish the strict concavity of ent in these settings?

As we mentioned earlier, given a lattice flow v on $\mathbb{Z}^3$ one can define a discrete "curl" that assigns to each oriented plaquette—which corresponds to an oriented edge of the dual lattice—the flow of v around that plaquette. One can then a define *vector potential function* A_τ on the dual lattice of $\mathbb{Z}^3$ whose curl corresponds to the flow f_τ on $\mathbb{Z}^3$, though A_τ is *a priori* only determined up to the addition of a vector field with curl zero. Restricting the flow f_τ to take values in $\{5/6, -1/6\}$ is then equivalent restricting the curl of A_τ to lie in $\{5/6, -1/6\}$.

Readers familiar with lattice gauge theory (see [Cha19] for a survey) can tell a similar story about a constrained lattice *connection* with gauge group $U(1)$ (the complex unit circle) as follows. Fix some small constant $\alpha \in (0, \pi)$ and constrain the holonomy around every plaquette (oriented clockwise as one looks from the even to the odd incident cube) to lie in $\{e^{5\alpha i/6}, e^{-\alpha i/6}\}$. Then define a domino to be a pair of cubes separated by a plaquette with holonomy $e^{5\alpha i/6}$. Since the product of oriented holonomies around a single cube is zero, each interior cube belongs to exactly one domino, and (up to boundary conditions) one expects a uniformly random constrained connection to correspond to a uniformly random 3D domino tiling.

Problem 2.9.14. Can our large deviation theory be extended to any other types of holonomy-constrained random connections, Abelian or otherwise? Are there other aspects of gauge theory for which this perspective is useful?

Although we have not explained this in detail, we believe that all the arguments of this paper will still apply to the setting where the edges are periodically "weighted" in the manner described in [KOS06]. For example, one might consider a weighting that strongly favors edges whose vertices have the form (x, y, z) and $(x, y, z+1)$ where z is even. If the weight is strong enough, one can use a standard Peierls argument to show if we are given two independent samples from the minimal-specific-free-energy ergodic Gibbs measure, then there are a.s. no infinite paths in their union.

Problem 2.9.15. If we allow periodic weights, as in [KOS06], what can we say about the phase diagram? Are there some choices of weights for which the double dimer model a.s. contains no infinite paths and others for which it a.s. contains infinitely many infinite path? Are there any other possibilities? Can one say, even on a rough qualitative level, how similar the function ent described here (and its periodically-edge-weighted analogs) will be to the surface tension functions described in [KOS06] (which are interesting algebraic geometry constructions with finitely many singular cusps)? In this generalized setting, can one say anything about the magnitude of the typical fluctuations of a random flow, or how such fluctuations might depend on the edge weights?

Chapter 3

ℓ^2 spaces of circle homeomorphisms in shear coordinates

Except for the final short Section 3.7, this chapter is based on [ŠWW22], which is joint work with Dragomir Šarić and Yilin Wang. Section 3.7 is about random shears, which is not part of the paper [ŠWW22] but uses the same set up.

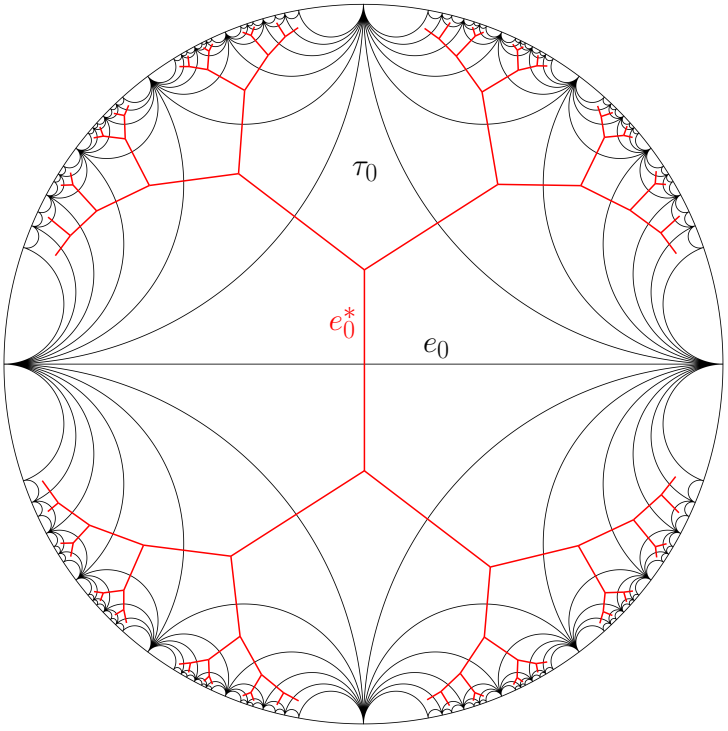


Figure 3.1: Farey tessellation in $\mathbb{D}$ (black) and the dual tree (red) up to generation 5.

3.1 Introduction

The Farey tessellation $\mathfrak{F}$ is an ideal triangulation of the disk $\mathbb{D}$, characterized by the modular invariance property that it is preserved by the action of $\mathrm{PSL}(2,\mathbb{Z})$ on the disk. Its vertices $V = \mathbb{Q}^2 \cap \mathbb{T}$ are the rational points on the circle $\mathbb{T}$, and we use E to denote its edges. The Farey tessellation has various connections to number theory [Hat22], but our motivation comes from Teichmüller theory, where the universal Teichmüller space

$$T(\mathbb{D}) = \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T})$$

can be identified with quasisymmetric circle homeomorphisms fixing three points $-1, \mathfrak{i}, 1$.

Any element h of the larger homogeneous space of orientation preserving circle homeomorphisms

$$\mathrm{Homeo}^+(S^1)/\mathrm{Möb}(\mathbb{T}) \simeq \{h \in \mathrm{Homeo}^+(\mathbb{T}) \ : \ h \text{ fixes } -1, \mathfrak{i}, 1\}$$

can be uniquely encoded by a *shear coordinate* on the edges of the Farey tessellation, namely a function $s : E \to \mathbb{R}$. However, not all functions $s : E \to \mathbb{R}$ encode circle homeomorphisms. Penner posed the question of classifying which functions $s : E \to \mathbb{R}$ encode various smoothness classes of homeomorphisms. In [Šar10; Šar21], the shear coordinates for homeomorphisms, symmetric homeomorphisms, and quasisymmetric homeomorphisms were characterized by the first author.

In this paper we take an opposite perspective. We define two ℓ^2 classes of shear functions, and then study the regularity properties these ℓ^2 conditions imply for circle homeomorphisms and the relation to the existing Hilbert manifold structure on the universal Teichmüller space. For precise descriptions of the Farey tessellation, shears and shear coordinates, see the preliminaries in Section 3.2 or e.g. [Bon09, Chapter 8] by Bonahon.

Naïvely, the first class to consider is the set of square summable shear functions, i.e.

$$\mathscr{S} := \{s : E \to \mathbb{R} : \sum_{e \in E} s(e)^2 < \infty\}.$$

However, we find that not all $s \in \mathscr{S}$ even encode circle homeomorphisms. Conversely, we also find that there are quasisymmetric homeomorphisms which are not in $\mathscr{S}$. See Proposition 3.5.11 for simple examples to illustrate both of these facts. In summary,

$$T(\mathbb{D}) \simeq \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \not\subset \mathscr{S} \qquad \text{and} \qquad \mathscr{S} \not\subset \mathrm{Homeo}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}).$$

These observations show that a basis of shear functions each supported on a single edge is "too large" to define an ℓ^2 space of circle homeomorphisms.

Motivated by this, we investigate shear functions supported on finitely many edges. Finitely supported shear functions always induce homeomorphisms, which in particular are *piecewise Möbius* with pieces bounded by rational points in V (see Lemma 3.3.1). This class of circle homeomorphisms has been studied in, e.g., [MRW22; BJM+23; MP98; FP22; Pen93]. We then show that a homeomorphism h with finitely supported shear s_h is piecewise Möbius *and* C^1 with breakpoints in V if and only if it belongs to a linear subspace of all

shear functions spanned by *diamond shears*. See Lemma 3.3.4 and Proposition 3.3.6. Combinatorially, this condition is equivalent to requiring that the shears on all edges incident to the same vertex sum up to zero (which we call the finite balanced condition).

To define *diamond shears* more precisely in terms of shears, choose an edge $e \in E$, and let $e_1 = (a, b)$, $e_2 = (b, c)$, $e_3 = (c, d)$, $e_4 = (d, a)$ in E be the boundary edges in counterclockwise order of quadrilateral $Q_e = (a, b, c, d)$ consisting of the two triangles from $\mathfrak{F}$ containing $e = (a, c) \in E$. A unit of diamond shear supported at the edge e, i.e. $\vartheta_h(e) = 1$ and $\vartheta_h(e') = 0$ for all $e' \in E$ and $e' \neq e$, is equivalent to four nonzero shears where $s_h(e_1) = s_h(e_3) = 1$ and $s_h(e_2) = s_h(e_4) = -1$. The name "diamond" comes from the picture that the support of one diamond shear corresponds to a quad/diamond of regular shears. See Section 3.3.2 for concrete examples of the correspondence between diamond shear coordinates and circle homeomorphisms.

When s_h has *infinite* support, we define the diamond shear coordinate ϑ_h combinatorially as an infinite sum denoted as $\Psi(s_h)$ whenever s_h is in a certain subclass $\mathscr{P}$ (which can be characterized analytically in terms of differentiability of h). See Section 3.3.3. It is often more convenient to define the diamond shear coordinate on the edges of the dual tree $\mathfrak{F}^*$. As the edges of $\mathfrak{F}$ and $\mathfrak{F}^*$ are in one-to-one correspondence, this identification should not add any ambiguity. See Figure 3.1.

We show that diamond shears can be described analytically:

Proposition 3.1.1 (See Proposition 3.3.25)**.** *Assume that $s_h \in \mathscr{P}_0 \subset \mathscr{P}$ (which implies that h is differentiable at all $v \in V$). The diamond shear coordinate ϑ_h of h is given by*

$$\vartheta_h(e) = \frac{1}{2} \log h'(a) h'(b) - \log \frac{h(a) - h(b)}{a - b} \tag{3.1}$$

for all $e = (a, b) \in E$.

This proposition connects diamond shears to the $\log \Lambda$-*lengths* defined by Penner on decorated Teichmüller space, which is a (trivial) bundle over $T(\mathbb{D})$, with fiber $\mathbb{R}^V_{>0}$ over $h \in T(\mathbb{D})$ corresponding to choosing a horocycle at each $h(v) \in h(V)$. See [Pen93; Pen02; MP98] or the book [Pen12]. Roughly speaking, the decoration allows one to truncate and define the "renormalized hyperbolic length" of an infinite geodesic $h(e) \in h(E)$. A homeomorphism h that is differentiable on V gives a canonical way to *fix* as in [MP98] a decoration on $h(V)$, and $\vartheta_h(e)$ is equal to $-1/2$ times the renormalized length of e.

Corollary 3.1.2 (See Lemma 3.3.30)**.** *If $s_h \in \mathscr{P}_0$, then for any $e = (a, b) \in E$,*

$$\vartheta_h(e) = -\log \Lambda_h(e) = -\frac{1}{2} \operatorname{length}(h(e))$$

where $\operatorname{length}(h(e))$ *is the signed hyperbolic length of the part of $h(e)$ between the horocycles centered at $h(a)$ and $h(b)$ chosen from the fixed decoration.*

Our main object of study in this paper is the set of shear functions with ℓ^2 summable diamond shear coordinates:

$$\mathscr{H} := \{s : E \to \mathbb{R} : \sum_{e \in E} \vartheta(e)^2 < \infty\}.$$

It is relatively straightforward to show that this corresponds to a class of circle homeomorphisms. Given this, we use the abuse of notation $h \in \mathscr{H}$ to mean $s_h \in \mathscr{H}$ throughout.

Proposition 3.1.3 (See Corollary 3.3.21)**.** *If $s \in \mathscr{H}$, then s induces a quasisymmetric circle homeomorphism. In other words, $\mathscr{H} \subset \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \simeq T(\mathbb{D})$.*

Our first main result is to characterize the Hölder classes of circle homeomorphisms that are contained in $\mathscr{H}$. Define for $\alpha \in (0, 1]$,

$$\mathscr{C}^{1,\alpha} := \{h : \mathbb{T} \to \mathbb{T} \text{ homeomorphism}\colon \log h' \text{ is } \alpha\text{-Hölder}\}. \tag{3.2}$$

In particular, the welding homeomorphisms of $C^{1,\alpha}$ Jordan curves belong to $\mathscr{C}^{1,\alpha}$.

Theorem 3.1.4 (See Theorem 3.4.1)**.** *If $\alpha > 1/2$, then $\mathscr{C}^{1,\alpha} \subset \mathscr{H}$.*

This result is sharp as Theorem 3.1.5 will show that $\mathscr{C}^{1,1/2}$ is not in $\mathscr{H}$. The proof of this result relies on Proposition 3.1.1 and the $\ell^{2\alpha}$ summability of the lengths of the shorter arcs in $\mathbb{T}$ between a and b for each $(a, b) \in E$. The ℓ^2 summability was studied and implied by results in [Hal70; Pen02], we improve it to $\ell^{2\alpha}$ summability. See Proposition 3.4.2.

Another main result of our work is an explicit construction of a quasiconformal extension $f : \mathbb{D} \to \mathbb{D}$ of $h \in \mathscr{H}$ inspired by a construction in [KM08] by Kahn and Markovic. The construction is adapted to the cell decomposition of the Farey tessellation, and is one of the places where its discrete structure is essential. This construction crucially uses the *generalized balanced condition* satisfied by shear functions that can be written in terms of diamond shears. While characterizations of shear functions for quasisymmetric homeomorphisms are known, analogous methods for constructing their quasiconformal extensions using the shear function are not known. We further find that if $h \in \mathscr{H}$, then the Beltrami differential $\mu_f = \bar{\partial} f / \partial f$ of the extension f is in $L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$.

This leads to a connection between $\mathscr{H}$ and the *Weil–Petersson class* of circle homeomorphisms $\mathrm{WP}(\mathbb{T})$. The *Weil–Petersson Teichmüller space* $\mathrm{WP}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) =: T_0(\mathbb{D})$ is a subspace of $T(\mathbb{D})$ defined as the completion of $\mathrm{Diff}(\mathbb{T})/\mathrm{Möb}(\mathbb{T})$ under its unique homogeneous Kähler metric (the Weil–Petersson metric) [TT06]. The space of Weil–Petersson homeomorphisms $\mathrm{WP}(\mathbb{T})$ is characterized analytically by Shen [She18], see Definition 3.2.7, and also by, e.g., [Cui00; Bis21; Wan19a; VW20b]. In particular one definition of *Weil–Petersson homeomorphisms* $\mathrm{WP}(\mathbb{T})$ is that they admit *a* quasiconformal extension to the disk whose Beltrami differential $\mu_f = \bar{\partial} f / \partial f$ is in $L^2(\mathbb{D}, \mathrm{d}_{\mathrm{hyp}})$ (see [TT06] or Theorem 3.2.8). Cui [Cui00] showed that the Douady–Earle quasiconformal extension of a Weil–Petersson homeomorphism satisfies this property, and we remark that it is notable that our construction using shears has the desired property for all $h \in \mathscr{H}$.

Ultimately we prove the following relationships between $\mathscr{H}, \mathscr{S}$ and the Weil–Petersson class.

Theorem 3.1.5. *We have $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$. Additionally if $h \in \mathrm{WP}(\mathbb{T})$, then $s_h \in \mathscr{S}$. Both inclusions are strict.*

See Theorem 3.5.5 for the first inclusion, and Section 3.5.3 for why it is strict. See Theorem 3.5.12 for the last inclusion. For comparison, note that this result implies that

Theorem 3.1.4 is sharp, since $\mathscr{C}^{1,1/2} \not\subseteq \mathscr{H}$ as otherwise it would also be in $\mathrm{WP}(\mathbb{T})$ (which contradicts Lemma 3.2.9). As smooth diffeomorphisms are dense in $\mathrm{WP}(\mathbb{T})$, so is $\mathscr{H}$.

In fact, our construction of the quasiconformal extension for functions in $\mathscr{H}$ can be adapted to show the following stronger result that convergence in $\mathscr{H}$ endowed with its ℓ^2 topology implies convergence in the Weil–Petersson metric.

Theorem 3.1.6 (See Corollary 3.5.10)**.** *Suppose that $h, (h_n)_{n\geq 1} \in \mathscr{H}$ with diamond shear coordinates ϑ, ϑ_n respectively. If*

$$\lim_{n\to\infty} \sum_{e\in E} (\vartheta_n(e) - \vartheta(e))^2 = 0,$$

then h_n converges to h in the Weil–Petersson metric.

We obtain immediately the following corollary.

Corollary 3.1.7. *The class of continuously differentiable and piecewise Möbius circle homeomorphisms (with break points in V) is dense in $\mathscr{H}$ and in $\mathrm{WP}(\mathbb{T})$.*

Indeed, this class is equal to the class of circle homeomorphisms with finitely supported diamond shear coordinates (Lemma 3.3.4 and Proposition 3.3.6) which is dense in $\mathscr{H}$ for the Weil–Petersson metric by the above theorem.

Finally, we study infinitesimal shear and diamond shear coordinates on the tangent spaces of $\mathscr{H}$. Since $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$, we compute the Weil–Petersson metric in terms of diamond shears.

Theorem 3.1.8 (See Corollary 3.6.5, Theorem 3.6.8 and Corollary 3.6.10)**.** *Each ℓ^2-summable infinitesimal diamond shear gives rise to a $H^{3/2}$ vector field on $\mathbb{T}$. Let u_1, u_2 be the $H^{3/2}$ vector fields corresponding to the ℓ^2-summable infinitesimal diamond shears $\dot\vartheta_1, \dot\vartheta_2 \in T_h\mathscr{H} \subset T_h\,\mathrm{WP}(\mathbb{T})$. Then*

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = \sum_{e_1\in E}\sum_{e_2\in E} \dot\vartheta_1(e_1)\dot\vartheta_2(e_2)\, g(h(Q_{e_1}), h(e_1), h(Q_{e_2}), h(e_2)),$$

where for $Q = (a_1, a_2, a_3, a_4)$, $e = (a_1, a_3)$, $Q' = (b_1, b_2, b_3, b_4)$, $e' = (b_1, b_3)$,

$$g(Q, e, Q', e') = \frac{2}{\pi}\,\mathrm{Re}\sum_{j,k=1}^{4} \frac{(-1)^{j+k} a_j^2 \bar b_k^2 (a_{j+1} - a_{j-1})(\bar b_{k+1} - \bar b_{k-1})}{(a_{j+1} - a_j)(a_j - a_{j-1})(\bar b_{k+1} - \bar b_k)(\bar b_k - \bar b_{k-1})}\sigma(a_j, b_k),$$

and for $a, b \in \mathbb{T}$,

$$\sigma(a, b) = \sum_{p=0}^{\infty} \frac{(a\bar b)^{p+1}}{(1+p)(2+p)(3+p)}.$$

The expression of the metric tensor is relatively complicated. In contrast, the symplectic form has a very simple expression first noticed by Penner in [Pen93; Pen92]. Using the formula in [Pen93, Thm. 5.5] and the relationship between diamond shears and log Λ-lengths that we describe in Section 3.3.5, we can rewrite the Weil–Petersson symplectic form in terms of a mixture of infinitesimal shears and diamond shears as follows.

Theorem 3.1.9 (See Theorem 3.6.11)**.** *Let ω denote the Weil–Petersson symplectic form on $\mathrm{WP}(\mathbb{T})$ and fix $h \in \mathscr{H}$. Suppose that u_1, u_2 are the $H^{3/2}$ vector fields corresponding to the ℓ^2-summable infinitesimal diamond shears $\dot\vartheta_1, \dot\vartheta_2 \in T_h\mathscr{H} \subset T_h\,\mathrm{WP}(\mathbb{T})$ with infinitesimal shear coordinates $\dot s_1, \dot s_2$ respectively. Then*

$$\omega(u_1, u_2) = \sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e) = -\sum_{e\in E} \dot s_1(e)\dot\vartheta_2(e).$$

We note the resemblance of this formula with the Weil–Petersson symplectic form on the finite dimensional Teichmüller spaces $T_{g,n}$ using the Fenchel-Nielson coordinates due to Wolpert [Wol83]:

$$\omega = -\frac{1}{2}\sum_{\gamma\in\Gamma} \mathrm{d}l \wedge \mathrm{d}\tau,$$

where Γ is a maximal multicurve on a Riemann surface of finite type. Here, one may draw the analogy by interpreting $\dot s$ as the deformation by twisting along closed geodesics corresponding to $\mathrm{d}\tau$, and $\dot\vartheta$ as the deformation by changing the length of geodesics corresponding to $-\frac{1}{2}\mathrm{d}l$ by Corollary 3.1.2.

Outline of the chapter. In Section 3.2, we recall definitions and basic results about the Farey tessellation, shears, and the classes of homeomorphisms of the circle that we consider. In Section 3.3, we relate the Weil–Petersson class to shears in the finite support case, and motivated by this define diamond shears coordinates combinatorially (on a class of shear functions called $\mathscr{P}$), analytically (in terms of h' on V), and in terms of $\log\Lambda$-lengths. We also define the classes $\mathscr{H}, \mathscr{S}$. In Section 3.4, we prove that $\mathscr{C}^{1,\alpha} \subset \mathscr{H}$ (Theorem 3.4.1). In Section 3.5, we prove the theorems relating $\mathrm{WP}(\mathbb{T})$, $\mathscr{H}$, and $\mathscr{S}$. Section 3.6 is devoted to the infinitesimal theory of the Weil–Petersson metric and symplectic form. We define infinitesimal shears and diamond shears and compute the Weil–Petersson metric tensor (Theorem 3.6.8, Corollary 3.6.10) and symplectic form (Theorem 3.6.11) in terms of shears and diamond shears on $\mathscr{H}$.

Finally, in Section 3.7, we show that the diamond shear function $\Theta : E \to \mathbb{R}$ where $(\Theta_e)_{e\in E}$ are i.i.d. standard Gaussians induces a homeomorphism almost surely.

3.2 Preliminaries

3.2.1 Farey tessellation

For $D = \mathbb{D}$ or $\mathbb{H}$, we say that a triangle $\tau \subset D$ is *geodesic* if all its edges are geodesics for the hyperbolic metric on D. All triangles in this discussion will be geodesic. An *ideal triangle* is a geodesic triangle with all its vertices on ∂D. An *ideal triangulation* of D is a (necessarily infinite) locally finite collection of non-overlapping ideal triangles that cover D. The data of an ideal triangulation is encoded in its edges and vertices. For $a \neq b \in \partial D$, we write (a, b) for the hyperbolic geodesic connecting a, b.

The *Farey tessellation* is an ideal triangulation of $\mathbb{D}$ with many natural symmetries and properties. Since the Farey tessellation will be ubiquitous throughout this paper, we denote

it just by $\mathfrak{F}$, its set of edges by E, and its set of vertices by V. Unless otherwise specified, the edges in E are unoriented and denoted $e = (a, b)$ where $a, b \in V$ are the endpoints of e.

Let τ_0 be the ideal triangle with vertices $(1, \mathfrak{i}, -1)$. From τ_0, all the other triangles in $\mathfrak{F}$ are images of τ_0 by reflections over edges which are hyperbolic (orientation reversing) isometries. The dual tree to the Farey tessellation, which we denote $\mathfrak{F}^*$, will also play a central role in our discussions. Every triangle $\tau \in \mathfrak{F}$ corresponds to a vertex τ^* in $\mathfrak{F}^*$, and every vertex in $\mathfrak{F}$ corresponds to a face v^* in $\mathbb{D} \smallsetminus \mathfrak{F}^*$. Every edge $e \in E$ corresponds to the dual edge $e^* \in E^*$. We call the edge e_0^* dual to $(-1, 1)$ the *root edge* of $\mathfrak{F}^*$.

The dual edges are sorted into *generations* based on their graph distance to the e_0^*. We write $E_n^* \subset E^*$ for the set of dual edges within distance n of e_0^*. From this we extend the definition of generations to the vertices, edges, and faces of $\mathfrak{F}$ and $\mathfrak{F}^*$. We define $E_n \subset E$ to be the collection of edges dual to E_n^*, $V_n \subset V$ to be the vertices which are endpoints of edges in E_n, and T_n^* to be the vertices of $\mathfrak{F}^*$ which are vertices of E_n^*. We then extend the definition of generation via duality to faces of $\mathfrak{F}$ and $\mathfrak{F}^*$. We say that a vertex, an edge or a face in $\mathfrak{F}$ or in $\mathfrak{F}^*$ *has generation* n if it belongs to the corresponding set of index n but not $n-1$, and we write $\mathrm{gen}(\cdot)$ for the generation function. It is easy to see:

Lemma 3.2.1. *If an edge $e = (a, b) \neq e_0$, then $\mathrm{gen}(a) \neq \mathrm{gen}(b)$.*

The Farey tessellation is nicely represented in the upper half plane $\mathbb{H}$. We choose the following Cayley map $\mathfrak{c}$ which maps $\mathbb{D}$ conformally onto $\mathbb{H}$:

$$\mathfrak{c} : z \mapsto -\mathfrak{i}\frac{z+1}{z-1}, \quad \text{which maps } -1 \mapsto 0,\ 1 \mapsto \infty,\ \mathfrak{i} \mapsto -1.$$

Under this identification, V is sent to $\mathbb{Q} \cup \{\infty\}$, therefore, $V = \mathbb{Q}^2 \cap \mathbb{T}$. In fact, the modular group

$$\mathrm{PSL}(2, \mathbb{Z}) = \{A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \colon a, b, c, d \in \mathbb{Z},\ ad - bc = 1\}_{/A \sim -A}$$

acts on $\mathbb{H}$ by fraction linear transformations

$$z \mapsto \frac{az+b}{cz+d} =: A(z). \tag{3.3}$$

The image of the Farey tessellation under $\mathfrak{c}$ contains the ideal triangle $\mathfrak{c}(\tau_0)$ of vertices $\{-1, 0, \infty\}$ and is preserved by the action of $\mathrm{PSL}(2, \mathbb{Z})$ on $\mathbb{H}$ which is generated by maps $z \mapsto z+1$ and $z \mapsto -1/z$. It is then not hard to see that $\mathrm{PSL}(2, \mathbb{Z})$ acts transitively on $\mathfrak{c}(E)$ (faithfully on oriented edges) and $\mathfrak{c}(V) = \mathbb{Q} \cup \{\infty\}$.

Each element in $\mathbb{Q} \smallsetminus \{0\}$ will be written in the form $\frac{p}{q}$, where $q \in \mathbb{Z}_{\geq 1}$, $p \in \mathbb{Z}$, and p, q are co-prime. We use the convention $0 = \frac{0}{1}$ and $\infty = \frac{1}{0} = \frac{-1}{0}$. We recall the following basic fact about the Farey tessellation in $\mathbb{H}$. For readers' convenience we also include the elementary proof of this classical result.

We say that $c \in V$ is a *child* of $(a, b) \in E$, if c has one generation larger than (a, b) and (a, b, c) is a triangle in $\mathfrak{F}$. Apart from $(-1, 1)$ which has two children $\{\mathfrak{i}, -\mathfrak{i}\}$, all other edges have only one child.

Lemma 3.2.2. *An edge $(\frac{p}{q}, \frac{r}{s}) \in \mathfrak{c}(E)$, if and only if $|ps - rq| = 1$. Moreover, $p + r$ and $q + s$ are co-prime and the child of $(\frac{p}{q}, \frac{r}{s})$ is $\frac{p+r}{q+s}$. Here, we choose the convention $\infty = \frac{1}{0}$ if $p \geq 0, q = 1$, and $\infty = \frac{-1}{0}$ if $p \leq 0, q = 1$ (for $p = 0$, we use both conventions).*

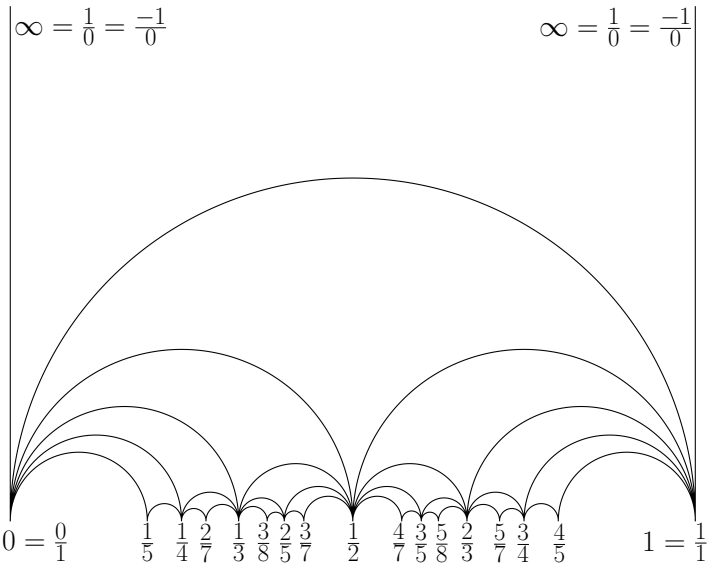


Figure 3.2: Farey tessellation in $\mathbb{H}$ between 0 and 1 up to generation 4 with vertices labeled.

Proof. Assume that $(\frac{p}{q}, \frac{r}{s}) \in \mathfrak{c}(E)$. Since $\mathrm{PSL}(2,\mathbb{Z})$ acts transitively on $\mathfrak{c}(E)$, there exists an element $A \in \mathrm{PSL}(2,\mathbb{Z})$, such that $A(\infty) = \frac{p}{q}$ and $A(0) = \frac{r}{s}$. Hence $A = \left(\begin{smallmatrix} \alpha p & \beta r \\ \alpha q & \beta s \end{smallmatrix}\right)$ for some $\alpha, \beta \in \mathbb{Z}$. Since $\det A = 1 = (ps - rq)\alpha\beta$ and $p, r, q, s \in \mathbb{Z}$, we have $\alpha, \beta \in \{1, -1\}$ and $|ps - rq| = 1$. Conversely, if $|ps - rq| = 1$, we let $A = \left(\begin{smallmatrix} \alpha p & r \\ \alpha q & s \end{smallmatrix}\right)$ with $\alpha \in \{1, -1\}$ such that $A \in \mathrm{PSL}(2,\mathbb{Z})$. Then $(\frac{p}{q}, \frac{r}{s})$ is the image of $(0, \infty)$ under the fractional linear transformation A. Therefore $(\frac{p}{q}, \frac{r}{s}) \in \mathfrak{c}(E)$.

Now we compute the child of $(\frac{p}{q}, \frac{r}{s})$. We treat first the case where s or $q = 0$. By symmetry, we assume that $s = 0$ and $q \neq 0$. Using our convention, this happens only if $p \geq 0$, $q = 1$, $r = 1$ (or resp. $p \leq 0$, $q = 1$, $r = -1$). The child of $(\infty = \frac{1}{0}, p)$ is $p + 1$ and child of $(\infty = \frac{-1}{0}, p)$ is $p - 1$ as claimed.

Now we consider the case where $s, q \neq 0$. By symmetry, we assume that $ps - rq = 1$. Note that

$$\frac{p}{q} = \frac{1}{sq} + \frac{r}{s},$$

$s, q \geq 0$ implies that $\frac{r}{s} < \frac{p}{q}$.

The matrix $A' = \left(\begin{smallmatrix} p+r & r \\ q+s & s \end{smallmatrix}\right) \in \mathrm{PSL}(2,\mathbb{Z})$. Therefore $p + r$ and $q + s$ are co-prime. The previous result shows that $\frac{p+r}{q+s}$ is adjacent to $\frac{r}{s}$. Consider similarly the matrix $A'' = \left(\begin{smallmatrix} p & p+r \\ q & q+s \end{smallmatrix}\right)$, we obtain that $\frac{p+r}{q+s}$ is adjacent to $\frac{p}{q}$. Moreover, we have the inequalities

$$\frac{r}{s} < \frac{p+r}{q+s} < \frac{p}{q},$$

which shows that $\frac{p+r}{q+s}$ is the child of the edge $(\frac{p}{q}, \frac{r}{s})$. □

3.2.2 Shear along an edge

Let e be a hyperbolic geodesic in the disk connecting $a, c \in \mathbb{T}$. A quad Q around e is an ideal quadrilateral in $\mathbb{D}$ with vertices $a, b, c, d \in \mathbb{T}$ in counterclockwise order, for some $b, d \in \mathbb{T}$. Recall the *cross ratio* of a, b, c, d is

$$\mathrm{cr}(a, b, c, d) = \frac{(b-a)(d-c)}{(c-b)(d-a)}.$$

Definition 3.2.3. The *shear* of $Q = (a, b, c, d)$ along $e = (a, c)$ is $s(Q, e) := \log \mathrm{cr}(a, b, c, d)$.

The shear of Q along a diagonal e does not depend on the orientation of e since $\mathrm{cr}(a, b, c, d) = \mathrm{cr}(c, d, a, b)$. The cross ratio is also invariant under Möbius transformations, and hence so is the shear of a quad around an edge. We can use the Cayley transform $\mathfrak{c}$ to easily compute the shear for quads around $e_0 = (-1, 1)$.

Example 3.2.4. Consider a quad around the edge $e_0 = (-1, 1)$ of the form $Q = \{1, \mathfrak{i}, -1, x_s\}$. Under the Cayley transform $\mathfrak{c} : (1, \mathfrak{i}, -1, x_s) \mapsto (\infty, -1, 0, \mathfrak{c}(x_s))$. Since the cross ratio is preserved by Möbius transformations, we get that

$$s(Q, e_0) = \log \mathrm{cr}(\infty, -1, 0, \mathfrak{c}(x_s)) = \log \mathfrak{c}(x_s).$$

While $s(Q, e)$ does not depend on the orientation of e, orienting e is useful for the following geometric interpretation of the shear (which also explains the name). The quad Q can be thought of as two triangles glued along e. Choosing an orientation of e, we call the triangle on the left of e (when e is pointing up) with respect to the orientation τ_L and the other τ_R. Geometrically, the shear of Q along e measures how τ_L, τ_R are glued together along e to construct Q.

