

# TOWARDS PATTERSON-SULLIVAN THEORY FOR TEICHMÜLLER SPACES OF NON-ORIENTABLE SURFACES

SAYANTAN KHAN

**ABSTRACT.** We show that a finite volume deformation retract  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)/\mathrm{MCG}(\mathcal{N}_g)$  of the moduli space  $\mathcal{M}(\mathcal{N}_g)$  of non-orientable surfaces  $\mathcal{N}_g$  behaves like the convex core of  $\mathcal{M}(\mathcal{N}_g)$ , despite not even being quasi-convex. We then show that geodesics in the convex core leave compact regions with exponentially low probabilities, showing that the action of  $\mathrm{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is *statistically convex-cocompact*. In combination with recent results of Coulon, Gekhtman, Ma, Tapie, and Yang, this leads to the construction of Patterson-Sullivan measures on the limit set, and a finite geodesic flow invariant measure for moduli spaces of non-orientable surfaces, with respect to which the geodesic flow is mixing. A key step of our proof relies on *complexity length*, developed by Dowdall and Masur, which is an alternative metric on Teichmüller space that accounts for geodesics that spend a considerable fraction of their time in the thin part.

## 1. INTRODUCTION

Mapping class groups  $\mathrm{MCG}(\mathcal{N}_g)$  of non-orientable surfaces  $\mathcal{N}_g$  have many similarities to infinite covolume geometrically finite Fuchsian groups, in a manner similar how mapping class groups of orientable surfaces sometimes behave like lattices in  $\mathrm{SL}_2(\mathbb{R})$ .

- (i) The mapping class groups  $\mathrm{MCG}(\mathcal{N}_g)$  are finitely presented.
- (ii) The action of  $\mathrm{MCG}(\mathcal{N}_g)$  on the Teichmüller space  $\mathcal{T}(\mathcal{N}_g)$ , has infinite  $\nu_N$ -covolume, where  $\nu_N$  is the generalization of the Weil-Petersson volume form on Teichmüller spaces of non-orientable surfaces ([Gen17, Theorem 17.1] and [Nor08]).
- (iii) The limit set of  $\mathrm{MCG}(\mathcal{N}_g)$  in the Thurston boundary of  $\mathcal{T}(\mathcal{N}_g)$  is  $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$ , i.e. the projective measured laminations that have no two-sided components ([EGPS23] and [Kha23]), which is a subset of the boundary with zero Lebesgue measure ([Nor08]).
- (iv) The Teichmüller geodesic flow is not ergodic with respect to any Borel measure on the unit cotangent bundle with full support ([Gen17, Proposition 17.5])
- (v) There exists an  $\mathrm{MCG}(\mathcal{N}_g)$ -equivariant deformation retract of  $\mathcal{T}(\mathcal{N}_g)$  to  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , the subset where no one-sided curve is shorter than  $\varepsilon_t > 0$ . The action of  $\mathrm{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is finite  $\nu_N$ -covolume [Gen17, Proposition 19.1].

With these similarities, one might expect that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  serves as a convex core of  $\mathcal{T}(\mathcal{N}_g)$ , and the geodesic flow restricted to tangent directions whose forward and backward end points lie in the limit set will be ergodic, with respect to a finite geodesic flow invariant measure supported only on these tangent directions.

A prior result of the author shows that this is not the case:  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is not even quasi-convex, i.e. geodesic segments whose endpoints lie in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  can travel arbitrarily far from  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

---

*Date:* Tuesday 19<sup>th</sup> March, 2024.

*2010 Mathematics Subject Classification.* 57K20.

*Key words and phrases.* mapping class group, Teichmüller space, Patterson-Sullivan theory.

**Theorem 1.1** (Theorem 5.2 of [Kha23]). *For all  $\varepsilon_t > 0$ , and all  $D > 0$ , there exists a Teichmüller geodesic segment whose endpoints lie in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , such that some point in the interior of the geodesic segment is more than distance  $D$  from  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .*

However, we show that this failure of quasi-convexity is not a serious obstruction to understanding geodesic segments whose endpoints lie in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

**Theorem 3.7.** *For any  $\varepsilon_d > 0$ , there exists constants  $\varepsilon'_t$  and  $c$ , such that any geodesic segment  $\gamma$ , whose length is more than  $c$ , with endpoints in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , for  $0 < \varepsilon_t < \varepsilon'_t$ , can be homotoped to a segment relative to endpoints to lie entirely within  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , such that the length of the homotoped segment  $\gamma'$  satisfies the following inequality.*

$$\ell(\gamma') \leq \ell(\gamma) \cdot (1 + \varepsilon_d)$$

Theorem 3.7 shows that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , despite not being convex, almost behaves like the convex core of  $\mathcal{T}(\mathcal{N}_g)$ : it is a metric subset of  $\mathcal{T}(\mathcal{N}_g)$  (with respect to the induced path metric) which is distorted by an arbitrarily small amount. We call  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  the *weak convex core* of  $\mathcal{T}(\mathcal{N}_g)$ , and focus our attention on this subspace as a metric space, where the metric is the induced path metric. If we restrict our attention to the cotangent directions in  $\mathcal{T}(\mathcal{N}_g)$  along which the geodesic flow does not eventually leave  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , we can use those cotangent directions to define a geodesic flow for  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ . We call this collection of restricted directions the restricted cotangent bundle over  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

In light of this, we restrict our attention to  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , and the  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ . Since the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is finite  $\nu_N$ -covolume (but not cocompact), one might try to prove that the action is like the action of lattices in  $\text{SL}_2(\mathbb{R})$  on  $\mathbb{H}$ . However, the results on lattices (and Teichmüller spaces of orientable surfaces) rely on having a measure preserving  $\text{SL}_2(\mathbb{R})$  action on the unit tangent bundle (respectively on the moduli space of quadratic differentials), and use the interplay between the geodesic flow and the horocycle flow.

For non-orientable surfaces, we do not have an analog of the horocycle flow on the space of quadratic differentials, so we cannot hope to directly import the techniques from the orientable case. However, Yang [Yan18] introduced a notion of *statistically convex-cocompact action*, which can replace the notion of a lattice-like action for our setting. In the setting of  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , proving statistical convex-cocompactness is equivalent to proving that geodesic segments between  $\text{MCG}(\mathcal{N}_g)$  orbit points in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  enter the thin part (i.e. the region in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  where some two-sided curve is short) with exponentially low probabilities.

Our main result is that this holds for the  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

**Theorem 1.2** (Corollary of Theorems 4.15 and 5.1). *The action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is statistically convex-cocompact.*

Gekhtman and Ma [GM23], and Coulon, Gekhtman, Tapie, and Yang [CGTY] (in upcoming work) prove the following theorems about statistically convex cocompact actions of subgroups of mapping class groups on Teichmüller space.

**Theorem 1.3** (To appear in Coulon, Gekhtman, Tapie, and Yang [CGTY]). *Let  $\Gamma$  be a subgroup of the mapping class group  $\text{MCG}(\mathcal{S}_g)$  whose action on  $\mathcal{T}(\mathcal{N}_g)$  is statistically convex-cocompact. Then the Bowen-Margulis measure  $\mu_{\text{BMS}}$  on the unit cotangent bundle of  $\mathcal{T}(\mathcal{S}_g)/\Gamma$  is finite.*

**Theorem 1.4** (Theorem 4.9 of [GM23]). *If  $\Gamma$  is a subgroup of the mapping class group such that the associated Bowen-Margulis measure is finite, then the geodesic flow is mixing with respect to  $\mu_{\text{BMS}}$ .*

As a corollary of the Theorem 1.4, we get the following results about the Patterson-Sullivan measures and the action of  $\Gamma$  on the limit set.

**Corollary 1.5.** *If  $\Gamma$  is a subgroup of  $\text{MCG}(\mathcal{S}_g)$  such that the Bowen-Margulis measure associated to  $\Gamma$  is finite, then the following results hold.*

- *There exists a unique family of conformal densities  $\{\mu_x\}_{x \in \mathcal{T}(\mathcal{S}_g)}$  on the limit set of  $\Gamma$ , given by the Patterson-Sullivan construction.*
- *The action of  $\Gamma$  on the limit set is ergodic.*

We believe Theorem 1.3 also works when we consider the  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  instead of  $\mathcal{T}(\mathcal{S}_{g-1})$  (i.e. the Teichmüller space of the orientation double cover), since the proof of Theorem 1.3 relies on considering the limit set of the group, which is the same for the action on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  as it is for the action on the entire Teichmüller space (we proved this in [Kha23, Theorem 4.2]).

Combining Theorem 1.2 with Theorems 1.3 and 1.4 lets us construct Patterson-Sullivan measures (with good dynamical properties) on the limit set, and a finite Bowen-Margulis measure on the unit cotangent bundle. In particular, letting  $\Gamma = \text{MCG}(\mathcal{N}_g) < \text{MCG}(\mathcal{S}_{g-1})$  gives us a version of Theorem 1.4 and Corollary 1.5 for  $\text{MCG}(\mathcal{N}_g)$ .

**Corollary 1.6.** *There exists a unique family of conformal densities  $\{\mu_x\}_{x \in \mathcal{T}(\mathcal{N}_g)}$  on the limit set  $\mathbb{PML}^+(\mathcal{N}_g)$ , given by the Patterson-Sullivan construction. Furthermore, the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathbb{PML}^+(\mathcal{N}_g)$  is ergodic, and the geodesic flow on the restricted cotangent bundle over  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is mixing with respect to the Bowen-Margulis measure.*

**Why we care about Patterson-Sullivan theory for  $\text{MCG}(\mathcal{N}_g)$ .**

*Counting functions.* Understanding the dynamics of the geodesic flow over the moduli space of *orientable surfaces* has led to solutions for two counting problems: one on the moduli space of hyperbolic surfaces, and one on hyperbolic surfaces themselves.

- (i) Counting closed curves in moduli space: Via techniques originally introduced to Margulis in his thesis [Mar04], one can reduce counting closed curves, which are conjugacy classes of mapping class group orbit points, to understanding the geodesic flow over the moduli space. The number of closed curves of length at most  $R$ , which we denote by  $N(R)$  has the following asymptotics (see [EM11]).

$$N(R) \sim \frac{\exp(hR)}{hR} \quad (1)$$

Here, the symbol  $\sim$  means that the ratio of the two quantities approaches a positive constant as  $R$  goes to  $\infty$ , and  $h$  is the volume growth entropy of  $\mathcal{T}(\mathcal{S}_g)$ , which is  $6g - 6$ .

- (ii) Counting *simple* closed curves on orientable hyperbolic surfaces: Mirzakhani [Mir08] proved that the counting function  $M(R)$  that counts *simple* closed curves satisfies a polynomial asymptotic.

$$M(R) \sim R^h \quad (2)$$

Here,  $h$  is again the volume growth entropy, i.e.  $6g - 6$ .

For non-orientable surfaces, the counting function does not behave like the orientable version. Gendulpe [Gen17] showed that the counting function  $N_{\text{no}}(R)$  and  $M_{\text{no}}(R)$ , which is the versions of the functions  $N(R)$  and  $M(R)$  for non-orientable surfaces satisfy the following asymptotic.

$$N_{\text{no}}(R) = o\left(\frac{\exp((3g-6)R)}{(3g-6)R}\right)$$

$$M_{\text{no}}(R) = o(R^{3g-6})$$

These asymptotics raise the question of whether there is an exponent  $h < 3g - 6$  for which the non-orientable versions of (1) and (2) continue to hold.

By establishing that the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}(\mathcal{N}_g)$  is statistically convex-cocompact, we can use the results of Coulon, Gekhtman, Ma, Tapie, and Yang to count lattice points and their conjugacy classes to obtain a non-orientable version of (1) where the role of  $h$  is played by the critical exponent for the group action of  $\text{MCG}(\mathcal{N}_g)$  with respect to the Teichmüller metric.

To explain how ergodicity of the  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$  might help count simple closed curves on  $\mathcal{N}_g$ , we outline Mirzakhani's original proof of the fact for orientable surfaces (see [Mir08] for the original proof, and [Ara22] for a gentler exposition).

*Sketch of simple closed curve counting in the orientable case.* The proof proceeds in 3 steps.

Step 1: For any simple closed curve  $\gamma$  and any  $L > 0$ , consider the measure  $\mu_L$  on the space  $\mathcal{ML}(\mathcal{S}_g)$  of measured laminations.

$$\mu_L := \frac{1}{L^{6g-6}} \sum_{\alpha \in \text{MCG}(\mathcal{S}_g)} \delta_{\frac{1}{L}\alpha\gamma}$$

Step 2: Letting  $L$  go to  $\infty$ ,  $\{\mu_L\}$  converges to measure  $\mu$  that is  $\text{MCG}(\mathcal{S}_g)$ -invariant. By ergodicity of the  $\text{MCG}(\mathcal{S}_g)$ -action on  $\mathcal{ML}(\mathcal{S}_g)$  with respect to the Thurston measure, we have that the limiting measure  $\mu$  is a constant multiple  $c$  times the Thurston measure.

Step 3: To show that the constant  $c$  is positive, one needs to average over the moduli space  $\mathcal{M}(\mathcal{S}_g)$ , using Mirzakhani's integration formula for the Weil-Petersson volume form.

□

To replicate this proof in the non-orientable setting, we pick the original simple closed curve  $\gamma$  to be a *two-sided* curve, and replace the exponent  $6g - 6$  with  $h + 1$ , where  $h$  is the critical exponent of  $\text{MCG}(\mathcal{N}_g)$ . With these replacements, we have that the measures  $\mu_L$  converge to an  $\text{MCG}(\mathcal{N}_g)$ -invariant measure supported on  $\mathcal{ML}^+(\mathcal{N}_g)$  (this is a result of Erlandsson, Gendulpe, Pasquinelli, and Souto [EGPS23]). Since our results imply ergodicity of the  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$  with respect to the Patterson-Sullivan measures, we can ask the following questions about the limiting measure  $\mu$  as well as the product of the Patterson-Sullivan measures with the Lebesgue measure.

**Question 1.7.** Is there an  $\text{MCG}(\mathcal{N}_g)$  ergodic measure on  $\mathcal{ML}^+(\mathcal{N}_g)$  that is absolutely continuous with the product of the Patterson-Sullivan measure and the Lebesgue measure, when  $\mathcal{ML}^+(\mathcal{N}_g)$  is identified with  $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g) \times \mathbb{R}$ ?

**Question 1.8.** Is the limiting measure  $\mu$  absolutely continuous with respect to an  $\text{MCG}(\mathcal{N}_g)$  ergodic measure on  $\mathcal{ML}^+(\mathcal{N}_g)$ ?

If the answer both the questions is yes, then one will have completed Step 2 of the proof for non-orientable surfaces.

Making Step 3 work for non-orientable surfaces is still open: this would require coming up with a recursive formula for volumes of  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  (or some other finite volume subset), and not all of  $\mathcal{T}(\mathcal{N}_g)$ , since that has infinite  $\nu_N$ -volume.

*Geometric finiteness for mapping class subgroups.* One can think of  $\text{MCG}(\mathcal{N}_g)$  as a subgroup of  $\text{MCG}(\mathcal{S}_{g-1})$  (where  $\mathcal{S}_{g-1}$  is the orientation double cover of  $\mathcal{N}_g$ ), where the embedding is obtained by lifting mapping classes on  $\mathcal{N}_g$  to orientation preserving mapping classes on  $\mathcal{S}_{g-1}$ . The image of  $\text{MCG}(\mathcal{N}_g)$  is an infinite-index subgroup, and stabilizes an isometrically embedded copy of  $\mathcal{T}(\mathcal{N}_g)$  inside  $\mathcal{T}(\mathcal{S}_{g-1})$ .

For subgroups of mapping class groups, the notion of convex-cocompactness was introduced by Farb and Mosher [FM02]: these groups have good properties with respect to their dynamics on the Teichmüller space. A natural generalization of these subgroups, inspired by the Kleinian setting, is the notion of geometric finiteness. While there is not universally agreed upon notion of geometric finiteness for mapping class subgroups, the following two classes are subgroups are considered to be geometrically finite by any reasonable definition.

- (i) Veech groups: These are stabilizers of Teichmüller discs in  $\mathcal{T}(\mathcal{S}_{g-1})$  which are finitely generated. They are lattices in  $\text{SL}_2(\mathbb{R})$ , and their action on the Teichmüller discs they stabilize is well understood via hyperbolic geometry.
- (ii) Combinations of Veech groups: Leininger and Reid [LR06] show that if two Veech groups  $H$  and  $K$  share a maximal parabolic subgroup  $A$ , the subgroup they generate is  $H *_A K$  (after possibly conjugating by a pseudo-Anosov).

The key emphasis with these two examples is that there are only finitely many cusps, i.e. finitely many conjugacy classes of reducible elements. However, that is not the case for  $\text{MCG}(\mathcal{N}_g)$ , it stabilizes an isometrically embedded sub-manifold, and yet there are infinitely many conjugacy classes of reducible elements. Despite having infinitely many “cusps”, our results show that it is still possible to do Patterson-Sullivan theory on  $\text{MCG}(\mathcal{N}_g)$ , which is a departure from the Fuchsian/Kleinian setting, where finite Bowen-Margulis measure requires finitely many cusps.

**Idea behind proofs of main theorems.** In this subsection, we outline the key ideas behind the proof of the main theorems.

*Weak convexity of  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .* We construct a projection map from  $\mathcal{T}(\mathcal{N}_g)$  to  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  which takes any one-sided curve of length less than  $\varepsilon_t$  and increases its length to  $\varepsilon_t$ , while keeping the lengths and twists of other curves constant. We then use Minsky’s product region theorem to show that this projection map increases distance by only a factor of  $(1 + \varepsilon_d)$ , where  $\varepsilon_d$  can be picked to be arbitrarily small.

*Statistical convexity of  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .* To show that geodesics in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  stay away from the thin part, we construct a random walk on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , and compute the probability of a single step of the random walk entering the thin part, and show that this probability is small. Estimating this probability reduces to computing an average over a ball in  $\mathbb{H}$  because of Minsky’s product region theorem. The random walk argument gives us that the number of geodesics of length at most  $R$  entering the thin part is at most  $\exp((h_{\text{NP}} - 1)R)$ , where  $h_{\text{NP}}$  is the discrete analog of the volume growth entropy of  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ . However, the total number of geodesics of length at most  $R$  grows like  $\exp(h_{\text{LP}}R)$ , where  $h_{\text{LP}}$  is the growth rate of the number of lattice



points. To show that the probability of a geodesic entering the thin part is exponentially small, we need to relate the two entropy terms, and show that  $h_{\text{LP}} > h_{\text{NP}} - 1$ .

*Showing  $h_{\text{LP}} = h_{\text{NP}}$ .* We prove entropy equality by inducting on the complexity of the surface. We first show it for surfaces with Euler characteristic equal to  $-1$  using direct methods, and reduce the inductive step to proving an estimate on complexity length for geodesic segments that spend a definite fraction of their time in thin part.

*Complexity length estimate.* In this section, we show that geodesic segments that spend a small but definite fraction of time near their end in the thin part are rare. We do this by showing that  $h_{\text{NP}}$  for a proper subsurface is strictly smaller than  $h_{\text{NP}}$  for the entire surface, and use the machinery of complexity length (due to Dowdall and Masur [DM23]), which builds upon Minsky's product region theorem and hierarchical hyperbolicity of Teichmüller space, to show that geodesic segments ending in the thin part are rare.

**Acknowledgements.** The author would like to thank his advisor Alex Wright for constant support for the duration of this project. The author would also like to thank Spencer Dowdall, Howard Masur, Kasra Rafi, and Jacob Russell for explaining various aspects of hierarchical hyperbolicity. The work done in this paper was supported by the Rackham Predoctoral Fellowship for the academic year 2022-2023.

## 2. PRELIMINARIES

**2.1. Non-orientable surfaces.** Similar to orientable surfaces, compact non-orientable surfaces with (possibly empty) boundary are classified by their *demigenus* and number of boundary components. The demigenus of a non-orientable surface is the number of copies of  $\mathbb{RP}^2$  that need to be connect-summed in order to get the non-orientable surface. An alternative way to construct non-orientable surfaces is to start with an orientable surface, and attach *crosscaps*: a crosscap is attached by deleting the interior of an embedded disc, and gluing the  $S^1$  boundary of that disc to itself via the antipodal map.