Lemma 3.2.5. *Let $\vec{e} = (c, a)$ be oriented from c to a so that $b \in \tau_L$ and $d \in \tau_R$. For any point $x \in \mathbb{T}$, define $m_x(e)$ to be the intersection between e and the hyperbolic geodesic through x perpendicular to e. Then*

$$s(Q, e) = \pm d_{hyp,\mathbb{D}}(m_b(e), m_d(e))$$

The sign is positive if m_b is before m_d along $\vec{e} = (c, a)$ and negative otherwise. See Figure 3.3.

Proof. Let $\vec{e} = (c, a)$ and $\vec{e}_0 = (-1, 1)$ as oriented edges. Let A be a Möbius transformation that sends a, b, c to $1, \mathfrak{i}, -1$ respectively. Under this map, $\vec{e}$ is sent to $\vec{e}_0$, the quad $Q = (a, b, c, d)$ around $\vec{e}$ is sent to the quad $Q' = (1, \mathfrak{i}, -1, A(d) =: x_s)$ around $\vec{e}_0$, and the intersection points $m_b(e), m_d(e)$ are sent to the intersection points $m_{\mathfrak{i}}(e_0), m_{x_s}(e_0)$ respectively. Since shears are preserved under Möbius transformations and using the calculation in Example 3.2.4,

$$s(Q, e) = s(Q', e_0) = \log \mathfrak{c}(x_s).$$

Since hyperbolic lengths are also preserved by Möbius transformations,

$$d_{\mathrm{hyp},\mathbb{D}}(m_b(e), m_d(e)) = d_{\mathrm{hyp},\mathbb{D}}(m_{\mathfrak{i}}(e_0), m_{x_s}(e_0)).$$

We can use the Cayley transform again to compute this distance. Under $\mathfrak{c}$, $Q' = \{1, \mathfrak{i}, -1, x_s\} \mapsto \{\infty, -1, 0, \mathfrak{c}(x_s)\}$, $m_{\mathfrak{i}}(e_0) \mapsto \mathfrak{i}$, and $m_{x_s}(e_0) \mapsto \mathfrak{c}(x_s)\mathfrak{i}$. Hence

$$d_{\mathrm{hyp},\mathbb{D}}(m_{\mathfrak{i}}(e_0), m_{x_s}(e_0)) = d_{\mathrm{hyp},\mathbb{H}}(\mathfrak{i}, \mathfrak{c}(x_s)\mathfrak{i}) = \left| \int_1^{\mathfrak{c}(x_s)} \frac{1}{y}\, \mathrm{d}y \right| = |\log \mathfrak{c}(x_s)|.$$

As for the sign, under the composition $\mathfrak{c} \circ A$, $\vec{e}$ is sent as an oriented edge to $(0, \infty)$. Hence $m_b(e)$ is before $m_d(e)$ along $\vec{e}$ if and only if $1 < \mathfrak{c}(x_s)$. □

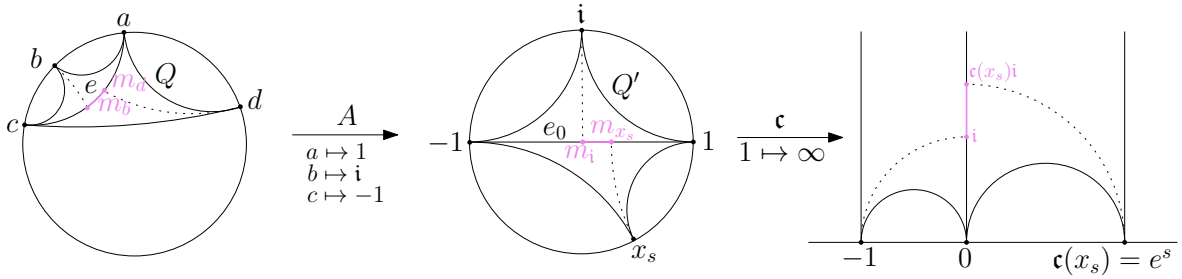


Figure 3.3: The quad $Q = \{a, b, c, d\}$ around $e = (c, a)$ with intersection points drawn, sent to $Q' = \{1, \mathfrak{i}, -1, x_s\}$ around $e_0 = (-1, 1)$ by A, and then the image under the Cayley transform $\mathfrak{c}$.

The pair of triangles around an edge e in a tessellation forms a quad around e. We define the *Farey quad*, denoted Q_e, to be the quad around e in $\mathfrak{F}$. The Farey tessellation is characterized by

$$s(Q_e, e) = 0, \qquad \text{for all } e \in E.$$

This follows from the construction of the Farey tessellation via reflection.

3.2.3 Shear coordinates and classes of circle homeomorphisms

We now introduce shear coordinates for orientation-preserving homeomorphisms of the circle $\mathbb{T}$. As one would often identify two homeomorphisms up to post-compositions by Möbius transformations, namely by the group $\mathrm{Möb}(\mathbb{T}) \simeq \mathrm{PSU}(1,1) \simeq \mathrm{PSL}(2, \mathbb{R})$, e.g., in the context of Teichmüller theory, we assume throughout the paper that all circle homeomorphisms fix $\{-1, \mathfrak{i}, 1\}$ unless otherwise specified.

If $h \colon \mathbb{T} \to \mathbb{T}$ is a circle homeomorphism, h induces a map $V \to V_h = h(V)$ by sending $v \mapsto h(v) \in \mathbb{T}$. For each edge $e \in E$ with end points v_1 and v_2, we write $h(e)$ for the hyperbolic geodesic connecting $h(v_1)$ and $h(v_2)$. The image of E under h forms a new tessellation $h(\mathfrak{F})$ of the unit disk where the ideal triangle τ_0 is fixed. We define similarly $h(Q_e)$ to be the quad around $h(e)$ in $h(\mathfrak{F})$.

Definition 3.2.6. The *shear coordinate* of h is the map $s_h \colon E \to \mathbb{R}$ such that $s_h(e)$ is the shear of the edge $h(e)$ in the quad $h(Q_e)$. Namely, $s_h(e) = s(h(Q_e), h(e))$.

Notice that s_h is determined by $h|_V$. However, not all functions $s : E \to \mathbb{R}$ arise as a shear coordinate for some circle homeomorphism. In fact, from s we recover a map $h : V \to \mathbb{T}$ which is strictly increasing and fixes $1, \mathbf{i}, -1$ such that $s_h = s$. However, $h(V)$ does not have to be dense in $\mathbb{T}$ and cannot always be extended continuously to a homeomorphism. In [Šar10] the first author characterized the class of shear functions which arise from a circle homeomorphism, as well as those from quasisymmetric and symmetric homeomorphisms. See also [Šar21]. In the present article, we are particularly interested in the following classes of circle homeomorphisms.

Definition 3.2.7 (See [TT06; She18])**.** A circle homeomorphism h is called *Weil–Petersson* if h is absolutely continuous (with respect to arclength measure) and $\log h'$ belongs to the Sobolev space $H^{1/2}$. In other words,

$$\iint_{\mathbb{T}\times\mathbb{T}} \left| \frac{\log h'(\zeta) - \log h'(\xi)}{\zeta - \xi} \right|^2 \mathrm{d}\zeta \mathrm{d}\xi < \infty. \tag{3.4}$$

We write $\mathrm{WP}(\mathbb{T})$ for the class of all Weil–Petersson homeomorphisms. The Weil–Petersson class has been studied extensively since the 80s because of its rich geometric structure and links to string theory [BR87; Wit88; NV90; Pek95], Teichmüller theory [Cui00; Guo00; TT06; She18; ST20; STW18], computer vision [SM06], periodic KdV equations [STZ99], and more recently, the discovery of links to SLE [Wan19a; Wan21; VW20a; VW20b], hyperbolic geometry [Bis19], Coulomb gases [Joh21; WZ22], etc. See, e.g., [Bis19] by Bishop for a survey as well as a number of new characterizations of the Weil–Petersson class.

Every quasisymmetric circle homeomorphism admits a quasiconformal extension $\mathbb{D} \to \mathbb{D}$, see, e.g, [Leh87]. For a quasisymmetric homeomorphism to be Weil–Petersson, it has to satisfy the following equivalent L^2 condition.

Theorem 3.2.8 (See [TT06; She18])**.** *A circle homeomorphism h is Weil–Petersson if and only if there is a quasiconformal extension $f : \mathbb{D} \to \mathbb{D}$ of h, such that the Beltrami coefficient $\mu = \overline{\partial} f / \partial f$ satisfies*

$$\|\mu\|_2^2 = \int_{\mathbb{D}} \frac{4|\mu(z)|^2}{(1-|z|^2)^2} \mathrm{d}A(z) < \infty,$$

where $\mathrm{d}A$ is the Euclidean area measure.

For $\alpha \in (0, 1]$, we let $\mathscr{C}^{1,\alpha}$ denote the class of circle homeomorphisms h such that $\log h'$ is α-Hölder continuous. Or equivalently, in terms of the Hölder classes $C^{1,\alpha}$,

$$\mathscr{C}^{1,\alpha} = \{h \in C^{1,\alpha} : h \text{ is a circle homeomorphism and } \inf_{\mathbb{T}} |h'| > 0\}. \tag{3.5}$$

In particular, it follows from the Kellogg theorem [GM05, Thm. II.4.3] that the welding homeomorphisms of $C^{1,\alpha}$ Jordan curves belong to $\mathscr{C}^{1,\alpha}$. It is easy to see from (3.4) the following lemma.

Lemma 3.2.9. *We have $\mathscr{C}^{1,\alpha} \subset \mathrm{WP}(\mathbb{T})$ for all $\alpha > 1/2$ and $\mathscr{C}^{1,1/2} \not\subseteq \mathrm{WP}(\mathbb{T})$.*

3.3 Diamond shear

3.3.1 Circle homeomorphisms with finite shear

We will introduce a new shear coordinate system (diamond shears), essential to describe the class $\mathscr{H}$ of circle homeomorphisms at the center of this work (see Definition 3.3.15). To motivate the definition of diamond shears, let us first consider circle homeomorphisms with finitely many nonzero shears (i.e. s_h has finite support).

We write

$$\mathrm{PSU}(1,1) = \left\{ \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C},\ |\alpha|^2 - |\beta|^2 = 1 \right\}_{/A \sim -A}$$

for the group of Möbius transformations preserving $\mathbb{T}$. This group is conjugate to $\mathrm{PSL}(2,\mathbb{R})$ via the Cayley transform $\mathfrak{c}$. For two distinct points $z, w \in \mathbb{T}$, we denote by $I(z,w) \subset \mathbb{T}$ the closed circular arc going counterclockwise from z to w.

Lemma 3.3.1. *A circle homeomorphism h has finitely many nonzero shears if and only if h is piecewise Möbius with rational breakpoints. Namely, there exist $k \geq 2$ distinct points $v_1, \dots, v_k \in V$ in counterclockwise order such that $h|_{I(v_i, v_{i+1})} \in \mathrm{PSU}(1,1)$, where $v_{k+1} = v_1$.*

Proof. Let h be a piecewise Möbius homeomorphism with break points $v_1, \dots, v_k \in V$. By possibly subdividing further, we assume that $(v_i, v_{i+1}) \in E$. We write $\mathscr{E}_i = \{(a,b) \in E : a, b \in I(v_i, v_{i+1})\}$. If $(a,b) \in \mathscr{E}_i$ and $(a,b) \neq (v_i, v_{i+1})$, then the four vertices of the Farey quad $Q_{(a,b)}$ are all in $I(v_i, v_{i+1})$. Since $h|_{I(v_i,v_{i+1})}$ is Möbius, which preserves the cross-ratio, $s_h(a,b) = 0$. We obtain that s_h has finite support since $E \smallsetminus \bigcup_{i=1}^k \mathscr{E}_i$ is finite.

Conversely, if s_h has finite support, then $\mathbb{T}$ can be partitioned to $\bigcup_{i=1}^k I(v_i, v_{i+1})$ such that $s_h|_{\mathscr{E}_i} \equiv 0$ for all $i = 1, \cdots, k$. By possibly further subdividing the intervals we may assume that $(v_i, v_{i+1}) \in E$ (and therefore in $\mathscr{E}_i$). Let $x_i \in I(v_i, v_{i+1})$ be the unique vertex which is adjacent to v_i and v_{i+1} and $\tilde{h}_i$ the unique Möbius map in $\mathrm{PSU}(1,1)$ which maps respectively v_i, x_i, v_{i+1} to $h(v_i), h(x_i), h(v_{i+1})$. The image of $V \cap I(v_i, v_{i+1})$ by h is determined by $s_h|_{\mathscr{E}_i}$ and the image of one triangle. Therefore $h = \tilde{h}_i$ on $V \cap I(v_i, v_{i+1})$. As $\tilde{h}_i$ is continuous and V is dense in $\mathbb{T}$, we obtain that $h = \tilde{h}_i$ on $I(v_i, v_{i+1})$. □

Remark 3.3.2. Notice that in the proof of the converse direction, we showed that the finitely supported shear coordinate defines a map $h : V \to \mathbb{R}$ which extends continuously to a piecewise Möbius homeomorphism (and we do not need to assume that h is a homeomorphism to start).

To state the next result, we organize the edges of $\mathfrak{F}$ into *fans*. We define

$$\mathrm{fan}(v) = \{e \in E : e \text{ is incident to } v\}.$$

We will index the edges in $\mathrm{fan}(v)$ as $(e_n)_{n \in \mathbb{Z}}$ in a way that n increases in the counterclockwise manner and e_0 is an arbitrary choice of edge in $\mathrm{fan}(v)$. The order is chosen such that after mapping $\mathbb{D}$ conformally onto $\mathbb{H}$ and v is sent to ∞, the image of $(e_n)_{n \in \mathbb{Z}}$ are equally spaced vertical lines with index increasing from left to right.

We write $v+$ (resp. $v-$) for a point on $\mathbb{T}$ approaching v infinitesimally counterclockwise (resp. clockwise), which corresponds to $x \to \infty$ (resp. $x \to -\infty$) in the upper half-plane

model. For instance, $f(1+)$ means the limit of $f(z)$ as $z \in \mathbb{T}$ approaches 1 from below, if it exists.

Definition 3.3.3. We say that $s : E \to \mathbb{R}$ satisfies the *finite balanced condition* if for all $v \in V$, $\{n \in \mathbb{Z} \colon s(e_n) \neq 0\}$ is finite and $\sum_{n\in\mathbb{Z}} s(e_n) = 0$, where $\operatorname{fan}(v) = (e_n)_{n\in\mathbb{Z}}$.

Lemma 3.3.4. *If a circle homeomorphism h has finitely many nonzero shears, the following are equivalent:*

i) h is Weil–Petersson;

ii) h is $\mathscr{C}^{1,1}$ with rational breakpoints;

iii) $s = s_h$ satisfies the finite balanced condition.

Proof. Since h has finitely many nonzero shears, Lemma 3.3.1 shows that h is piecewise Möbius with rational breakpoints that we denote by $v_1, \dots, v_k \in V$. One and only one of the following is true:

- For all $i = 1, \dots, k$, $h'(v_i+) = h'(v_i-)$. In this case, $h \in \mathscr{C}^{1,1}$. Lemma 3.2.9 shows that $h \in \mathrm{WP}(\mathbb{T})$.
- There exists i such that $h'(v_i+) \neq h'(v_i-)$. In this case, $\log h'(v_i+) \neq \log h'(v_i-)$. We see from (3.4) that $h \notin \mathrm{WP}(\mathbb{T})$.

This proves the equivalence between i) and ii).

Now we show that $h'(v_i+) = h'(v_i-)$ is equivalent to $\sum_{n\in\mathbb{Z}} s(e_n) = 0$, where $(e_n)_{n\in\mathbb{Z}} = \operatorname{fan}(v_i)$. We define $\varphi : \mathbb{R} \to \mathbb{R}$ to be the homeomorphism $\varphi = \mathfrak{c}_1 \circ h \circ \mathfrak{c}_2^{-1}$, where $\mathfrak{c}_1, \mathfrak{c}_2$ are two Möbius transformations $\mathbb{D} \to \mathbb{H}$ such that $\mathfrak{c}_1(h(v_i)) = \infty$, $\mathfrak{c}_2(v_i) = \infty$, $\mathfrak{c}_2(e_0) = (0, \infty)$, and $\mathfrak{c}_2(e_1) = (1, \infty)$. Given this, φ fixes ∞, and $\mathfrak{c}_2(e_n) = (n, \infty)$ for all $n \in \mathbb{Z}$.

From Lemma 3.2.5, we know that

$$\frac{\varphi(n+1) - \varphi(n)}{\varphi(n) - \varphi(n-1)} = \exp(s(e_n)). \tag{3.6}$$

Since s has finite support, there exists $n_0 \geq 0$ such that $s(e_n) = 0$ if $|n| \geq n_0$. Therefore, there exists $\ell, \ell' > 0$ such that for all $n, m \geq n_0$, $\varphi(m+1) - \varphi(m) = \ell$, $\varphi(-n) - \varphi(-n-1) = \ell'$, and

$$\frac{\varphi(m+1) - \varphi(m)}{\varphi(-n) - \varphi(-n-1)} = \frac{\ell}{\ell'} = \exp\left(\sum_{n\in\mathbb{Z}} s(e_n)\right).$$

We have

$$\begin{aligned}\varphi(m+1) - \varphi(m) &= \int_m^{m+1} |\varphi'(x)| \, \mathrm{d}x \\ &= \int_m^{m+1} \left|\mathfrak{c}_1'(h \circ \mathfrak{c}_2^{-1}(x))\right| \left|h'(\mathfrak{c}_2^{-1}(x))\right| \left|(\mathfrak{c}_2^{-1})'(x)\right| \, \mathrm{d}x.\end{aligned}$$

Since $\mathfrak{c}_1, \mathfrak{c}_2$ are Möbius transformations $\mathbb{D} \to \mathbb{H}$ sending $v_i, h(v_i)$ to ∞ respectively, there exist $\alpha, \beta > 0$ such that

$$|\mathfrak{c}_1'(z)| = \frac{\alpha}{|z - h(v_i)|^2} + O\Big(\frac{1}{|z - h(v_i)|}\Big), \quad |\mathfrak{c}_2'(z)| = \frac{\beta}{|z - v_i|^2} + O\Big(\frac{1}{|z - v_i|}\Big).$$

On the other hand, since h is piecewise Möbius, we have

$$|h'(z)| = |h'(v_i\pm)| + O(|z - v_i|),$$

depending on which side of v_i does z approaches from. Therefore, for $x \in [n_0, \infty)$,

$$\begin{aligned}
&\left|\mathfrak{c}_1'(h \circ \mathfrak{c}_2^{-1}(x))\right| \left|h'(\mathfrak{c}_2^{-1}(x))\right| \left|(\mathfrak{c}_2^{-1})'(x)\right| \\
&= \left(\alpha \left|h \circ \mathfrak{c}_2^{-1}(x) - h(v_i)\right|^{-2} + O\left(\left|h \circ \mathfrak{c}_2^{-1}(x) - h(v_i)\right|^{-1}\right)\right) \left(|h'(v_i+)| + O(|\mathfrak{c}_2^{-1}(x) - v_i|)\right) \\
&\quad \left(\frac{1}{\beta}\left|\mathfrak{c}_2^{-1}(x) - v_i\right|^2 + O(|\mathfrak{c}_2^{-1}(x) - v_i|^3)\right) \\
&= \frac{\alpha}{\beta}\frac{1}{|h'(v_i+)|} + O(|\mathfrak{c}_2^{-1}(x) - v_i|).
\end{aligned}$$

We obtain

$$\ell = \lim_{m\to\infty} \varphi(m+1) - \varphi(m) = \frac{\alpha}{\beta}\frac{1}{|h'(v_i+)|}.$$

Similarly,

$$\ell' = \lim_{n\to\infty} \varphi(-n) - \varphi(-n-1) = \frac{\alpha}{\beta}\frac{1}{|h'(v_i-)|}.$$

Equation (3.6) then shows that $h'(v_i+) = h'(v_i-)$ if and only if $\sum_{n\in\mathbb{Z}} s(e_n) = 0$ which concludes the proof. □

We now introduce the diamond shear coordinates which are well-adapted to describe finite shears with the finite balanced condition. We say that $e, e' \in E$ are *adjacent* if e and e' share a vertex v and are consecutive in $\mathrm{fan}(v)$ (their indices in $\mathrm{fan}(v)$ differ by exactly 1). We say that $e^* \in E^*$ and $e' \in E$ are *adjacent* if the dual edge e of e^* is adjacent to e'. Note that e^* is adjacent to exactly 4 edges in E, that we denote by (e_1, e_2, e_3, e_4), in counterclockwise order, such that e_1 is the edge on the left of $\vec{e}$ and has the same head as $\vec{e}$. We do not distinguish between (e_1, e_2, e_3, e_4) and (e_3, e_4, e_1, e_2) which results from inverting the orientation of e.

A *diamond shear function* is a map $\vartheta : E^* \to \mathbb{R}$. The space of diamond shear functions has a basis $\{\vartheta_{e^*}(\cdot)\}_{e^*\in E^*}$, where $\vartheta_{e^*}(\cdot)$ equals 1 at e^* and 0 elsewhere. For each $e^* \in E^*$, the corresponding shear function of ϑ_{e^*} has $s(e_1) = s(e_3) = 1$ and $s(e_2) = s(e_4) = -1$. See Figure 3.4. More generally, the diamond shear functions translate to the shear functions via the map $\Phi : \mathbb{R}^{E^*} \to \mathbb{R}^E$:

$$s(e) = \Phi(\vartheta)(e) = -\vartheta(e_1^*) + \vartheta(e_2^*) - \vartheta(e_3^*) + \vartheta(e_4^*), \quad \forall e \in E. \tag{3.7}$$

Here, $(e_i^*)_{i=1,\dots,4}$ are the dual edges of $(e_i)_{i=1,\dots,4}$ ordered as described above.

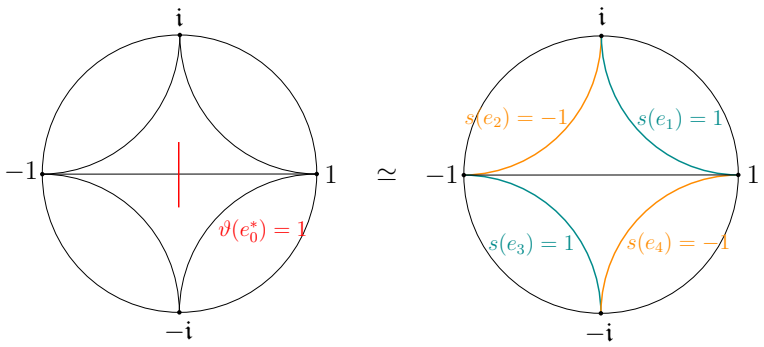


Figure 3.4: A single diamond shear along the edge $(-1, 1)$ (left) is equivalent to four shears with alternating sign (right).

Remark 3.3.5.

- We note that Φ is linear and if $\vartheta \equiv 1$, then $\Phi(\vartheta) \equiv 0$. In other words, constant diamond shear coordinates belong to the kernel of Φ.
- If ϑ has finite support, then $\Phi(\vartheta)$ has finite support and satisfies the finite balanced condition. The converse follows from the next proposition.

Proposition 3.3.6. *Assume that $s : E \to \mathbb{R}$ has finite support and satisfies the finite balanced condition. There exists a unique $\vartheta : E^* \to \mathbb{R}$ with finite support such that $\Phi(\vartheta) = s$.*

Proof. Every $e \in E$ belongs to two triangles which are dual to two vertices in $\mathfrak{F}^*$. Let $\tau^*(e)$ be the dual vertex that has the lower generation, if $e \neq e_0$. We define $\mathfrak{a}_s$ to be the union of e_0^* and the geodesic path from $\tau^*(e)$ to e_0^* in $\mathfrak{F}^*$ for all $e \neq e_0$ such that $s(e) \neq 0$. (We call $\mathfrak{a}_s$ the *convex hull* of $\{\tau^*(e) \colon s(e) \neq 0\} \cup e_0^*$.) Since s has finite support, $\mathfrak{a}_s$ is a finite tree.

We prove the existence of ϑ by induction on $\mathfrak{a}_s$. If $\mathfrak{a}_s$ contains only e_0^*, then $\{e \colon s(e) \neq 0\} \subset \{(-1, 1), (\mathfrak{i}, 1), (\mathfrak{i}, -1), (-\mathfrak{i}, 1), (-\mathfrak{i}, -1)\}$. The finite balanced condition shows that the only possibility is

$$s(\mathfrak{i}, 1) = -s(\mathfrak{i}, -1) = s(-\mathfrak{i}, -1) = -s(-\mathfrak{i}, 1) = \alpha$$

for some $\alpha \in \mathbb{R}$. For convenience, we write $s(a, b)$ for $s((a, b))$. Therefore, $s = \Phi(\alpha\vartheta_{e_0^*})$.

Now assume that $\mathfrak{a}_s$ is a general finite tree containing e_0^* and e^* is a leaf of $\mathfrak{a}_s$. Assume that e^* has generation n. The dual edge $e \in E$ has two vertices $\{a, c\}$, their child $b \in V$ has generation $n+1$. We assume that a, b, c are in the counterclockwise order. $\mathrm{Fan}(b)$ contains at most two edges, (a, b) and (b, c), on which s is nonzero. From the finite balanced condition, there is $\alpha = \alpha(e^*) \in \mathbb{R}$ such that

$$\alpha(e^*) = s(a, b) = -s(b, c).$$

Therefore $s' := s - \Phi(\alpha(e^*)\vartheta_{e^*})$ is a shear coordinate with finite support, and $\mathfrak{a}_{s'} = \mathfrak{a}_s \setminus e^*$. By the assumption of induction, let ϑ' be a finite support diamond shear coordinate such that $\Phi(\vartheta') = s'$. The linearity of Φ shows that $\Phi(\vartheta' + \alpha(e^*)\vartheta_{e^*}) = s$.

Now we show the uniqueness. Assume that ϑ and ϑ' have finite support and $\Phi(\vartheta) = \Phi(\vartheta')$. Then $\Phi(\vartheta - \vartheta') \equiv 0$. Let $\mathfrak{a}$ be the convex hull of $\{e^* \in E^* \colon \vartheta(e^*) \text{ or } \vartheta'(e^*) \neq 0\}$. The above argument shows that for any leaf e^* of $\mathfrak{a}$, $\vartheta(e^*) - \vartheta'(e^*) = 0$. By induction, we have $\vartheta = \vartheta'$ which concludes the proof. □

Definition 3.3.7. If a circle homeomorphism h satisfies the conditions in Lemma 3.3.4, the *diamond shear coordinate* of h is the unique finitely supported diamond shear function ϑ_h such that $\Phi(\vartheta_h) = s_h$.

3.3.2 Examples and developing algorithm

In this section we provide a few explicit examples to show concretely how circle homeomorphisms are related to shear and diamond shear coordinates. Recall that for $a \neq b \in \mathbb{T}$, $I(a,b) \subset \mathbb{T}$ denotes the circular arc going counterclockwise from a to b.

Example 3.3.8 (Single shear). For $t \in \mathbb{R}$, let h_t be the normalized circle homeomorphism with shear coordinate $s_t(e_0) = t$, and $s_t(e) = 0$ for all $e \in E$, $e \neq e_0 = (-1, 1)$. Then

$$\mathfrak{c} \circ h_t \circ \mathfrak{c}^{-1}(x) = \begin{cases} x & \forall x \leq 0 \\ e^t x & \forall x \geq 0. \end{cases}$$

In other words,

$$h_t(z) = \begin{cases} z & \forall z \in I(1,-1) \\ \dfrac{\alpha_t z + \beta_t}{\overline{\beta_t} z + \overline{\alpha_t}} & \forall z \in I(-1,1) \end{cases}$$

with $\alpha_t = \cosh(t/2)$ and $\beta_t = \sinh(t/2)$. In particular $(h_t)_{t\in\mathbb{R}}$ forms a one-dimensional subgroup of the group of piecewise $\mathrm{PSU}(1,1)$ circle homeomorphisms.

For $e = (a,b)$, not necessarily an edge of $\mathfrak{F}$, there exists $A \in \mathrm{PSU}(1,1)$ such that $A(a) = -1$, $A(b) = 1$. Then $h_{e,t} := A^{-1} \circ h_t \circ A$ is also a one-dimensional subgroup of (non-normalized) circle homeomorphisms (and independent of the choice of A) fixing the circular arc $I(b,a)$. Explicitly,

$$h_{e,t} = \begin{cases} z & \forall z \in I(b,a) = A^{-1}I(1,-1) \\ A_t(z) & \forall z \in I(a,b) = A^{-1}I(-1,1) \end{cases}$$

where $A_t = A^{-1}\begin{pmatrix} \alpha_t & \beta_t \\ \overline{\beta_t} & \overline{\alpha_t} \end{pmatrix} A$. We note that $h_{e,t}$ is not C^1 if $t \neq 0$ and we find

$$h'_{e,t}(a+) = 1, \quad h'_{e,t}(a-) = e^t, \qquad h'_{e,t}(b+) = e^{-t}, \quad h'_{e,t}(b-) = 1,$$

where $a+$ means approaching a counterclockwisely, and $a-$ clockwisely.

Example 3.3.9 (Standard single diamond shear). For $t \in \mathbb{R}$, let H_t be a normalized circle homeomorphism satisfying the conditions in Lemma 3.3.4 (we can hence talk about its diamond shear coordinates ϑ_t). In particular, suppose it has diamond shear coordinates such that $\vartheta_t(e_0^*) = t$, and $\vartheta_t(e^*) = 0$ for all $e^* \in E^*$, $e^* \neq e_0^*$. Then the corresponding shear coordinate $S_t = \Phi(\vartheta_t)$ of H_t is given by

$$S_t(1,\mathfrak{i}) = t, \quad S_t(\mathfrak{i},-1) = -t, \quad S_t(-1,-\mathfrak{i}) = t, \quad S_t(-\mathfrak{i},1) = -t.$$

It is easy to see that H_t fixes $1, \mathfrak{i}, -1, -\mathfrak{i}$. We obtain the following explicit expression of $H_t(z)$ (by symmetry it suffices to compute H_t on $I(1,\mathfrak{i})$):

$$\begin{cases} h_{(1,\mathfrak{i}),t}(z) = \dfrac{\alpha_{1,t} z + \beta_{1,t}}{\overline{\beta_{1,t}} z + \overline{\alpha_{1,t}}} & \text{with } \begin{cases} \alpha_{1,t} = \cosh(\frac{t}{2}) - \mathfrak{i}\sinh(\frac{t}{2}) \\ \beta_{1,t} = (\mathfrak{i}-1)\sinh(\frac{t}{2}), \end{cases} \quad \forall z \in I(1,\mathfrak{i}); \\ h_{(\mathfrak{i},-1),-t}(z) = \dfrac{\alpha_{2,t} z + \beta_{2,t}}{\overline{\beta_{2,t}} z + \overline{\alpha_{2,t}}} & \text{with } \begin{cases} \alpha_{2,t} = \overline{\alpha_{1,t}} \\ \beta_{2,t} = -\overline{\beta_{1,t}}, \end{cases} \quad \forall z \in I(\mathfrak{i},-1); \\ h_{(-1,-\mathfrak{i}),t}(z) = \dfrac{\alpha_{3,t} z + \beta_{3,t}}{\overline{\beta_{3,t}} z + \overline{\alpha_{3,t}}} & \text{with } \begin{cases} \alpha_{3,t} = \alpha_{1,t} \\ \beta_{3,t} = -\beta_{1,t}, \end{cases} \quad \forall z \in I(-1,-\mathfrak{i}); \\ h_{(-\mathfrak{i},1),-t}(z) = \dfrac{\alpha_{4,t} z + \beta_{4,t}}{\overline{\beta_{4,t}} z + \overline{\alpha_{4,t}}} & \text{with } \begin{cases} \alpha_{4,t} = \overline{\alpha_{1,t}} \\ \beta_{4,t} = \overline{\beta_{1,t}}, \end{cases} \quad \forall z \in I(-\mathfrak{i},1). \end{cases} \tag{3.8}$$

We observe

$$H_t'(1) = H_t'(-1) = e^t, \qquad H_t'(\mathfrak{i}) = H_t'(-\mathfrak{i}) = e^{-t}$$

and that $(H_t)_{t\in\mathbb{R}}$ is a one-dimensional subgroup of the group of $C^{1,1}$ and piecewise $\mathrm{PSU}(1,1)$ circle homeomorphisms.

The circle homeomorphism satisfying the conditions in Lemma 3.3.4 whose diamond shear is supported on a single dual edge $e^* \in E^*$ can be obtained by $A^{-1} \circ H_t \circ A$ for some $A \in \mathrm{PSU}(1,1)$ up to normalization. We can also define the homeomorphism associated to a diamond shear on a *non-standard* quad.

Definition 3.3.10 (Single diamond shear on non-standard quad)**.** Let Q be a quad with vertices $a, b, c, d \in \mathbb{T}$ in counterclockwise order. (We do not require Q is a quad in $\mathfrak{F}$, in particular, $\mathrm{cr}(a,b,c,d)$ might not be zero.) We define $H_{Q,(a,c),t} \in C^{1,1}$ to be the (non-normalized) circle homeomorphism which fixes the vertices of Q, that is piecewise $\mathrm{PSU}(1,1)$ with break points at the vertices, and

$$H'_{Q,(a,c),t}(a) = e^t. \tag{3.9}$$

Remark 3.3.11. Since for any Möbius transform A and $x \neq y$, $A'(x)A'(y) = \frac{(A(x)-A(y))^2}{(x-y)^2}$, the $C^{1,1}$ condition and (3.9) uniquely determine $H_{Q,(a,c),t}$ on $\mathbb{T}$ and give us

$$H'_{Q,(a,c),t}(c) = e^t, \qquad H'_{Q,(a,c),t}(b) = H'_{Q,(a,c),t}(d) = e^{-t}.$$

From this we obtain

$$\begin{aligned} H_{Q,(a,c),t}|_{I(a,b)} &= h_{(a,b),t}|_{I(a,b)}, & H_{Q,(a,c),t}|_{I(b,c)} &= h_{(b,c),-t}|_{I(b,c)}, \\ H_{Q,(a,c),t}|_{I(c,d)} &= h_{(c,d),t}|_{I(c,d)}, & H_{Q,(a,c),t}|_{I(d,a)} &= h_{(d,a),-t}|_{I(d,a)}. \end{aligned}$$

This relation similar to (3.8) justifies the name of *homeomorphism associated to non-standard diamond shear* and also shows that $(H_{Q,(a,c),t})_{t\in\mathbb{R}}$ is a one-dimensional subgroup of $C^{1,1}$ and piecewise $\mathrm{PSU}(1,1)$ circle homeomorphisms. Note that this definition coincides with the one for the single diamond shear supported on an edge (as diagonal of a Farey quad) as in Example 3.3.9.