To unify notation between orientable and non-orientable surfaces, we will denote a compact surface with boundary using  $\mathcal{S}_{g,b,c}$ , which denotes a surface of genus  $g$ , with  $b$  boundary components, and  $c$  crosscaps attached. With this notation, a non-orientable surface  $\mathcal{N}_{g,b}$  of demigenus  $g$  with  $b$  boundary components is  $\mathcal{S}_{\frac{g-1}{2},b,1}$  if  $g$  is odd, and  $\mathcal{S}_{\frac{g-2}{2},b,2}$  if  $g$  is even.

One can classify simple closed curves on a non-orientable surface into two categories based on the topology of their tubular neighbourhoods.

**Two-sided curves:** These are curves whose tubular neighbourhoods are homeomorphic to cylinders.

**One-sided curves:** These are curves whose tubular neighbourhoods are homeomorphic to Möbius bands.

The orientation double cover of  $\mathcal{N}_g$  is  $\mathcal{S}_{g-1}$ , where  $p$  denotes the covering map: the one-sided curves on  $\mathcal{N}_g$  lift to a single curve on  $\mathcal{S}_{g-1}$  that is twice as long, and the two-sided curves on  $\mathcal{N}_g$  lift to two disjoint curves on  $\mathcal{S}_{g-1}$ , both of which are the same length as the original curve. We also have an orientation reversing deck transformation  $\iota$  on  $\mathcal{S}_{g-1}$  corresponding to the covering map. The map  $\iota$  swaps the lifts of the two-sided curves, and leaves the lifts of the one-sided curves invariant.

The subgroup  $\pi_1(\mathcal{S}_{g-1}) < \pi_1(\mathcal{N}_g)$  is a *characteristic* subgroup, i.e. left invariant by an automorphism of  $\pi_1(\mathcal{N}_g)$  induced by a homeomorphism, and consequently, self-homeomorphisms

of  $\mathcal{N}_g$  have a unique orientation preserving lift to self-homeomorphisms of  $\mathcal{S}_{g-1}$ , giving us an embedding  $p^*$  of mapping class groups, induced by the covering map  $p$ .

$$p^* : \text{MCG}(\mathcal{N}_g) \hookrightarrow \text{MCG}(\mathcal{S}_{g-1})$$

The image of  $\text{MCG}(\mathcal{N}_g)$  is an infinite-index subgroup of  $\text{MCG}(\mathcal{S}_{g-1})$ . The lifting map also induces an embedding of the corresponding Teichmüller spaces, where the image of  $\mathcal{T}(\mathcal{N}_g)$  is the locus left invariant by  $\iota^*$ , where  $\iota^*$  is the deck transformation induced map on  $\mathcal{T}(\mathcal{S}_{g-1})$ .

$$p^* : \mathcal{T}(\mathcal{N}_g) \hookrightarrow \mathcal{T}(\mathcal{S}_{g-1})$$

This embedding is isometric, i.e. Teichmüller geodesics joining points in the image of  $\mathcal{T}(\mathcal{N}_g)$  stay within the image of  $\mathcal{T}(\mathcal{N}_g)$ .

These facts present an alternative way of thinking about mapping class groups and Teichmüller spaces of non-orientable surfaces. They can be thought of as a special infinite index subgroup of  $\text{MCG}(\mathcal{S}_{g-1})$ , and a isometrically embedded totally real submanifold of  $\mathcal{T}(\mathcal{S}_{g-1})$ . We will use this point of view to prove some of the metric properties of  $\mathcal{T}(\mathcal{N}_g)$  we will require, but for most other applications, we prefer to think of  $\mathcal{T}(\mathcal{N}_g)$  and  $\text{MCG}(\mathcal{N}_g)$  as independent objects, without embedding them in other spaces.

The Teichmüller space for non-orientable surfaces can be given Fenchel-Nielsen coordinates using a pants decomposition for  $\mathcal{N}_g$ : the only difference from the orientable setting is that for all the one-sided curves in the pants decomposition, there is only one coordinate, associated to the length of the one-sided curve, rather than both the twist and length. This means that Teichmüller spaces of non-orientable surfaces can have odd  $\mathbb{R}$ -dimension.

Since these Teichmüller spaces of non-orientable surfaces can have odd dimension, we no longer have a symplectic structure, and a corresponding volume form. However, the image of  $\mathcal{T}(\mathcal{N}_g)$  in  $\mathcal{T}(\mathcal{S}_{g-1})$  is a Lagrangian submanifold, and consequently a Lagrangian volume form. This Lagrangian volume form  $\nu_N$  has a particularly nice description in terms of a pants decomposition  $\mathcal{P}$ , due to Norbury [Nor08].

Let  $\mathcal{P}$  be a pants decomposition for  $\mathcal{N}_g$ :  $\nu_N$  is defined in terms of the lengths and twists of curves in  $\mathcal{P}$ .

$$\nu_N = \left( \bigwedge_{\gamma_i \text{ one-sided}} \coth(\ell(\gamma_i)) d\ell(\gamma_i) \right) \wedge \left( \bigwedge_{\gamma_i \text{ two-sided}} d\tau(\gamma_i) \wedge d\ell(\gamma_i) \right)$$

Here  $\ell(\gamma_i)$  denotes the length of the curve  $\gamma_i$ , and  $\tau(\gamma_i)$  denotes the twist, when  $\gamma_i$  is two-sided.

Similar to Wolpert's magic formula, the  $\mu_N$  has the following properties.

- The form  $\nu_N$  does not depend on the choice of pants decomposition.
- $\nu_N$  is  $\text{MCG}(\mathcal{N}_g)$  invariant, up to sign.

This lets us use the absolute value of  $\nu_N$  as a volume form on the quotient  $\mathcal{T}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$ . We will, for notational convenience, use  $\nu_N$  to mean  $|\nu_N|$ .

With respect to  $\nu_N$ , the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}(\mathcal{N}_g)$  is infinite covolume: the same also holds for the geodesic flow invariant volume on the full<sup>1</sup> unit cotangent bundle. Furthermore, the set of cotangent directions in which the geodesic flow recurs to the  $\text{MCG}(\mathcal{N}_g)$ -cocompact part of  $\mathcal{T}(\mathcal{N}_g)$  has  $\nu_N$ -measure 0: this is due to Norbury [Nor08] (see Gendulphe [Gen17] for more analogies with infinite covolume Fuchsian groups).

<sup>1</sup>Full referring to the entire unit cotangent bundle as opposed to the restricted unit cotangent bundle.

**2.2. Critical exponents and Patterson-Sullivan theory.** In this section, we outline techniques that are used to deal with infinite-covolume group actions on non-positively curved metric spaces, i.e. Patterson-Sullivan theory. For the sake of concreteness, we will state most results in this section for infinite-covolume geometrically finite Fuchsian groups, and specify a generalized theorem/conjecture for the setting of mapping class groups.

Let  $\Gamma$  be an infinite-covolume geometrically finite Fuchsian group. Geometric finiteness in this context means that the surface  $\mathbb{H}/\Gamma$  is composed of *finitely* many components outside of a large enough compact set, where each component is isometric to one of the following regions.

- (i) **Cusps:** A cusp is the quotient of a horoball (i.e.  $\{\text{Im}(z) > t_0\}$  with the upper half plane model) with respect to an isometry of the form  $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ .
- (ii) **Flares:** A flare is quotient of the region  $\{\text{Re}(z) > 0\}$  with respect to an isometry of the form  $\begin{pmatrix} q & 0 \\ 0 & \frac{1}{q} \end{pmatrix}$ .

Note that it is the flares of the hyperbolic surface that make its volume infinite: each of the cusps has finite hyperbolic volume.

The presence of flares also means that the limit set of  $\Gamma$ , i.e. the set  $\overline{\Gamma p} \cap \partial\mathbb{H}$  (for any  $p \in \mathbb{H}$ ) is a measure 0 subset of the boundary (with respect to the usual Lebesgue measure on  $S^1$ ), as well as forcing the Liouville measure, which is a geodesic flow invariant measure on the unit tangent bundle  $S^1\Gamma/\mathbb{H}$  to be infinite.

Since most results from ergodic theory need a finite flow-invariant measure, the Liouville measure does not work for these infinite-covolume groups. The fix to this problem is to construct a new (family of) measure(s)  $\{\mu_q\}$  on the boundary, which replaces the Lebesgue measure, with respect to which the limit set has full measure, and then use that measure to construct a finite geodesic flow invariant measure on the unit tangent bundle.

The family of measures on the boundary is called the Patterson-Sullivan measure, and the corresponding measure on the unit tangent bundle is called the Bowen-Margulis-Sullivan measure.

**2.2.1. Construction of Patterson-Sullivan measures.** We begin by picking a basepoint  $p \in \mathbb{H}$ , and a parameter  $h > 0$ , and consider the measure  $\mu_q^h$ , for  $q \in \mathbb{H}$ .

$$\mu_q^h := \frac{\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma q)) \delta_{\gamma q}}{\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma p))}$$

Here,  $\delta_{\gamma q}$  denotes the Dirac mass at  $\gamma q$ , and  $d(p, \gamma q)$  denotes the hyperbolic distance between  $p$  and  $\gamma q$ .

For large enough  $h$ , the denominator of the expression is a convergent sum, and the resulting measure has total mass that only depends on the choice of  $p$  and  $q$ .

Conversely, for small enough values of  $h > 0$ , the sum in the denominator diverges, and the measure  $\mu_q^h$  is not well defined. To see this, one can use the ping pong lemma to embed a copy of the free group  $F_2$  in  $\Gamma$ , and show that for this copy of  $F_2$ , there is a small enough  $h$  to make the sum diverge. We can now define the critical exponent  $h_\Gamma$  of the group  $\Gamma$ .



**Definition 2.1** (Critical exponent). The critical exponent  $h_\Gamma$  is the infimum of all the values of  $h$  for which the following infinite sum converges.

$$\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma p))$$

Note that for  $h = h_\Gamma$ , it is possible for the exponential sum to converge or diverge. If the sum converges at the critical exponent, the group  $\Gamma$  is said to be of *convergent type*, and if it diverges, the group  $\Gamma$  is of *divergent type*.

Since the measures  $\mu_q^h$  are well-defined for  $h > h_\Gamma$ , and their mass is uniformly bounded (where the bound only depends on  $q$ ), we have that for some sequence of  $h \searrow h_\Gamma$ , the sequence of measures  $\mu_q^h$  converges to some limiting measure  $\mu_q$ . This family of limiting measures  $\{\mu_q\}$  is called a Patterson-Sullivan measure. The Patterson-Sullivan measure  $\{\mu_q\}$  is not unique *a priori*, since picking different sequences  $h \searrow h_\Gamma$  might lead to different limiting measures.

In practice, the uniqueness of the Patterson-Sullivan measure follows from the ergodicity of the geodesic flow with respect to the Bowen-Margulis-Sullivan measure constructed from a given Patterson-Sullivan measure. We will skip the construction of the Bowen-Margulis-Sullivan measure  $\mu_{\text{BMS}}$ , since the specifics of the construction are not relevant for the remainder of the paper. We refer the reader to Quint [Qui06] for the construction of  $\mu_{\text{BMS}}$ .

**2.2.2. Some results in Patterson-Sullivan theory.** The question of finiteness and ergodicity of the Bowen-Margulis-Sullivan measure is equivalent to several other conditions, some of which are easier to check in some examples.

**Theorem 2.2** (Hopf-Tsuji-Sullivan dichotomy; Sullivan [Sul79]). *For a geometrically finite group  $\Gamma$ , the following conditions are equivalent.*

- (i) *The group is of divergent type.*
- (ii) *The Bowen-Margulis-Sullivan measure is finite.*
- (iii) *The geodesic flow is ergodic with respect to the Bowen-Margulis-Sullivan measure.*

The ergodicity of the geodesic flow with respect to  $\mu_{\text{BMS}}$  can be upgraded to mixing if the length spectrum of  $\mathbb{H}/\Gamma$  generates a dense subgroup of  $\mathbb{R}$ .

**Theorem 2.3** (Babillot [Bab02]). *If  $\mu_{\text{BMS}}$  is finite, and the lengths of the closed geodesics on  $\mathbb{H}/\Gamma$  generate a dense subgroup of  $\mathbb{R}$ , then the geodesic flow is mixing with respect to  $\mu_{\text{BMS}}$ .*

One can then combine Theorem 2.3 with the following result of Roblin to count lattice points where the logarithmic error goes to 0.

**Theorem 2.4** (Roblin [Rob03]). *Let  $B_p(R)$  denote the lattice point counting function.*

$$B_p(R) := \#(\gamma \in \Gamma \mid d(p, \gamma p) \leq R)$$

*Then there exists a constant  $C$ , which is the  $\mu_{\text{BMS}}$ -volume of the unit tangent bundle of  $\mathbb{H}/\Gamma$ , such that  $B_p(R)$  can be approximated in the following manner.*

$$\lim_{R \rightarrow \infty} \log \left( \frac{C \exp(h_\Gamma R)}{B_p(R)} \right) = 0$$

**2.2.3. Extending these results to subgroups of mapping class groups.** In [Yan18], Yang outlined a criterion for a *non-elementary group with contracting element* acting on metric space to be of divergent type: the action must be *statistically convex-cocompact*. In the context of subgroups of mapping class groups, a subgroup is non-elementary if it contains two non-commuting pseudo-Anosov elements.

To explain what a statistically convex-cocompact action is, we first need to describe what it means for a subset of a metric space to be statistically convex.

Let  $X$  be a metric space with a group  $G$  acting on it, and let  $Y$  be a subset of  $X$  which is invariant under the  $G$ -action, i.e. we have a  $G$ -action on  $Y$  as well, and let  $p$  be a point in  $Y$ . One can consider two kinds of counting functions for the  $G$ -action on  $Y$ .

$$N_p(R) := \{\gamma \in G \mid d(p, \gamma p) \leq R\}$$

The function  $N_p$  is the standard lattice point counting function. We also want to look at those lattice points that detect a failure of convexity of  $Y$ : we call these points *concave lattice points*.

**Definition 2.5** (*s-Concave lattice points*). A lattice point  $\gamma p$  is *s-concave* if some geodesic segment  $\kappa$  starting in a ball of radius  $s$  centered at  $p$  and ending in ball of radius  $s$  centered at  $\gamma p$  stays outside the set  $Y$ .

The path obtained by joining  $p$  to the starting point of  $\kappa$ , then following  $\kappa$ , and then joining the end point of  $\kappa$  to  $\gamma p$  is called the *concavity detecting path* for  $\gamma p$ .

For our applications, the precise value of  $s$  will not be very important: we fix it to be twice the diameter of the compact set  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$  (any value larger than the diameter of  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$  will work though).

Let  $M_p(R)$  denote the counting function for concave lattice points. Let  $h$  and  $h_c$  be the exponential growth rates for  $N_p(R)$  and  $M_p(R)$ .

$$h := \lim_{R \rightarrow \infty} \frac{\log(N_p(R))}{R}$$

$$h_c := \lim_{R \rightarrow \infty} \frac{\log(M_p(R))}{R}$$

**Definition 2.6** (Statistically convex subset). The subset  $Y$  is said to be statistically convex if  $h_c < h$ .

**Definition 2.7** (Statistically convex-cocompact action). The action of  $G$  on  $X$  is statistically convex-cocompact if there exists some  $G$ -invariant subset  $Y$  such that  $Y$  is statistically convex, and the action of  $G$  on  $Y$  is cocompact.

In [Yan18], Yang shows that when a non-elementary group acts statistically convex-cocompactly on a space, the group is of divergent type.

In an upcoming paper, Coulon, Gekhtman, Tapie, and Yang [CGTY] show that for subgroups of mapping class groups that act statistically convex-cocompactly, a version of the Hopf-Tsuji-Sullivan dichotomy (Theorem 2.2) holds. Combining this with Yang's result of the group being of divergence type, one can conclude that the Bowen-Margulis-Sullivan measure on the unit cotangent bundle has finite mass.

In Gekhtman and Ma [GM23], they show that assuming  $\mu_{\text{BMS}}$  has finite mass, the analogs of Theorems 2.3 and 2.4 are proven, establishing lattice point counting results for the actions of these mapping class subgroups.

In the remainder of this paper, we show that the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  (the subset of  $\mathcal{T}(\mathcal{N}_g)$  where the one-sided curves cannot be shorter than  $\varepsilon_t$ ) is statistically convex-cocompact.

This result, combined with the results of Yang [Yan18], Coulon, Gekhtman, Tapie, and Yang [CGTY], and Gekhtman and Ma [GM23], give us precise counting results for the action of  $\text{MCG}(\mathcal{N}_g)$ .

### List of notation.

- $\mathcal{S}_g$ : An orientable surface of genus  $g$ .
- $\mathcal{S}_{g,b,c}$ : A surface of genus  $g$  with  $b$  boundary components, and  $c$  crosscaps attached.
- $\mathcal{N}_g$ : A non-orientable surface of genus  $g$ : this is the same as  $\mathcal{S}_{\frac{g-1}{2},0,1}$  if  $g$  is odd, and  $\mathcal{S}_{\frac{g-2}{2},0,2}$  if  $g$  is even.
- $\mathcal{T}(S)$ : The Teichmüller space of the surface  $S$ .
- $\mathcal{T}_{\varepsilon_t}^-(S)$ : The one-sided systole superlevel set in  $\mathcal{T}(S)$ .
- $\nu_N$ : The Lagrangian volume form on  $\mathcal{T}(\mathcal{N}_g)$ .
- $B_\tau(x)$ : A ball of radius  $\tau$  (with respect to the Teichmüller metric) centered at  $x$ .
- $B_\tau^{\varepsilon_t}(x)$ : A ball of radius  $\tau$  (with respect to the induced path metric on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ ) centered at  $x$ .
- $A_\tau$ : The averaging operator on a ball of radius  $\tau$ .
- $h_{\text{LP}}(\mathcal{T}(S))$ : The exponential growth rate for the mapping class group orbit of a point  $x$  in  $\mathcal{T}(S)$ .
- $h_{\text{LP}}(H)$ : For a subgroup  $H$  of  $\text{MCG}(S)$ , this is the exponential growth rate of for the  $H$ -orbit of a point  $x$  in  $\mathcal{T}(S)$ .
- $\mathfrak{N}$ : An  $(\varepsilon_n, 2\varepsilon_n)$ -net.
- $h_{\text{NP}}(\text{core}(\mathcal{T}(S)))$ : This is the exponential growth rate for the net points in an  $(\varepsilon_n, 2\varepsilon_n)$ -net in the weak convex core of  $\mathcal{T}(S)$ . The value of  $\varepsilon_n$  is usually clear from the context.
- $\pitchfork$ :  $U \pitchfork V$  denotes that the surfaces  $U$  and  $V$  are transverse.
- $\pitchfork_W$ :  $U \pitchfork_W V$  denotes that  $U$  and  $V$  are transverse when restricted to any subsurface of  $W$  which intersects both  $U$  and  $V$  non-trivially.
- $U \leqslant V$ : The Behrstock partial order for transverse subsurfaces  $U$  and  $V$ .
- $\asymp$ : We say  $a \asymp b$  if  $a$  and  $b$  are equal up to a multiplicative error of  $k$  and an additive error of  $c$ , where  $k$  and  $c$  are some fixed constants.

## 3. THE WEAK CONVEX CORE OF $\mathcal{T}(\mathcal{N}_g)$

**3.1. Issues with geometric finiteness and statistical convex-cocompactness.** In order to show that the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}(\mathcal{N}_g)$  is geometrically finite (in the sense of Fuchsian groups), we need to exhibit a *convex core*, i.e. a convex subset of  $\mathcal{T}(\mathcal{N}_g)$  on which the action of  $\text{MCG}(\mathcal{N}_g)$  is finite covolume. Similarly, to show that the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}(\mathcal{N}_g)$  is statistically convex-cocompact, we need to exhibit a *statistical convex core*, which is a *statistically convex subset* (see Definition 2.6) of  $\mathcal{T}(\mathcal{N}_g)$  on which  $\text{MCG}(\mathcal{N}_g)$  acts cocompactly.