Non-standard diamond shears are useful in the following developing algorithm for finding the associated circle homeomorphism given the diamond shear coordinate inductively when it has finite support.

Proposition 3.3.12. *Let h be a circle homeomorphism satisfying the conditions in Lemma 3.3.4 and ϑ its (finitely supported) diamond shear coordinates. Let $t \in \mathbb{R}$ and $e \in E$. The homeomorphism with diamond shear coordinate $\vartheta + t\vartheta_{e^*}$ is $H_{h(Q_e),h(e),t} \circ h$ after normalizing to fix $-1, \mathbf{i}, 1$.*

Proof. Let $s = \Phi(\vartheta)$ denote the shear coordinate of h. Let $h_t := H_{h(Q_e),h(e),t} \circ h$ and s_t its shear coordinate. We write the Farey quad $Q_e = (a, b, c, d)$ such that $e = (a, c)$. Let $e_1 = (a, b)$, $e_2 = (b, c)$, $e_3 = (c, d)$, and $e_4 = (d, a)$ be the adjacent edges in $\mathfrak{F}$. We need to show that

$$S_t(e_j) := s(h_t(Q_{e_j}), h_t(e_j)) = s(e_j) + (-1)^{j-1}t, \qquad j = 1, \cdots, 4. \tag{3.10}$$

We see it from the geometric interpretation of the shear (Lemma 3.2.5). In fact, $s(e_1) = s(h(Q_{e_1}), h(e_1))$ is the signed distance between the geodesics normal to $h(e_1)$ and starting from the third vertex of the two ideal triangles that we call τ_L, τ_R, where $\tau_L \cup \tau_R = h(Q_{e_1})$ and $\tau_R \subset h(Q_e)$. Since $H_{h(Q_e),h(e),t}$ fixes $h(Q_e)$, it also fixes τ_R. On the arc $I_1 \subset \mathbb{T}$ which has the same vertices as $h(e_1) = (h(a), h(b))$ and contains the vertices of τ_L, $H_{h(Q_e),h(e),t}$ shears further the normal starting from any point of I_1 by hyperbolic distance t in the direction from $h(a)$ to $h(b)$. We obtain (3.10) for $j = 1$. See Figure 3.5. The same argument works for other j. □

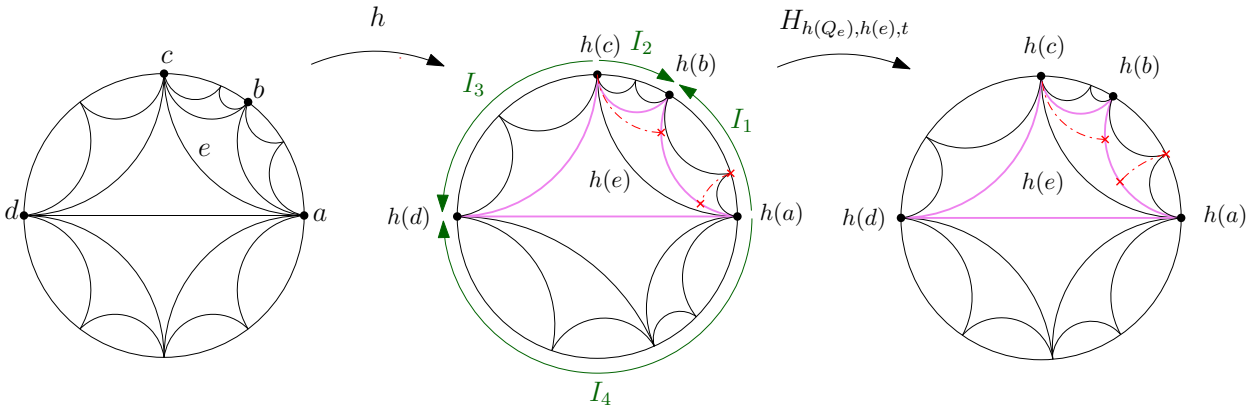


Figure 3.5: Illustration of the maps in the proof of Proposition 3.3.12. Left: Farey tessellation $\mathfrak{F}$ with an edge $e \in E$ and Q_e with vertices a, b, c, d marked. Middle: $h(\mathfrak{F})$ with the edges of $h(Q_e)$ in pink, the green arrows indicate the direction in which of the piecewise Möbius circle homeomorphism $H_{h(Q_e),h(e),t}$ moves the points on each arcs, when $t > 0$. Red dashed lines indicate the normals to $h(e_1) = (h(a), h(b))$. Right: the tessellation $h_t(\mathfrak{F})$.

3.3.3 Combinatorial definition of diamond shear

The goal of this section is to extend the definition of diamond shear coordinates to a more general class of circle homeomorphisms. Definition 3.3.7 suggests that ϑ should be defined as the image of s by the inverse of Φ defined in (3.7). However, the map Φ does not map onto $\mathbb{R}^E$, nor is it injective by Remark 3.3.5. Therefore, we will restrict to the following family of shear coordinates to define a right-inverse map of Φ:

$$\mathscr{P} = \{s \in \mathbb{R}^E \colon \forall\, v \in V,\, \mathrm{fan}(v) = (e_n)_{n\in\mathbb{Z}},\, \lim_{n\to\infty} \sum_{k=-n}^{-1} s(e_k) \in \mathbb{R} \text{ and } \lim_{n\to\infty} \sum_{k=0}^{n} s(e_k) \in \mathbb{R}\}.$$

Similar to Definition 3.3.3, we say that $s \in \mathscr{P}$ satisfies the (generalized) *balanced condition*, if

$$s \in \mathscr{P}_0 = \{s \in \mathscr{P} \colon \forall\, v \in V,\, \mathrm{fan}(v) = (e_n)_{n\in\mathbb{Z}},\, \sum_{k=-\infty}^{\infty} s(e_k) = 0\}. \tag{3.11}$$

Note that if s satisfies the finite balanced condition, then $s \in \mathscr{P}_0$.

Definition 3.3.13. We define for $s \in \mathscr{P}$, $v \in V$, and $e \in \mathrm{fan}(v)$,

$$p_{s,v}(e_+) = \sum_{e' >_v e} s(e'), \quad p_{s,v}(e_-) = \sum_{e' <_v e} s(e')$$

where $e' >_v e$ means that $e' \in \mathrm{fan}(v)$ and has strictly larger index than e, and similarly, $e' <_v e$ for strictly smaller index than e. Define $\Psi : \mathscr{P} \to \mathbb{R}^{E^*}$ by

$$\Psi(s)(e^*) = \frac{1}{4}\Big(p_{s,a}(e_-) - p_{s,a}(e_+) + p_{s,b}(e_-) - p_{s,b}(e_+)\Big) \tag{3.12}$$

where $e = (a, b)$ is dual to e^*. See Figure 3.6 for an illustration of $\Psi(s)(e_0^*)$.

Proposition 3.3.14. *The function Ψ is a right-inverse of Φ, namely, $\Phi \circ \Psi = \mathrm{Id}_{\mathscr{P}}$.*

Proof. The maps Φ and Ψ are both linear. Combining Equations (3.7) and (3.12), $\Phi(\Psi(s))(e)$ is a sum of $p_{s,a}(e_\pm)$ for sixteen different choices of $a, e, \pm$. Since $s \in \mathscr{P}$, the limits $p_{s,a}(e_\pm)$ are well-defined for all $a \in V, e \in \mathrm{fan}(a)$, and hence we can switch the finite sum with sixteen terms with the limits defining $p_{s,a}(e_\pm)$. Therefore $\Phi \circ \Psi$ is linear on $\mathscr{P}$ even for infinite linear combinations, so it is enough to show that $\Phi \circ \Psi(s_e) = s_e$ where s_e takes value 1 on the edge $e = (a, b) \in E$ and 0 elsewhere.

Let e_1, e_2, e_3, e_4 be the edges around Q_e in counterclockwise order starting from a. We denote the four half-fans around Q_e in counterclockwise order by $\mathscr{E}_1 = \{e' \in \mathrm{fan}(a) : e' \leq_a e_1\}$, $\mathscr{E}_2 = \{e' \in \mathrm{fan}(b) : e' \geq_b e_2\}$, $\mathscr{E}_3 = \{e' \in \mathrm{fan}(b) : e' \leq_b e_3\}$, $\mathscr{E}_4 = \{e' \in \mathrm{fan}(a) : e' \geq_a e_4\}$. See Figure 3.6. To simplify notation, we identify the dual edges $(e')^*$ with the corresponding edge e'. By Equation (3.12)

$$\Psi(s_e)(e') = \begin{cases} -1/4 & e' \in \mathscr{E}_1 \cup \mathscr{E}_3 \\ +1/4 & e' \in \mathscr{E}_2 \cup \mathscr{E}_4 \\ 0 & \text{otherwise.} \end{cases}$$

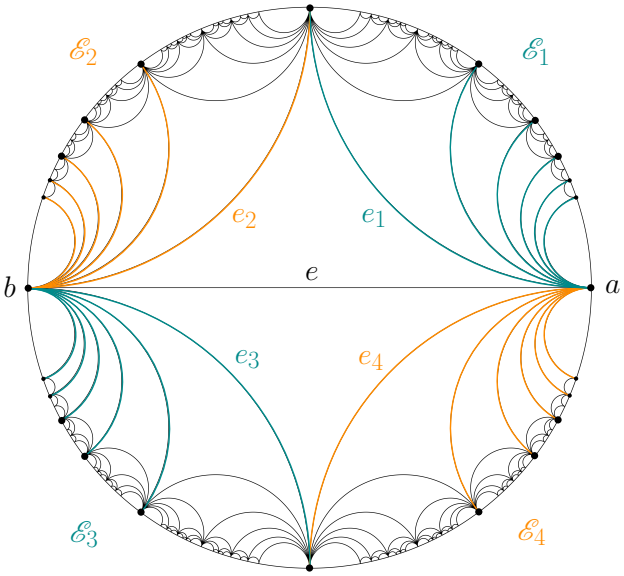


Figure 3.6: The shears on blue edges are counted as positive and the shears on orange edges are counted as negative in $\Psi(s)(e^*)$.

To compute $\Phi(\Psi(s_e))(e')$ we look at the the values of $\Psi(s_e)$ on the edges e'_1, e'_2, e'_3, e'_4 around $Q_{e'}$ and use Equation (3.7).

If $e' = e$,

$$\Phi(\Psi)(s_e)(e') = -\left(-\frac{1}{4}\right) + \frac{1}{4} - \left(-\frac{1}{4}\right) + \frac{1}{4} = 1.$$

For $e' \neq e$, we check that $\Phi(\Psi(s_e))(e') = 0$.

If $e' \neq e_i$ for $i = 1, 2, 3, 4$, the edges around $Q_{e'}$ either all have diamond shear $\Psi(s_e)(\cdot) = 0$, or there are two consecutive edges with the same nonzero diamond shear followed by two edges with zero diamond shear. In both cases, $\Phi(\Psi(s_e))(e') = 0$.

For $e' = e_i$ for $i = 1, 2, 3, 4$, one can check that two non-consecutive edges around $Q_{e'}$ have nonzero diamond shear coordinates of opposite sign, and the other two edges have diamond shear coordinate 0. □

Definition 3.3.15. We define the *diamond shear coordinate* $\vartheta_h := \Psi(s_h)$ of a circle homeomorphism h if the shear coordinate $s_h \in \mathscr{P}$. We let

$$\mathscr{S} = \{s \in \mathbb{R}^E : \sum_{e \in E} s(e)^2 < \infty\} \text{ and } \mathscr{H} = \{s \in \mathscr{P} : \sum_{e^* \in E^*} \vartheta(e^*)^2 < \infty \text{ where } \vartheta = \Psi(s)\}.$$

We say that a circle homeomorphism h has *a square summable diamond shear coordinate* if $\vartheta_h \in \mathscr{H}$. We endow $\mathscr{S}$ and $\mathscr{H}$ with the topology of ℓ^2 convergence in s and ϑ respectively.

In the finite support case, it follows from Proposition 3.3.14 that $\Psi(s_h)$ is the diamond shear coordinate defined in Definition 3.3.7. Here and in the rest of the paper we identify E with E^* to simplify the notation.

Lemma 3.3.16. *Assume that $s \in \mathscr{H}$ and $\vartheta = \Psi(s)$. Then we have $s \in \mathscr{S}$ and*

$$\sum_{v\in V}\sum_{e\in\mathrm{fan}(v)} p_{s,v}(e_+)^2 = 2\sum_{e\in E}\vartheta(e)^2 + \sum_{e\sim e'}(\vartheta(e)-\vartheta(e'))^2 < \infty, \tag{3.13}$$

where $e \sim e'$ means that e and e' are adjacent in the same fan. In particular, $s \in \mathscr{P}_0$.

Proof. The Cauchy-Schwarz inequality, $s = \Phi(\vartheta)$, and the assumption of ϑ is square summable show that $s \in \mathscr{S}$.

Now we fix $v \in V$ and let $\mathrm{fan}(v) = (e_n)_{n\in\mathbb{Z}}$. For $e_k \in \mathrm{fan}(v)$, we compute $p_{s,v}(e_{k+}) = \sum_{n=k+1}^{\infty} s(e_n)$. For this, we write the edge of $\mathfrak{F}$ connecting the endpoints of e_n and e_{n+1} other than v as e'_n. Since $s = \Phi(\vartheta)$, (3.7) shows that for $m > k$,

$$\begin{aligned}\sum_{n=k+1}^{m} s(e_n) &= \sum_{n=k+1}^{m} \vartheta(e_{n-1}) - \vartheta(e'_{n-1}) + \vartheta(e'_n) - \vartheta(e_{n+1}) \\ &= \vartheta(e_k) + \vartheta(e_{k+1}) - \vartheta(e'_k) + \vartheta(e'_m) - \vartheta(e_m) - \vartheta(e_{m+1}).\end{aligned}$$

Since ϑ is square summable, we have $\vartheta(e_m)$ and $\vartheta(e'_m)$ converging to 0 as $m \to \infty$. Hence,

$$p_{s,v}(e_{k+}) = \lim_{m\to\infty}\sum_{n=k+1}^{m} s(e_n) = \vartheta(e_k) + \vartheta(e_{k+1}) - \vartheta(e'_k).$$

When we sum $p_{s,v}(e_+)^2$ over all $v \in V$ and $e \in \mathrm{fan}(v)$, the triplet $(\vartheta(e_k), \vartheta(e_{k+1}), \vartheta(e'_k))$ in the above identity appears three times but with different signs, once in each fan at the vertices of the triangle formed by e_k, e_{k+1} and e'_k. Using the identity

$$(a+b-c)^2 + (b+c-a)^2 + (c+a-b)^2 = a^2+b^2+c^2 + (a-b)^2 + (a-c)^2 + (b-c)^2,$$

we obtain

$$\sum_{v\in V}\sum_{e\in\mathrm{fan}(v)} p_{s,v}(e_+)^2 = 2\sum_{e\in E}\vartheta(e)^2 + \sum_{e\sim e'}(\vartheta(e)-\vartheta(e'))^2 < \infty$$

as claimed. Here we note that the constant in front of the first sum is 2 since every edge appears in two triangles, while the constant in front of the second is 1 because every pair of adjacent edges appears in only one triangle. Finally, (3.13) implies that $p_{s,v}(e_{k+}) \to 0$ as $k \to -\infty$. This shows $s \in \mathscr{P}_0$. □

Summarizing the above results, we obtain the following inclusions.

Corollary 3.3.17. *We have $\mathscr{H} \subset \mathscr{P}_0 \cap \mathscr{S} \subset \mathscr{P}$.*

Shear functions in $\mathscr{H}$ also satisfy another boundedness condition.

Lemma 3.3.18. *If $s \in \mathscr{H}$, then there exists a constant $M = M(s) \geq 0$ such that for all $v \in V$ and all $n, m \in \mathbb{Z}$, $n \leq m$,*

$$\left|\sum_{i=n}^{m} s(e_i)\right| \leq M, \tag{3.14}$$

where $\mathrm{fan}(v) = (e_i)_{i\in\mathbb{Z}}$.

Proof. By Lemma 3.3.16, $s \in \mathscr{H}$ implies that $\{p_{s,v}(e_+) : v \in V, e \in \text{fan}(v)\}$ is square summable, so there is a constant $C > 0$ such that $|p_{s,v}(e_+)| < C$ for all $v \in V, e \in \text{fan}(v)$. Choose any $v \in V$, and let $\text{fan}(v) = (e_i)_{i\in\mathbb{Z}}$. For any n, m,

$$\left|\sum_{i=n}^{m} s(e_i)\right| = |p_{s,v}(e_{m+}) - p_{s,v}(e_{n+})| \leq 2C.$$

Therefore (3.14) holds with $M = 2C$. □

Remark 3.3.19. The class of shear functions satisfying Equation (3.14) does not include, nor is contained in $\mathscr{P}_0$ or $\mathscr{P}$.

- (3.14) does not imply that $s \in \mathscr{P}$. For example, the map that has shears alternating ± 1 along a fan satisfies (3.14), but is not in $\mathscr{P}$.
- $s \in \mathscr{P}_0$ does not imply (3.14) either. The condition $s \in \mathscr{P}_0$ implies that for each $v \in V$ there exists a constant M_v such that the sums in $\text{fan}(v)$ are bounded by M_v, but these M_v constants may not be the same and the collection $\{M_v : v \in V\}$ may be unbounded for $s \in \mathscr{P}_0$.

The condition (3.14) helps us show that any $s \in \mathscr{H}$ induces a quasisymmetric homeomorphism h of the circle. Shears for quasisymmetric homeomorphisms have been characterized.

Theorem 3.3.20 (See [Šar10],[Šar21])**.** *A shear function $s : E \to \mathbb{R}$ is induced by a quasisymmetric map if and only if there exists $C \geq 1$ such that for all $v \in V$ with $\text{fan}(v) = (e_i)_{i\in\mathbb{Z}}$ and for all $k \in \mathbb{Z}$, $n \in \mathbb{N}$,*

$$\frac{1}{C} \leq s(k, n; v) \leq C.$$

Here $s(k, n; v)$ is

$$s(k, n; v) = \frac{e^{s(e_k)} + e^{s(e_k)+s(e_{k+1})} + \cdots + e^{s(e_k)+\cdots+s(e_{k+n})}}{1 + e^{-s(e_{k-1})} + \cdots + e^{-s(e_{k-1})-\cdots-s(e_{k-n})}}.$$

We obtain from this theorem the following corollary which is also considered in the paper of Parlier and the first author [PŠ22].

Corollary 3.3.21. *If $s : E \to \mathbb{R}$ satisfies (3.14), then s induces a quasisymmetric homeomorphism $h : \mathbb{T} \to \mathbb{T}$. In particular, if $s_h \in \mathscr{H}$, then $h \in \text{QS}(\mathbb{T})$.*

Remark 3.3.22. Given the result above, in later sections we often abuse notation and write $h \in \mathscr{H}$ to mean that the homeomorphism h has shear coordinates $s_h \in \mathscr{H}$. Despite the fact that not all shear functions in $\mathscr{P}, \mathscr{P}_0$ induce homeomorphisms, we also sometimes write $h \in \mathscr{P}$ or $h \in \mathscr{P}_0$ to mean that h has shear function $s_h \in \mathscr{P}$ or $s_h \in \mathscr{P}_0$ respectively.

3.3.4 Analytic definition of diamond shear

In the section we show that the diamond shear coordinate of a circle homeomorphism can be described directly using derivatives of h. This description of diamond shears also leads to a relationship with coordinates called $\log\Lambda$-*lengths* for decorated Teichmüller space studied in [Pen93; Pen02; Pen12], see Section 3.3.5. The following lemma is reminiscent of Lemma 3.3.4 for finite support shears.

Lemma 3.3.23. *i) If $h \in \mathscr{P}$, then h admits left and right derivatives at all rational points, i.e., $\forall v \in V$, $h'(v_+)$ and $h'(v_-)$ exist.*

ii) If $h \in \mathscr{P}_0$, then h is differentiable at all rational points, i.e., $\forall v \in V$, $h'(v_+) = h'(v_-)$.

iii) Conversely, if $h \in \mathscr{C}^1$, i.e., h is continuously differentiable and $h' \neq 0$ everywhere, then $h \in \mathscr{P}_0$.

Proof. Assume that $s = s_h \in \mathscr{P}$. We fix $v \in V = \mathbb{Q}^2 \cap \mathbb{T}$ and $\mathrm{fan}(v) = (e_k)_{k\in\mathbb{Z}}$. As in Lemma 3.3.4, we define $\varphi : \mathbb{R} \to \mathbb{R}$ to be the homeomorphism $\varphi = \mathfrak{c}_1 \circ h \circ \mathfrak{c}_2^{-1}$, where $\mathfrak{c}_1, \mathfrak{c}_2$ are two Möbius transformations $\mathbb{D} \to \mathbb{H}$ such that $\mathfrak{c}_1(h(v)) = \infty$, $\mathfrak{c}_2(v) = \infty$, $\mathfrak{c}_2(e_0) = (0, \infty)$, and $\mathfrak{c}_2(e_1) = (1, \infty)$. We have φ fixes ∞, and $\mathfrak{c}_2(e_n) = (n, \infty)$ for all $n \in \mathbb{Z}$. Since the limits as $n \to \infty$ of $\sum_{k=-n}^{-1} s(e_k)$ and $\sum_{k=0}^{n} s(e_k)$ exist by definition of $\mathscr{P}$,

$$\ell := \lim_{n\to\infty} \varphi(n+1) - \varphi(n) = (\varphi(0) - \varphi(-1)) \exp\Big(\sum_{k=0}^{\infty} s(e_k)\Big) \in (0, \infty) \tag{3.15}$$

and

$$\ell' := \lim_{n\to\infty} \varphi(-n) - \varphi(-n-1) = (\varphi(0) - \varphi(-1)) \exp\Big(-\sum_{k=-\infty}^{-1} s(e_k)\Big) \in (0, \infty) \tag{3.16}$$

also exist. In particular, $\varphi(n) = n\ell + o(n)$ and $\varphi(-n) = -n\ell' + o(n)$ as $n \to \infty$ by Cesàro summation.

To show the left and right derivatives of h at v exist, it suffices to show that $\tilde{\varphi} = \iota\circ\varphi\circ\iota$ has left and right derivatives at 0, where $\iota(x) = -1/x$. Note that $\tilde{\varphi}$ fixes 0 and is an increasing function. We have

$$\frac{\tilde{\varphi}(-n^{-1})}{-n^{-1}} = \frac{n}{\varphi(n)} \xrightarrow{n\to\infty} \ell^{-1}.$$

From the monotonicity of $\tilde{\varphi}$, we have

$$\frac{\tilde{\varphi}(-(n+1)^{-1})}{-n^{-1}} \leq \frac{\tilde{\varphi}(x)}{x} \leq \frac{\tilde{\varphi}(-n^{-1})}{-(n+1)^{-1}}, \qquad \forall x \in [-n^{-1}, -(n+1)^{-1}].$$

Hence, $\tilde{\varphi}$ admits the left derivative ℓ^{-1} at 0. Similarly, we can show that $\tilde{\varphi}$ admits the right derivative $(\ell')^{-1}$. This concludes the proof of i).

Now we assume that $h \in \mathscr{P}_0$. The equations (3.15), (3.16) show that $\ell = \ell'$. Hence, the left and right derivatives of h coincide, which shows ii).

For iii), if $h \in \mathscr{C}^1$, then φ is continuously differentiable with $\varphi'(0) \neq 0$. We have

$$\begin{aligned}\varphi(n+1) - \varphi(n) &= \frac{1}{\tilde{\varphi}(-n^{-1})} - \frac{1}{\tilde{\varphi}(-(n+1)^{-1})} = \frac{\tilde{\varphi}(-(n+1)^{-1}) - \tilde{\varphi}(-n^{-1})}{\tilde{\varphi}(-(n+1)^{-1})\tilde{\varphi}(-n^{-1})} \\ &= \frac{n^{-1} - (n+1)^{-1}}{(n+1)^{-1}n^{-1}} \frac{\tilde{\varphi}'(x)}{\tilde{\varphi}'(y)\tilde{\varphi}'(z)} = \frac{\tilde{\varphi}'(x)}{\tilde{\varphi}'(y)\tilde{\varphi}'(z)}\end{aligned}$$

for some $x \in [-n^{-1}, -(n+1)^{-1}]$, $y \in [-(n+1)^{-1}, 0]$, $z \in [-n^{-1}, 0]$. Therefore,

$$\varphi(n+1) - \varphi(n) \xrightarrow{n\to\infty} \tilde{\varphi}'(0)^{-1}, \tag{3.17}$$

which shows that $\sum_{k=0}^{n} s(e_k)$ converges. Similarly, $\sum_{k=-n}^{-1} s(e_k)$ converges as well. We conclude with (3.15), (3.16) that $h \in \mathscr{P}_0$. □

Remark 3.3.24. The converse statement iii) is slightly weaker and we do not have an equivalent description for circle homeomorphisms whose shear coordinate satisfies the generalized balanced condition $\mathscr{P}_0$. The naive converse of ii) is not true. This lemma is trickier than Lemma 3.3.4 as the set of vertices is dense in $\mathbb{T}$ and we do not have the a priori smoothness of piecewise Möbius maps.

Proposition 3.3.25. *If a circle homeomorphism $h \in \mathscr{P}_0$, then ϑ_h is given by*

$$\vartheta_h(e) = \frac{1}{2}\log h'(a)h'(b) - \log\frac{h(a) - h(b)}{a - b} \tag{3.18}$$

for all $e = (a, b) \in E$.

Proof. We assume first that $e = (-1, 1)$ and recall that h fixes $\pm 1, \mathfrak{i}$. The Cayley map $\mathfrak{c}$ sends $1 \mapsto \infty$, $-1 \mapsto 0$, $\mathfrak{i} \mapsto -1$. We index edges $(e_n)_{n\in\mathbb{Z}}$ in fan(1) such that $\mathfrak{c}(e_n)$ is the geodesic (n, ∞) in $\mathbb{H}$. In this way, $e = e_0$. We write also $(e'_n)_{n\in\mathbb{Z}} = \mathrm{fan}(-1)$, such that $e'_0 = e$. Let $\varphi = \mathfrak{c} \circ h \circ \mathfrak{c}^{-1}$, and $\tilde{\varphi} = \iota \circ \varphi \circ \iota$, where $\iota(x) = -1/x$. We note that φ fixes $-1, 0, \infty$ and $\tilde{\varphi}$ fixes $0, 1, \infty$.

Let $s = s_h$ and $\vartheta = \vartheta_h = \Psi(s_h)$. Since $s \in \mathscr{P}_0$, we have from (3.12) that

$$\vartheta_h(e_0) = -\frac{1}{2}\Big(p_{s,1}(e_+) + p_{s,-1}(e_+) + s(e)\Big).$$

It follows from the proof of Lemma 3.3.23, (3.15), and (3.17) that

$$p_{s,1}(e_+) + s(e) = \sum_{k=0}^{\infty} s(e_k) = -\log\tilde{\varphi}'(0). \tag{3.19}$$

Similarly, applying the same proof to $\mathrm{fan}(-1)$ with the homeomorphism $\tilde{\varphi}$, and $\varphi = \iota \circ \tilde{\varphi} \circ \iota$, we obtain

$$\varphi'(0)^{-1} = \lim_{n\to\infty} \tilde{\varphi}(n+1) - \tilde{\varphi}(n) = (\tilde{\varphi}(1) - \tilde{\varphi}(0)) \exp\Big(\sum_{k=1}^{\infty} s(e'_k)\Big) = \exp\Big(\sum_{k=1}^{\infty} s(e'_k)\Big).$$

Hence

$$p_{s,-1}(e_+) = \sum_{k=1}^{\infty} s(e'_k) = -\log \varphi'(0).$$

On the other hand,

$$\varphi'(0) = \mathfrak{c}'(-1)h'(-1)(\mathfrak{c}^{-1})'(0) = h'(-1), \qquad \tilde{\varphi}'(0) = (\iota \circ \mathfrak{c})'(1)h'(1)(\iota \circ \mathfrak{c})^{-1\prime}(0) = h'(1).$$

We obtain (3.18) since in this case $\frac{h(1)-h(-1)}{1-(-1)} = 1$.

For a general edge $e = (a, b)$, h might not fix a, b. We choose $\gamma, \delta \in \mathrm{PSU}(1,1)$, such that γ maps $\mathfrak{F}$ to $\mathfrak{F}$, sending $-1 \mapsto a$ and $1 \mapsto b$ (see Section 3.2.1); and δ is such that the homeomorphism $\tilde{h} = \delta \circ h \circ \gamma$ fixes $\pm 1, \mathfrak{i}$. In particular, δ maps $h(a) \mapsto -1$ and $h(b) \mapsto 1$. The homeomorphism $\tilde{h}$ has the shear coordinate $\tilde{s} = s \circ \gamma$ and therefore the diamond shear coordinate $\tilde{\vartheta} = \vartheta \circ \gamma$. Applying the previous result, we have

$$\vartheta(e) = \tilde{\vartheta}((-1,1)) = \frac{1}{2}\log \tilde{h}'(-1)\tilde{h}'(1).$$

We use the fact that for any Möbius transformation A, as long as it is well defined, $A'(a)A'(b) = \frac{(A(a)-A(b))^2}{(a-b)^2}$. We obtain

$$\tilde{h}'(-1)\tilde{h}'(1) = [\delta'(h(a))\delta'(h(b))]h'(a)h'(b)[\gamma'(-1)\gamma'(1)] = h'(a)h'(b)\frac{(a-b)^2}{(h(a)-h(b))^2}$$

which concludes the proof. □

Remark 3.3.26. We can see directly that the right-hand side of (3.18) is real-valued. In fact, for $e = (a, b)$, consider two Möbius transformations $\mathfrak{c}_1, \mathfrak{c}_2$ sending $\mathbb{D}$ onto $\mathbb{H}$, such that

- If $x = \mathfrak{c}_1(a) \in \mathbb{R}$, $y = \mathfrak{c}_1(b) \in \mathbb{R}$, and $\mathfrak{c}_2(h(a)), \mathfrak{c}_2(h(b)) \in \mathbb{R}$, then $\varphi := \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}$ satisfies

$$\vartheta_h(e) = \frac{1}{2}\log h'(a)h'(b) - \log\frac{h(a)-h(b)}{a-b} = \frac{1}{2}\log \varphi'(x)\varphi'(y) - \log\frac{\varphi(x)-\varphi(y)}{x-y} \in \mathbb{R};$$

- If $\mathfrak{c}_1(a) = \mathfrak{c}_2(h(a)) = \infty$, and $y = \mathfrak{c}_1(b) \in \mathbb{R}$, then $\varphi := \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}$ satisfies

$$\vartheta_h(e) = \frac{1}{2}\log \varphi'(y)\varphi'(\infty),$$

where $\varphi'(\infty) := \tilde{\varphi}'(0)$ for $\tilde{\varphi} = \iota \circ \varphi \circ \iota$.

3.3.5 Diamond shear in terms of log Λ-length

In this section we show a simple relation (Lemma 3.3.30) between diamond shear coordinates of a circle homeomorphism and the log Λ-lengths, which are coordinates on the *decorated Teichmüller space*, denoted $\widetilde{T(\mathbb{D})}$ introduced by Penner. See [Pen93; Pen02; Pen12]. This relation will play an essential role in Section 3.6.4.

We view the Teichmüller space $T(\mathbb{D})$ as a space of *tessellations* by identifying

$$h \in T(\mathbb{D}) \quad \Longleftrightarrow \quad h(\mathfrak{F}).$$

A point in $\widetilde{T(\mathbb{D})}$ is a tessellation $h(\mathfrak{F})$ plus a "decoration", namely a choice of horocycle at each vertex $h(v) \in h(V)$. The $\log\Lambda$-*length* along an edge (a, b) decorated with horocycles ρ_a, ρ_b at $a, b \in \mathbb{T}$ is

$$\log\Lambda(\rho_a, \rho_b) := \delta/2$$

where δ is the signed hyperbolic distance between $\rho_a \cap e$ and $\rho_b \cap e$ with the convention that if $\rho_a \cap \rho_b = \emptyset$, then δ is positive. Since hyperbolic distances are invariant under Möbius transformations, so are $\log\Lambda$-lengths. Moreover, the $\log\Lambda$-length can be computed in terms of the Euclidean diameters of the horocycles in $\mathbb{H}$.

Lemma 3.3.27 (See [Pen12, Chap. 1, Sec. 1.4, Cor. 4.6])**.** *Let $x = \mathfrak{c}(a), y = \mathfrak{c}(b)$, $c_x = \mathfrak{c}(\rho_a)$, and $\rho_y = \mathfrak{c}(\rho_b)$. If $x, y \neq \infty$,*

$$\log\Lambda(\rho_a, \rho_b) = \log|x - y| - \frac{1}{2}\log d_x d_y$$

where d_x, d_y are the Euclidean diameters of ρ_x, ρ_y respectively. If $x = \infty$, then

$$\log\Lambda(\rho_a, \rho_b) = \frac{1}{2}\log H - \frac{1}{2}\log d_y$$

where $\rho_\infty = \{z \in \mathbb{H} \colon \mathrm{Im}(z) = H\}$. (Recall that horocycles at ∞ are horizontal lines, and we call H the Euclidean height of ρ_∞.) Further, for any Möbius transformation $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{R})$, if $A(x) \in \mathbb{R}$ then $A(\rho_x)$ has diameter $d_x/(cx+d)^2 = A'(x)\, d_x$. If $A(x) = \infty$, then $A(\rho_x)$ has height $1/(c^2 d_x)$.

The Farey tessellation admits a very special decoration by a collection of horocycles called the *Ford circles*. In the Farey tessellation $\mathfrak{c}(\mathfrak{F})$ of $\mathbb{H}$, the Ford circle $\rho_{p/q}$ is the horocycle centered at p/q with Euclidean diameter $1/q^2$. The Ford circle at infinity is the line $\{\mathrm{Im}(z) = 1\}$. See Figure 3.7. This collection of horocycles has the property that:

- $\rho_{p/q}$ is tangent to $\rho_{r/s}$ if $(p/q, r/s) \in \mathfrak{c}(E)$,
- $\rho_{p/q}$ is disjoint from $\rho_{r/s}$ if $(p/q, r/s) \notin \mathfrak{c}(E)$.

The Ford circles in the disk are the pullback of the Ford circles in the upper half plane by $\mathfrak{c}$. For all $e = (a, b) \in E$, the Farey tessellation decorated by the Ford circles has $\log\Lambda(\rho_a, \rho_b) = 0$. Starting from the Ford circle decoration of $\mathfrak{F}$ for the identity map, one can define a section σ from homeomorphisms $h \in \mathscr{P}_0$ to $\widetilde{T(\mathbb{D})}$ motivated by Lemma 3.3.27.

Definition 3.3.28 (See [Pen12, p. 119-120])**.** If $h \in \mathscr{P}_0$, we define a section σ to $\widetilde{T(\mathbb{D})}$ as follows:

- $\sigma(\mathrm{Id}_{\mathbb{T}})$ assigns the Ford circle as the horocycle at each $v \in V$. In $\mathbb{H}$, this means the horocycle at $p/q \in \mathbb{R}$ has diameter q^{-2} and the horocycle at ∞ has height 1.

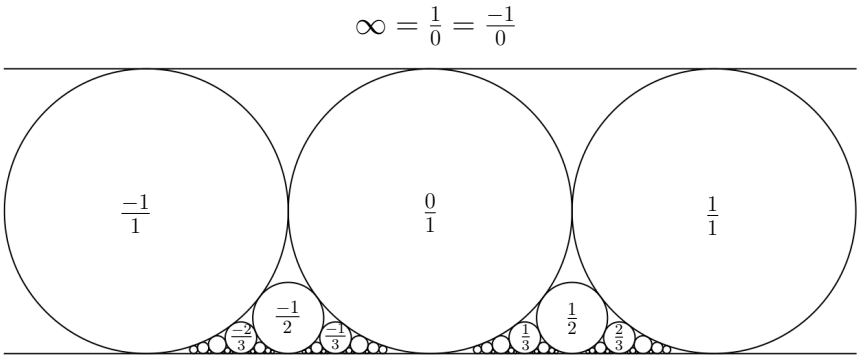


Figure 3.7: A few generations of Ford circles between -1 and $+1$ labeled by their center points in $\mathfrak{c}(\mathfrak{F})$.

- For any other $h \in \mathscr{H}$, let $\varphi = \mathfrak{c} \circ h \circ \mathfrak{c}^{-1} : \mathbb{R} \to \mathbb{R}$. At each point $\varphi(p/q) \in \mathbb{R}$, $\sigma(h)$ assigns the horocycle with diameter $|\varphi'(p/q)|q^{-2}$. At $\varphi(\infty) = \infty$, σ assigns the horocycle of height $|\varphi'(\infty)|^{-1}$ at ∞. Here recall that $\varphi'(\infty) = \tilde{\varphi}'(0)$ for $\tilde{\varphi}(x) = -1/\varphi(-1/x)$.

Remark 3.3.29. The section σ was first introduced in [MP98] for diffeomorphisms and is defined for any homeomorphism $h : \mathbb{T} \to \mathbb{T}$ fixing $\pm 1, \mathfrak{i}$ such that h is differentiable at all rational points of the circle. By Lemma 3.3.23 ii), σ is well-defined if $h \in \mathscr{P}_0$. It is also not hard to see that the decoration does not depend on the choice of conformal map $\mathbb{D} \to \mathbb{H}$, we choose $\mathfrak{c}$ for simplicity.

When $e = (a, b) \in E$ and $h \in \mathscr{P}_0$, we define the notation

$$\log \Lambda_h(e) := \log \Lambda(\rho_a, \rho_b)$$

where ρ_a, ρ_b are the horocycles at $h(a), h(b)$ chosen by the section σ. Using this, we can describe the relationship between diamond shear coordinates and $\log \Lambda$-lengths.

Lemma 3.3.30. *If $h \in \mathscr{P}_0$, then for any $e = (a, b) \in E$,*

$$\vartheta_h(e) = -\log \Lambda_h(e).$$

Proof. Let $\varphi := \mathfrak{c} \circ h \circ \mathfrak{c}^{-1}$ which is a homeomorphism of $\mathbb{R}$ fixing ∞. Choose $(a, b) \in E$ and let $\mathfrak{c}(a) = x, \mathfrak{c}(b) = y$. If $x, y \neq \infty$, let d_x, d_y denote the diameters of the Ford circles at x, y. We have

$$\log \Lambda_{\mathrm{Id}}(e) = \log |x - y| - \frac{1}{2} \log d_x d_y = 0,$$

which follows from direct computation or the fact that the Ford circles at x and y are tangent. Using Definition 3.3.28 and Lemma 3.3.27,

$$\begin{aligned}
\log \Lambda_h(e) &= \log \Lambda_h(e) - \log \Lambda_{\mathrm{Id}}(e) \\
&= \log |\varphi(x) - \varphi(y)| - \frac{1}{2} \log |\varphi'(x)||\varphi'(y)| d_x d_y - \log |x - y| + \frac{1}{2} \log d_x d_y \\
&= -\frac{1}{2} \log |\varphi'(x)\varphi'(y)| + \log \frac{|\varphi(x) - \varphi(y)|}{|x - y|}.
\end{aligned}$$

If $x = \infty = \varphi(x)$, then the horocycle at $\varphi(x)$ has height $H = \varphi'(\infty)^{-1}$, $y \in \mathbb{Z}$ and hence $d_y = 1$. We have

$$\log \Lambda_h(e) = \frac{1}{2} \log |\varphi'(\infty)|^{-1} - \frac{1}{2} \log |\varphi'(y)| d_y = -\frac{1}{2} \log \varphi'(y)\varphi'(\infty).$$

In both cases, we have $\vartheta_h(e) = -\log \Lambda_h(e)$ by Remark 3.3.26. □

3.4 Relation to Hölder classes

Here we relate the class $\mathscr{H}$ of homeomorphisms with square-summable diamond shears to the Hölder class $\mathscr{C}^{1,\alpha}$ defined in (3.2).

Theorem 3.4.1. *We have $\mathscr{C}^{1,\alpha} \subset \mathscr{H}$ if and only if $\alpha > 1/2$.*

For comparison, recall Lemma 3.2.9, which says analogously that $\mathscr{C}^{1,\alpha} \subset \mathrm{WP}(\mathbb{T})$ if and only if $\alpha > 1/2$. The "only if" direction of Theorem 3.4.1 will follow from the fact that $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$ and the latter does not contain $\mathscr{C}^{1,1/2}$, see Theorem 3.5.5. In this section, we show that $\mathscr{C}^{1,\alpha} \subset \mathscr{H}$ for $\alpha > 1/2$.

For $(a, b) \in E$, let $\ell(a, b)$ denote the arclength of the arc in $\mathbb{T}$ from a to b containing the child of a, b (which is the shorter of the two arcs between a and b, we call it a *Farey segment*). We call these lengths the *Farey lengths*.

Proposition 3.4.2. *The Farey lengths are ℓ^r-summable if and only if $r > 1$, e.g.*

$$\sum_{(a,b)\in E} \ell(a, b)^r < \infty$$

if and only if $r > 1$.

Proof. Since $\sum_{(a,b)\in E_n \smallsetminus E_{n-1}} \ell(a, b) = 2\pi$ for all n, the sum diverges for $r \leq 1$.

Now we show the sum converges when $r > 1$. We sort the sum over edges (a, b) by the endpoint of the earlier generation. Note that by Lemma 3.2.1 there are no edge between vertices of the same generation except e_0. Therefore,

$$\sum_{(a,b)\in E} \ell(a, b)^r = \pi^r + \sum_{a\in V} \sum_{b\in \mathrm{child}(a)} \ell(a, b)^r.$$

The π^r term corresponds to $e_0 = (-1, 1)$. We refer to the set of vertices

$$\mathrm{child}(a) = \{b \in V : (a, b) \in E,\ \mathrm{gen}(b) > \mathrm{gen}(a)\}$$

as the *children of a*.

Let $\Gamma_1, ..., \Gamma_4$ denote the closed quarter circles with vertices $1, \mathfrak{i}, -1, -\mathfrak{i}$. All the Farey segments (other than the one corresponding to $e_0 = (-1, 1)$) are contained in one of these closed quarter circles, so it suffices to show that the lengths in $\Gamma := \Gamma_3$, the arc from -1 to $-\mathfrak{i}$, are ℓ^r-summable. The inverse Cayley transform $\mathfrak{c}^{-1}$ sends $[0, 1]$ onto Γ and is Lipschitz

on $[0,1]$. The image $\mathfrak{c}(V \cap \Gamma)$ consists of the rational points between 0 and 1. If λ is the Lipschitz constant of $\mathfrak{c}^{-1}|_{[0,1]}$, then

$$\sum_{a \in V \cap \Gamma} \sum_{\text{child}(a)} \ell(a,b)^r \leq \lambda^r \sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} \sum_{\frac{p'}{q'} \in \text{child}(\frac{p}{q})} \left| \frac{p}{q} - \frac{p'}{q'} \right|^r.$$

If $k_1/m_1 < p/q < k_2/m_2$ are the parents of p/q, the children of p/q are of the form

$$\frac{p'}{q'} = \frac{k_i + np}{m_i + nq} \qquad i = 1,2 \quad \text{and} \quad n \in \mathbb{N}_{\geq 1}$$

by Lemma 3.2.2. Hence the distances we must bound are of the form

$$\left| \frac{p}{q} - \frac{np + k_i}{nq + m_i} \right|^r = \left| \frac{pm_i - qk_i}{q(nq + m_i)} \right|^r.$$

Lemma 3.2.2 also shows that $|pm_i - qk_i|^r = 1$, hence

$$\left| \frac{pm_i - qk_i}{q(nq + m_i)} \right|^r \leq \frac{1}{q^{2r} n^r}.$$

Therefore

$$\begin{aligned} \sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} \sum_{\frac{p'}{q'} \in \text{child}(\frac{p}{q})} \left| \frac{p}{q} - \frac{p'}{q'} \right|^r &= \sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} \sum_{\substack{n \geq 1 \\ i=1,2}} \left| \frac{p}{q} - \frac{np + k_i}{nq + m_i} \right|^r \leq \sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} \sum_{n \geq 1} \frac{2}{q^{2r} n^r} \\ &= \sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} 2q^{-2r} \zeta(r). \end{aligned}$$

There are $\phi(q)$ rational points in $[0,1]$ with denominator q, where ϕ is Euler's ϕ function. Since $\phi(q) \leq q$,

$$\sum_{\frac{p}{q} \in \mathfrak{c}(V \cap \Gamma)} 2q^{-2r} \zeta(r) = 2\zeta(r) \sum_{q \geq 1} \phi(q) q^{-2r} \leq 2\zeta(r) \sum_{q \geq 1} q^{1-2r}$$

which is finite exactly if $r > 1$. □

Theorem 3.4.1 follows straightforwardly from Proposition 3.4.2 and the analytic description of diamond shear coordinates.

Proof of Theorem 3.4.1. By Lemma 3.3.23 iii), $h \in \mathscr{C}^{1,\alpha}$ implies $s_h \in \mathscr{P}_0$. For any closed interval $I \subset \mathbb{T}$ which is a proper subset of $\mathbb{T}$, we can find Möbius transformations $\mathfrak{c}_1, \mathfrak{c}_2 : \mathbb{D} \to \mathbb{H}$ such that $\mathfrak{c}_1(I), \mathfrak{c}_2(h(I))$ are bounded intervals of $\mathbb{R}$.

Define $\varphi_I := \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1} : \mathbb{R} \to \mathbb{R}$. By Remark 3.3.26, for all $(a,b) \in E$ such that $a, b \in I$,

$$\vartheta_h(e) = \frac{1}{2} \log \varphi_I'(x) \varphi_I'(y) - \log \frac{\varphi_I(x) - \varphi_I(y)}{x - y}$$

where $x = \mathfrak{c}_1(a), y = \mathfrak{c}_1(b)$.

By the mean value theorem applied to φ_I, there exists $z \in (x, y)$ such that $\varphi_I'(z) = \frac{\varphi_I(x) - \varphi_I(y)}{x - y}$. Thus

$$\vartheta_h(e) = \frac{1}{2}(\log \varphi_I'(x) - \log \varphi_I'(z)) + \frac{1}{2}(\log \varphi_I'(y) - \log \varphi_I'(z)).$$

Further, $c := \mathfrak{c}_1^{-1}(z)$ is contained in the interval $I(a, b) \subset \mathbb{T}$.

Since $\log h'$ is α-Hölder and since $\mathfrak{c}_2(h(I))$ and $\mathfrak{c}_1(I)$ are bounded intervals, $\log \varphi_I'$ is α-Hölder. Thus there is a constant $C > 0$ such that for all $s, t \in \mathfrak{c}_1(I)$,

$$|\log \varphi_I'(s) - \log \varphi_I'(t)| \leq C|s - t|^\alpha.$$

Given also that $\mathfrak{c}_1 : I \to \mathfrak{c}_1(I)$ is Lipschitz, it follows that

$$|\vartheta(e)| \leq \frac{C}{2}\left(|x - z|^\alpha + |z - y|^\alpha\right) \leq \frac{K}{2}\left(\ell(a, c)^\alpha + \ell(b, c)^\alpha\right) \leq K\ell(a, b)^\alpha \tag{3.20}$$

for a constant $K = K(\mathfrak{c}_1, I)$ and all $e = (a, b) \in E$ with $a, b \in I$.

We cover $\mathbb{T}$ by the intervals $I_1 = \{e^{\mathrm{i}\vartheta} : \vartheta \in [0, \pi]\}$ and $I_2 = \{e^{\mathrm{i}\vartheta} : \vartheta \in [\pi, 2\pi]\}$. For every $(a, b) \in E$, there exists j such that $a, b \in I_j$. Summing the bounds given in (3.20) using an appropriate interval for each $e \in E$,

$$\sum_{e \in E} \vartheta(e)^2 \leq K^2 \sum_{(a,b)=e \in E} \ell(a, b)^{2\alpha},$$

where $K^2 = \max(K_1^2, K_2^2)$ and K_j is the constant in (3.20) for I_j. By Proposition 3.4.2, this bound is finite if and only if $\alpha > 1/2$. □

3.5 Relation to Weil–Petersson homeomorphisms

We now describe relationships between two classes of homeomorphisms defined in terms of shears, namely $\mathscr{H}$ and $\{h : s_h \in \mathscr{S}\}$, and the Weil–Petersson class $\mathrm{WP}(\mathbb{T})$. To summarize, the main results of this section are that $\mathscr{H} \subset \mathrm{WP}(\mathbb{T}) \subset \{h : s_h \in \mathscr{S}\}$, and the inclusions are strict.

3.5.1 Cell decomposition of $\mathbb{D}$ or $\mathbb{H}$ along $\mathfrak{F}^*$

We say that an embedding of the dual tree of an ideal tessellation of $D = \mathbb{D}$ or $\mathbb{H}$ is *centered* if the vertices of the tree are at the centers of the triangles in the tessellation and *geodesic* if the edges of the tree are geodesics for the hyperbolic metric on D. Given a tessellation, there is a well-defined centered geodesic embedding of its dual tree. Further, since Möbius transformations preserve angles and hyperbolic distances, if $\mathscr{T} = h(\mathfrak{F})$ is a tessellation embedded in D with dual tree $\mathscr{T}^*$ embedded as a centered geodesic tree, then for any Möbius transformation $A : D \to D'$, $A(\mathscr{T}^*)$ is a centered geodesic embedding of the dual tree of $A(\mathscr{T})$ in D'. The complementary region $D \setminus \mathscr{T}$ is a union of disjoint

simply connected regions with piecewise-geodesic boundary, each of which contains exactly one vertex of $V(\mathscr{T})$. We call these regions *cells*.

We embed the dual tree $\mathfrak{F}^*$ as a centered geodesic tree in $\mathbb{D}$. Applying the Cayley map $\mathfrak{c} : \mathbb{D} \to \mathbb{H}$, the centered geodesic embedding of $\mathfrak{F}^*$ in $\mathbb{D}$ is sent to a centered geodesic tree $\mathfrak{c}(\mathfrak{F}^*)$ in $\mathfrak{c}(\mathfrak{F})$ in $\mathbb{H}$. We denote the cells of $\mathfrak{F}$ or $\mathfrak{c}(\mathfrak{F})$ simply by C_v or $C_{\mathfrak{c}(v)}$ for $v \in V$. The group $\mathrm{PSL}(2,\mathbb{Z})$ acts transitively on the set of cells. Further, for any $v \in V$, there is $A \in \mathrm{PSL}(2,\mathbb{Z})$ such that $A \circ \mathfrak{c}$ sends C_v to C_∞. See the pink area of the left-hand side of Figure 3.8 for an illustration of C_∞.

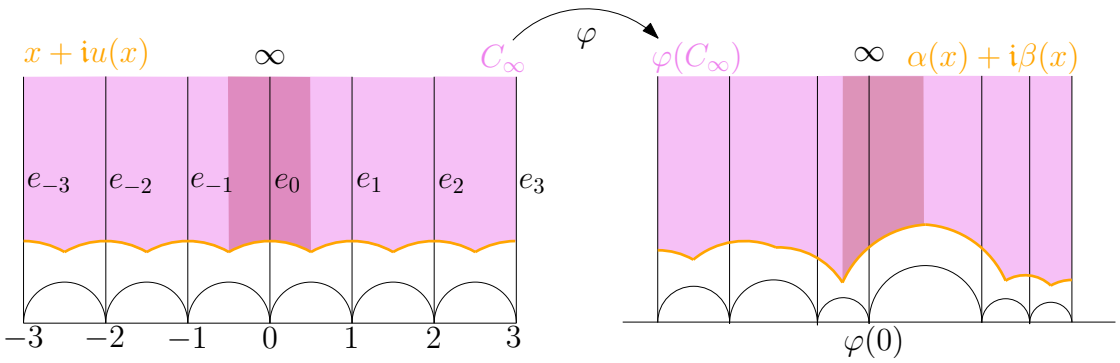


Figure 3.8: The cell at ∞ in the Farey tessellation between -3 and 3 with boundary given by $x + \mathrm{i}u(x)$, and an example of its image under a map φ with boundary given by $\alpha(x) + \mathrm{i}\beta(x)$. The strip A_0 and its image are illustrated in dark pink.

We now describe the cell C_∞ more explicitly. We denote $e_0 = (0, \infty)$, and let $\mathrm{fan}(\infty) = (e_n = (n, \infty))_{n \in \mathbb{Z}}$. Let $a_n = (n, n+1) \in \mathfrak{c}(E)$ and let τ_n be the ideal triangle bounded by $\{e_n, a_n, e_{n+1}\}$. The center of τ_n is

$$c_n = \frac{1}{2} + n + \mathrm{i}\frac{\sqrt{3}}{2}.$$

The geodesic arc connecting c_n, c_{n+1} is an arc of the circle centered at $n+1$ of radius 1. We define $u(x)$ to be the function whose graph is these arcs. Explicitly, on the interval $[n - 1/2, n + 1/2]$, $u(x) = \sqrt{1 - (x-n)^2}$. In terms of $u(x)$, the cell at ∞ is

$$C_\infty = \{x + \mathrm{i}y : y \geq u(x)\}.$$

Note that $u(x)$ is continuous everywhere and differentiable everywhere except the half-integers. We further split C_∞ into infinitely many *strips* A_n, broken at the half-integers

$$A_n = \left\{x + \mathrm{i}y : y \geq u(x),\ n - \frac{1}{2} \leq x \leq n + \frac{1}{2}\right\}.$$

Homeomorphisms $h \in \mathbb{T} \to \mathbb{T}$ act naturally on geodesic centered dual trees. Since h determines the tessellation $h(\mathfrak{F})$, it determines the images of the centers of the triangles and thus determines the centered geodesic tree $h(\mathfrak{F}^*)$. We write $h(C_v)$ for the connected component of $\mathbb{D} \setminus h(\mathfrak{F}^*)$ containing v. We define the corresponding *hyperbolic stretching map* F_h, which is a homeomorphism from $\mathfrak{F}^*$ to $h(\mathfrak{F}^*)$ defined as follows:

- F_h maps the center of a triangle τ of the Farey tessellation to the center of $h(\tau)$;
- F_h linearly stretches the hyperbolic length along each geodesic edge.

Similarly, for $\varphi : \mathbb{R} \to \mathbb{R}$ fixing ∞, let $\varphi(C_\infty)$ denote the image of the cell at ∞ under φ and define similarly the hyperbolic stretching map $F_\varphi : \mathfrak{c}(\mathfrak{F}^*) \to \varphi(\mathfrak{c}(\mathfrak{F}^*))$. We will give an explicit expression for F_φ on ∂C_∞. For this, we define functions $\alpha(x), \beta(x)$ such that

$$F_\varphi(x + \mathfrak{i}u(x)) = \alpha(x) + \mathfrak{i}\beta(x),$$

so that the image cell is

$$\varphi(C_\infty) = \{\alpha(x) + \mathfrak{i}y : y \geq \beta(x)\}$$

and the image strips are

$$\varphi(A_n) = \left\{\alpha(x) + \mathfrak{i}y : y \geq \beta(x),\, n - \frac{1}{2} \leq x \leq n + \frac{1}{2}\right\}.$$

The maps α, β are continuous for all $x \in \mathbb{R}$ and differentiable everywhere except the half-integers. Restricted to $[n-1/2, n+1/2]$, $\alpha(x)+\mathfrak{i}\beta(x)$ is a parametrization of the hyperbolic geodesic connecting $F_\varphi(c_n)$ to $F_\varphi(c_{n+1})$. In particular, $F_\varphi(c_n), F_\varphi(c_{n+1})$ are the centers of the triangles $\{\varphi(n-1), \varphi(n), \infty\}$, $\{\varphi(n), \varphi(n+1), \infty\}$ respectively, meaning that

$$\begin{aligned} F_\varphi(c_n) &= \varphi(n) - \frac{\rho}{2} + \mathfrak{i}\frac{\rho\sqrt{3}}{2}, \qquad \rho = \varphi(n) - \varphi(n-1), \\ F_\varphi(c_{n+1}) &= \varphi(n) + \frac{\lambda}{2} + \mathfrak{i}\frac{\lambda\sqrt{3}}{2}, \qquad \lambda = \varphi(n+1) - \varphi(n). \end{aligned}$$

Using this, we now explicitly compute the functions α, β on a single strip. Translating, it suffices to compute the following:

Lemma 3.5.1. *For $\rho, \lambda > 0$, let $\gamma_{\rho,\lambda}$ denote the geodesic connecting $-\frac{\rho}{2} + \mathfrak{i}\frac{\rho\sqrt{3}}{2}$ to $\frac{\lambda}{2} + \mathfrak{i}\frac{\lambda\sqrt{3}}{2}$. We have $\gamma_{\rho,\lambda}$ is an arc of the circle centered at $\lambda - \rho$ with Euclidean radius $r = \sqrt{\lambda^2 - \lambda\rho + \rho^2}$ and hyperbolic length $\ell = \log\frac{\rho+\lambda+r}{\rho+\lambda-r}$. Let $F_{\rho,\lambda} : \gamma_{1,1} \to \gamma_{\rho,\lambda}$ be the hyperbolic stretching. If $\gamma_{1,1}$ is parametrized by $\gamma_{1,1}(x) = x + \mathfrak{i}u(x)$, then $\gamma_{\rho,\lambda}(x) = F_{\rho,\lambda}(\gamma_{1,1}(x)) = \alpha_{\rho,\lambda}(x) + \mathfrak{i}\beta_{\rho,\lambda}(x)$ where*

$$\alpha_{\rho,\lambda}(x) = \frac{(r+\lambda-\rho)e^\ell\left(\frac{1+x}{1-x}\right)^{\ell/\log 3} - (r-\lambda+\rho)K(\rho,\lambda)^2}{e^\ell\left(\frac{1+x}{1-x}\right)^{\ell/\log 3} + K(\rho,\lambda)^2},$$

$$\beta_{\rho,\lambda}(x) = \frac{2re^{\ell/2}\left(\frac{1+x}{1-x}\right)^{\ell/2\log 3} K(\rho,\lambda)}{e^\ell\left(\frac{1+x}{1-x}\right)^{\ell/\log 3} + K(\rho,\lambda)^2},$$

and $K(\rho,\lambda) = \frac{2r+2\lambda-\rho}{\sqrt{3}\rho}$.

Proof. Direct computations show that $\gamma_{\rho,\lambda}$ is an arc of the circle centered at $\lambda-\rho$ of Euclidean radius $\sqrt{\lambda^2-\lambda\rho+\rho^2}$. On the imaginary axis, the hyperbolic stretching map $C_{a_1,a_2}:\mathfrak{i}\mathbb{R}_+\to\mathfrak{i}\mathbb{R}_+$ that sends $(\mathfrak{i},a_1\mathfrak{i})$ to $(\mathfrak{i},a_2\mathfrak{i})$ is given by

$$C_{a_1,a_2}(z)=\mathfrak{i}\,(z/\mathfrak{i})^{\log a_2/\log a_1}.$$

To compute $F_{\rho,\lambda}$, we conjugate by Möbius transformations that send $\gamma_{1,1}$ and $\gamma_{\rho,\lambda}$ to segments of the form $(\mathfrak{i},a_1\mathfrak{i})$ and $(\mathfrak{i},a_2\mathfrak{i})$ respectively, where $\log a_1=\text{length}(\gamma_{1,1})$ and $\log a_2=\text{length}(\gamma_{\rho,\lambda})$.

Let $B_{\rho,\lambda}:\mathbb{H}\to\mathbb{H}$ be the Möbius transformation such that $B_{\rho,\lambda}(\gamma_{\rho,\lambda})\subset\mathfrak{i}\mathbb{R}_+$ (with negative endpoint of the geodesic containing $\gamma_{\rho,\lambda}$ sent to 0 and positive endpoint sent to ∞) and $B_{\rho,\lambda}(-\frac{\rho}{2}+\mathfrak{i}\frac{\rho\sqrt{3}}{2})=\mathfrak{i}$. To compute the length of $\gamma_{1,1}$, we note that

$$B_{1,1}(z)=\sqrt{3}\,\frac{1+z}{1-z},\qquad B_{1,1}(\frac{1}{2}+\mathfrak{i}\frac{\sqrt{3}}{2})=3\mathfrak{i}.$$

Thus $\text{length}(\gamma_{1,1})=\log 3$. Given this and denoting $\ell=\text{length}(\gamma_{\rho,\lambda})$,

$$F_{\rho,\lambda}(x+\mathfrak{i}u(x))=B_{\rho,\lambda}^{-1}\circ C_{3,e^\ell}\circ B_{1,1}(x+\mathfrak{i}u(x)).$$

To find ℓ, we compute

$$B_{\rho,\lambda}(z)=K(\rho,\lambda)\frac{r-\lambda+\rho+z}{r+\lambda-\rho-z}$$

where $K(\rho,\lambda)$ is the multiplicative constant so that $B_{\rho,\lambda}$ sends $-\frac{\rho}{2}+\frac{\rho\sqrt{3}}{2}\mathfrak{i}$ to $\mathfrak{i}$. For $x+\mathfrak{i}y\in\gamma_{\rho,\lambda}$, we get that

$$\frac{B_{\rho,\lambda}(x+\mathfrak{i}y)}{K(\rho,\lambda)}=\frac{2ry\mathfrak{i}}{(r+\lambda-\rho-x)^2+y^2}.$$

Thus

$$K(\rho,\lambda)=\frac{(r+\lambda-\rho+\rho/2)^2+(\sqrt{3}\rho/2)^2}{2r\sqrt{3}\rho/2}=\frac{2r+2\lambda-\rho}{\sqrt{3}\rho}.$$

Similarly,

$$B_{\rho,\lambda}(\frac{\lambda}{2}+\frac{\sqrt{3}\lambda}{2}\mathfrak{i})=K(\rho,\lambda)\frac{\sqrt{3}\lambda}{2r+\lambda-2\rho}=\frac{\lambda(2r+2\lambda-\rho)}{\rho(2r+\lambda-2\rho)}.$$

From this and simplification, we find

$$\ell=\log\frac{\lambda(2r+2\lambda-\rho)}{\rho(2r+\lambda-2\rho)}=\log\frac{\rho+\lambda+r}{\rho+\lambda-r}.$$

Putting these together we compute $F_{\rho,\lambda}(x+\mathfrak{i}u(x))$. First

$$B_{1,1}(x+\mathfrak{i}u(x))=\sqrt{3}\left(\frac{1+x}{1-x}\right)^{1/2}\mathfrak{i}$$

so then

$$C_{3,e^\ell} \circ B_{1,1}(x + \mathfrak{i}u(x)) = \mathfrak{i}\sqrt{3}^{\ell/\log 3}\left(\frac{1+x}{1-x}\right)^{\ell/2\log 3} = \mathfrak{i}e^{\ell/2}\left(\frac{1+x}{1-x}\right)^{\ell/2\log 3}.$$

Note that

$$B^{-1}_{\rho,\lambda}(z) = \frac{(r+\lambda-\rho)z - K(\rho,\lambda)(r-\lambda+\rho)}{z + K(\rho,\lambda)}$$

so finally

$$F_{\rho,\lambda}(x + \mathfrak{i}u(x)) = \frac{(r+\lambda-\rho)\mathfrak{i}e^{\ell/2}\left(\frac{1+x}{1-x}\right)^{\ell/2\log 3} - K(\rho,\lambda)(r-\lambda+\rho)}{\mathfrak{i}e^{\ell/2}\left(\frac{1+x}{1-x}\right)^{\ell/2\log 3} + K(\rho,\lambda)}.$$

Taking real and imaginary parts completes the proof. □

The following observation follows directly from the explicit formulas of α and β.

Corollary 3.5.2. *The functions $(s,t,x) \mapsto \alpha_{e^s,e^t}(x)$, $\alpha'_{e^s,e^t}(x)$, $\beta_{e^s,e^t}(x)$, and $\beta'_{e^s,e^t}(x)$ are real analytic and bounded on $[-M,M]\times[-M,M]\times[-1/2,1/2]$ for all $M>0$. Moreover, $\alpha'_{\rho,\lambda}(x) > 0$ for all $\rho,\lambda > 0$ and all $x \in [-1/2,1/2]$.*

3.5.2 Proof of $\mathscr{H} \subset \mathbf{WP}(\mathbb{T})$

In this section, we fix a circle homeomorphism $h \in \mathscr{H}$. We explicitly construct a homeomorphism $f : \mathbb{D} \to \mathbb{D}$ which extends h using the cell structure described above. The extension coincides with the hyperbolic stretching map along the centered geodesic tree. We show that f is quasiconformal (Theorem 3.5.3) and then show that the Beltrami coefficient μ of f is L^2-integrable on the disk with respect to the hyperbolic metric (Theorem 3.5.5) which shows that h is Weil–Petersson. The construction is an adaption of a construction of Kahn–Markovic from [KM08] to the infinite tessellation setting.

Construction of an extension $f : \mathbb{D} \to \mathbb{D}$ of $h : \mathbb{T} \to \mathbb{T}$.

Recall that we embed the Farey dual tree $\mathfrak{F}^*$ as a centered geodesic tree. We first define $f|_{\mathfrak{F}^*}$ to be the hyperbolic stretching map from $\mathfrak{F}^* \to h(\mathfrak{F}^*)$, where $h(\mathfrak{F}^*)$ is the centered geodesic tree associated to the tessellation induced by h. Since f sends $\mathfrak{F}^*$ to the centered geodesic tree in $h(\mathfrak{F})$, for any $v \in V$, $f(v) = h(v)$. Since V is dense in $\mathbb{T}$ and h is continuous, if f is also continuous then f extends h.

We now define f on $\mathbb{D} \smallsetminus \mathfrak{F}^*$ cell-by-cell. Choose $v \in V$ and let $\mathrm{fan}(v) = (e_k)_{k\in\mathbb{Z}}$ centered on an arbitrary but fixed middle edge e_0. By Lemma 3.3.16, $h \in \mathscr{P}_0$, so

$$M := \sum_{i=0}^{\infty} s_h(e_i) = -\sum_{i=-1}^{-\infty} s_h(e_i) < \infty.$$

If we choose Möbius transformations $\mathfrak{c}_1, \mathfrak{c}_2 : \mathbb{D} \to \mathbb{H}$ as in Lemma 3.3.23 so that $\mathfrak{c}_1(v) = \infty$, $\mathfrak{c}_2(h(v)) = \infty$, $\mathfrak{c}_1(e_0) = \mathfrak{c}_2(e_0) = (0, \infty)$, and $\mathfrak{c}_1(e_{-1}) = \mathfrak{c}_2(e_{-1}) = (-1, \infty)$, then as in (3.15), (3.16), for all $n \geq 0$, $\mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}$ satisfies

$$\lim_{n\to\infty} \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}(n+1) - \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}(n) = \lim_{n\to\infty} \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}(-n) - \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1}(-n-1) = e^M.$$

We construct an extension ψ on C_∞ for

$$\varphi := e^{-M} \mathfrak{c}_2 \circ h \circ \mathfrak{c}_1^{-1} : \mathbb{R} \to \mathbb{R}.$$

This way, $\varphi(0) = 0$ and $\varphi(\infty) = \infty$, and φ is asymptotic to the identity near ∞. The restriction of ψ to $\partial C_\infty \subset \mathfrak{c}_1(\mathfrak{F}^*)$ is given by the hyperbolic stretching, already studied in the previous section, and we denoted the map by

$$\psi(x + \mathfrak{i}u(x)) = \alpha(x) + \mathfrak{i}\beta(x).$$

We extend ψ to C_∞ by

$$\psi(x + \mathfrak{i}y) := \alpha(x) + \mathfrak{i}(\beta(x) - u(x) + y), \qquad \forall x \in \mathbb{R}, \quad y \geq u(x).$$

With this definition, ψ sends the strip A_n onto $\varphi(A_n)$ and is a homeomorphism $C_\infty \to \varphi(C_\infty)$. Conjugating back to $\mathbb{D}$, we obtain a continuous extension $\mathfrak{c}_2^{-1} \circ (e^M \psi) \circ \mathfrak{c}_1$ of $f|_{\mathfrak{F}^*}$ to $C_v \cup \mathfrak{F}^*$. This construction applied to all $v \in V$ gives a continuous map $f : \mathbb{D} \to \mathbb{D}$ extending h.