A candidate for the convex core was suggested by Gendulphie [Gen17], namely the *one-sided systole superlevel set*  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

**Definition 3.1** (One-sided systole superlevel set). The one-sided systole superlevel set is the subset of  $\mathcal{T}(\mathcal{N}_g)$  where no one-sided curve is shorter than  $\varepsilon_t$ . This set is denoted  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

The subset  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  has several properties that suggest it should be the convex core for the  $\text{MCG}(\mathcal{N}_g)$  action.

- The space  $\mathcal{T}(\mathcal{N}_g)$   $\text{MCG}(\mathcal{N}_g)$ -equivariantly deformation retracts onto the subset  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  (Proposition 19.2 of [Gen17]).
- The  $\text{MCG}(\mathcal{N}_g)$  action on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  has finite  $\nu_N$ -covolume, where  $\nu_N$  is the non-orientable analog of the Weil-Petersson volume form (Proposition 19.1 of [Gen17]).

However, the subset  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  fails to be convex, in a very strong sense, as we show in a prior paper.

**Theorem 3.2** (Theorem 5.2 of [Kha23]). *For all  $\varepsilon_t > 0$ , and all  $D > 0$ , there exists a Teichmüller geodesic segment whose endpoints lie in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  such that some point in the interior of the geodesic is more than distance  $D$  from  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .*

With this obstruction, we see that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  will not work as a convex core for a geometrically finite action. There are two directions one could go in with this obstruction in mind, which we phrase as open questions.

If we wish to show that  $\text{MCG}(\mathcal{N}_g)$  acts geometrically finitely, this is the question we need to answer.

**Question 3.3.** Does there exist some other subset of  $\mathcal{T}(\mathcal{N}_g)$  that is finite  $\nu_N$ -covolume, convex, and an  $\text{MCG}(\mathcal{N}_g)$ -equivariant deformation retract of  $\mathcal{T}(\mathcal{N}_g)$ ?

Alternatively, if we wish to show that  $\text{MCG}(\mathcal{N}_g)$  acts statistically convex-cocompactly on  $\mathcal{T}(\mathcal{N}_g)$ , where  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  acts as the statistical convex core, this is the question we need to answer.

**Question 3.4.** Is  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  statistically convex?

We suspect the answer to Question 3.4 is no, since our methods for proving statistical convexity rely on proving recurrence of random walks on  $\mathcal{T}(\mathcal{N}_g)$ , and random walks on all of  $\mathcal{T}(\mathcal{N}_g)$  have poor recurrence properties when they enter regions where one-sided curves are short. When the random walks enter these one-sided thin regions, they behave like symmetric random walks on  $\mathbb{Z}$ , which we know do not have very strong recurrence properties. We explain this in more detail in Section 4.3.2, when we set up the machinery of Foster-Lyapunov-Margulis functions.

**3.2. A weaker notion of convexity.** Rather than directly answering questions 3.3 or 3.4, we define an even weaker notion of convexity as an intermediate goal. In the next subsection, we will show that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  satisfies this weaker notion of convexity. In this section, we define the notion, and explain why this weaker notion of convexity is still sufficient for the purposes of Patterson-Sullivan theory.

**Definition 3.5** (Weak convexity). A subset  $S$  of a geodesic metric space  $X$  is said to be  $\varepsilon_d$ -weak convex (for  $\varepsilon_d > 0$ ) if there exists a constant  $t > 0$  such that for any pair of points  $x$  and  $y$  in  $S$ , any geodesic path  $\gamma$  joining  $x$  and  $y$  longer than  $t$  can be homotoped to a path  $\gamma'$  joining  $x$  and  $y$  such that  $\gamma'$  lies entirely within  $S$ , and the lengths of  $\gamma$  and  $\gamma'$  satisfy the following inequality.

$$\ell(\gamma') \leq (1 + \varepsilon_d)\ell(\gamma)$$

*Remark.* Strictly speaking, we should call a subset  $(\varepsilon_d, t)$ -weak convex, as the constant  $t$  is part of the data that makes a set weak convex. However, the constant  $t$  will not matter for us, so we suppress it in all mentions of weak convexity.

An  $\varepsilon_d$ -weak convex subset is a subset which, while not entirely undistorted, has bounded distortion with respect to the ambient metric space at large enough scale. Weak convexity also interacts well with results from Patterson-Sullivan theory. Suppose we have a discrete group  $G$  acting properly discontinuously on a metric space  $X$ , and let  $X_{\varepsilon_d}$  be an  $\varepsilon_d$ -weak convex subset of  $X$  upon which  $G$  also acts. If the critical exponent for the  $G$  action on  $X$  is  $\delta$ , and the corresponding exponent for  $X_{\varepsilon_d}$  is  $\delta_{\varepsilon_d}$ , we immediately get the following estimate for  $\delta$ .

$$\delta \leq \delta_{\varepsilon_d}(1 + \varepsilon_d)$$

For tangent directions along which the Teichmüller geodesic stays in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  for arbitrarily large times, we can consider the Teichmüller geodesic flow, and reparameterize the flow speed such that the following equation holds.

$$d_{\varepsilon_t}(v, g_\tau v) = \tau$$

If we can establish mixing results for this reparameterized geodesic flow, we can use the techniques of Roblin [Rob03] to count lattice points in  $X_{\varepsilon_d}$ , the counting results translate into estimates for lattice points in  $X$ .

$$\#(\text{Lattice points in } X \text{ within distance } R) \leq \#(\text{Lattice points in } X_{\varepsilon_d} \text{ within distance } R(1 + \varepsilon_d))$$

We can get even better estimates if rather than having a single  $\varepsilon_d$ -weak convex subset, we have an family of subsets, such that the  $\varepsilon_d$  goes to 0, and the union of the weak convex subsets is the entire space.

**Definition 3.6** (Exhaustion by weak convex subsets). A metric space  $X$  is said to be exhausted by weak convex subsets if there exists a nested family of subsets  $\{X_i\}$ , such that  $X_i$  is  $\varepsilon_i$ -weak convex, where  $\varepsilon_i$  goes to 0, and  $\bigcup_{i=1}^{\infty} X_i = X$ .

When there is an exhaustion by weak convex subsets, one can get arbitrarily good bounds for the critical exponent for the  $G$  action on  $X$ .

**3.3. Weak convexity for  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .** In this section, we will show that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is an  $\varepsilon_d$ -weak convex subset of  $\mathcal{T}(\mathcal{N}_g)$ , and that  $\mathcal{T}(\mathcal{N}_g)$  can be exhausted by the subsets  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  as  $\varepsilon_t$  goes to 0.

**Theorem 3.7.** *For any  $\varepsilon_d > 0$ , there exists a  $\varepsilon_t > 0$  such that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is a  $\varepsilon_d$ -weak convex subset of  $\mathcal{T}(\mathcal{N}_g)$ .*

The key ingredient in the proof of Theorem 3.7 is a version of Minsky's product region theorem [Min96, Theorem 6.1] for non-orientable surfaces, which we prove in Section A.1.

Let  $\gamma = \{\gamma_1, \dots, \gamma_j, \dots, \gamma_k\}$  be a multicurve on a non-orientable surface  $\mathcal{N}_g$ , where for  $i \leq j$ ,  $\gamma_i$  is a two-sided curve, and for  $i > j$ ,  $\gamma_i$  is a one-sided curve. Let  $X_\gamma$  denote the metric space obtained as the sup-product  $\mathcal{T}(\mathcal{N}_g \setminus \gamma) \times \mathbb{H}_1 \times \dots \times \mathbb{H}_j \times (\mathbb{R}_{>0})_{j+1} \times \dots \times (\mathbb{R}_{>0})_k$ , where the  $\mathbb{H}_i$  are copies of the upper half plane with the hyperbolic metric, and  $\mathbb{R}_{>0}$  is the set of positive real numbers, where the distance between  $x$  and  $y$  is  $\left| \log \left( \frac{x}{y} \right) \right|$ . For any pants decomposition that contains  $\gamma$ , we consider the Fenchel-Nielsen coordinate systems

associated to the pants decomposition. We have a map  $\Pi$  from  $\mathcal{T}(\mathcal{N}_g)$  to  $X_\gamma$ , which is called the *product region projection map*.

**Definition 3.8** (Product region projection map). The product region projection map  $\Pi : \mathcal{T}(\mathcal{N}_g) \rightarrow X_\gamma$  is defined in the following manner.

- The  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ -coordinate is obtained by setting the lengths to 0 of all the curves in  $\gamma$  to get a punctured hyperbolic surface.
- The  $\mathbb{H}_i$ -coordinate is  $(t, \frac{1}{\ell})$ , where  $t$  is the *twist* (i.e. the twist coordinate in the Fenchel-Nielsen coordinate system) of the two-sided curve  $\gamma_i$ , and  $\ell$  is the hyperbolic length.
- The  $(\mathbb{R}_{>0})_i$  coordinate is  $\frac{1}{\ell}$ , where  $\ell$  is the hyperbolic length of the one-sided curve  $\gamma_i$ .

We define a metric on the product  $X_\gamma$  as the supremum of the metrics on each of the components, where the metric on  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$  is the Teichmüller metric, the metric on the  $\mathbb{H}_i$  components is the hyperbolic metric, and the metric on the  $(\mathbb{R}_{>0})_i$  is given by  $d(x, y) = \left| \log \left( \frac{x}{y} \right) \right|$ , i.e. the restriction of the hyperbolic metric in  $\mathbb{H}$  to a vertical line.

We consider the restriction of  $\Pi$  to the thin region of Teichmüller space, denoted  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ , which is the region where all curves in  $\gamma$  have hyperbolic length at most  $\varepsilon_t$ .

**Theorem A.3** (Product region theorem for non-orientable surfaces). *For any  $c > 0$ , there exists  $\varepsilon'_t > 0$ , such that for any  $\varepsilon_t < \varepsilon'_t$ , the restriction of  $\Pi$  to  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$  is an isometry with additive error at most  $c$ , i.e. the following holds for any  $x$  and  $y$  in  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ .*

$$|d(x, y) - d_{X_\gamma}(\Pi(x), \Pi(y))| \leq c$$

We can now prove Theorem 3.7.

*Proof of Theorem 3.7.* We begin by picking a small constant  $\varepsilon'_t > 0$  and  $\delta > 0$ . We will fix the values of these constants at the end of the proof. Let  $[x, y]$  be a geodesic segment that starts and ends in  $\mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$ . Let  $\{p_i\}$  be points on  $[x, y]$ , such that  $x = p_0$ ,  $d(p_i, p_{i+1}) = \delta$ , and  $d(p_n, y) \leq \delta$ , where  $p_n$  is the last of the  $p_i$ 's.

The first step of our proof is modifying the path  $[x, y]$  and estimating the length of the modified path. We do so by constructing new points  $p'_i$ , where  $p'_i$  is obtained from  $p_i$  by increasing the length of any one-sided curve that is shorter than  $\varepsilon'_t$  to  $\varepsilon'_t$ . This ensures that the endpoints of the segments  $[p'_i, p'_{i+1}]$  are in  $\mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$ . Estimating  $d(p'_i, p'_{i+1})$  splits up into two cases.

- When  $p_i = p'_i$  and  $p_{i+1} = p'_{i+1}$ : In this case  $d(p'_i, p'_{i+1}) = \delta$ , by construction.
- When at least one of  $p_i$  and  $p_{i+1}$  are not equal to  $p'_i$  and  $p'_{i+1}$ : In this case, we can assume without loss of generality that both  $p_i \neq p'_i$  and  $p_{i+1} \neq p'_{i+1}$ . If that is not the case, and say  $p_i \neq p'_i$  and  $p_{i+1} = p'_{i+1}$ , we replace  $p_{i+1}$  with the last point  $y$  on  $[p_i, p_{i+1}]$  that is outside  $\mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$ . The interval  $[y, p_{i+1}]$  can be treated as in case (i), and we focus on  $[p_i, y]$ .

We have that the interior of  $[p_i, p_{i+1}]$  and  $[p'_i, p'_{i+1}]$  both lie in the region where some one-sided curve  $\gamma$  is shorter than  $\varepsilon'_t$ . We invoke Theorem A.3 to estimate distances in this region: we have a constant  $c(\varepsilon'_t)$  that depends on  $\varepsilon'_t$  such that following holds.

$$\left| d(p_i, p_{i+1}) - \sup \left( d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p_i), \Pi(p_{i+1})), \left| \log \left( \frac{\ell_{p_i}(\gamma)}{\ell_{p_{i+1}}(\gamma)} \right) \right| \right) \right| \leq c(\varepsilon'_t) \quad (3)$$



Observe that when we replace  $p_i$  by  $p'_i$  and  $p_{i+1}$  by  $p'_{i+1}$ , the first argument sup stays the same, and the second argument becomes 0.

$$|d(p'_i, p'_{i+1}) - \sup(d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p'_i), \Pi(p'_{i+1})), 0)| \leq c(\varepsilon'_t) \quad (4)$$

This leads to the following estimate for  $d(p'_i, p'_{i+1})$ .

$$d(p'_i, p'_{i+1}) \leq \delta + 2c(\varepsilon'_t) \quad (5)$$

We construct a new path  $\lambda$  by joining  $p'_i$ 's, and  $p_n$  to  $y$ . If we let  $l$  denote the length of  $[x, y]$ , we get the following estimate for  $\ell(\lambda)$  using (5).

$$\ell(\lambda) \leq l \left( 1 + \frac{2c(\varepsilon'_t)}{\delta} \right)$$

We now pick a value of  $\delta$  small enough such that along each of the segments  $[p_i, p_{i+1}]$ , there is at least one one-sided curve that stays short throughout, and then we pick  $\varepsilon'_t$  small enough so that  $c(\varepsilon'_t)$  is small enough to make  $\frac{2c(\varepsilon'_t)}{\delta} < \varepsilon_d$ .

We now need to show that this new path stays within  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  for some  $\varepsilon_t < \varepsilon'_t$ . We already have that  $x, y$  and all the  $p'_i$  are in  $\mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$  and thus in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ . For the interior of the geodesic segments  $[p'_i, p'_{i+1}]$ , since the endpoints are in  $\mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$ , and the length of the segments is no more than  $\delta(1 + \varepsilon_d)$ , we have that there exists some  $\varepsilon_t$  such that  $[p'_i, p'_{i+1}]$  lies in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .

Finally, we have to deal with geodesic segments  $[w, z]$  which start or end in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g) \setminus \mathcal{T}_{\varepsilon'_t}^-(\mathcal{N}_g)$ . We do so by increasing the lengths of short one-sided curves on  $w$  and  $z$  to  $\varepsilon'_t$  if there are any curves shorter than  $\varepsilon'_t$ . Let the modified points be  $w'$  and  $z'$ : we first construct a path joining  $w'$  and  $z'$  that stays within  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  using our construction, and then prepend that path with a path joining  $w$  with  $w'$  and append a path joining  $z'$  to  $z$ . This new path joining  $w$  to  $z$  stays entirely within  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , but we now incur a fixed additive error along with our multiplicative error as well. However, if the path is long enough, the additive error can be absorbed in the multiplicative error, with a slightly worse constant. We do that, and the threshold for the path being long enough is the constant  $t$  that appears in our definition of weak convexity. This proves the result.  $\square$

*Remark.* We emphasize that the key step in the above proof is going from (3) to (4), where the  $\log \left( \frac{\ell_{p_i}(\gamma)}{\ell_{p'_i}(\gamma)} \right)$  term becomes 0. This is only possible because there cannot be any twisting around a one-sided curve  $\gamma$ , so the projection map that sends  $p_i$  to  $p'_i$  and  $p_{i+1}$  to  $p'_{i+1}$  is distance reducing. If one tried to use the same proof strategy to show that the thick part of  $\mathcal{T}(\mathcal{S})$ , for any orientable or non-orientable surface  $\mathcal{S}$  is weak convex in  $\mathcal{T}(\mathcal{S})$ , the step we described would be the point of failure. In particular, if there's a twist along  $\gamma$ , going from (3) to (4) will not be distance reducing, and will exponentially increase the distance, leading the estimate to fail.

Now that we have established that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is  $\varepsilon_d$ -weak convex, we can justifiably call it the weak convex core of  $\mathcal{T}(\mathcal{N}_g)$ . For the remainder of this paper, we fix  $\varepsilon_d < \left( \frac{1}{6g-12} \right)^2$ , and a value of  $\varepsilon_t$  such that  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  is  $\varepsilon_d$ -weak convex.

**Definition 3.9** (Weak convex core of  $\mathcal{T}(\mathcal{N}_g)$ ). We call  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  the weak convex core of  $\mathcal{T}(\mathcal{N}_g)$ , and denote it  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ .

#### 4. GEODESICS IN THE THIN PART OF $\text{core}(\mathcal{T}(\mathcal{N}_g))$

Inspired by Theorem 3.7, we will focus our attention on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  instead of the entirety of  $\mathcal{T}(\mathcal{N}_g)$ . In this section, we will begin a proof of the fact that the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  is statistically convex-cocompact (we abbreviate that to SCC for the remainder of the paper).

To show that the  $\text{MCG}(\mathcal{N}_g)$  action is SCC, we need to exhibit a subset of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  which has the following two properties.

- (i) The action of  $\text{MCG}(\mathcal{N}_g)$  on the subset is cocompact.
- (ii) The subset is statistically convex.

We claim that the subset  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  satisfies these properties.

$$\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g) := \{z \in \mathcal{T}(\mathcal{N}_g) \mid \text{No curve on } z \text{ is shorter than } \varepsilon_t\}$$

Although we have defined  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  as a subset of  $\mathcal{T}(\mathcal{N}_g)$ , it is also a subset of  $\text{core}(\mathcal{T}(\mathcal{N}_g)) = \mathcal{T}_{\varepsilon_t}^{-}(\mathcal{N}_g)$ : this follows from its very definition, which is a more restrictive version of the definition of  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ . The action of  $\text{MCG}(\mathcal{N}_g)$  on  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  is also cocompact, because the quotient is the thick part of the moduli space, which is known to be compact.

Showing that  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  is a statistically convex subset of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  requires more work. We begin by rephrasing what it means for  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  to be statistically convex in a form that's more convenient for our methods.

Consider the metric  $d_{\varepsilon_t}$  on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ , defined by the following formula.

$$d_{\varepsilon_t}(x, y) := \inf(\ell(\lambda) \mid \lambda \text{ is a path in } \text{core}(\mathcal{T}(\mathcal{N}_g)) \text{ joining } x \text{ and } y)$$

This metric is not the same as the usual Teichmüller metric  $d$ , but by Theorem 3.7, we can make the ratio of these two metrics arbitrarily close to 1 by picking  $\varepsilon_t$  small enough.

We now define lattice point entropy, and entropy for concave lattice points.

**Definition 4.1** (Lattice point entropy for  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ ). Let  $p$  be a point in  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ , and let  $N_p(R, \varepsilon_t)$  be the lattice point counting function.

$$N_p(R, \varepsilon_t) := \#(\gamma \in \text{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t}(p, \gamma p) \leq R)$$

The lattice point entropy  $h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$  is the following quantity.

$$h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log N_p(R, \varepsilon_t)}{R}$$

*Remark.* The lattice point entropy is a well-defined quantity since we have that  $N_p(R, \varepsilon_t)$  is a sub-multiplicative function, and therefore  $\log N_p(R, \varepsilon_t)$  is sub-additive, and the limit is well defined by Fekete's lemma.

We define a variable  $s = 2\text{diam}(\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$ , and recall the definition of  $s$ -concave lattice points.

**Definition 4.2** (Concave lattice points). A lattice point  $\gamma p$  is  $s$ -concave if some geodesic segment  $\kappa$  starting in a ball of radius  $s$  centered at  $p$  and ending in ball of radius  $t$  centered at  $\gamma p$  stays outside the set  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ .

The path obtained by joining  $p$  to the starting point of  $\kappa$ , then following  $\kappa$ , and then joining the end point of  $\kappa$  to  $\gamma p$  is called the *concavity detecting path* for  $\gamma p$ .