Theorem 3.5.3. *If $h \in \mathscr{H}$, then the extension $f : \mathbb{D} \to \mathbb{D}$ constructed above is quasiconformal.*

Proof. Choose $v \in V$ and consider the map $\psi = e^{-M} \mathfrak{c}_2 \circ f \circ \mathfrak{c}_1^{-1}$. On $C_\infty = \mathfrak{c}_1(C_v)$,

$$\psi(x + \mathfrak{i}y) := \alpha(x) + \mathfrak{i}(\beta(x) - u(x) + y), \qquad y \geq u(x).$$

Since ψ differs from f by Möbius transformations, ψ being K-quasiconformal on C_∞ is equivalent to f being K-quasiconformal on C_v (See e.g. [Nag88, Sec. 1.2.8]). Since u, α, β are differentiable almost everywhere, so is ψ. A direct computation shows that the Beltrami coefficient of ψ on C_∞ is given by

$$\mu(x + \mathfrak{i}y) = \frac{\psi_{\bar{z}}}{\psi_z}(x + \mathfrak{i}y) = \frac{\alpha'(x) - 1 + \mathfrak{i}(\beta'(x) - u'(x))}{\alpha'(x) + 1 + \mathfrak{i}(\beta'(x) - u'(x))}.$$

Another direct computation shows that

$$|\mu(x + \mathfrak{i}y)|^2 = 1 - \frac{4\alpha'(x)}{(\alpha'(x) + 1)^2 + (\beta'(x) - u'(x))^2}. \tag{3.21}$$

It is clear that the ratio in (3.21) takes value in $(0, 1]$, we now show it to be uniformly bounded away from 0. For $x \in [n - 1/2, n + 1/2]$, the infimum of the ratio is a continuous function of $\rho = \varphi(n) - \varphi(n-1)$ and $\lambda = \varphi(n+1) - \varphi(n)$ by Lemma 3.5.1. Moreover,

$$\varphi(n+1) - \varphi(n) = \exp(-M) \exp(\sum_{i=0}^{n} s_h(e_i)) = \exp(\sum_{i=-\infty}^{n} s_h(e_i)) = \exp(-p_{s_h, v}(e_n+)).$$

Here we use the convention that if $n \leq -1$ then $\sum_{i=0}^{n} s_h(e_i) = -\sum_{i=n}^{-1} s_h(e_i)$. Since $p_{s_h,v}(e_{n+})$ is uniformly bounded for all $v \in V$, $n \in \mathbb{Z}$ by Lemma 3.3.16, from the continuity, we obtain that there exists $k < 1$ independent of the cell chosen, such that $\|\mu\|_{\infty, C_\infty} \leq k$. This shows that f is K-quasiconformal in $\mathbb{D} \setminus \mathfrak{F}^*$, where $K = (1+k)/(1-k)$. Points and C^1 Jordan curves are quasiconformally removable, see [FM07, Thm. 3.1.3], thus $\mathfrak{F}^*$ is quasiconformally removable. Since f is a homeomorphism of $\mathbb{D}$, we obtain that f is a K-quasiconformal extension of h to $\mathbb{D}$. □

To show that $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$, we need the following lemma.

Lemma 3.5.4. *Let $\psi_{s,t}(x+\mathfrak{i}y) = \alpha_{e^s,e^t}(x) + \mathfrak{i}(\beta_{e^s,e^t}(x) - u(x) + y)$ for $x \in [-1/2, 1/2]$ and $y \geq u(x)$, and let $\mu_{s,t}$ be the Beltrami coefficient of $\psi_{s,t}$. For any $\varepsilon > 0$ small, there is a constant $C = C(\varepsilon) > 0$ such that if $|s|, |t| < \varepsilon$,*

$$|\mu_{s,t}(x+\mathfrak{i}y)| \leq C(|s| + |t|)$$

for all $x \in [-1/2, 1/2]$ and all $y \geq u(x)$.

Proof. Recall that

$$\mu_{s,t}(x+\mathfrak{i}y) = \frac{\alpha'_{e^s,e^t}(x) - 1 + \mathfrak{i}(\beta'_{e^s,e^t}(x) - u'(x))}{\alpha'_{e^s,e^t}(x) + 1 + \mathfrak{i}(\beta'_{e^s,e^t}(x) - u'(x))}. \tag{3.22}$$

Using the explicit formulas for α, β in Lemma 3.5.1, notice that $\mu_{0,0}(x+\mathfrak{i}y) \equiv 0$. Since the modulus of the denominator of μ is bounded below by 1,

$$|\mu_{s,t}(x+\mathfrak{i}y)| \leq |\alpha'_{e^s,e^t}(x) - 1| + |\beta'_{e^s,e^t(x)} - u'(x)|.$$

By Corollary 3.5.2, the right hand side of (3.22) is analytic in s, t, x on the appropriate interval. Further, when $(s,t) = (0,0)$, the right hand side is 0. Fixing $x \in [-1/2, 1/2]$ and expanding around $(s,t) = (0,0)$, we find that for $|s|, |t| < \varepsilon$,

$$|\mu_{s,t}(x+\mathfrak{i}y)| \leq C(\varepsilon, x)(|s| + |t|).$$

By Corollary 3.5.2, $C(\varepsilon, x)$ can be chosen to be a continuous function of $x \in [-1/2, 1/2]$ and hence achieves a maximum value $C = C(\varepsilon)$, which completes the proof. □

Theorem 3.5.5. *If $h \in \mathscr{H}$, then the extension f constructed above has Beltrami coefficient μ such that*

$$\int_{\mathbb{D}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\, \mathrm{d}A(z) \lesssim \sum_{v \in V} \sum_{e \in \mathrm{fan}(v)} p^2_{s_h,v}(e_+) < \infty.$$

In particular, $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$.

Proof. Choose $v \in V$ and again consider $\psi = e^{-M}\mathfrak{c}_2 \circ f \circ \mathfrak{c}_1^{-1}$. For $x + \mathfrak{i}y \in C_\infty = \mathfrak{c}_1(C_v)$, recall that

$$\mu(x+\mathfrak{i}y) = \frac{\alpha'(x) - 1 + \mathfrak{i}(\beta'(x) - u'(x))}{\alpha'(x) + 1 + \mathfrak{i}(\beta'(x) - u'(x))}.$$

Further, for $x \in [n-1/2, n+1/2]$, $\alpha(x) = \alpha_{\lambda_{n-1},\lambda_n}(x-n)$ and $\beta(x) = \beta_{\lambda_{n-1},\lambda_n}(x-n)$, where

$$\lambda_n = \varphi(n+1) - \varphi(n) = \exp(-p_{s_h,v}(e_n+)).$$

Fix a small threshold $\varepsilon > 0$. By Lemma 3.3.16, there are only finitely many (v,e) such that $|p_{s_h,v}(e+)| > \varepsilon$. We say that a strip A_n is bad if $|p_{s_h,v}(e+)| > \varepsilon$ for $e = e_n$ or $e = e_{n-1}$. Let $N(\varepsilon) < \infty$ be the number of bad strips across all cells. Since each strip A_n is contained in an ideal triangle, $\mathrm{Area}_{\mathrm{hyp}}(A_n) \le \pi$. Since $|\mu(x+\mathfrak{i}y)| < 1$,

$$\int_{\text{bad strips}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) < \pi N(\varepsilon).$$

On the other hand, if $|p_{s_h,v}(e_n+)|, |p_{s_h,v}(e_{n-1}+)| < \varepsilon$, then by Lemma 3.5.4,

$$|\mu(x+\mathfrak{i}y)| \le C(\varepsilon)(|p_{s_h,v}(e_n+)| + |p_{s_h,v}(e_{n-1}+)|)$$

for all $x+\mathfrak{i}y \in A_n$. Therefore

$$\int_{C_v \smallsetminus \text{bad strips}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \le 4\pi C(\varepsilon)^2 \sum_{e \in \mathrm{fan}(v)} p_{s_h,v}(e+)^2.$$

Summing over $v \in V$, adding back the bad strips, and applying Lemma 3.3.16, we get that

$$\int_{\mathbb{D}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \le N(\varepsilon)\pi + 4\pi C(\varepsilon)^2 \sum_{v \in V} \sum_{e \in \mathrm{fan}(v)} p^2_{s_h(e+)} < \infty.$$

By Theorem 3.2.8, $h \in \mathrm{WP}(\mathbb{T})$. □

3.5.3 Counterexample: an element of WP(𝕋) which is not in $\mathscr{H}$

We saw in the last section that $\mathscr{H} \subset \mathrm{WP}(\mathbb{T})$. It is straightforward to see that the reverse inclusion does not hold.

Proposition 3.5.6. $\mathrm{WP}(\mathbb{T}) \not\subset \mathscr{P}$. *As a consequence,* $\mathrm{WP}(\mathbb{T}) \not\subset \mathscr{H}$.

Proof. By Lemma 3.3.23 i), if $h \in \mathscr{P}$ then h has left and right derivatives at all rational points. A Weil–Petersson homeomorphism may not have left or right derivatives since functions in the Sobolev space $H^{1/2}$ may not have left or right limits everywhere, so $\mathrm{WP}(\mathbb{T}) \not\subset \mathscr{P}$ from Definition 3.2.7. By Corollary 3.3.17, $\mathrm{WP}(\mathbb{T}) \not\subset \mathscr{H}$. □

We illustrate an explicit example of a homeomorphism $\varphi : \mathbb{R} \to \mathbb{R}$ which is Weil–Petersson but not in $\mathscr{P}$.

Example 3.5.7. Consider $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi(x) = x \log|x| - x$ for $|x| > 2$, and is smooth on $\mathbb{R}$. On one hand, $\varphi(x)$ grows faster than linear functions as $x \to \infty$ or $x \to -\infty$, and hence $\varphi \notin \mathscr{P}$.

To show that φ is Weil–Petersson, we show[1] that $u(x) := \log \varphi'(x)$ is in $H^{1/2}(\mathbb{R})$ by showing that it is the trace of a map $f : \mathbb{H} \to \mathbb{H}$ which has finite Dirichlet energy (by the classical Douglas formula, see, e.g. [VW20a, Eq.(2.2), (2.3)]). We compute

$$u(x) = \log \log |x|$$

for x outside $(-2, 2)$. Hence u is the trace of a smooth function f on $\overline{\mathbb{H}}$ which takes the values $f(z) = \log \log |z|$ on $\overline{\mathbb{H}} \smallsetminus \mathbb{D}(0, 2)$. The gradient of f satisfies $|\nabla f(z)| = \frac{1}{r \log r}$ if $|z| = r > 2$. Thus,

$$\int_{\mathbb{H} \smallsetminus \mathbb{D}(0,2)} |\nabla f(z)|^2 \, \mathrm{d}A(z) = \int_0^\pi \int_2^\infty \frac{r}{r^2 (\log r)^2} \, \mathrm{d}r \mathrm{d}\theta = \pi \Big[-\frac{1}{\log r} \Big]_{r=2}^\infty = \frac{\pi}{\log 2} < \infty,$$

and φ is Weil–Petersson.

We can also see explicitly that $\varphi \notin \mathscr{P}$ by computing its shears $s_\varphi(e_n)$ for $e_n = (n, \infty) \in \mathrm{fan}(\infty)$. For $n \geq 0$,

$$\begin{aligned} \varphi(n+1) - \varphi(n) &= (n+1) \log(n+1) - n \log n - 1 \\ &= (n+1)[\log n + \log(1 + 1/n)] - \log n - 1 \\ &= \log n + (n+1) \left(\frac{1}{n} - \frac{1}{2n^2} + O(\frac{1}{n^3}) \right) - 1 \\ &= \log n + \frac{1}{2n} + O(\frac{1}{n^2}). \end{aligned}$$

Analogously,

$$\varphi(n) - \varphi(n-1) = \log n - \frac{1}{2n} + O(\frac{1}{n^2}).$$

Therefore

$$s_\varphi(e_n) = \log \frac{\varphi(n+1) - \varphi(n)}{\varphi(n) - \varphi(n-1)} = \log \left(1 + \frac{1}{n \log n} + O(\frac{1}{n^2}) \right) = \frac{1}{n \log n} + O(\frac{1}{n^2}).$$

In particular, $s_\varphi(e_n)$ is not summable. Note however that $s_\varphi(e_n)$ is square-summable (as it must be by Theorem 3.5.12).

3.5.4 Convergence in $\mathscr{H}$ implies convergence in WP($\mathbb{T}$)

In this section we show that convergence in $\mathscr{H}$ is stronger than convergence in the Weil–Petersson metric (Corollary 3.5.10).

Theorem 3.5.8. *Suppose that $g, h \in \mathscr{H}$, and let $q = g \circ h^{-1}$. Then there exists $C(\varepsilon, h) > 0$ such that q has a quasiconformal extension f_q with Beltrami coefficient μ satisfying*

$$\int_{\mathbb{D}} \frac{|\mu(z)|^2}{(1 - |z|^2)^2} \, \mathrm{d}A(z) \leq C(\varepsilon, h) \sum_{e \in E} (\vartheta_g(e) - \vartheta_h(e))^2,$$

for any $g \in \mathscr{H}$ such that $\sum_{e \in E} (\vartheta_g(e) - \vartheta_h(e))^2 \leq \varepsilon$.

[1] This characterization of the Weil–Petersson class is obtained in [ST20, Thm. 2.2].

Proof. Let f_g, f_h be the quasiconformal extensions of g, h respectively constructed in the section above. We will show that the extension $f_q := f_g \circ f_h^{-1}$ of q has Beltrami coefficient satisfying the bound above cell by cell.

Choose $v \in V$, and let $\mathrm{fan}(v) = (e_n)_{n\in\mathbb{Z}}$. For g, choose Cayley maps $\mathfrak{c}_1, \mathfrak{c}_2$ as in Section 3.5.2 so that $\mathfrak{c}_1(v) = \infty$, $\mathfrak{c}_2(g(v)) = \infty$, and $\mathfrak{c}_1(e_0) = \mathfrak{c}_2(e_0) = (0, \infty)$, $\mathfrak{c}_1(e_{-1}) = \mathfrak{c}_2(e_{-1}) = (-1, \infty)$. Define $\psi_g = e^{-M_g}\mathfrak{c}_2 \circ f_g \circ \mathfrak{c}_1^{-1}$, where $M_g = \sum_{n=0}^{\infty} s_g(e_n)$, and let α_g, β_g be such that $\psi_g(x + \mathfrak{i}u(x)) = \alpha_g(x) + \mathfrak{i}\beta_g(x)$. Analogously choose Cayley maps $\mathfrak{c}_3, \mathfrak{c}_4$ to define $\psi_h, \alpha_h, \beta_h$. Since ψ_g, ψ_h both fix ∞, $\psi_q := \psi_g \circ \psi_h^{-1}$ fixes ∞. In particular, ψ_q maps $C_\infty(h(\mathfrak{F}))$ onto $C_\infty(g(\mathfrak{F}))$.

The boundary curve of $C_\infty(h(\mathfrak{F}))$ is given by

$$x + \mathfrak{i}u_h(x), \qquad u_h(x) = \beta_h \circ \alpha_h^{-1}(x).$$

We define α_q, β_q so that

$$\psi_q(x + \mathfrak{i}u_h(x)) = \alpha_q(x) + \mathfrak{i}\beta_q(x).$$

Since $\psi_q = \psi_g \circ \psi_h^{-1}$,

$$\begin{aligned}\alpha_q(x) &= \alpha_g \circ \alpha_h^{-1}(x)\\ \beta_q(x) &= \beta_g \circ \alpha_h^{-1}(x).\end{aligned}$$

Hence the Beltrami coefficient μ of ψ_q on $C_\infty(h(\mathfrak{F}))$ is

$$\mu(x + \mathfrak{i}y) = \frac{\alpha_g' - \alpha_h' + \mathfrak{i}(\beta_g' - \beta_h')}{\alpha_g' + \alpha_h' + \mathfrak{i}(\beta_g' - \beta_h')} \circ \alpha_h^{-1}(x).$$

For $w = \alpha_h^{-1}(x) \in [n - 1/2, n + 1/2]$, $\alpha_h'(w)$ is a continuous function of $p_{s_h,v}(e_{n+})$ and $p_{s_h,v}(e_{n-1+})$. By Lemma 3.3.16, $p_{s_h,v}(e_{n+})$ is uniformly bounded for all $v \in V, n \in \mathbb{Z}$. Further, by Corollary 3.5.2, $\alpha_h'(w) > 0$. Combining this with the fact that α_h' is continuous, there exists a constant $K(h) > 0$ independent of cell such that $\alpha_h'(w) \geq K(h)$. Therefore

$$|\mu(x + \mathfrak{i}y)| \leq \frac{|\alpha_g' - \alpha_h'| + |\beta_g' - \beta_h'|}{K(h)} \circ \alpha_h^{-1}(x).$$

Fix a threshold $\varepsilon > 0$. From the assumption and Lemma 3.3.16, there are only finitely many (v, e) such that $|p_{s_h,v}(e+) - p_{s_g,v}(e+)| > \varepsilon$. Let $N(\varepsilon)$ be the number of strips A_n across all cells where $|p_{s_g,v}(e_{n+}) - p_{s_h,v}(e_{n+})|$ or $|p_{s_g,v}(e_{n-1+}) - p_{s_h,v}(e_{n-1+})|$ is larger than ε. Any other strip A_n we call a good strip.

By Corollary 3.5.2, α', β' are analytic functions of the (s, t, w). Expanding around any (s_0, t_0), there is a constant C depending on ε and (s_0, t_0) such that if $|s - s_0|, |t - t_0| < \varepsilon$, then

$$\begin{aligned}|\alpha'_{e^s,e^t}(w) - \alpha'_{e^{s_0},e^{t_0}}(w)| &\leq C(|s - s_0| + |t - t_0|)\\ |\beta'_{e^s,e^t}(w) - \beta'_{e^{s_0},e^{t_0}}(w)| &\leq C(|s - s_0| + |t - t_0|)\end{aligned}$$

for all $w \in [-1/2, 1/2]$. For each good strip A_n, we use this expansion for $(s_0, t_0) = (p_{s_h,v}(e_{n-1}+), p_{s_h,v}(e_n+))$. By Lemma 3.3.16 applied to h these are uniformly bounded for all $v \in V, n \in \mathbb{Z}$, so we can take the constant C to depend only on ε and h to find

$$|\alpha_g'(w) - \alpha_h'(w)| \le C(|p_{s_g,v}(e_{n-1}+) - p_{s_h,v}(e_{n-1}+)| + |p_{s_g,v}(e_n+) - p_{s_h,v}(e_n+)|)$$
$$|\beta_g'(w) - \beta_h'(w)| \le C(|p_{s_g,v}(e_{n-1}+) - p_{s_h,v}(e_{n-1}+)| + |p_{s_g,v}(e_n+) - p_{s_h,v}(e_n+)|).$$

Therefore there is another constant $C(\varepsilon, h)$ such that for all $x + \mathrm{i}y \in A_n$,

$$|\mu(x+\mathrm{i}y)| \le C(\varepsilon, h)\Big(\big|p_{s_g,v}(e_{n-1}+) - p_{s_h,v}(e_{n-1}+)\big| + \big|p_{s_g,v}(e_n+) - p_{s_h,v}(e_n+)\big| \Big).$$

Every strip $h(A_n)$ is a geodesic triangle, so it is contained in an ideal triangle and its hyperbolic area is bounded by π. Summing over $v \in V, n \in \mathbb{Z}$, adding back the bad strips, and integrating, we find

$$\int_{\mathbb{D}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \le \pi N(\varepsilon) + 4\pi C(\varepsilon, h) \sum_{v\in V} \sum_{e \in \mathrm{fan}(v)} (p_{s_g,v}(e+) - p_{s_h,v}(e+))^2.$$

Note that

$$N(\varepsilon) \le \frac{1}{\varepsilon^2} \sum_{v\in V} \sum_{e \in \mathrm{fan}(v)} (p_{s_g,v}(e+) - p_{s_h,v}(e+))^2.$$

Applying Lemma 3.3.16 with $s = s_g - s_h$, we obtain using Cauchy-Schwarz

$$\begin{aligned}
&\sum_{v\in V} \sum_{e \in \mathrm{fan}(v)} (p_{s_g,v}(e+) - p_{s_h,v}(e+))^2 \\
&\le 2\sum_{e\in E} (\vartheta_g(e) - \vartheta_h(e))^2 + \sum_{e\sim e'} 2(\vartheta_g(e) - \vartheta_h(e))^2 + 2(\vartheta_g(e') - \vartheta_h(e'))^2 \\
&= 6\sum_{e\in E} (\vartheta_g(e) - \vartheta_h(e))^2,
\end{aligned}$$

which completes the proof. □

Lemma 3.5.9. *Suppose that $h, (h_n)_{n\ge1}$ are Weil–Petersson homeomorphisms fixing $\pm1, \mathrm{i}$, and let μ_n be the Beltrami coefficient of a quasiconformal extension of $h_n \circ h^{-1}$ in $\mathbb{D}$. If*

$$\lim_{n\to\infty} \int_{\mathbb{D}} \frac{|\mu_n(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) = 0,$$

then h_n converges to h in the Weil–Petersson metric.

This result must be well-known. For readers' convenience we sketch the proof using several lemmas from [TT06]. The results in [TT06] are stated using μ defined in the outer disk $\mathbb{D}^* = \{z \in \mathbb{C} \colon |z| > 1\}$, by pre-composing quasiconformal maps by $z \mapsto 1/\overline{z}$ we can easily translate those results to $\mathbb{D}$.

Proof. Theorem I.3.8 in [TT06] shows that $\mathrm{WP}(\mathbb{T})$ is a topological group. Therefore, it suffices to show the claim when $h = \mathrm{Id}_{\mathbb{T}}$.

We write

$$\|\mu\|_2^2 := \int_{\mathbb{D}} \frac{|\mu(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \text{ and } \|\mu\|_\infty = \sup_{z\in\mathbb{D}} |\mu(z)|.$$

We note that two measurable Beltrami differentials μ, ν with $\|\mu\|_\infty < 1$ and $\|\nu\|_\infty < 1$ are said to be *equivalent*, if they are the Beltrami coefficients of a quasiconformal extension of the same circle homeomorphism fixing $\pm 1, \mathrm{i}$.

If $\|\mu\|_\infty < 1$, the Bers embedding of the equivalence class of μ, denoted as $B([\mu])$, is a holomorphic function $\phi \in \mathscr{Q}(\mathbb{D}^*)$. See Section 3.6.2 for more details. If furthermore $\|\mu\|_2 < \infty$, [TT06] shows that

$$\|\phi\|^2_{A_2(\mathbb{D}^*)} := \int_{\mathbb{D}^*} |\phi|^2(1-|z|^2)^2\,\mathrm{d}A(z) < \infty.$$

Now let μ_n be a family of Beltrami differentials such that $\lim_{n\to\infty}\|\mu_n\|_2 = 0$, [TT06, Lem. I.2.9] implies that there exists $C > 0$, such that

$$\|B([\mu_n])\|_{A_2(\mathbb{D}^*)} \le C\|\mu_n\|_2 \to 0.$$

Since $B|_{T_0(\mathbb{D})}$ is a biholomorphic mapping of Hilbert manifolds [TT06, Thm. I.2.13], where

$$T_0(\mathbb{D}) = \{[\mu]\colon \|\mu\|_2 < \infty, \|\mu\|_\infty < 1\} \simeq \mathrm{WP}(\mathbb{T})$$

and the identification $\mathrm{WP}(\mathbb{T}) \to T_0(\mathbb{D})$ is the map from a circle homeomorphism h to the equivalence class of Beltrami coefficients of any quasiconformal extension $[\mu]$, it shows $[\mu_n]$ converges to $[0]$ in $T_0(\mathbb{D})$ which is by definition equivalent to h_n converges to $\mathrm{Id}_{\mathbb{T}}$ for the Weil–Petersson metric. □

Lemma 3.5.9 and Theorem 3.5.8 applied to $g = h_n$ combine to give the following corollary.

Corollary 3.5.10. *Suppose that $h, (h_n)_{n\ge1} \in \mathscr{H}$ with diamond shear coordinates ϑ, ϑ_n respectively. If $\lim_{n\to\infty}\sum_{e\in E}(\vartheta_n(e)-\vartheta(e))^2 = 0$, then h_n converges to h in the Weil–Petersson metric.*

3.5.5 Square summable shears

In this section we prove three results about square summable shear functions $\mathscr{S}$. First we show that $\mathscr{S}$ is not contained in $\mathrm{Homeo}(\mathbb{T})$, nor does it contain $\mathrm{QS}(\mathbb{T})$.

Proposition 3.5.11. *There exists $s \in \mathscr{S}$ such that s does not induce a homeomorphism. Conversely, there exists $h \in \mathrm{QS}(\mathbb{T})$ such that its shear $s_h \notin \mathscr{S}$.*

Proof. To prove this result, we just need to exhibit two examples and apply the suitable conditions from [Šar10; Šar21].

For $\mathscr{S} \not\subset \mathrm{Homeo}(\mathbb{T})$: let $(e_n)_{n\in\mathbb{Z}}$ denote the fan of edges incident to 1 in counterclockwise order. By [Šar10, Theorem C], if $s : E \to \mathbb{R}$ has

$$\sum_{n=1}^{\infty} \exp(s(e_1) + \cdots + s(e_n)) < \infty,$$

then s does *not* induce a homeomorphism. Given this, we choose $1/2 < \alpha < 1$, and define the shear function $s : E \to \mathbb{R}$ so that $s(e_n) = -\frac{1}{n^\alpha}$ for $n \geq 1$, and is 0 on all other edges in E. On one hand, since $\alpha > 1/2$,

$$\sum_{e\in E} s(e)^2 = \sum_{n\geq 1} \frac{1}{n^{2\alpha}} < \infty,$$

so $s \in \mathscr{S}$. On the other hand, $s(e_1)+\ldots s(e_n) = -H(n,\alpha)$ is minus the generalized harmonic number with parameter α, and since $\alpha < 1$,

$$\sum_{n=1}^{\infty} \exp(s(e_1) + \cdots + s(e_n)) = \sum_{n=1}^{\infty} \exp(-H(n,\alpha)) < \infty.$$

Therefore s does not induce a homeomorphism.

For $\mathrm{QS}(\mathbb{T}) \not\subset \mathscr{S}$: if $s : E \to \mathbb{R}$ has $s(e) = 1$ for infinitely many edges $e \in E$, then $s \notin \mathscr{S}$. In particular, consider $s : E \to \mathbb{R}$ where for each n, there is one edge e connecting vertices of generations $2n$ and $2n+1$ of $\mathfrak{F}$ where $s(e) = 1$, and all other shears are 0. Clearly this includes infinitely many edges e where $s(e) = 1$. Note also that this has the property that every fan contains either zero or one edge with nonzero shear. One can check the condition for a shear to induce a quasisymmetric homeomorphism (from [Šar21; Šar10], and included here as Theorem 3.3.20) is satisfied with $C = e$. □

Finally we show the following inclusion:

Theorem 3.5.12. *If $h \in \mathrm{WP}(\mathbb{T})$, then $s_h \in \mathscr{S}$.*

Note that the reverse statement is not true: a circle homeomorphism h with shear coordinate s_h supported on a single edge is has $s_h \in \mathscr{S}$ but does not satisfied the finite balanced condition from Definition 3.3.3, so Lemma 3.3.4 shows that such a homeomorphism is not Weil–Petersson. To show the inclusion we use the following property of Weil–Petersson homeomorphisms due to Wu.

Theorem 3.5.13 (See [Wu11])**.** *Suppose $h \in \mathrm{WP}(\mathbb{T})$. Then there is a constant $C = C(h) > 0$ such that for any pairwise disjoint collection of quads $Q_1, \ldots, Q_n$ with vertices on $\mathbb{T}$,*

$$\sum_{i=1}^{n} d^2(\mathrm{cr}(Q_i), \mathrm{cr}(h(Q_i)))\lambda(\mathrm{cr}(Q_i)) < C$$

where $d(\cdot,\cdot)$ is the hyperbolic metric on $\mathbb{C} \smallsetminus \{-1, 0\}$ and $\lambda(x) = \exp(d(1,x) - |\log x|/2)$.

We clarify that quads are considered as open sets bounded by hyperbolic geodesics. In other words, quads sharing only boundary edges or vertex are also considered as disjoint. Note that $\lambda(1) = 1$ and since the hyperbolic metric has smooth conformal factor with respect to the Euclidean metric, there exists $a > 0$ such that

$$d(1, e^s) = a|s| + O(s^2), \quad s \to 0. \tag{3.23}$$

Proof of Theorem 3.5.12. Let $h \in \mathrm{WP}(\mathbb{T})$. If Q is a Farey quad, then $\mathrm{cr}(Q) = 1$. Therefore for an infinite sequence $Q_1, Q_2, \ldots$ of pairwise disjoint Farey quads, Theorem 3.5.13 implies that

$$\sum_{i=1}^{\infty} d^2(1, \mathrm{cr}(h(Q_i))) < C, \tag{3.24}$$

as C is independent of the number of quads.

Farey quads Q_e are in one-to-one correspondence with dual edges $e^* \in E^*$. If two dual edges e^*, f^* are disjoint and do not share a vertex, then the quads Q_e and Q_f are disjoint. Since $\mathfrak{F}^*$ is a trivalent tree, the dual edges E^* can be colored red, blue, or green so that no two edges of the same color intersect. Let R, B, G be the collections of dual edges colored red, blue, or green respectively. Each of R, B, G corresponds to a collection of disjoint Farey quads, so (3.24) applies. On the other hand, $E^* = R \cup B \cup G$, and hence

$$\sum_{e \in E} d^2(1, \mathrm{cr}(h(Q_e))) < 3C. \tag{3.25}$$

Since this sum converges, for any $\varepsilon > 0$ there are only finitely many $e \in E$ such that $|\mathrm{cr}(h(Q_e)) - 1| > \varepsilon$. Recall that $s_h(e) = \log \mathrm{cr}(h(Q_e))$. Hence, there are only finitely many edges $e \in E$ such that $|s_h(e)| > \varepsilon$. We now choose ε such that $|s| < \varepsilon$ implies $|s| < 2a^{-1} d(1, e^s)$ by (3.23). In particular, if $|s_h(e)| < \varepsilon$, $s_h(e)^2 < 4a^{-2} d(1, \mathrm{cr}(h(Q_e)))^2$. We obtain $s_h \in \mathscr{S}$ from (3.25). □

3.6 Weil–Petersson metric tensor and symplectic form

3.6.1 Finite shears and Zygmund functions

The tangent space to the universal Teichmüller space

$$T(\mathbb{D}) := \mathrm{QS}(\mathbb{T})/\mathrm{Möb}(\mathbb{T}) \simeq \{h : \mathbb{T} \to \mathbb{T}, \text{ quasisymmetric and fixing } -1, \mathfrak{i}, 1\}$$

at the origin $\mathrm{Id}_{\mathbb{T}}$ consists of all Zygmund functions on the unit circle that vanish at 1, $\mathfrak{i}$ and -1 (see [GL00, Sec. 16.6]). More precisely, consider a differentiable path (for the Banach manifold structure of $T(\mathbb{D})$) $t \mapsto h_t$ with $t \in (-\varepsilon, \varepsilon)$ of quasisymmetric maps such that $h_0 = \mathrm{Id}_{\mathbb{T}}$. Then $\frac{\mathrm{d}}{\mathrm{d}t} h_t(x)|_{t=0} = u(x)$ is a Zygmund function on the unit circle and conversely, every Zygmund function on the unit circle is the tangent vector to a differentiable path of quasisymmetric maps at $\mathrm{Id}_{\mathbb{T}} \in T(\mathbb{D})$.

For any $h \in T(\mathbb{D})$ we can identify $T_h T(\mathbb{D})$ with the space of Zygmund functions on the circle by pullback, meaning if $t \mapsto h_t$, $t \in (-\varepsilon, \varepsilon)$ is a differentiable path of quasisymmetric

maps fixing 1, $\mathfrak{i}$, and -1 with $h_0 = h$, then we identify it with the Zygmund function $u(x) = \frac{\mathrm{d}}{\mathrm{d}t} h_t \circ h^{-1}(x)|_{t=0}$.

The set of *finitely supported shear functions* is

$$\mathscr{F} := \{\dot{s} : E \to \mathbb{R} : \dot{s}(e) \neq 0 \text{ for finitely many } e \in E\}.$$

For any $h \in T(\mathbb{D})$ with shear coordinate s_h and $\dot{s} \in \mathscr{F}$, the path of shear functions $s_h + t \cdot \dot{s}$ for $t \in (-\varepsilon, \varepsilon)$ induces a path of homeomorphisms $(h_t)_{t\in(-\varepsilon,\varepsilon)} \subset T(\mathbb{D})$. Using the developing algorithm of Section 3.3.2, h_t is of the form $h_t = H_t \circ h$, where H_t is a piecewise-Möbius homeomorphism with breakpoints in $h(V)$. Another way to view this is that the shear function of H_t on $h(\mathfrak{F})$ (instead of $\mathfrak{F}$)

$$S_t(h(e)) := s(H_t \circ h(Q_e), h(e)) - s(h(Q_e), h(e)) = s_h(e) + t \cdot \dot{s}(e) - s_h(e) = t \cdot \dot{s}(e).$$

is finitely supported. The first author [Šar06] proved that $h_t \circ h^{-1}$ is a differentiable path in t for the Banach manifold structure of $T(\mathbb{D})$. We obtain

$$u = \frac{\mathrm{d}(h_t \circ h^{-1})}{\mathrm{d}t}\Big|_{t=0} = \frac{\mathrm{d}H_t}{\mathrm{d}t}\Big|_{t=0} \in T_{\mathrm{Id}}T(\mathbb{D})$$

is a piecewise $\mathfrak{psu}(1,1)$ vector field with break points in $h(V)$.