**Definition 4.3** (Entropy for concave lattice points). Let  $M_p(R, \varepsilon_t)$  be the counting function for concave lattice points.

$$M_p(R, \varepsilon_t) := \# (\gamma \in \text{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t}(p, \gamma p) \leq R \text{ and } \gamma p \text{ is concave})$$

The entropy for concave lattice points  $h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$  is the following quantity.

$$h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log M_p(R, \varepsilon_t)}{R}$$

The statistical convexity of  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  is equivalent to the following statement, which states that the entropy for concave lattice points is strictly lower than entropy for all lattice points.

**Theorem 4.4** (Statistical convexity). *For  $\varepsilon_t > 0$  small enough, the following inequality holds.*

$$h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) < h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$$

We prove Theorem 4.4 by constructing a random walk on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  and proving a similar entropy gap between all random walk trajectories and the random walk trajectories that spend their time outside  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ .

**4.1. Construction of random walk.** Let  $\mathbf{p}$  be the projection map from  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  to  $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$ , and  $\varepsilon_n > 0$  be a fixed constant.

**Definition 4.5** ( $(\varepsilon_n, 2\varepsilon_n)$ -net in  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ ). Let  $\mathfrak{M}$  be a subset of  $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$  satisfying the following two conditions.

- (i) If  $z_1$  and  $z_2$  lie in  $\mathfrak{M}$ , then  $d_{\varepsilon_t}(z_1, z_2) \geq \varepsilon_n$ .
- (ii) For any  $z_1$  in  $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$ , there exists  $z_2 \in \mathfrak{M}$  such that  $d_{\varepsilon_t}(z_1, z_2) \leq 2\varepsilon_n$ .

An  $(\varepsilon_n, 2\varepsilon_n)$ -net  $\mathfrak{N}$  in  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  is  $\mathbf{p}^{-1}(\mathfrak{M})$  for any subset  $\mathfrak{M}$  satisfying the above conditions.

The random walk is defined in terms of a net  $\mathfrak{N}$  and a parameter  $\tau > 0$ : we pick a starting point  $r_0$  (which we call step 0) for the random walk from one of the net points, and  $r_n$  is picked uniformly at random amongst all the net points that are within distance  $\tau$  of  $r_{n-1}$ .

We will be interested in counting the number of  $n$ -step trajectories of the random walk as a function of  $n$  and  $\tau$ . The count will also involve the exponential growth rate of the number of net points in a ball of radius  $R$ , which we call the *net point entropy*  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ .

**Definition 4.6** (Net point entropy). Let  $K_p(R, \varepsilon_t)$  be the counting function for net points, where  $p \in \text{core}(\mathcal{T}(\mathcal{N}_g))$ .

$$K_p(R, \varepsilon_t) := \# (y \in \mathfrak{N} \mid d_{\varepsilon_t}(p, y) \leq R)$$

The net point entropy  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$  is the following function defined in terms of  $K_p$ .

$$h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

Note that  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$  does not depend on the choice of the actual net, nor does it depend on the parameter  $\varepsilon_n$ . Two different nets with different choices of  $\varepsilon_n$  will have counting functions that differ by at most a constant multiplicative term, which will not change the value of  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ . This follows from Proposition A.4: let  $n_1$  and  $n_2$  be the number of net points of two different nets contained in a ball  $B_R(z)$  of radius  $R$

centered at a point  $z$ . We get a lower bound for  $\nu_N(B_{R+\varepsilon_n}(z))$  by adding up the  $\nu_N$  volumes of balls of radius  $\varepsilon_n$  around each point in the first net.

$$n_1 c_1(\varepsilon_n) \leq \nu_N(B_{R+\varepsilon_n}(z)) \quad (6)$$

We get an upper bound for  $\nu_N(B_{R+\varepsilon_n}(z))$  by adding up the  $\nu_N$  volumes of radius  $2\varepsilon_n$  around each point in the second net.

$$\nu_N(B_{R+\varepsilon_n}(z)) \leq n_2 c_2(2\varepsilon_n) \quad (7)$$

Here,  $c_1$  and  $c_2$  are functions that appear in the statement of Proposition A.4. Combining (6) and (7), as well as using the fact that  $c_1(\varepsilon_n)$  is positive gives us the claim.

The above argument also shows that the number of net points in a ball of radius  $R$  is equal (up to multiplicative errors) to the Norbury measure  $\nu_N$ . In Section 4.2, we will focus our attention on averaging functions with respect to this measure, instead of the uniform measure obtained via the net points.

We will replicate the proof of Theorem 1.2 of Eskin and Mirzakhani [EM11], where they construct a random walk on a net, and use that to count concave trajectories. The key difficulty that comes up in our proof and which does not come up in their proof is the fact that they get an estimate for the cardinality of concave trajectories (and therefore concave lattice points) in terms of  $h_{\text{NP}}(\mathcal{T}(\mathcal{S}_g))$ , which they know is the same as  $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_g))$  (i.e.  $6g - 6$ ) by Theorem 1.2 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12].

Since we are working with non-orientable surfaces, we cannot invoke Theorem 1.2 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12], and instead need to relate  $h_{\text{LP}}$  and  $h_{\text{NP}}$  more directly: this is what we do in Sections 5 and 6.

**4.2. Construction of the Foster-Lyapunov-Margulis function.** One of the ways to show that a random walk on a non-compact space avoids the complement of a compact region with high probability is to construct a proper function  $f_{\text{FLM}}$  on the space which satisfies a certain inequality when averaged over one step of the random walk. See Eskin and Mozes [EM22] for an exposition on the construction of these functions as well as some applications to dynamics and random walks.

**Definition 4.7** (Averaging operator). Let  $\tau > 0$  be the parameter associated to the random walk, and  $f$  be any real valued function  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ . Then the action of the averaging operator  $A_\tau$  on  $f$  is given by the following formula.

$$(A_\tau f)(x) := \frac{1}{\nu_N(B_\tau^{\varepsilon_t}(x))} \left( \int_{B_\tau^{\varepsilon_t}(x)} f(z) d\nu_N(z) \right)$$

Here,  $B_\tau^{\varepsilon_t}(x)$  is a ball of radius  $\tau$  around  $x$  with respect to the metric  $d_{\varepsilon_t}$ .

A Foster-Lyapunov-Margulis function is a function that has strong decay properties when the operator  $A_\tau$  is applied to it.

**Definition 4.8** (Foster-Lyapunov-Margulis function). A proper function  $f$  on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  quotiented by the  $\text{MCG}(\mathcal{N}_g)$ -action is called a Foster-Lyapunov-Margulis function if, if there exists a polynomial  $p$ , and a compact subset  $W_0$  of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ , and functions  $b(x)$  and  $c(x)$ , such that  $A_\tau f$  satisfies the following inequality.

$$(A_\tau f)(x) \leq c(x)f(x) + b(x)$$

Furthermore,  $b(x)$  is a bounded function that is supported within a compact set  $W_0$ , and  $c(x)$  satisfies the following inequality for all  $x$  outside of  $W_0$ .

$$c(x) \leq p(\tau) \cdot \exp(-\tau)$$

Consider the function  $f_{\text{FLM}}$ , defined on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  in terms of the length of the shortest two-sided curve on the surface.

$$f_{\text{FLM}}(x) := \sqrt{\frac{1}{\inf_{\gamma \text{ two-sided}} \ell_{\gamma}(x)}}$$

This function is a proper function on  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$ , since the sub-level sets of this function are regions in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  where the hyperbolic lengths of all curves are bounded from below.

**Proposition 4.9.** *The function  $f_{\text{FLM}}$  is a Foster-Lyapunov-Margulis function on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  with respect to  $A_{\tau}$ , for large values of  $\tau$ .*

*Proof.* Let  $W_0$  be the region of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  where all two-sided curves are longer than  $\varepsilon_t$ . We divide  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  into three regions, and prove the estimate for  $(A_{\tau} f_{\text{FLM}})(x)$  for  $x$  in these three regions. The regions  $R_1$ ,  $R_2$  and  $R_3$  are defined in the following manner.

-  $R_1$ : The subset  $R_1$  is defined in the following manner.

$$R_1 := \{x \mid \text{Shortest curve for any } z \in B_{\tau}^{\varepsilon_t}(x) \text{ is } \gamma \text{ and } \ell_z(\gamma) < \varepsilon_t \text{ for } z \in B_{\tau}^{\varepsilon_t}(x)\}$$

This is the set of points  $x$  such that there exists a unique curve  $\gamma$  which is the shortest curve at all points in  $B_{\tau}^{\varepsilon_t}(x)$ , and the length of  $\gamma$  is less than  $\varepsilon_t$  for all points in  $B_{\tau}^{\varepsilon_t}(x)$ .

-  $R_2$ : This subset is  $R_0 \setminus R_1$ , where  $R_0$  is defined in the following manner.

$$R_0 := \{x \mid \text{Shortest curve for any } z \in B_{\tau}^{\varepsilon_t}(x) \text{ shorter than } \varepsilon_t\}$$

$R_2$  is the region where multiple curves can be simultaneously short.

-  $R_3$ : This is all of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  with  $R_1$  and  $R_2$  removed.

(i) Proof for  $x \in R_1$ : In this case, the entire ball  $B_{\tau}^{\varepsilon_t}$  is contained in the product region where  $\gamma$  stays short. By Theorem A.3, there exists a constant  $c(\varepsilon_t)$  such that the ball  $B_{\tau}^{\varepsilon_t}(x)$  contains, and is contained inside a product of balls in  $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$  and  $\mathbb{H}$  (which corresponds to length and twist around  $\gamma$ ).

$$\begin{aligned} B_{\tau-c(\varepsilon_t)}(x, \text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))) \times B_{\tau-c(\varepsilon_t)}(\mathbb{H}) &\subset B_{\tau}^{\varepsilon_t}(x) \\ &\subset B_{\tau+c(\varepsilon_t)}(x, \text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))) \times B_{\tau+c(\varepsilon_t)}(\mathbb{H}) \end{aligned}$$

Instead of computing the average of  $f$  over  $B_{\tau}^{\varepsilon_t}(x)$ , we can compute it over the product of the balls as described above. To do so, we need to verify that the measure on the product of the two balls is the product of the measures on the individual balls: we do this in the proof of Proposition A.4: specifically (41). Since the volumes of these balls grow exponentially with respect to radius, computing the average over the product of balls will give us an average that differs from the true average by a bounded multiplicative constant. This constant will be one of the terms that contribute to  $c'$  in Definition 4.8. Furthermore, note that the function  $f_{\text{FLM}}$  is constant along the  $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$  component, since  $\gamma$  is the shortest curve in the product of balls. It thus suffices to compute the average of  $f_{\text{FLM}}$  on a ball in  $\mathbb{H}$ .