We now compute u explicitly in terms of the shear and the computation is often simpler in the half plane model. By conjugating by the Cayley transform $\mathfrak{c}$, the differentiable path of quasisymmetric maps of $\mathbb{T}$ that fixes 1, $\mathfrak{i}$ and -1 is transformed to a differentiable path $t \mapsto \varphi_t$ with $t \in (-\varepsilon, \varepsilon)$ of quasisymmetric maps of $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ that fix -1, 0 and ∞, namely in

$$T(\mathbb{H}) := \{\varphi : \widehat{\mathbb{R}} \to \widehat{\mathbb{R}}, \text{ quasisymmetric and fixing } -1, 0, \infty\}.$$

If $\varphi_0 = \varphi$, then $\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t \circ \varphi^{-1}(x)|_{t=0} = u(x)$ is a Zygmund function on $\mathbb{R}$ that vanishes at -1 and 0 and satisfies $|u(x)| = O(|x| \log |x|)$ as $|x| \to \infty$.

Let $e \in E$ and $h \in T(\mathbb{D})$. Let $a, b \in \widehat{\mathbb{R}}$ such that $(a, b) = \mathfrak{c}(h(e))$. Let $\varphi_t : \widehat{\mathbb{R}} \to \widehat{\mathbb{R}}$ be the path of normalized homeomorphisms conjugate to the circle homeomorphism H_t with shear coordinates $t \cdot \dot{s}_e$ on $h(\mathfrak{F})$, where $\dot{s}_e(e) = 1$ and $\dot{s}_e(e') = 0$ for all $e' \in E$, $e' \neq e$. We define

$$u_{(a,b)} := \frac{\mathrm{d}(\mathfrak{c} \circ H_t \circ \mathfrak{c}^{-1})}{\mathrm{d}t}\Big|_{t=0}. \tag{3.26}$$

Example 3.3.8 or [Šar06] gives the following explicit formulas for $u_{(a,b)}$.

When $a \geq 0$ and $b = \infty$

$$u_{(a,\infty)}(x) = \begin{cases} x - a, & \text{for } x > a \\ 0, & \text{for } x \leq a; \end{cases}$$

for $a \leq -1$ and $b = \infty$

$$u_{(-\infty,a)}(x) = \begin{cases} -(x - a), & \text{for } x < a \\ 0, & \text{for } x \geq a; \end{cases}$$

and for $a < b$, such that the open interval $(a, b) \subset \mathbb{R}$ does not contain -1 or 0,

$$u_{(a,b)}(x) = \begin{cases} \frac{(x-a)(x-b)}{a-b}, & \text{for } a < x < b \\ 0, & \text{otherwise.} \end{cases}$$

Since we assumed that $h(\mathfrak{F})$ contains the triangle $(-1, \mathfrak{i}, 1)$ and $(a, b) = \mathfrak{c}(h(e))$, the above cases cover all possible scenarios.

More generally, suppose that $\dot{s} \in \mathscr{F}$ is supported on $\{e_1, ..., e_n\} \subset E$ and let $\varphi_t : \widehat{\mathbb{R}} \to \widehat{\mathbb{R}}$ be the homeomorphism conjugate to the circle homeomorphism of shear coordinate $s_h + t \cdot \dot{s}$. By developing and the chain rule

$$u = \frac{\mathrm{d}(\varphi_t \circ \varphi^{-1})}{\mathrm{d}t}\Big|_{t=0} = \sum_{j=1}^{n} \dot{s}(e_j) u_{\mathfrak{c}(h(e_j))}. \tag{3.27}$$

Note that the above formula which gives a Zygmund function in terms of a shear function does not extend to the case of a shear function with infinite support. The first author [Šar13] proved that a summation along each fan followed by the sum over all fans is a correct notion for extending the above formula.

Definition 3.6.1. For each $h \in T(\mathbb{D})$, we define the linear operator $\Omega_h : \mathscr{F} \to T_{\mathrm{Id}} T(\mathbb{H})$ by

$$\Omega_h(\dot{s}) := \sum_{j=1}^{n} \dot{s}(e_j) u_{\mathfrak{c}(h(e_j))}.$$

3.6.2 Finite shears and harmonic differentials

A point $h \in T(\mathbb{D})$ (or its conjugate $\varphi \in T(\mathbb{H})$) can be represented by an equivalence class of Beltrami coefficients in $\mathbb{H}$, which consists of $\mu \in L_1^\infty(\mathbb{H})$ satisfying $\|\mu\|_\infty < 1$ and $\mu = \bar{\partial} w / \partial w$ for some quasiconformal extension $w : \mathbb{H} \to \mathbb{H}$ of φ. A differentiable path in $T(\mathbb{H})$ is represented by a differentiable path of Beltrami coefficients with respect to the L^∞-norm. By taking derivative of this path with respect to L^∞-norm, we conclude that a tangent vector to $T(\mathbb{H})$ at the identity is represented by an equivalence class of Beltrami differentials of $\dot{\mu} \in L^\infty(\mathbb{H})$ (for example, see [GL00]). A special representative of the equivalence class is given by the Ahlfors-Weill section.

More precisely, let $\dot{\mu} \in L^\infty(\mathbb{H})$ and $|\varepsilon| < 1/\|\dot{\mu}\|_\infty$. We define w_ε to be the solution of the Beltrami equation

$$\bar{\partial} w_\varepsilon(z) = \begin{cases} \varepsilon \dot{\mu}(z) \partial w_\varepsilon(z), & z \in \mathbb{H} \\ 0, & z \in \mathbb{H}^* \end{cases}$$

normalized to fix $-1, 0, \infty$. Here $\mathbb{H}^*$ denotes the lower half plane. Then,

$$\widehat{u}(z) := \frac{\mathrm{d} w_\varepsilon(z)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} = -\frac{z(z+1)}{\pi} \iint\limits_{\mathbb{H}} \frac{\dot{\mu}(\zeta) \mathrm{d}\xi \mathrm{d}\eta}{\zeta(\zeta+1)(\zeta-z)} \tag{3.28}$$

for all $z \in \mathbb{C}$ (see [GL00, Sec. 6.5]). Note that $\widehat{u}(z)$ is holomorphic in the lower half-plane.

The *Bers embedding*

$$B : T(\mathbb{H}) \to \mathscr{Q}(\mathbb{H}^*) = \{\phi : \mathbb{H}^* \to \mathbb{C} \text{ holomorphic and } \sup_{z\in\mathbb{H}^*} |\phi(z)|\,\mathrm{Im}(z)^2 < \infty\}$$

is given by $[\mu] \mapsto S[w|_{\mathbb{H}^*}]$, where w solves the Beltrami equation

$$\bar\partial w(z) = \begin{cases} \mu(z)\partial w(z), & z \in \mathbb{H} \\ 0, & z \in \mathbb{H}^* \end{cases}$$

and $S[w] = \frac{w'''}{w'} - \frac{3}{2}(\frac{w''}{w'})^2$ is the Schwarzian derivative of w. The Nehari bound shows that $S[w] \in \mathscr{Q}(\mathbb{H}^*)$. Moreover, μ and ν represent the same element in $T(\mathbb{H})$ if and only if they give the same $S[w|_{\mathbb{H}^*}]$. Therefore the map B is well-defined and is an embedding.

The derivative of B at the origin of $T(\mathbb{H})$ evaluated at the tangent vector represented by an infinitesimal Beltrami differential $\dot\mu$ is given by

$$\phi(z) := (\mathrm{d}B)_{\mathrm{Id}}([\dot\mu])(z) = \frac{\mathrm{d}S[w_\varepsilon](z)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} = \widehat{u}'''(z), \qquad z \in \mathbb{H}^*. \tag{3.29}$$

The Ahlfors-Weill section is the *harmonic Beltrami differential* $\dot\mu_u$ in the equivalence class $[\dot\mu]$ representing a tangent vector u (a Zygmund vector field on $\mathbb{R}$) at the origin of $T(\mathbb{H})$ and is given by

$$\dot\mu_u(z) := -2y^2\phi(\bar z) \tag{3.30}$$

where $z = x + \mathrm{i}y \in \mathbb{H}$.

Our goal is to express $\phi(z)$ in terms of the infinitesimal shear function. Since $(\mathrm{d}B)_{\mathrm{Id}}$ is linear and by equation (3.27), it is enough to compute $(\mathrm{d}B)_{\mathrm{Id}}([\dot\mu_{(a,b)}])$ where $\dot\mu_{(a,b)}$ is the harmonic Beltrami differential corresponding to $u_{(a,b)}$ defined in (3.26) and computed explicitly. Extend $\dot\mu_{(a,b)}$ to $\mathbb{C}$ such that $\dot\mu_{(a,b)}(z) = \overline{\dot\mu_{(a,b)}(\bar z)}$. Then we have for $x \in \mathbb{R}$

$$u_{(a,b)}(x) = -\frac{1}{\pi}\iint_{\mathbb{C}} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta = -\frac{2}{\pi}\,\mathrm{Re}\iint_{\mathbb{H}} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta,$$

where $R(x,\zeta) = \frac{x(x+1)}{\zeta(\zeta+1)(\zeta-x)}$ and $\zeta = \xi + \mathrm{i}\eta$. The Hilbert transform for $u_{(a,b)}(x)$ on $\mathbb{R}$ is given by the formula

$$Hu_{(a,b)}(x) = \frac{1}{\pi}\,\mathrm{p.v.}\int_{-\infty}^{\infty} u_{(a,b)}(\xi)R(x,\xi)\mathrm{d}\xi.$$

An application of Stokes' theorem gives, for $x \in \mathbb{R}$,

$$Hu_{(a,b)}(x) = \frac{2\mathrm{i}}{\pi}\iint_{\mathbb{H}} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta + \mathrm{i}u_{(a,b)}(x)$$

and

$$Hu_{(a,b)}(x) = -\frac{2\mathrm{i}}{\pi}\iint_{\mathbb{H}^*} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta - \mathrm{i}u_{(a,b)}(x).$$

By adding the above two equations we obtain

$$Hu_{(a,b)}(x) = \frac{\mathrm{i}}{\pi}\iint_{\mathbb{H}} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta - \frac{\mathrm{i}}{\pi}\iint_{\mathbb{H}^*} \dot\mu_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta$$

and together with the above formula for $u_{(a,b)}$ gives

$$u_{(a,b)}(x) + \mathfrak{i}Hu_{(a,b)}(x) = -\frac{2}{\pi}\iint_{\mathbb{H}} \dot{\mu}_{(a,b)}(\zeta)R(x,\zeta)\,\mathrm{d}\xi\mathrm{d}\eta.$$

By replacing x with $z \in \mathbb{C}$ in the above integral, we obtain the function $2\widehat{u}_{(a,b)}(x)$ where $\widehat{u}_{(a,b)}$ is defined in (3.28) with $\dot{\mu} = \dot{\mu}_{(a,b)}$, that is holomorphic in $\mathbb{H}^*$ and whose $\bar{\partial}$ derivative in $\mathbb{H}$ is $2\dot{\mu}_{(a,b)}$. A direct computation of the Hilbert transform (see [Šar13, Section 3]) and extending $u_{(a,b)}(x) + \mathfrak{i}Hu_{(a,b)}(x)$ to a holomorphic function in $\mathbb{H}^*$ gives (up to an addition of a linear polynomial)

$$\widehat{u}_{(a,b)}(z) = \frac{\mathfrak{i}}{2\pi}\frac{(z-a)(z-b)}{a-b}\log\frac{z-b}{z-a}$$

for (a,b) with $a < b \neq \infty$, and

$$\widehat{u}_{(a,b)}(z) = -\frac{\mathfrak{i}}{2\pi}(z-a)\log(z-a)$$

for $e_j = (a,\infty)$.

In the formulas above, $\frac{z-b}{z-a} \in \mathbb{H}^*$ when $z \in \mathbb{H}^*$. The natural logarithm $\log z$ for $z \in \mathbb{H}^*$ has the imaginary part in $[-\pi, 0]$ with $-\pi$ corresponding to the negative axis. When $x \in [a,b]$, then $\frac{x-b}{x-a}$ is on the negative real axis and the imaginary part of the logarithm is $-\pi$. Using (3.29) we obtained the following formula.

Theorem 3.6.2. *Let $\dot{s} \in \mathscr{F}$ with support $\{e_1, \dots e_n\} \subset E$ and $h \in T(\mathbb{D})$. Let $\dot{\mu}$ be any Beltrami differential representing the Zygmund vector field $\Omega_h(\dot{s})$ (Definition 3.6.1). The infinitesimal Bers embedding of $\dot{\mu}$ is given by*

$$(\mathrm{d}B)_{\mathrm{Id}}(\dot{\mu})(z) = \frac{\mathfrak{i}}{2\pi}\sum_{j=1}^{n}\dot{s}(e_j)\frac{(a_j-b_j)^2}{(z-a_j)^2(z-b_j)^2}, \qquad z \in \mathbb{H}^* \tag{3.31}$$

where $(a_j, b_j) = \mathfrak{c}(h(e_j))$.

Note that the right-hand side of (3.31) is symmetric in a_j and b_j, therefore we do not require $a_j < b_j$. When a_j or $b_j = \infty$, the ratio is understood as the limit as a_j or $b_j \to \infty$.

3.6.3 Weil–Petersson metric on $\mathscr{H}$

In this section we compute the Weil–Petersson metric tensor on $\mathscr{H}$ (see Theorem 3.6.8, Corollary 3.6.10). Recall that $\mathscr{H}$ is equipped with the topology induced by the ℓ^2 norm, so for any $h \in \mathscr{H}$, the tangent space at h is

$$\mathfrak{h} := T_h\mathscr{H} = \{\dot{\vartheta} : \sum_{e\in E}\dot{\vartheta}(e)^2 < \infty\}.$$

For $\dot{\vartheta} \in \mathfrak{h}$, the path h_t defined by

$$\vartheta_{h_t}(e) = \vartheta_h(e) + t\cdot\dot{\vartheta}(e) \qquad \forall e \in E$$

is contained in $\mathscr{H}$ for $t \in (-\varepsilon, \varepsilon)$, and has tangent vector $\dot{\vartheta} \in \mathfrak{h}$. Since the coordinate-change map Φ from diamond shears to shears is linear (see Equation (3.7)), $\mathfrak{h}$ can also be written in terms of infinitesimal shears as $\{\dot{s} = \Phi(\dot{\vartheta}) : \sum_{e\in E}\dot{\vartheta}(e)^2 < \infty\}$.

Remark 3.6.3. We have seen that the tangent space for quasisymmetric homeomorphisms consists of Zygmund functions, where the notion of *a differentiable path* uses the Teichmüller metric on $T(\mathbb{D})$. It is known that quasisymmetric homeomorphisms and Zygmund functions have *different* characterizations in terms of shears [Šar13]. In particular, when an infinitesimal shear $\dot{s}$ corresponding to a Zygmund vector field has infinite support, the one-parameter family $\{t\dot{s}(e) : e \in E\}$ is not necessarily contained in $T(\mathbb{D})$. (In fact, some might not even be induced by circle homeomorphisms.) On the other hand, $\mathscr{H}$ is defined in terms of diamond shears, and we are using its ℓ^2 topology in diamond shears, so the identification of $\mathfrak{h}$ with $\mathscr{H}$ is automatic.

In the half-plane model $\mathbb{H}$, the Weil–Petersson Riemannian pairing of two Zygmund vector fields u_1 and u_2 is given by

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = \operatorname{Re} \iint_{\mathbb{H}} \dot{\mu}_{u_1}(z)\overline{\dot{\mu}_{u_2}(z)}\frac{1}{y^2}\,\mathrm{d}x\mathrm{d}y, \qquad z = x + \mathfrak{i}y. \tag{3.32}$$

We say that $u \in T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$ if $\|u\|^2_{\mathrm{WP}} = \langle u, u\rangle_{\mathrm{WP}} < \infty$. In terms of Fourier coefficients [NV90]

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = 2\pi \operatorname{Re}\sum_{n\geq 2}(n^3 - n)\tilde{u}_{1,n}\tilde{u}_{2,-n}, \qquad \text{where } u_j = \sum_{n\in\mathbb{Z}} \tilde{u}_{j,n}e^{\mathfrak{i}n\theta}\frac{\partial}{\partial\theta}, \quad \tilde{u}_{j,n} = \overline{\tilde{u}_{j,-n}} \in \mathbb{C}.$$

This shows that $T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T}) = H^{3/2}(\mathbb{T})$, the $H^{3/2}$ Sobolev space of vector fields on $\mathbb{T}$.

We show first that for $h \in \mathscr{H}$, $\dot{\vartheta} \in \mathfrak{h}$ induces a vector in $T_h\,\mathrm{WP}(\mathbb{T}) \simeq T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$ (where the identification is the isometry given by the right-composition by h^{-1}).

Lemma 3.6.4. *Fix $h \in \mathscr{H}$. Let $\dot{\vartheta}$ be a function $E \to \mathbb{R}$ with finite support and h_t be the Weil–Petersson homeomorphism induced by the diamond shear function $\vartheta_h + t\cdot\dot{\vartheta}$. We write $u = \mathrm{d}(h_t \circ h^{-1})/\mathrm{d}t|_{t=0} \in T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$. There exists $C(h) > 0$, such that*

$$\|u\|_{\mathrm{WP}} \leq C(h)\|\dot{\vartheta}\|_{\mathfrak{h}}.$$

From Proposition 3.3.12 it is clear that u is a piecewise $\mathfrak{psu}(1,1)$ vector field, $C^{1,1}$ regular, with finitely many break points all in $h(V)$. This implies that $\|u\|_{\mathrm{WP}} < \infty$ as $C^{1,1} \subset H^{3/2}$. The point of the lemma is the quantitative bound of in terms of $\dot{\vartheta}$ which implies the following:

Corollary 3.6.5. *The linear map $\dot{\vartheta} \mapsto u$ in Lemma 3.6.4 extends by continuity to a bounded linear operator $\Xi_h : \mathfrak{h} \simeq T_h\mathscr{H} \to T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})(\simeq T_h\,\mathrm{WP}(\mathbb{T}))$.*

Remark 3.6.6. By linearity, if $\Phi(\dot{\vartheta})$ is a finitely supported infinitesimal shear function, then $\Xi_h(\dot{\vartheta}) = \mathfrak{c}^*\Omega_h(\Phi(\dot{\vartheta}))$, where Ω_h is as in Definition 3.6.1 and $\mathfrak{c}^*$ is the pull-back map sending Zygmund vector fields on $\mathbb{R}$ to Zygmund vector fields on $\mathbb{T}$.

Proof of Lemma 3.6.4. We use the quasiconformal extension of $h_t \circ h^{-1}$ as in Section 3.5 and let μ_t be the associated Beltrami differential. By fixing $\varepsilon = 1$ and t small enough, Theorem 3.5.8 shows that there exists $C(h) > 0$,

$$\iint_{\mathbb{D}} \frac{4|\mu_t(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \leq C(h)t^2\sum_{e\in E}\dot{\vartheta}(e)^2.$$

From the explicit expression of our quasiconformal extension we see that μ_t depends on t analytically, and therefore we have $\mu_t(z) = t \cdot \dot{\mu}(z) + O(t^2)$. Letting $t \to 0$ we obtain the bound

$$\iint_{\mathbb{D}} \frac{4|\dot{\mu}(z)|^2}{(1-|z|^2)^2}\,\mathrm{d}A(z) \leq C(h) \sum_{e \in E} \dot{\vartheta}(e)^2 = C(h)\|\dot{\vartheta}\|_{\mathfrak{h}}^2.$$

However, $\dot{\mu}$ is not a harmonic Beltrami differential. Let $\dot{\mu}_u$ be the corresponding harmonic Beltrami differential (in the half-plane model) as defined in (3.30), we have $\dot{\mu}_u(z) = -2y^2(\mathrm{d}B)_{\mathrm{Id}}([\dot{\mu}])(\bar{z}) =: -2y^2\phi(\bar{z})$. By (3.29), denoting the pushforward of $\dot{\mu}$ to $\mathbb{H}$ also by $\dot{\mu}$,

$$\|u\|_{\mathrm{WP}}^2 = \iint_{\mathbb{H}} |\dot{\mu}_u(z)|^2 y^{-2}\,\mathrm{d}x\mathrm{d}y = 4\iint_{\mathbb{H}} |\phi(\bar{z})|^2 y^2\,\mathrm{d}x\mathrm{d}y = 4\iint_{\mathbb{H}^*} |\hat{u}'''(z)|^2 y^2\,\mathrm{d}x\mathrm{d}y,$$

where $\hat{u}$ is defined in Equation (3.28) and we have

$$\hat{u}'''(z) = -\frac{6}{\pi}\iint_{\mathbb{H}} \frac{\dot{\mu}(\zeta)\,\mathrm{d}\xi\mathrm{d}\eta}{(\zeta - z)^4}, \qquad \zeta = \xi + \mathrm{i}\eta.$$

By Cauchy-Schwarz,

$$\|u\|_{\mathrm{WP}}^2 \leq \frac{4 \cdot 6^2}{\pi}\iint_{\mathbb{H}^*} y^2 \left(\iint_{\mathbb{H}} \frac{\mathrm{d}\xi_1\mathrm{d}\eta_1}{|\zeta_1 - z|^4} \cdot \iint_{\mathbb{H}} \frac{|\dot{\mu}(\zeta_2)|^2\,\mathrm{d}\xi_2\mathrm{d}\eta_2}{|\zeta_2 - z|^4} \right) \mathrm{d}x\mathrm{d}y.$$

Using the identities

$$\iint_{\mathbb{H}} \frac{\mathrm{d}\xi\mathrm{d}\eta}{|\zeta - z|^4} = \frac{\pi}{4y^2}, \qquad z = x + \mathrm{i}y \in \mathbb{H}^*,$$

$$\iint_{\mathbb{H}^*} \frac{\mathrm{d}x\mathrm{d}y}{|\zeta - z|^4} = \frac{\pi}{4\eta^2} \qquad \zeta = \xi + \mathrm{i}\eta \in \mathbb{H},$$

we get that

$$\|u\|_{\mathrm{WP}}^2 \leq 9\int_{\mathbb{H}} \eta_2^{-2}|\dot{\mu}(\zeta_2)|^2\,\mathrm{d}\xi_2\mathrm{d}\eta_2 \leq C'(h)\|\dot{\vartheta}\|_{\mathfrak{h}}^2$$

as claimed.

□

Now we explicitly compute the Weil–Petersson metric tensor on $\mathscr{H}$. Let $\{\dot{\vartheta}_e\}_{e \in E}$ denote a basis of $\mathfrak{h}$, where $\dot{\vartheta}_e(e) = 1$ and 0 otherwise. It suffices to compute for all $h \in \mathscr{H}$, $e_1, e_2 \in E$,

$$\left\langle \Xi_h(\dot{\vartheta}_{e_1}), \Xi_h(\dot{\vartheta}_{e_2}) \right\rangle_{\mathrm{WP}},$$

namely, the inner product between two unit infinitesimal diamond shears on $h(\mathfrak{F})$.

More precisely, assume that a quad Q has vertices $a_1, a_2, a_3, a_4 \in \mathbb{T}$ in counterclockwise order. Then *the unit infinitesimal diamond shear on Q with diagonal (a_1, a_3)* is the infinitesimal shear with coordinates $\dot{s}$ which is only non-zero at the edges of Q with

$$\dot{s}(a_1, a_2) = \dot{s}(a_3, a_4) = 1$$

and

$$\dot{s}(a_2, a_3) = \dot{s}(a_4, a_1) = -1.$$

Lemma 3.6.7. *The unit infinitesimal diamond shear on $h(Q_e)$ with diagonal $h(e)$ is $\Xi_h(\dot{\vartheta}_e)$.*

Proof. This follows directly from Proposition 3.3.12. □

Theorem 3.6.2 implies that in the half-plane model, the corresponding quadratic differential is

$$\phi(z) := \frac{\mathfrak{i}}{2\pi}\Big[\frac{(\mathfrak{c}(a_1)-\mathfrak{c}(a_2))^2}{(z-\mathfrak{c}(a_1))^2(z-\mathfrak{c}(a_2))^2} - \frac{(\mathfrak{c}(a_2)-\mathfrak{c}(a_3))^2}{(z-\mathfrak{c}(a_2))^2(z-\mathfrak{c}(a_3))^2} + \frac{(\mathfrak{c}(a_3)-\mathfrak{c}(a_4))^2}{(z-\mathfrak{c}(a_3))^2(z-\mathfrak{c}(a_4))^2} - \frac{(\mathfrak{c}(a_4)-\mathfrak{c}(a_1))^2}{(z-\mathfrak{c}(a_4))^2(z-\mathfrak{c}(a_1))^2}\Big] \tag{3.33}$$

for all $z \in \mathbb{H}^*$, and we have

$$\phi(z) = \frac{\mathfrak{i}}{\pi}\sum_{j=1}^{4}(-1)^{j-1}\left(\frac{1}{\mathfrak{c}(a_{j+1})-\mathfrak{c}(a_j)} + \frac{1}{\mathfrak{c}(a_j)-\mathfrak{c}(a_{j-1})}\right)\frac{1}{z-\mathfrak{c}(a_j)}, \tag{3.34}$$

where the subscripts are considered modulo 4.

Theorem 3.6.8. *If u is the Zygmund vector field associated with the unit infinitesimal diamond shear on the quad with vertices (a_1,a_2,a_3,a_4) in $\mathbb{T}$ with diagonal (a_1,a_3), then*

$$\|u\|^2_{\mathrm{WP}} = \frac{2}{\pi}\sum_{j,k=1}^{4}\frac{(-1)^{j+k}a_j^2\bar{a}_k^2(a_{j+1}-a_{j-1})(\bar{a}_{k+1}-\bar{a}_{k-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})(\bar{a}_{k+1}-\bar{a}_k)(\bar{a}_k-\bar{a}_{k-1})}\sigma(a_j,a_k).$$

where for $a,b\in\mathbb{T}$,

$$\sigma(a,b) = \sum_{p=0}^{\infty}\frac{(a\bar{b})^{p+1}}{(1+p)(2+p)(3+p)}. \tag{3.35}$$

Further, let $u_1,u_2 : E\to\mathbb{R}$ be two unit infinitesimal diamond shears on quads with vertices $Q_1=(a_1,a_2,a_3,a_4)$ and $Q_2=(b_1,b_2,b_3,b_4)$ in $\mathbb{T}$ and diagonals $e_1=(a_1,a_3)$ and $e_2=(b_1,b_3)$ respectively. Then

$$\langle u_1,u_2\rangle_{\mathrm{WP}} = \frac{2}{\pi}\operatorname{Re}\sum_{j,k=1}^{4}\frac{(-1)^{j+k}a_j^2\bar{b}_k^2(a_{j+1}-a_{j-1})(\bar{b}_{k+1}-\bar{b}_{k-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})(\bar{b}_{k+1}-\bar{b}_k)(\bar{b}_k-\bar{b}_{k-1})}\sigma(a_j,b_k). \tag{3.36}$$

Proof. Let $\zeta := \mathfrak{c}^{-1}(z) = \xi+\mathfrak{i}\eta$. We have from change of variables, (3.34), and (3.30) that in the disk model

$$\dot{\mu}_u(\zeta) = -\frac{\mathfrak{i}}{2\pi}(1-|\zeta|^2)^2\sum_{j=1}^{4}\frac{(-1)^{j-1}}{\bar{\zeta}^3}\left(\frac{1}{a_{j+1}-a_j}+\frac{1}{a_j-a_{j-1}}\right)\frac{1}{1-a_j\bar{\zeta}}.$$

We notice that

$$\sum_{j=1}^{4}(-1)^j\left(\frac{1}{a_{j+1}-a_j}+\frac{1}{a_j-a_{j-1}}\right)a_j^p = 0$$

for $p = 0, 1, 2$. Therefore,

$$\begin{aligned}
\dot{\mu}_u(\zeta) &= \frac{\mathrm{i}}{2\pi}(1-|\zeta|^2)^2 \sum_{j=1}^{4} \frac{(-1)^j}{\bar{\zeta}^3} \left(\frac{1}{a_{j+1}-a_j} + \frac{1}{a_j - a_{j-1}} \right) \sum_{p\geq 0} (a_j \bar{\zeta})^p \\
&= \frac{\mathrm{i}}{2\pi}(1-|\zeta|^2)^2 \sum_{j=1}^{4} \frac{(-1)^j}{\bar{\zeta}^3} \left(\frac{1}{a_{j+1}-a_j} + \frac{1}{a_j - a_{j-1}} \right) \sum_{p\geq 3} (a_j \bar{\zeta})^p \\
&= \frac{\mathrm{i}}{2\pi}(1-|\zeta|^2)^2 \sum_{j=1}^{4} (-1)^j a_j^2 \left(\frac{1}{a_{j+1}-a_j} + \frac{1}{a_j - a_{j-1}} \right) \frac{a_j}{1-a_j\bar{\zeta}}.
\end{aligned}$$

Define for $a, b \in \mathbb{T}$,

$$\sigma(a,b) = \frac{1}{2\pi} \iint\limits_{\mathbb{D}} \frac{a}{1-a\bar{z}} \frac{\bar{b}}{1-\bar{b}z} (1-|z|^2)^2 \,\mathrm{d}x\mathrm{d}y. \tag{3.37}$$

Using polar coordinates we find

$$\begin{aligned}
\sigma(a,b) &= \frac{1}{2\pi} \iint\limits_{\mathbb{D}} \frac{a}{1-a\bar{z}} \frac{\bar{b}}{1-\bar{b}z} (1-|z|^2)^2 \,\mathrm{d}x\mathrm{d}y \\
&= \frac{a\bar{b}}{2\pi} \int_0^1 \int_0^{2\pi} \sum_{p,q\geq 0} (are^{-\mathrm{i}\theta})^p (\bar{b}re^{\mathrm{i}\theta})^q (1-r^2)^2 r \,\mathrm{d}\theta\mathrm{d}r \\
&= \sum_{p\geq 0} (a\bar{b})^{p+1} \int_0^1 r^{2p+1}(1-r^2)^2 \,\mathrm{d}r \\
&= \sum_{p\geq 0} \frac{(a\bar{b})^{p+1}}{(p+1)(p+2)(p+3)}.
\end{aligned}$$

The square of the Weil–Petersson norm of u equals

$$\begin{aligned}
\|u\|_{\mathrm{WP}}^2 &= \mathrm{Re} \iint\limits_{\mathbb{D}} |\dot{\mu}_u|^2 \frac{4}{(1-|\zeta|^2)^2} \,\mathrm{d}\xi\mathrm{d}\eta \\
&= \frac{1}{\pi^2} \sum_{j,k=1}^{4} (-1)^{j+k} a_j^2 \bar{a}_k^2 \left(\frac{1}{a_{j+1}-a_j} + \frac{1}{a_j - a_{j-1}} \right) \left(\frac{1}{\bar{a}_{k+1}-\bar{a}_k} + \frac{1}{\bar{a}_k - \bar{a}_{k-1}} \right) \\
&\qquad \iint\limits_{\mathbb{D}} \frac{a_j}{1-a_j\bar{\zeta}} \frac{\bar{a}_k}{1-\bar{a}_k\zeta} (1-|\zeta|^2)^2 \,\mathrm{d}\xi\mathrm{d}\eta \\
&= \frac{2}{\pi} \sum_{j,k=1}^{4} \frac{(-1)^{j+k} a_j^2 \bar{a}_k^2 (a_{j+1}-a_{j-1})(\bar{a}_{k+1}-\bar{a}_{k-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})(\bar{a}_{k+1}-\bar{a}_k)(\bar{a}_k-\bar{a}_{k-1})} \sigma(a_j, a_k).
\end{aligned}$$

Notice that the first integral is real so we omit Re in the second equality. The same computation gives the claimed formula for $\langle u_1, u_2 \rangle_{\mathrm{WP}} = \mathrm{Re} \iint_{\mathbb{D}} \dot{\mu}_{u_1} \overline{\dot{\mu}_{u_2}} \frac{4}{(1-|\zeta|^2)^2} \,\mathrm{d}\xi\mathrm{d}\eta$. □

Remark 3.6.9. Since the Weil–Petersson metric is invariant under the adjoint action of $\mathrm{PSL}(2,\mathbb{R})$, therefore also under $\mathrm{PSL}(2,\mathbb{Z})$, the Weil–Petersson norm of a unit infinitesimal diamond shear is constant for all Farey quads. To compute it, we consider the case where $a_1 = 1$, $a_2 = \mathfrak{i}$, $a_3 = -1$, and $a_4 = -\mathfrak{i}$, with diagonal $(-1, 1)$. We obtain

$$\begin{aligned}
\sigma(1,1) = \sigma(\mathfrak{i},\mathfrak{i}) = \sigma(-1,-1) = \sigma(-\mathfrak{i},-\mathfrak{i}) &= \sum_{p=1}^{\infty} \frac{1}{p(1+p)(2+p)} = \frac{1}{4} \\
\sigma(1,-\mathfrak{i}) = \sigma(\mathfrak{i},1) = \sigma(-1,\mathfrak{i}) = \sigma(-\mathfrak{i},-1) &= \sum_{p=1}^{\infty} \frac{\mathfrak{i}^p}{p(1+p)(2+p)} = \frac{3-\pi}{4} - \mathfrak{i}\frac{\log 2 - 1}{2}. \\
\sigma(-\mathfrak{i},1) = \sigma(1,\mathfrak{i}) = \sigma(\mathfrak{i},-1) = \sigma(-1,-\mathfrak{i}) &= \overline{\sigma(1,-\mathfrak{i})} = \frac{3-\pi}{4} + \mathfrak{i}\frac{\log 2 - 1}{2} \\
\sigma(1,-1) = \sigma(\mathfrak{i},-\mathfrak{i}) = \sigma(-1,1) = \sigma(-\mathfrak{i},\mathfrak{i}) &= \sum_{p=1}^{\infty} \frac{(-1)^p}{p(1+p)(2+p)} = \frac{5}{4} - 2\log 2.
\end{aligned}$$

Therefore Theorem 3.6.8 implies that

$$\|u\|_{\mathrm{WP}}^2 = \frac{8}{\pi}\log 2$$

for u the Zygmund vector field corresponding the unit single infinitesimal diamond shear supported on $e \in E$.

Corollary 3.6.10. *We define $g(Q_1, e_1, Q_2, e_2)$ to be the right-hand side of (3.36). For $j = 1, 2$, let $\dot\vartheta_j \in \mathfrak{h} \simeq T_h\mathscr{H}$ and $u_j := \Xi_h(\dot\vartheta_j) \in T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T}) \simeq T_h\,\mathrm{WP}(\mathbb{T})$ be the corresponding vector field. Then*

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = \sum_{e_1\in E}\sum_{e_2\in E} \dot\vartheta_1(e_1)\dot\vartheta_2(e_2)\, g(h(Q_{e_1}), h(e_1), h(Q_{e_2}), h(e_2)).$$

Proof. This follows from Lemma 3.6.7 and Theorem 3.6.8 if $\dot\vartheta_1$ and $\dot\vartheta_2$ are finitely supported. Lemma 3.6.4 extends the result to all $\dot\vartheta_j \in \mathfrak{h}$. □

3.6.4 Weil–Petersson symplectic form on $\mathscr{H}$

We give an expression for the Weil–Petersson symplectic form ω restricted to $\mathscr{H}$ in terms of a mixture of shears and diamond shears. Similar to the computation of the metric, the Weil–Petersson symplectic form is also right-invariant on $\mathrm{WP}(\mathbb{T}) \simeq T_0(\mathbb{D})$, so we simply use ω to denote its alternating bilinear form on $T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$ (and compute only for pull-backs to $T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$).

Theorem 3.6.11. *Let $h \in \mathscr{H} \subset \mathrm{WP}(\mathbb{T})$. For $j = 1, 2$, let $\dot\vartheta_j \in \mathfrak{h}$, $\dot s_j := \Phi(\dot\vartheta_j)$ be the corresponding infinitesimal shears, and let $u_j := \Xi_h(\dot\vartheta_j) \in T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T}) = H^{3/2}(\mathbb{T})$ be the pull-back by h^{-1} of the tangent vector in $T_h\,\mathrm{WP}(\mathbb{T})$ represented by the differentiable path $t \mapsto \vartheta_h + t\cdot\dot\theta_j$ as in Corollary 3.6.5. Then*

$$\omega(u_1, u_2) = \sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e) = -\sum_{e\in E} \dot s_1(e)\dot\vartheta_2(e).$$

Remark 3.6.12. It is remarkable that unlike the expression of the metric tensor (Corollary 3.6.10), the expression of the symplectic form in shear and diamond shear coordinates is very simple and independent of h.

Concretely, the above theorem shows if $u_1 = \Xi_h(\dot{\vartheta}_{e_1}), u_2 = \Xi_h(\dot{\vartheta}_{e_2})$ are the vector fields given by the unit infinitesimal diamond shears on quads $Q_1 = h(Q_{e_1}), Q_2 = h(Q_{e_2}) \in h(\mathfrak{F})$ of diagonals $h(e_1), h(e_2)$ respectively, then

- $\omega(u_1, u_2) = 1$ if Q_1, Q_2 overlap in one triangle and e_1, e_2 are adjacent in a fan and the index of e_2 is the index of e_1 plus 1;
- $\omega(u_1, u_2) = -1$ if Q_1, Q_2 overlap in one triangle and e_1, e_2 are adjacent in a fan and the index of e_2 is the index of e_1 minus 1;
- $\omega(u_1, u_2) = 0$ if $Q_1 = Q_2$ or if Q_1, Q_2 are disjoint. Note that two quads are considered disjoint if they overlap in only a vertex or edge. See Figure 3.9.

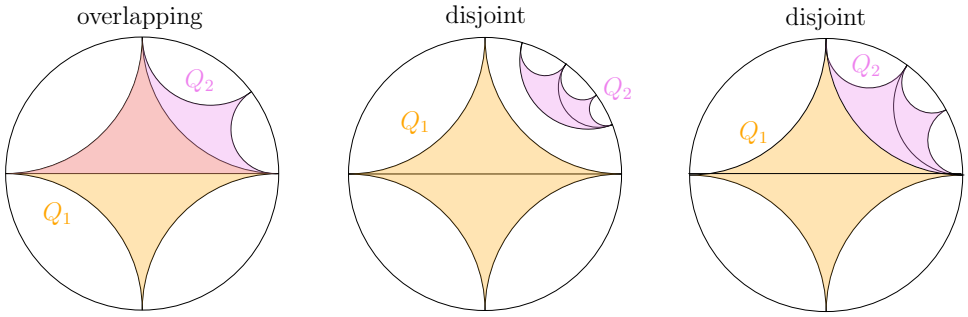


Figure 3.9: An example of quads overlapping in a triangle and gives value $\omega(u_1, u_2) = -1$ (left) and two examples of disjoint quads (middle and right).

The most direct way to prove Theorem 3.6.11 would be to use the same computations as in Theorem 3.6.8, as in the half-plane model $\mathbb{H}$,

$$\omega(u_1, u_2) = -\operatorname{Im} \iint_{\mathbb{H}} \dot{\mu}_{u_1}(z)\overline{\dot{\mu}_{u_2}(z)} \frac{1}{y^2} \, \mathrm{d}x\mathrm{d}y, \qquad z = x + \mathrm{i}y. \tag{3.38}$$

However, we have not been able to prove in general that ω has such a simple formula directly from Equation (3.38) (see Remark 3.6.17 and Lemma 3.6.18 below for discussion). Instead, we use an expression for a symplectic form $\tilde{\omega}$ on decorated Teichmüller space from [Pen93], and the relationship between diamond shears and $\log \Lambda$-lengths described in Section 3.3.5.

Recall from Definition 3.3.28 the section $\sigma : \mathscr{P}_0 \to \widetilde{T(\mathbb{D})}$ which gives a canonical way to choose the decoration for each $h \in \mathscr{H} \subset \mathscr{P}_0$. This allows us to compare the diamond shear coordinate of h with the $\log \Lambda$-coordinate of $\sigma(h)$. For $e = (a, b) \in E$, if ρ_a, ρ_b are the horocycles chosen by σ at $h(a), h(b)$, recall the notation

$$\log \Lambda_h(e) := \log \Lambda(\rho_a, \rho_b).$$

Lemma 3.3.30 shows that

$$\vartheta_h(e) = -\log \Lambda_h(e).$$

As a corollary, the same relationship passes to $\mathfrak{h}$.

Corollary 3.6.13. *Choose $h \in \mathscr{H}$ with diamond shear coordinate ϑ_h and $\dot{\vartheta} \in \mathfrak{h}$ and let h_t be the Weil–Petersson homeomorphism induced by the diamond shear function $\vartheta_h + t \cdot \dot{\vartheta}$. Then we have*

$$\dot{\vartheta}(e) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} - \log \Lambda_{h_t}(e).$$

The following result from [Pen93, Sec. 5.1] shows that the symplectic form $\widetilde{\omega}$ defined below on $\widetilde{T(\mathbb{D})}$ in terms of $\log\Lambda$-lengths projects to the Weil–Petersson symplectic form restricted to $\mathrm{Diff}(\mathbb{T})/\mathrm{Möb}$.

Theorem 3.6.14 (See [Pen93, Thm. 5.5])**.** *Let τ be a triangle in $\mathfrak{F}$, and let $\{e_1, e_2, e_3\}$ be its edges in counterclockwise order. For any $h \in \mathrm{Diff}(\mathbb{T})/\mathrm{Möb}(\mathbb{T})$, define*

$$\begin{aligned}(\widetilde{\omega}_\tau)_h :=& \mathrm{d}\log\Lambda_h(e_1) \wedge \mathrm{d}\log\Lambda_h(e_2) + \mathrm{d}\log\Lambda_h(e_2) \wedge \mathrm{d}\log\Lambda_h(e_3) \\ &+ \mathrm{d}\log\Lambda_h(e_3) \wedge \mathrm{d}\log\Lambda_h(e_1).\end{aligned}$$

Then

$$\widetilde{\omega} := -\sum_{\tau\in\mathfrak{F}} \widetilde{\omega}_\tau$$

projects to the Weil–Petersson symplectic form ω under forgetting the decoration.

Remark 3.6.15. The expression of the Weil–Petersson symplectic form differs from the one in [Pen93] by a factor 2 due to a different choice of scalar factor. Indeed, a direct computation shows that our symplectic form (3.38) can be expressed in terms of the Fourier coefficients of the vector fields on the circle as

$$\omega(u_1, u_2) = \mathfrak{i}\pi \sum_{n\in\mathbb{Z}} (n^3 - n)\tilde{u}_{1,n}\tilde{u}_{2,-n},$$

first proved in [NV90], where

$$u_j = \sum_{n\in\mathbb{Z}} \tilde{u}_{j,n} e^{\mathfrak{i}n\theta}\frac{\partial}{\partial\theta}, \qquad \tilde{u}_{j,n} = \overline{\tilde{u}_{j,-n}} \in \mathbb{C}.$$

Whereas Penner uses the symplectic form $2\mathfrak{i}\pi\sum_{n\in\mathbb{Z}}(n^3 - n)\tilde{u}_{1,n}\tilde{u}_{2,-n}$. We also verify Theorem 3.6.14 in a special case by direct computation in terms of the shears in Lemma 3.6.18.

Proof of Theorem 3.6.11. First suppose that $\dot{s}_1, \dot{s}_2$ are finitely supported shear functions with the end points of the support edges in $\{a_1, ..., a_n\} \subset V$. By Remark 3.6.6 and Definition 3.6.1, $u_1 = \Xi_h(\dot{\vartheta}_1)$ and $u_2 = \Xi_h(\dot{\vartheta}_2)$ depend only on the points $h(a_1), ..., h(a_n)$. In particular, if we replace h by any function g which agrees with h at $a_1, ..., a_n$, then we still have $\Xi_g(\dot{\vartheta}_j) = \Xi_h(\dot{\vartheta}_j)$ for $j = 1, 2$. Therefore we can replace h with $g \in \mathrm{Diff}(\mathbb{T})/\mathrm{Möb}(\mathbb{T})$. We use Corollary 3.6.13 to identify infinitesimal diamond shears with infinitesimal $\log\Lambda$-lengths and then apply Theorem 3.6.14.

Indeed, for each $v \in V$, let $(e_n)_{n\in\mathbb{Z}}$ be a labelling of $\mathrm{fan}(v)$ in counterclockwise order. We rewrite the expression for ω from Theorem 3.6.14 as a sum over fans instead of triangles.

$$\omega(u_1, u_2) = \sum_{v\in V}\sum_{e_n\in\mathrm{fan}(v)} \dot{\vartheta}_1(e_n)\dot{\vartheta}_2(e_{n+1}) - \dot{\vartheta}_1(e_{n+1})\dot{\vartheta}_2(e_n). \tag{3.39}$$

On the other hand, if $e = (a, b)$, we can label the edges around Q_e in counterclockwise order so that the counterclockwise order in $\mathrm{fan}(a)$ is e_1, e, e_4 and the counterclockwise order in $\mathrm{fan}(b)$ is e_3, e, e_2. Using Equation (3.7),

$$\begin{aligned}\sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e) &= \sum_{e\in E} \dot\vartheta_1(e)(-\dot\vartheta_2(e_1) + \dot\vartheta_2(e_2) - \dot\vartheta_2(e_3) + \dot\vartheta_2(e_4)) \\ &= \sum_{v\in V}\sum_{e_n\in\mathrm{fan}(v)} \dot\vartheta_1(e_n)\dot\vartheta_2(e_{n+1}) - \dot\vartheta_1(e_{n+1})\dot\vartheta_2(e_n).\end{aligned}$$

Switching u_1, u_2, it is clear that $\sum_{e\in E} \dot s_1(e)\dot\vartheta_2(e) = -\sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e)$. This completes the proof in the case that $\dot s_1, \dot s_2$ are finitely supported.

In general, given $\dot\vartheta_1, \dot\vartheta_2 \in \mathfrak{h}$ for which $\dot s_1, \dot s_2$ are not finitely supported, we can find sequences $(\dot\vartheta_j^n)_{n\geq 1}$, $j = 1, 2$ such that

- $\dot s_j^n = \Phi(\dot\vartheta_j^n)$ is finitely supported for $j = 1, 2$ and all $n \geq 1$.
- $\dot\vartheta_j^n$ converges to $\dot\vartheta_j$ in ℓ^2 in diamond shears for $j = 1, 2$.

(For example, the sequences $(\dot\vartheta_j^n)_{n\geq 1}$ where $\dot\vartheta_j^n$ is $\dot\vartheta_j$ restricted to edges with generation less than n for $j = 1, 2$ satisfy these properties.)

For each finite n, we have

$$\sum_{e\in E} \dot\vartheta_1^n(e)\dot s_2^n(e) = \omega(u_1^n, u_2^n),$$

where $u_j^n := \Xi_h(\dot\vartheta_j^n)$ for $j = 1, 2$. It remains to compute the limits of both sides as $n \to \infty$.

By Lemma 3.6.4, $u_j^n \xrightarrow{n\to\infty} u_j$ in $H^{3/2}$ for $j = 1, 2$. By Remark 3.6.16,

$$\omega(u_1^n, u_2^n) = \langle u_1^n, J(u_2^n)\rangle_{\mathrm{WP}},$$

where $J : T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T}) \to T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$ is the (almost) complex structure and an isometry, hence

$$\lim_{n\to\infty} \omega(u_1^n, u_2^n) = \omega(u_1, u_2).$$

On the other hand, since ℓ^2 convergence in diamond shears implies ℓ^2 convergence in shears by the expression $\Phi(\dot\vartheta) = \dot s$, (3.7), and Cauchy-Schwarz inequality, we have

$$\lim_{n\to\infty} \sum_{e\in E} \dot\vartheta_1^n(e)\dot s_2^n(e) = \sum_{e\in E} \dot\vartheta_1(e)\dot s_2(e).$$

This completes the proof for general $\dot\vartheta_1, \dot\vartheta_2 \in \mathfrak{h}$. □

Remark 3.6.16. The Weil–Petersson metric $\langle\cdot,\cdot\rangle_{\mathrm{WP}}$ (computed in Theorem 3.6.8, Corollary 3.6.10), Weil–Petersson symplectic form ω (computed in Theorem 3.6.11), and a complex structure J form a Kähler structure on Weil–Petersson Teichmüller space, meaning that

$$\langle u_1, u_2\rangle_{\mathrm{WP}} = \omega(u_1, J(u_2)).$$

The complex structure J is the Hilbert transform [NV90] on the space of Zygmund vector fields, and was computed explicitly in terms of infinitesimal $\log\Lambda$-lengths in [Pen02, Thm. 6.8]. Combining the symplectic form ω and complex structure J from [Pen02] gives us another way to compute the metric explicitly. Using this, we independently verify that $\|u\|^2_{\mathrm{WP}} = 8\log 2/\pi$ when u is the Zygmund vector field associated to a single infinitesimal diamond shear which coincides with our result in Remark 3.6.9.

Remark 3.6.17. Starting from the definition in Equation (3.38), the same computation (but taking the imaginary part) and the same notation as in Theorem 3.6.8 gives that

$$\omega(u_1,u_2) = -\frac{2}{\pi}\operatorname{Im}\sum_{j,k=1}^{4}\frac{(-1)^{j+k}a_j^2\bar{b}_k^2(a_{j+1}-a_{j-1})(\bar{b}_{k+1}-\bar{b}_{k-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})(\bar{b}_{k+1}-\bar{b}_k)(\bar{b}_k-\bar{b}_{k-1})}\sigma(a_j,b_k), \tag{3.40}$$

where u_1 and u_2 are the vector fields representing two unit infinitesimal diamond shears.

Can one recover Theorem 3.6.11 directly from (3.40)? We are unable to prove this in general, but check it in a special case (Lemma 3.6.18). The general case to check, after normalization, would be when $a_1 = 1, a_2 = \mathfrak{i}, a_3 = -1$, but $a_4 \neq -\mathfrak{i}$ and u_2 is an arbitrary diamond shear of vertices (b_1,b_2,b_3,b_4), such that (a_1,a_2,a_3,a_4) and (b_1,b_2,b_3,b_4) are both quads in the same tessellation $h(\mathfrak{F})$ of the disk.

Lemma 3.6.18. *Let $u_1 \in T_{\mathrm{Id}}\,\mathrm{WP}(\mathbb{T})$ be the vector field associated with the unit infinitesimal diamond shear on the quad $(a_1,a_2,a_3,a_4) = (1,\mathfrak{i},-1,\mathfrak{i})$ of diagonal $(1,-1)$. Let u_2 be the unit infinitesimal diamond shear associated with $(b_1,b_2,b_3,b_4) = (1,e^{\mathfrak{i}\theta},\mathfrak{i},-1)$ of diagonal $(1,\mathfrak{i})$. Then $\omega(u_1,u_2) = -1$ for all $\theta \in (0,\pi/2)$.*

Proof. Rearranging (3.40) gives

$$-\frac{2}{\pi}\operatorname{Im}\sum_{p\geq 1}\frac{1}{p(p+1)(p+2)}\left(\sum_{j=1}^{4}(-1)^j\frac{a_j^{p+2}(a_{j+1}-a_{j-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})}\right)\left(\sum_{k=1}^{4}(-1)^k\frac{\bar{b}_k^{p+2}(\bar{b}_{k+1}-\bar{b}_{k-1})}{(\bar{b}_{k+1}-\bar{b}_k)(\bar{b}_k-\bar{b}_{k-1})}\right).$$

Since $(a_1,a_2,a_3,a_4) = (1,\mathfrak{i},-1,\mathfrak{i})$,

$$\sum_{j=1}^{4}(-1)^j\frac{a_j^{p+2}(a_{j+1}-a_{j-1})}{(a_{j+1}-a_j)(a_j-a_{j-1})} = 4\mathfrak{i} \tag{3.41}$$

if $p = 1 \bmod 4$, and 0 otherwise.

Using the identity

$$e^{\mathfrak{i}a} - e^{\mathfrak{i}b} = e^{\mathfrak{i}\frac{a+b}{2}}(2\mathfrak{i}\sin(\frac{a-b}{2})),$$

and writing $\alpha_j = e^{\mathfrak{i}\theta_j}$, we find that

$$\begin{aligned}\frac{\alpha_j(\alpha_{j+1}-\alpha_{j-1})}{(\alpha_{j+1}-\alpha_j)(\alpha_j-\alpha_{j-1})} &= e^{\mathfrak{i}(\theta_j+\frac{\theta_{j+1}+\theta_{j-1}}{2}-\frac{\theta_{j+1}+\theta_j}{2}-\frac{\theta_j+\theta_{j-1}}{2})}\frac{\sin(\frac{\theta_{j+1}-\theta_{j-1}}{2})}{2\mathfrak{i}\sin(\frac{\theta_{j+1}-\theta_j}{2})\sin(\frac{\theta_j-\theta_{j-1}}{2})}\\ &= -\frac{\mathfrak{i}}{2}\frac{\sin(\frac{\theta_{j+1}-\theta_{j-1}}{2})}{\sin(\frac{\theta_{j+1}-\theta_j}{2})\sin(\frac{\theta_j-\theta_{j-1}}{2})}.\end{aligned}$$

Therefore, for $k = 1, 3, 4$ and $p = 1 \bmod 4$,

$$(-1)^k \frac{\bar{b}_k^{p+2}(\bar{b}_{k+1} - \bar{b}_{k-1})}{(\bar{b}_{k+1} - \bar{b}_k)(\bar{b}_k - \bar{b}_{k-1})} = (-1)^k \frac{\bar{b}_k^{3}(\bar{b}_{k+1} - \bar{b}_{k-1})}{(\bar{b}_{k+1} - \bar{b}_k)(\bar{b}_k - \bar{b}_{k-1})}$$

is purely imaginary and does not contribute to $\omega(u_1, u_2)$ given (3.41). The remaining term is

$$\frac{\bar{b}_2^{p+2}(\bar{b}_3 - \bar{b}_1)}{(\bar{b}_3 - \bar{b}_2)(\bar{b}_2 - \bar{b}_1)} = \frac{\mathfrak{i}}{2} \frac{e^{-\mathfrak{i}(p+1)\theta} \sin(\pi/4)}{\sin(\pi/4 - \theta/2)\sin(\theta/2)} = \frac{\mathfrak{i}}{2\sqrt{2}} \frac{e^{-\mathfrak{i}(p+1)\theta}}{\sin(\pi/4 - \theta/2)\sin(\theta/2)}.$$

This gives that

$$\begin{aligned}\omega(u_1, u_2) &= -\frac{2}{\pi} \operatorname{Im}\left(4\mathfrak{i} \sum_{n=0}^{\infty} \frac{\mathfrak{i}}{2\sqrt{2}} \frac{e^{-\mathfrak{i}(4n+2)\theta}}{\sin(\pi/4 - \theta/2)\sin(\theta/2)(4n+1)(4n+2)(4n+3)}\right) \\ &= \frac{2\sqrt{2}}{\pi \sin(\pi/4 - \theta/2)\sin(\theta/2)} \operatorname{Im}\left(\sum_{n=0}^{\infty} \frac{e^{-\mathfrak{i}(4n+2)\theta}}{(4n+1)(4n+2)(4n+3)}\right)\end{aligned}$$

A simple trigonometry gives

$$\sin(\pi/4 - \theta/2)\sin(\theta/2) = \frac{\sqrt{2}}{4}(\sin\theta + \cos\theta - 1).$$

Simplify the imaginary part of the series for $\theta \in (0, \pi/2)$ gives

$$f(\theta) := \operatorname{Im}\left(\sum_{n=0}^{\infty} \frac{e^{-\mathfrak{i}(4n+2)\theta}}{(4n+1)(4n+2)(4n+3)}\right) = -\frac{\pi}{8}(\sin(\theta) + \cos(\theta) - 1).$$

Indeed, it suffices to check that $f''(\theta) + f(\theta) = \pi/8$, and $\lim_{\theta \to 0+} f(\theta) = 0$ and $\lim_{\theta \to (\pi/2)-} f(\theta) = 0$. This gives $\omega(u_1, u_2) = -1$ as claimed. □

3.7 Random diamond shears

This section is not part of the paper [ŠWW22], but uses the same notation and set up. We denote the Gaussian random variable with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$.

The natural "random object" in $\mathscr{H}$ is the diamond shear function $\Theta : E \to \mathbb{R}$ such that $(\Theta_e)_{e \in E}$ are i.i.d. standard Gaussians $\mathcal{N}(0, 1)$. Here we show that this corresponds to a homeomorphism almost surely.

Theorem 3.7.1. *The diamond shear function $\Theta : E \to \mathbb{R}$ induces a homeomorphism almost surely.*

Remark 3.7.2. In fact, this theorem holds for Θ with any law such that i) the collection $(\Theta(e))_{e \in E}$ is i.i.d., ii) the law of $\Theta(e)$ is centered, symmetric, and infinitely divisible.

Recall that a diamond shear function $\vartheta : E \to \mathbb{R}$ can be written as a shear function $s : E \to \mathbb{R}$ as follows. Namely, if $e \in E$ and e_1, e_2, e_3, e_4 are the edges around the quad of e in counterclockwise order starting from an endpoint of e, then

$$s(e) = -\vartheta(e_1) + \vartheta(e_2) - \vartheta(e_3) + \vartheta(e_4). \tag{3.42}$$

We let $s_\Theta : E \to \mathbb{R}$ denote the random shear function corresponding to the random diamond shear function $\Theta : E \to \mathbb{R}$.

Shear functions $s : E \to \mathbb{R}$ which induce homeomorphisms were characterized in [Šar10] by a condition on *chains* of edges. A chain is a sequence of edges $(e_n)_{n\geq 1} \subset E$ such that e_n, e_{n-1} share an endpoint for all $n \geq 1$. If e, e' are in the same fan (i.e. they are incident to the same vertex $v \in V$), then we say that $e' < e$ if e' comes before e in the counterclockwise ordering of edges in the fan at v. The following condition on chains characterizes shear functions that encode homeomorhpisms:

Theorem 3.7.3 ([Šar10])**.** *The shear function $s : E \to \mathbb{R}$ induces a homeomorphism if and only if for all chains $(e_n)_{n\geq 1}$,*

$$\sum_{n\geq 1} \exp(s_1^n + ... + s_n^n) = \infty. \tag{3.43}$$

Here $s_i^n = \pm s(e_i)$. The sign is determined by

$$s_i^n = (-1)^{a_i}(-1)^{k_{i,n}} s(e_i)$$

where $a_i = 0$ if $e_i < e_{i+1}$ and $a_i = 1$ if $e_i > e_{i+1}$, and where $k_{i,n}$ is the number of times that the chain changes fan between e_i and e_{n+1}.

Recall that $\mathfrak{F}^*$ denotes the dual tree of the Farey tessellation, which has a vertex for each triangle $\mathfrak{F}$. To prove our result, we first show that it suffices to check a smaller class of chains which have a nice description in terms of the dual tree $\mathfrak{F}^*$. There is a correspondence between dual edges E^* and edges E, and for $e \in E$, we denote the corresponding dual edge e^*. The *root dual edge* e_0^* is the edge connecting the dual vertices corresponding to the triangles $\{-1, \mathbf{i}, 1\}$ and $\{-1, -\mathbf{i}, 1\}$.

A dual edge e^* has *generation* n if its distance from e_0^* is n (we count this as the number of edges between e^* and e_0^*, including e^* but not including e_0^*). A *branch of the dual tree* is a sequence $(e_n^*)_{n\geq 1}$ which forms a connected set of edges and has $\text{gen}(e_n^*) > \text{gen}(e_{n-1}^*)$ for all n. We say that a chain $(e_n)_{n\geq 1}$ is a branch of the dual tree if $(e_n^*)_{n\geq 1}$ is.

Lemma 3.7.4. *A shear function $s : E \to \mathbb{R}$ induces a homeomorphism if and only if (3.43) holds for all chains which are branches of the dual tree.*

Proof of Lemma 3.7.4. Any $s : E \to \mathbb{R}$ induces a piecewise-Möbius map $H_s : \mathbb{D} \to \mathbb{D}$ (which is discontinuous on E), where on each triangle of $\mathfrak{F}$, H_s is defined to the Möbius transformation given by composing the Möbius transformations for the shears along the path in the dual tree from the root triangle to it (the developing map for the shear). A key step in the original proof of [Šar10, Theorem C] is that $s : E \to \mathbb{R}$ fails to induce a circle homeomorphism if and only if $H_s : \mathbb{D} \to \mathbb{D}$ fails to be surjective [Šar10, Proposition 4.1],

which is equivalent to saying there exists a sequence of edges $(e_n)_{n\geq 1} \subset E$ such that $H_s(e_n)$ accumulates on a geodesic $\gamma = (x, y) \subset \mathbb{D}$ for $x, y \in \mathbb{T}$ distinct as $n \to \infty$. The remainder of the original proof is to show a sequence of geodesics accumulates on a geodesic inside the disk if and only if the corresponding sum in (3.43) is finite.

For us, we first note that it suffices to check (3.43) only for chains that contain infinitely many distinct edges; let $(e_n)_{n\geq 1}$ be any chain containing infinitely many distinct edges. Since the chain is connected and infinite, it must contain a branch of the dual tree. On the other hand since e_n is a sequence such that e_n, e_{n+1} share a vertex for all n, it contains at most one infinite end. Thus it contains a unique branch of the dual tree which we denote $(e_{n_k})_{k\geq 1} \subseteq (e_n)_{n\geq 1}$. If $(e_n)_{n\geq 1}$ converges to a geodesic γ in the disk then so does $(e_{n_k})_{k\geq 1}$, hence (3.43) is finite for $(e_n)_{n\geq 1}$ only if it is also finite for $(e_{n_k})_{n\geq 1}$. Therefore to show that s induces a circle homeomorphism, it suffices to check (3.43) is infinite along branches of the dual tree.

□

Branches of the dual tree can be encoded in a simple way. After choosing one of the four half-spaces given by the complement of the triangles $\{-1, \mathbf{i}, 1\} \cup \{-1, -\mathbf{i}, 1\}$, a branch of the tree is encoded by a binary sequence $b \in \{0, 1\}^{\mathbb{N}}$, which says whether the branch has e^*_{i+1} to the left (i.e. $b_i = 0$) or the e^*_{i+1} to the right (i.e. $b_i = 1$).

If $b_i = b_{i-1}$ (i.e. subsequence 00 or 11), then e_{i-1}, e_i, e_{i+1} are all contained in the same fan. If $b_i \neq b_{i-1}$ (i.e. subsequence 01 or 10), then the branch switches fans between e_{i-1}, e_i and e_i, e_{i+1}.

For the two positive half-spaces, if $b_i = 1$, then $e_{i+1} < e_i$ (corresponding to clockwise orientation of the fan) and if $b_i = 0$ then $e_{i+1} > e_i$ (corresponding to counterclockwise orientation of the fan). For the lower to half-spaces, this is reversed. However note that a branch of the dual tree is always contained in only one of the four half-spaces.

Lemma 3.7.5. *If $(e_n)_{n\geq 1}$ is a branch of the dual tree that changes fans at most finitely many times, then (3.43) holds for $s = s_\Theta$ along $(e_n)_{n\geq 1}$.*

Proof. If the branch switches fans at most finitely many times, there exists N such that $(e_n)_{n\geq N}$ is contained in one fan. To show that (3.43) holds for $(e_n)_{n\geq 1}$, it suffices to show that

$$\sum_{n\geq N} \exp(s^n_N + ... + s^n_n) = \infty.$$

Since $(e_n)_{n\geq N}$ is contained in a single fan, the sign of s^n_i is the same for all $i, n \geq N$. Without loss of generality we can assume that $(e_n)_{n\geq N} = ((n, \infty))_{n\geq 1}$, i.e. that it is the fan at ∞ in $\mathbb{H}$. The corresponding edges of the quad around e_n are labeled by the values of Θ, denoted $\alpha_n, \vartheta_{n+1}, \vartheta_{n-1}, \alpha_{n-1}$. See Figure 3.10. Edges are colored blue if the diamond shear coordinate on the edge is counted with $-$ sign and orange if it counted with $+$ sign in $\sum_{n=1}^5 s(e_n)$. Notice that

$$\begin{aligned}\sum_{i=N}^{n} s(e_i) &= \sum_{i=N}^{n} -\alpha_i + \vartheta_{i+1} - \vartheta_{i-1} + \alpha_{i-1} \\ &= -\alpha_n + \vartheta_n + \vartheta_{n+1} + \alpha_N - \vartheta_{N-1} - \vartheta_N.\end{aligned}$$

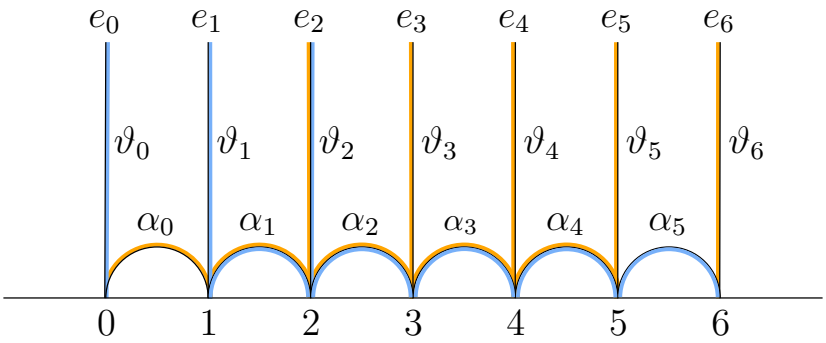


Figure 3.10: Diamond shears corresponding to shears on the fan $\{(n, \infty)\}_{n\in\mathbb{Z}}$, colored blue or orange if they contribute with a $-$ sign or $+$ sign respectively.

Since the last three terms are the same for all values of n, we have that

$$\sum_{n\geq N} \exp(s_N^n + ... + s_n^n) \geq \exp(\alpha_N - \vartheta_{N-1} - \vartheta_N) \sum_{n\geq N} \exp(-\alpha_n + \vartheta_n + \vartheta_{n+1}). \tag{3.44}$$

Since ϑ_n, α_n are i.i.d. standard Gaussians, $X_n := \exp(-\alpha_n + \vartheta_n + \vartheta_{n+1})$ is distributed as $\exp(\mathcal{N}(0, 3))$ for all n. Further, $(X_{2k})_{2k\geq N}$ are independent. Therefore by the strong law of large numbers,

$$\sum_{n\geq N} \exp(-\alpha_n + \vartheta_n + \vartheta_{n+1}) \geq \sum_{k\,:\,2k\geq N} X_{2k} = \infty \text{ a.s.} \tag{3.45}$$

Combining (3.44) and (3.45) completes the proof. □

Now we address the slightly more complicated case where the chain $(e_n)_{n\geq 1}$ is a branch of the dual tree, but switches fans infinitely many times. First we relate this to a particular form of sum, and then we show that this sum is infinite almost surely.

Lemma 3.7.6. *Suppose that $s = s_\Theta$ and that $(e_n)_{n\geq 1}$ is a branch of the dual tree which switches fan infinitely many times. Then*

$$\sum_{k\geq 1} (\prod_{i=1}^{k} A_i) X_k = \infty \qquad \Longrightarrow \qquad \sum_{n\geq 1} \exp(s_1^n + ... + s_n^n) = \infty,$$

where for all i, $\log A_i \sim 2\mathcal{N}(0, \sigma^2)$, where σ^2 is equal to 1 or 2. For all k, $\log X_k \sim \mathcal{N}(0, \sigma^2)$ where σ^2 is either 2 or 3.

Further, $(A_i)_{i\geq 1}$ is an independent collection and X_k is independent of $(A_i)_{i=1}^k$ (though X_k may depend on A_{k+1}).

Remark 3.7.7. We saw above that when all the edges are in the same fan, we have cancellation and only see contribution from a first three and last three terms in each sum. Here, the random variables A_i correspond to terms that are "accumulated" (i.e., that double instead of cancelling) when we change fans, and therefore are factors for all later terms. The random variables X_k are correspond to the diamond shears on a last few edges.

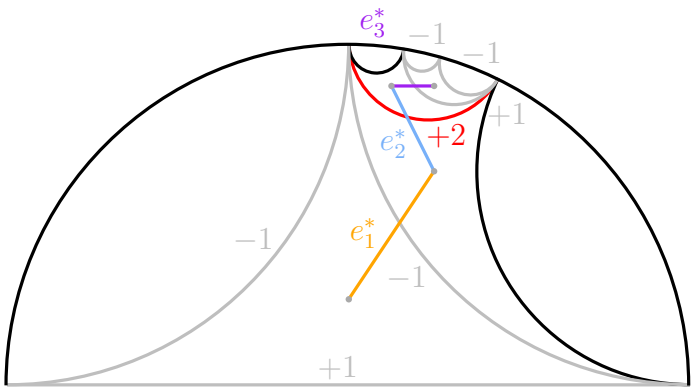


Figure 3.11: Example of a sequence of dual edges with a fan switch (the b sequence is 101). Edges with nonzero diamond shear corresponding to $s(e_1) + s(e_2) + s(e_3)$ are colored grey or red; red means diamond shear accumulated instead of cancelling.

Proof. This argument is completely combinatorial. Let $(e_{n_k})_{k\geq 1} \subseteq (e_n)_{n\geq 1}$ be the sequence of all edges where we switch fans, i.e. where the fan shared by $\{e_{n_k-1}, e_{n_k}\}$ is different from the fan shared by $\{e_{n_k}, e_{n_k+1}\}$. By assumption this is an infinite sequence.

Notice that since $(e_n)_{n\geq 1}$ is a branch of the dual tree, if an edge $e \in E$ is along a quad around some edge in the chain, then it is an edge along the quad around exactly two edges in the chain. In particular, for some n it is an edge along either i) the quad around e_n and the quad around e_{n+1} or ii) the quad around e_n and the quad around e_{n+2}. Further, note that two distinct quads share either zero or one edges.

If $e_n, ..., e_{n+m}$ are consecutive edges in the same fan, we saw in the previous lemma that the diamond shears for all but six edges cancel (corresponding to three terms at the beginning of the sequence and three at the end).

At n_k, where $\{e_{n_k-1}, e_{n_k}\}$ and $\{e_{n_k}, e_{n_k+1}\}$ are contained in different fans, some of the diamond shears terms could double instead of cancelling. The contributions of all edges will double or cancel eventually because every edge is along exactly two quads for edges in the chain. We let $A_k = \exp(2N_k)$, where N_k is the sum of the diamond shear variables that are doubled when we add the shear on e_{n_k}. The quad around e_{n_k} shares at most two total edges with the quads around e_m for $m < n_k$. Therefore N_k is distributed as either $\mathcal{N}(0,1)$ or $\mathcal{N}(0,2)$ (here we are using that diamond shears on different edges are independent and that the law is symmetric). By construction, for any collection $\{k_1, ..., k_m\}$, $N_{k_1}, ..., N_{k_m}$ are sums of disjoint collections of diamond shear variables, so $(A_k)_{k\geq 1}$ is independent.

Excluding a few terms at the beginning of the chain, the variable $\log X_k$ is the sum of the diamond shear variables for edges appearing at the end of the sequence at level n_k, which up to level n_k have appeared only once. There are either two or three such edges for each k, so X_k is distributed as either $\mathcal{N}(0,2)$ or $\mathcal{N}(0,3)$. The edges defining X_k have appeared in only once up to n_k steps whereas the edges defining $\{A_1, \ldots, A_k\}$ have appeared twice, so these collections of edges are disjoint and hence X_k is independent of $\{A_1, ..., A_k\}$.

Since all terms in (3.43) are positive, it follows that up to a multiplicative factor C

(corresponding to the first three terms in the chain),

$$\sum_{n\geq 1}\exp(s_1^n+\ldots+s_n^n)\geq C\sum_{k\geq 1}(\prod_{i=1}^{k}A_i)X_k.$$

This completes the proof. □

Finally, we use a random walk argument to show that the sum of products in the previous lemma is infinite for $s=s_\Theta$.

Lemma 3.7.8. *For X_k and A_i as in Lemma 3.7.6,*

$$\sum_{k\geq 1}(\prod_{i=1}^{k}A_i)X_k=\infty \quad a.s.$$

Proof. Define the notation

$$S_k:=\sum_{i=1}^{k}\log A_i+\log X_k$$

$$Z_k:=S_k-\log X_k=\sum_{i=1}^{k}A_i.$$

Since the normal distribution is infinitely divisible, and since $4\leq \mathrm{Var}(\log A_i)\leq 8$, there is $4k\leq t(k)\leq 8k$ such that Z_k is a sum of $t(k)$ independent copies of $\mathcal{N}(0,1)$. By the central limit theorem, $Z(k)/\sqrt{t(k)}$ converges in distribution to $\mathcal{N}(0,1)$ and therefore

$$\lim_{k\to\infty}\mathbb{P}\left(\frac{Z_k}{\sqrt{k}}>1\right)\geq\lim_{k\to\infty}\mathbb{P}\left(\frac{Z_k}{\sqrt{t(k)}}>\frac{1}{4}\right)>0.$$

On the other hand,

$$\mathbb{P}\left(\frac{S_k}{\sqrt{k}}>1\right)=\mathbb{P}\left(\frac{(S_k-Z_k)+Z_k}{\sqrt{k}}\geq 1\right)\geq\mathbb{P}\left(\frac{S_k-Z_k}{\sqrt{k}}>0,\ \frac{Z_k}{\sqrt{k}}>1\right).$$

Note that $S_k-Z_k=\log X_k$ is distributed as a centered normal random variable with variance 2 or 3 and is independent of Z_k. Hence for all k,

$$\mathbb{P}\left(\frac{S_k}{\sqrt{k}}>1\right)=\mathbb{P}\left(\frac{S_k-Z_k}{\sqrt{k}}>0\right)\cdot\mathbb{P}\left(\frac{Z_k}{\sqrt{k}}>1\right)=\frac{1}{2}\mathbb{P}\left(\frac{Z_k}{\sqrt{k}}>1\right).$$

Therefore

$$\lim_{k\to\infty}\mathbb{P}\left(\frac{S_k}{\sqrt{k}}>1\right)>0.$$

By the Kolmogorov 0-1 law,

$$\mathbb{P}(\limsup_{k\to\infty}S_k=\infty)=1.$$

Therefore $\{S_k > 0\}$ occurs for infinitely many k a.s. and hence

$$\sum_{k\geq 1}(\prod_{i=1}^{k} A_i)X_k = \sum_{k\geq 1}\exp(S_k) = \infty \text{ a.s.}$$

□

Proof of Theorem 3.7.1. By Lemma 3.7.4, it suffices to check chains that are branches of the dual tree. By Lemma 3.7.5, (3.43) is satisfied for all branches which switch fan finitely many times. By Lemma 3.7.6 and Lemma 3.7.8, (3.43) is satisfied for all branches which switch fan infinitely many times. □

References

[Ada94] C. Adams, *The Knot Book*. W.H. Freeman, 1994.

[ADS15] M. Aizenman, H. Duminil-Copin, and V. Sidoravicius, "Random currents and continuity of Ising model's spontaneous magnetization," *Comm. Math. Phys.*, vol. 334, no. 2, pp. 719–742, 2015.

[Alm86] F. Almgren, "Optimal isoperimetric inequalities," *Indiana University mathematics journal*, vol. 35, no. 3, pp. 451–547, 1986.

[AMS11] L. Ambrosio, E. Mainini, and S. Serfaty, "Gradient flow of the Chapman-Rubinstein-Schatzman model for signed vortices," *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, vol. 28, no. 2, pp. 217–246, 2011.

[ACHS21a] O. Angel, D. A. Croydon, S. Hernandez-Torres, and D. Shiraishi, "Scaling limits of the three-dimensional uniform spanning tree and associated random walk," *Ann. Probab.*, vol. 49, no. 6, pp. 3032–3105, 2021.

[ACHS21b] O. Angel, D. A. Croydon, S. Hernandez-Torres, and D. Shiraishi, "The number of spanning clusters of the uniform spanning tree in three dimensions," in *Stochastic analysis, random fields and integrable probability—Fukuoka 2019*, ser. Adv. Stud. Pure Math. Vol. 87, Math. Soc. Japan, Tokyo, [2021] ©2021, pp. 403–414.

[AriBC] Aristotle, *On the Heavens, Book III, Part 8*, trans. by J. L. Stocks, 350 BC.

[Bed19a] G. Bednik, "Hopfions in a lattice dimer model," *Physical Review B*, vol. 100, no. 2, p. 024 420, 2019.

[Bed19b] G. Bednik, "Probing topological properties of a three-dimensional lattice dimer model with neural networks," *Physical Review B*, vol. 100, no. 18, p. 184 414, 2019.

[BNR23] T. Berggren, M. Nicoletti, and M. Russkikh, "Perfect t-Embeddings of Uniformly Weighted Aztec Diamonds and Tower Graphs," *International Mathematics Research Notices*, vol. 2024, no. 7, pp. 5963–6007, Dec. 2023.

[Bis21] C. J. Bishop, "Function-theoretic characterization of weil-petersson curves," *Preprint https://www.math.stonybrook.edu/~bishop/papers/wp_fcnthy.pdf*, 2021.

[Bis19] C. J. Bishop, "Weil-petersson curves, conformal energies, β-numbers, and minimal surfaces," *Preprint http://www.math.stonybrook.edu/~bishop/papers/wpbeta.pdf*, 2019.

[Bon09] F. Bonahon, *Low-dimensional geometry* (Student Mathematical Library). American Mathematical Society, Providence, RI; Institute for Advanced Study (IAS), Princeton, NJ, 2009, vol. 49, pp. xvi+384, From Euclidean surfaces to hyperbolic knots, IAS/Park City Mathematical Subseries.

[BJM+23] M. Bonk, J. Junnila, D. Marshall, S. Rohde, and Y. Wang, "Piecewise geodesic Jordan curves II: Loewner energy, complex projective structure, and new accessory parameters," *in preparation*, 2023+.

[BR87] M. J. Bowick and S. G. Rajeev, "The holomorphic geometry of closed bosonic string theory and $\mathrm{Diff}(S^1)/S^1$," *Nuclear Physics B*, vol. 293, pp. 348–384, 1987.

[Bro86] A. Z. Broder, "How hard is it to marry at random? (On the approximation of the permanent)," in *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, ser. STOC '86, Berkeley, California, USA: Association for Computing Machinery, 1986, pp. 50–58.

[BK18] A. Bufetov and A. Knizel, "Asymptotics of random domino tilings of rectangular Aztec diamonds," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 54, no. 3, pp. 1250–1290, 2018.

[BS94] R. Burton and J. E. Steif, "Non-uniqueness of measures of maximal entropy for subshifts of finite type," *Ergodic Theory and Dynamical Systems*, vol. 14, no. 2, pp. 213–235, 1994.

[CSW23] N. Chandgotia, S. Sheffield, and C. Wolfram, "Large deviations for the 3d dimer model," *aXiv preprint arXiv:2304.08468*, 2023.

[Cha19] S. Chatterjee, "Yang–mills for probabilists," in *Probability and Analysis in Interacting Physical Systems: In Honor of SRS Varadhan, Berlin, August, 2016*, Springer, 2019, pp. 1–16.

[CEP96] H. Cohn, N. Elkies, and J. Propp, "Local statistics for random domino tilings of the Aztec diamond," *Duke Mathematical Journal*, vol. 85, no. 1, pp. 117–166, 1996.

[CKP01] H. Cohn, R. Kenyon, and J. Propp, "A variational principle for domino tilings," *Journal of the American Mathematical Society*, vol. 14, no. 2, pp. 297–346, 2001.

[Cui00] G. Cui, "Integrably asymptotic affine homeomorphisms of the circle and Teichmüller spaces," *Sci. China Ser. A*, vol. 43, no. 3, pp. 267–279, 2000.

[DS10] D. De Silva and O. Savin, "Minimizers of convex functionals arising in random surfaces," *Duke Math. J.*, vol. 151, no. 3, pp. 487–532, 2010.

[DZ09] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Stochastic Modelling and Applied Probability). Springer Berlin Heidelberg, 2009.

[DK95] C. Dress and W. Krauth, "Cluster algorithm for hard spheres and related systems," *Journal of Physics A: Mathematical and General*, vol. 28, no. 23, p. L597, 1995.

[Dum17] H. Duminil-Copin, *Random currents expansion of the ising model*, 2017.

[FM07] A. Fletcher and V. Markovic, *Quasiconformal Maps and Teichmüller Theory* (Oxford Graduate Texts in Mathematics). Oxford University Press, 2007, pp. viii+189.

[FF56] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[Fou96] J. Fournier, "Pavage des figures planes sans trous par des dominos: Fondement graphique de l'algorithme de Thurston, parallélisation, unicité et décomposition," *Theoretical Computer Science*, vol. 159, no. 1, pp. 105–128, 1996, Selected Papers from the "GASCOM '94" and the "Polynominoes and Tilings" workshops.

[FHNQ11] M. Freedman, M. B. Hastings, C. Nayak, and X.-L. Qi, "Weakly coupled non-Abelian anyons in three dimensions," *Physical Review B*, vol. 84, no. 24, p. 245 119, 2011.

[FKMS22] J. Freire, C. J. Klivans, P. H. Milet, and N. C. Saldanha, "On the connectivity of spaces of three-dimensional tilings," *Transactions of the American Mathematical Society*, vol. 375, pp. 1579–1605, 2022.

[FP22] I. Frenkel and R. Penner, "Sketch of a program for automorphic functions from universal Teichmüller theory to capture monstrous moonshine," *Comm. Math. Phys.*, vol. 389, no. 3, pp. 1525–1567, 2022.

[GL00] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory* (Mathematical Surveys and Monographs). American Mathematical Society, Providence, RI, 2000, vol. 76, pp. xx+372.

[GM05] J. B. Garnett and D. E. Marshall, *Harmonic measure* (New Mathematical Monographs). Cambridge University Press, Cambridge, 2005, vol. 2, pp. xvi+571.

[Geo11] H.-O. Georgii, *Gibbs measures and phase transitions* (De Gruyter Studies in Mathematics), Second. Walter de Gruyter & Co., Berlin, 2011, vol. 9, pp. xiv+545.

[GV89] I. Gessel and X. Viennot, "Determinants, paths, and plane partitions," 1989.

[GMT20] A. Giuliani, V. Mastropietro, and F. L. Toninelli, "Non-integrable dimers: Universal fluctuations of tilted height profiles," *Comm. Math. Phys.*, vol. 377, no. 3, pp. 1883–1959, 2020.

[GRT22] A. Giuliani, B. Renzi, and F. Toninelli, "Weakly non-planar dimers," *arXiv preprint arXiv:2207.10428*, 2022.

[Gor21] V. Gorin, *Lectures on random lozenge tilings*. Cambridge University Press, 2021, vol. 193.

[Guo00] H. Guo, "Integrable Teichmüller spaces," *Sci. China Ser. A*, vol. 43, no. 1, pp. 47–58, 2000.

[Hal35] P. Hall, "On representatives of subsets," *Journal of the London Mathematical Society*, vol. s1-10, no. 1, pp. 26–30, 1935.

[Hal70] R. R. Hall, "A note on Farey series," *J. London Math. Soc. (2)*, vol. 2, pp. 139–148, 1970.

[HLT23] I. Hartarsky, L. Lichev, and F. Toninelli, "Local dimer dynamics in higher dimensions," *arXiv preprint arXiv:2304.10930*, 2023.

[Hat22] A. Hatcher, *Topology of Numbers*. American Math Society, online preliminary version https://pi.math.cornell.edu/~hatcher/TN/TNpage.html, 2022.

[HKMS03] D. A. Huse, W. Krauth, R. Moessner, and S. L. Sondhi, "Coulomb and liquid dimer models in three dimensions," *Phys. Rev. Lett.*, vol. 91, p. 167 004, 16 Oct. 2003.

[JPS98] W. Jockusch, J. Propp, and P. Shor, "Random domino tilings and the Arctic circle theorem," *arXiv preprint arXiv:math/9801068*, 1998.

[Joh21] K. Johansson, *Strong Szegö theorem on a Jordan curve*, 2021.

[KM08] J. Kahn and V. Markovic, *Random ideal triangulations and the Weil-Petersson distance between finite degree covers of punctured Riemann surfaces*, 2008.

[Kas61] P. Kasteleyn, "The statistics of dimers on a lattice: I. the number of dimer arrangements on a quadratic lattice," *Physica*, vol. 27, no. 12, pp. 1209–1225, 1961.

[Kel98] G. Keller, *Equilibrium states in ergodic theory*. Cambridge University Press, 1998, vol. 42.

[Ken00] R. Kenyon, "Conformal invariance of domino tiling," *Annals of probability*, pp. 759–795, 2000.

[Ken14] R. Kenyon, "Conformal invariance of loops in the double-dimer model," *Communications in Mathematical Physics*, vol. 326, no. 2, pp. 477–497, 2014.

[Ken01] R. Kenyon, "Dominos and the Gaussian Free Field," *The Annals of Probability*, vol. 29, no. 3, pp. 1128–1137, 2001.

[Ken09] R. Kenyon, "Lectures on dimers," *aXiv preprint arXiv:0910.3129*, 2009.

[KO07] R. Kenyon and A. Okounkov, "Limit shapes and the complex burgers equation," *Acta Mathematica*, vol. 199, no. 2, pp. 263–302, 2007.

[KOS06] R. Kenyon, A. Okounkov, and S. Sheffield, "Dimers and amoebae," *Annals of mathematics*, pp. 1019–1056, 2006.

[KP22a] R. Kenyon and C. Pohoata, "The multinomial tiling model," *Ann. Probab.*, vol. 50, no. 5, pp. 1986–2012, 2022.

[KP22b] R. Kenyon and I. Prause, "Gradient variational problems in $\mathbb{R}^2$," *Duke Mathematical Journal*, vol. 171, no. 14, pp. 3003–3022, 2022.

[KP24] R. Kenyon and I. Prause, "Limit shapes from harmonicity: Dominos and the five vertex model," *Journal of Physics A: Mathematical and Theoretical*, vol. 57, no. 3, p. 035 001, Jan. 2024.

[KW09] R. Kenyon and P. Winkler, "Branched polymers," *Amer. Math. Monthly*, vol. 116, no. 7, pp. 612–628, 2009.

[KS04] R. W. Kenyon and S. Sheffield, "Dimers, tilings and trees," *J. Combin. Theory Ser. B*, vol. 92, no. 2, pp. 295–317, 2004.

[KS22] C. J. Klivans and N. C. Saldanha, "Domino tilings and flips in dimensions 4 and higher," *Algebr. Comb.*, vol. 5, no. 1, pp. 163–185, 2022.

[Koz07] G. Kozma, "The scaling limit of loop-erased random walk in three dimensions," *Acta Math.*, vol. 199, no. 1, pp. 29–152, 2007.

[KM03] W. Krauth and R. Moessner, "Pocket Monte Carlo algorithm for classical doped dimer models," *Phys. Rev. B*, vol. 67, p. 064 503, 6 Feb. 2003.

[KMT20] A. Krieger, G. Menz, and M. Tassy, "Deducing a variational principle with minimal *a priori* assumptions," *Electron. J. Combin.*, vol. 27, no. 4, Paper No. 4.1, 53, 2020.

[Kuc22] N. Kuchumov, "A variational principle for domino tilings of multiply connected domains," *arXiv preprint arXiv:2110.06896*, 2022.

[LZ12] J. C. Lagarias and C. Zong, "Mysteries in packing regular tetrahedra," *Notices Amer. Math. Soc.*, vol. 59, no. 11, pp. 1540–1549, 2012.

[Lam21] P. Lammers, "A generalisation of the honeycomb dimer model to higher dimensions," *Ann. Probab.*, vol. 49, no. 2, pp. 1033–1066, 2021.

[LR69] O. E. Lanford III and D. Ruelle, "Observables at infinity and states with short range correlations in statistical mechanics," *Comm. Math. Phys.*, vol. 13, pp. 194–215, 1969.

[LSW04] G. F. Lawler, O. Schramm, and W. Werner, "Conformal invariance of planar loop-erased random walks and uniform spanning trees," *Ann. Probab.*, vol. 32, no. 1B, pp. 939–995, 2004.

[Leh87] O. Lehto, *Univalent functions and Teichmüller spaces* (Graduate Texts in Mathematics). Springer-Verlag, New York, 1987, vol. 109, pp. xii+257.

[LS22] X. Li and D. Shiraishi, "The Hölder continuity of the scaling limit of three-dimensional loop-erased random walk," *Electron. J. Probab.*, vol. 27, Paper No. 144, 37, 2022.

[LMN01] J. Linde, C. Moore, and M. G. Nordahl, "An n-dimensional generalization of the rhombus tiling," in *Discrete models: combinatorics, computation, and geometry (Paris, 2001)*, ser. Discrete Math. Theor. Comput. Sci. Proc., AA, Maison Inform. Math. Discrèt. (MIMD), Paris, 2001, pp. 023–042.

[Lin73] B. Lindström, "On the Vector Representations of Induced Matroids," *Bulletin of the London Mathematical Society*, vol. 5, no. 1, pp. 85–90, Mar. 1973.

[Lit75] C. Little, "A characterization of convertible (0, 1)-matrices," *Journal of Combinatorial Theory, Series B*, vol. 18, no. 3, pp. 187–208, 1975.

[Loe23] K. Loewner, "Untersuchungen über schlichte konforme Abbildungen des Einheitskreises. I," *Math. Ann.*, vol. 89, no. 1-2, pp. 103–121, 1923.

[MP98] F. Malikov and R. C. Penner, "The Lie algebra of homeomorphisms of the circle," *Adv. Math.*, vol. 140, no. 2, pp. 282–322, 1998.

[MRW22] D. Marshall, S. Rohde, and Y. Wang, "Piecewise geodesic Jordan curves I: Weldings, explicit computations, and Schwarzian derivatives," *preprint*, 2022.

[Mil15] P. H. Milet, "Domino tilings of three-dimensional regions," *PhD thesis, arXiv preprint arXiv:1503.04617*, 2015.

[MS14a] P. H. Milet and N. C. Saldanha, "Domino tilings of three-dimensional regions: Flips, trits and twists," *arXiv preprint arXiv:1410.7693*, 2014.

[MS] P. H. Milet and N. C. Saldanha, *Enumeration of tilings for the 4x4x4 box*, http://mat.puc-rio.br/~nicolau/multiplex/example444.html.

[MS15] P. H. Milet and N. C. Saldanha, "Flip invariance for domino tilings of three-dimensional regions with two floors," *Discrete Comput. Geom.*, vol. 53, no. 4, pp. 914–940, 2015.

[MS14b] P. H. Milet and N. C. Saldanha, "Twists for duplex regions," *arXiv preprint arXiv:1411.1793*, 2014.

[Nag88] S. Nag, *The complex analytic theory of Teichmüller spaces* (Canadian Mathematical Society Series of Monographs and Advanced Texts). John Wiley & Sons, Inc., New York, 1988, pp. xiv+427, A Wiley-Interscience Publication.

[NV90] S. Nag and A. Verjovsky, "Diff(S^1) and the Teichmüller spaces," *Comm. Math. Phys.*, vol. 130, no. 1, pp. 123–138, 1990.

[PŠ22] H. Parlier and D. Šarić, *Quasisymmetric maps, shears, lambda lengths and flips*, 2022.

[Pek95] O. Pekonen, "Universal Teichmüller space in geometry and physics," *J. Geom. Phys.*, vol. 15, no. 3, pp. 227–251, 1995.

[PW21] E. Peltola and Y. Wang, "Large deviations of multichordal SLE_{0+}, real rational functions, and zeta-regularized determinants of Laplacians," *To appear in J. Eur. Math. Soc.*, 2021.

[Pen02] R. C. Penner, "On Hilbert, Fourier, and wavelet transforms," *Comm. Pure Appl. Math.*, vol. 55, no. 6, pp. 772–814, 2002.

[Pen93] R. C. Penner, "Universal constructions in Teichmüller theory," *Adv. Math.*, vol. 98, no. 2, pp. 143–215, 1993.

[Pen92] R. C. Penner, "Weil-Petersson volumes," *J. Differential Geom.*, vol. 35, no. 3, pp. 559–608, 1992.

[Pen12] R. C. Penner, *Decorated Teichmüller theory* (QGM Master Class Series). European Mathematical Society (EMS), Zürich, 2012, pp. xviii+360, With a foreword by Yuri I. Manin.

[PR14] B. Piccoli and F. Rossi, "Generalized Wasserstein distance and its application to transport equations with source," en, *Arch. Ration. Mech. Anal.*, vol. 211, no. 1, pp. 335–358, Jan. 2014.

[PR16] B. Piccoli and F. Rossi, "On properties of the generalized Wasserstein distance," en, *Arch. Ration. Mech. Anal.*, vol. 222, no. 3, pp. 1339–1365, Dec. 2016.

[PRT19] B. Piccoli, F. Rossi, and M. Tournus, "A Wasserstein norm for signed measures, with application to nonlocal transport equation with source term," *arXiv preprint arXiv:1910.05105*, 2019.

[QT22] A. Quitmann and L. Taggi, "Macroscopic loops in the $3d$ double-dimer model," *arXiv preprint arXiv:2206.08284*, 2022.

[QT23] A. Quitmann and L. Taggi, "Macroscopic loops in the Bose gas, spin $O(N)$ and related models," *Communications in Mathematical Physics*, pp. 1–56, 2023.

[RT00] D. Randall and P. Tetali, "Analyzing Glauber dynamics by comparison of Markov chains," *Journal of Mathematical Physics*, vol. 41, no. 3, pp. 1598–1615, 2000.

[RY00] D. Randall and G. Yngve, "Random three-dimensional tilings of Aztec octahedra and tetrahedra: An extension of domino tilings," in *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 2000)*, ACM, New York, 2000, pp. 636–645.

[RST99] N. Robertson, P. D. Seymour, and R. Thomas, "Permanents, Pfaffian orientations, and even directed circuits," *Annals of Mathematics*, vol. 150, no. 3, pp. 929–975, 1999.

[RS05] S. Rohde and O. Schramm, "Basic properties of SLE," *Ann. of Math. (2)*, vol. 161, no. 2, pp. 883–924, 2005.

[RW19] S. Rohde and Y. Wang, "The Loewner energy of loops and regularity of driving functions," *International Mathematics Research Notices*, published online 2019.

[Rus18] M. Russkikh, "Dimers in piecewise temperley domains," *Communications in Mathematical Physics*, vol. 359, Apr. 2018.

[Sal20] N. C. Saldanha, "Domino tilings in dimension 3," in *2020 Fall Central Sectional Meeting*, AMS, 2020.

[Sal21] N. C. Saldanha, "Domino tilings of cylinders: Connected components under flips and normal distribution of the twist," *Electron. J. Combin.*, vol. 28, no. 1, Paper No. 1.28, 23, 2021.

[Sal22] N. C. Saldanha, "Domino tilings of cylinders: The domino group and connected components under flips," *Indiana Univ. Math. J.*, vol. 71, no. 3, pp. 965–1002, 2022.

[Šar10] D. Šarić, "Circle homeomorphisms and shears," *Geom. Topol.*, vol. 14, no. 4, pp. 2405–2430, 2010.

[Šar06] D. Šarić, "Real and complex earthquakes," *Trans. Amer. Math. Soc.*, vol. 358, no. 1, pp. 233–249, 2006.

[Šar21] D. Šarić, "Shears for quasisymmetric maps," *Proc. Amer. Math. Soc.*, vol. 149, no. 6, pp. 2487–2499, 2021.

[Šar13] D. Šarić, "Zygmund vector fields, Hilbert transform and Fourier coefficients in shear coordinates," *Amer. J. Math.*, vol. 135, no. 6, pp. 1559–1600, 2013.

[ŠWW22] D. Šarić, Y. Wang, and C. Wolfram, "Circle homeomorphisms with square summable diamond shears," *aXiv preprint arXiv:2211.11497*, 2022.

[STZ99] M. E. Schonbek, A. N. Todorov, and J. P. Zubelli, "Geodesic flows on diffeomorphisms of the circle, Grassmannians, and the geometry of the periodic KdV equation," *Adv. Theor. Math. Phys.*, vol. 3, no. 4, pp. 1027–1092, 1999.

[Sch07] O. Schramm, "Conformally invariant scaling limits: An overview and a collection of problems," in *International Congress of Mathematicians. Vol. I*, Eur. Math. Soc., Zürich, 2007, pp. 513–543.

[Sch00] O. Schramm, "Scaling limits of loop-erased random walks and uniform spanning trees," *Israel J. Math.*, vol. 118, pp. 221–288, 2000.

[SS09] O. Schramm and S. Sheffield, "Contour lines of the two-dimensional discrete Gaussian free field," *Acta Math.*, vol. 202, no. 1, pp. 21–137, 2009.

[SM06] E. Sharon and D. Mumford, "2D-Shape analysis using Conformal Mapping," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 55–75, 2006.

[She05] S. Sheffield, "Random surfaces," *Astérisque*, vol. 304, 2005.

[SY14] S. Sheffield and A. Yadin, "Tricolor percolation and random paths in 3D," *Electron. J. Probab.*, vol. 19, no. 4, 23, 2014.

[She18] Y. Shen, "Weil-Petersson Teichmüller space," *Amer. J. Math.*, vol. 140, no. 4, pp. 1041–1074, 2018.

[ST20] Y. Shen and S. Tang, "Weil-Petersson Teichmüller space II: Smoothness of flow curves of $H^{\frac{3}{2}}$-vector fields," *Adv. Math.*, vol. 359, pp. 106891, 25, 2020.

[STW18] Y. Shen, S. Tang, and L. Wu, "Weil-Petersson and little Teichmüller spaces on the real line," *Ann. Acad. Sci. Fenn. Math.*, vol. 43, no. 2, pp. 935–943, 2018.

[Shi18] D. Shiraishi, "Growth exponent for loop-erased random walk in three dimensions," *Ann. Probab.*, vol. 46, no. 2, pp. 687–774, 2018.

[Shi19] D. Shiraishi, "Hausdorff dimension of the scaling limit of loop-erased random walk in three dimensions," *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 55, no. 2, pp. 791–834, 2019.

[Smi10] S. Smirnov, "Conformal invariance in random cluster models. I. Holomorphic fermions in the Ising model," *Ann. of Math. (2)*, vol. 172, no. 2, pp. 1435–1467, 2010.

[Smi06] S. Smirnov, "Towards conformal invariance of 2D lattice models," in *International Congress of Mathematicians*, vol. 2, Eur. Math. Soc., Zürich, 2006, pp. 1421–1451.

[Str10] G. Strang, "Maximum flows and minimum cuts in the plane," *J. Global Optimization*, vol. 47, pp. 527–535, Jul. 2010.

[Tag22] L. Taggi, "Uniformly positive correlations in the dimer model and macroscopic interacting self-avoiding walk in $\mathbb{Z}^d$, $d \geq 3$," *Communications on Pure and Applied Mathematics*, vol. 75, no. 6, pp. 1183–1236, 2022.

[TT06] L. A. Takhtajan and L.-P. Teo, "Weil-Petersson metric on the universal Teichmüller space," *Mem. Amer. Math. Soc.*, vol. 183, no. 861, pp. viii+119, 2006.

[TF61] H. N. V. Temperley and M. E. Fisher, "Dimer problem in statistical mechanics-an exact result," *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, vol. 6, no. 68, pp. 1061–1063, 1961.

[Thu90] W. P. Thurston, "Conway's tiling groups," *The American Mathematical Monthly*, vol. 97, no. 8, pp. 757–773, 1990.

[Thu22] W. P. Thurston, "Earthquakes in 2-dimensional hyperbolic geometry," in *Collected works of William P. Thurston with commentary. Vol. I. Foliations, surfaces and differential geometry*, Reprint of [2235712], Amer. Math. Soc., Providence, RI, [2022] ©2022, pp. 511–531.

[vdBer93] J. van den Berg, "A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet," *Comm. Math. Phys.*, vol. 152, no. 1, pp. 161–166, 1993.

[vdBS94] J. van den Berg and J. E. Steif, "Percolation and the hard-core lattice gas model," *Stochastic Process. Appl.*, vol. 49, no. 2, pp. 179–197, 1994.

[Var16] S. Varadhan, *Large deviations* (Courant Lecture Notes). Courant Institute of Mathematical Sciences, 2016.

[VW20a] F. Viklund and Y. Wang, "Interplay between Loewner and Dirichlet energies via conformal welding and flow-lines," *Geom. Funct. Anal.*, vol. 30, no. 1, pp. 289–321, 2020.

[VW20b] F. Viklund and Y. Wang, "The Loewner-Kufarev Energy and Foliations by Weil-Petersson Quasicircles," *arXiv preprint: 2012.05771*, 2020.

[Vil09] C. Villani, "The Wasserstein distances," in *Optimal Transport: Old and New*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 93–111.

[Wan21] Y. Wang, "A note on Loewner energy, conformal restriction and Werner's measure on self-avoiding loops," *Ann. Inst. Fourier (Grenoble)*, vol. 71, no. 4, pp. 1791–1805, 2021.

[Wan19a] Y. Wang, "Equivalent descriptions of the Loewner energy," *Invent. Math.*, vol. 218, no. 2, pp. 573–621, 2019.

[Wan22] Y. Wang, "Large deviations of Schramm-Loewner evolutions: A survey," *Probab. Surv.*, vol. 19, pp. 351–403, 2022.

[Wan19b] Y. Wang, "The energy of a deterministic Loewner chain: Reversibility and interpretation via SLE_{0+}," *J. Eur. Math. Soc. (JEMS)*, vol. 21, no. 7, pp. 1915–1941, 2019.

[Wan24] Y. Wang, *Two optimization problems for the loewner energy*, 2024.

[WMDB02] M. Widom, R. Mosseri, N. Destainville, and F. Bailly, "Arctic octahedron in three-dimensional rhombus tilings and related integer solid partitions," *J. Statist. Phys.*, vol. 109, no. 5-6, pp. 945–965, 2002.

[WZ22] P. Wiegmann and A. Zabrodin, "Dyson gas on a curved contour," *J. Phys. A*, vol. 55, no. 16, Paper No. 165202, 34, 2022.

[Wil10] D. B. Wilson, "Dimension of the loop-erased random walk in three dimensions," *Phys. Rev. E*, vol. 82, p. 062 102, 6 Dec. 2010.

[Wit88] E. Witten, "Coadjoint orbits of the Virasoro group," *Comm. Math. Phys.*, vol. 114, no. 1, pp. 1–53, 1988.

[Wol83] S. Wolpert, "On the symplectic geometry of deformations of a hyperbolic surface," *Ann. of Math. (2)*, vol. 117, no. 2, pp. 207–234, 1983.

[Wu11] C. Wu, "The cross-ratio distortion of integrably asymptotic affine homeomorphism of unit circle," *Science China Mathematics*, vol. 55, no. 3, pp. 625–632, Dec. 2011.