Parameterizing  $\mathbb{H}$  as the upper half plane with coordinates  $z = (z_{\text{real}}, z_{\text{im}})$ , the function  $f_{\text{FLM}}(z)$  is the square root of the second coordinate, i.e.  $f_{\text{FLM}}(z) = \sqrt{z_{\text{im}}}$ . The

average of this function over a sphere is well-understood (see [EM22, Lemma 4.2]). We recall the estimate here for the reader's convenience: denoting the averaging operator over a sphere of radius  $\tau$  in  $\mathbb{H}$  by  $B_\tau$ , the estimate is as follows.

$$(B_\tau f_{\text{FLM}})(z) \leq c'' \exp(-\tau) f_{\text{FLM}}(z)$$

We use the spherical average to compute the average over a ball by taking a weighted average of the spherical averages. Doing so gives the following estimate for  $(A_\tau f_{\text{FLM}})(z)$  (where  $c'$  is some fixed constant).

$$(A_\tau f_{\text{FLM}})(z) \leq c' \tau \exp(-\tau) f_{\text{FLM}}(z)$$

Since we have already established that the value  $f_{\text{FLM}}(x)$  only depends on what happens in the  $\mathbb{H}$ -coordinate, namely  $z$ , we get a corresponding inequality for  $x$ , which proves the result in this case.

$$(A_\tau f_{\text{FLM}})(x) \leq c' \tau \exp(-\tau) f_{\text{FLM}}(x)$$

- (ii) Proof for  $x \in R_2$ : In this case, let  $\{\gamma_1, \dots, \gamma_k\}$  be the two-sided curves that become shorter at some point in  $B_\tau^{\varepsilon_t}(x)$ . We have that the ball lies in the product region where all the curves  $\{\gamma_1, \dots, \gamma_k\}$  are short simultaneously. We have that there exists some constant  $c_g$ , depending only on  $g$ , such that  $k \leq c_g$ , since we cannot have too many curves being short simultaneously.

Similar to the previous case, changing only the  $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \bigcup_{i=1}^k \gamma_i))$  coordinate will not change the value of the function  $f_{\text{FLM}}$ , so it suffices to focus our attention on the coordinates  $\prod_{i=1}^k \mathbb{H}_i$ , where each  $\mathbb{H}_i$  corresponds to the length and twist around  $\gamma_i$ .

Let  $z_{i,\text{im}}$  be the imaginary part of the  $i^{\text{th}}$  copy of  $\mathbb{H}$  in  $\prod_{i=1}^k \mathbb{H}_i$ . The function  $f_{\text{FLM}}$  on  $\prod_{i=1}^k \mathbb{H}_i$  is given by the following formula.

$$f_{\text{FLM}}(x) = \max_i \sqrt{z_{i,\text{im}}}$$

Since averaging this function over a product of balls is somewhat tedious, we relate it to a different function  $f'_{\text{FLM}}$  that is easier to average.

$$f'_{\text{FLM}}(x) := \sum_i \sqrt{z_{i,\text{im}}}$$

These two functions are equal, up to a constant multiplicative error.

$$\frac{f'_{\text{FLM}}(x)}{c_g} \leq f_{\text{FLM}}(x) \leq f'_{\text{FLM}}(x)$$

This means we can prove the averaging estimate for  $f'_{\text{FLM}}$ , and the same estimate will hold for  $f_{\text{FLM}}$ , with a slightly worse multiplicative constant.

Furthermore, since  $z_{i,\text{im}}$  is constant along balls in  $\mathbb{H}_j$  for  $j \neq i$ , it suffices to average just each term of the sum in the corresponding  $\mathbb{H}_i$ . We do so, using the same estimate from the proof in the  $R_1$  case.

$$(A_\tau f'_{\text{FLM}})(x) \leq c' \tau \exp(-\tau) f'_{\text{FLM}}(x)$$

Replacing  $f'_{\text{FLM}}$  with  $f_{\text{FLM}}$  gives us the inequality we want, and proves the result in this case.

$$(A_\tau f_{\text{FLM}})(x) \leq (c_g \cdot c') \tau \exp(-\tau) f_{\text{FLM}}(x)$$



- (iii) Proof for  $x \in R_3$ : Note that the region  $R_3$  is compact, which means the function  $f_{\text{FLM}}$  is bounded in this region, and consequently, there exists a uniform upper bound for  $A_\tau f_{\text{FLM}}$  as well. Let us denote the uniform upper bounding function by  $b(x)$ : we can modify this function to be compactly supported by multiplying it with a bump function that is 1 on a  $\tau$ -neighbourhood of  $R_3$ , and decays to 0 outside. By construction, we have for  $x \in R_3$ ,  $(A_\tau f_{\text{FLM}})(x) \leq b(x)$ . This proves the result for  $x \in R_3$ .

Putting together the estimate from the three cases, we get the standard form of the inequality (which holds for any  $x \in \text{core}(\mathcal{T}(\mathcal{N}_g))$ ).

$$(A_\tau f_{\text{FLM}})(x) \leq c(x)f_{\text{FLM}}(x) + b(x)$$

Here,  $c(x) := (c_g \cdot c')\tau \exp(-\tau)$ , and  $b(x)$  is the function from the proof in case  $R_3$ .  $\square$

#### 4.3. Recurrence for random walks and geodesic segments.

4.3.1. *Recurrence for random walks.* In this section, we will count the number of random walk trajectories that are  $\mathfrak{s}$ -concave, where  $\mathfrak{s} = \lceil \frac{\text{diam}(\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g))}{\tau} \rceil + 1$ .

**Definition 4.10** (Concave trajectories). A trajectory  $(r_0, r_1, \dots, r_{n-1})$  is said to be  $\mathfrak{s}$ -concave if all the points in the trajectory except the first  $\mathfrak{s}$  points and the last  $\mathfrak{s}$  points are at least  $\tau$ -distance away from  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ .

We call these middle points the *concave trajectory points*.

From Proposition 4.9, we have that for any of the concave trajectory points  $r_i$ , the following decay estimate for  $A_\tau f_{\text{FLM}}(r_i)$  holds.

$$(A_\tau f_{\text{FLM}})(r_i) \leq c'\tau \exp(-\tau)f_{\text{FLM}}(r_i)$$

If  $r_i$  is not a concave trajectory point, we only have a weaker estimate in terms of a uniformly bounded function  $b(x)$ .

$$(A_\tau f_{\text{FLM}})(r_i) \leq b(x)$$

Fix an  $r_0$  in  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ , and let  $\mathcal{P}(n, \tau)$  denote the collection of  $n$ -step concave random walk trajectories which start at  $r_0$ . We will use the term  $r$  to denote trajectories in  $\mathcal{P}(n, \tau)$ , and  $r_i$  to denote the  $i^{\text{th}}$  step of the trajectory  $r$ .

**Proposition 4.11.** *For any  $\varepsilon_{\text{err}} > 0$ , there exists a  $\tau > 0$  large enough, and a constant  $C \gg 0$ , such that the following bound on  $|\mathcal{P}(n, \tau)|$  holds.*

$$|\mathcal{P}(n, \tau)| \leq C \exp((h_{\text{NP}} - 1 + \varepsilon_{\text{err}})n\tau)$$

Here,  $h_{\text{NP}} = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ .

*Proof.* Since the  $f_{\text{FLM}}(x)$  has a positive lower bound  $C_l$  as  $x$  varies over  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  (which comes from the Bers constant associated to  $\mathcal{N}_g$ ), we can estimate  $|\mathcal{P}(n, \tau)|$  by summing up  $f_{\text{FLM}}(r_{n-1})$  over all the trajectories in  $\mathcal{P}(n, \tau)$ .

$$|\mathcal{P}(n, \tau)| \leq \frac{1}{C_l} \sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-1})$$

It therefore will suffice to estimate  $\sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-1})$ : we do so by conditioning on the previous step of the random walk over and over again until we get to the first step  $r_0$ .

We have the following recursive inequality for  $\sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-i})$ .

$$\sum_{r \in \mathcal{P}(n-i+1, \tau)} f_{\text{FLM}}(r_{n-1}) = \sum_{r \in \mathcal{P}(n-i, \tau)} \left( \sum_{\substack{y \in \mathfrak{N} \\ d_{\varepsilon_t}(y, r_{n-i-1}) \leq \tau}} f_{\text{FLM}}(y) \right) \quad (8)$$

$$\leq \sum_{r \in \mathcal{P}(n-i, \tau)} C \int_{B_{\tau}^{\varepsilon_t}(r_{n-i-1})} f_{\text{FLM}}(y) d\nu_N(y) \quad (9)$$

$$= \sum_{r \in \mathcal{P}(n-i, \tau)} C \nu_N(B_{\tau}^{\varepsilon_t}(r_{n-i-1})) (A_{\tau} f_{\text{FLM}})(r_{n-i-1}) \quad (10)$$

Here, we go from (8) to (9) by integrating the indicator function supported in a ball of radius  $\frac{\varepsilon_n}{2}$  around each net point, and using Proposition A.4 to uniformly bound the integral of the indicator independent of the basepoint.

If  $n-i-1$  is one of the first  $\mathfrak{s}$  or last  $\mathfrak{s}$  indices, we have the following inequality.

$$(A_{\tau} f_{\text{FLM}})(r_{n-i-1}) \leq b(r_{n-i-1}) f_{\text{FLM}}(r_{n-i-1}) \quad (11)$$

Otherwise  $r_{n-i-1}$  is a concave trajectory point and we have strong bounds on  $(A_{\tau} f_{\text{FLM}})(r_{n-i-1})$ .

$$(A_{\tau} f_{\text{FLM}})(r_{n-i-1}) \leq c' \tau \exp(-\tau) f_{\text{FLM}}(r_{n-i-1}) \quad (12)$$

Combining (8) and (11) for the case of non concave trajectory points, we get an estimate we need to repeat  $2\mathfrak{s}$  times.

$$\sum_{r \in \mathcal{P}(n-i+1, \tau)} f_{\text{FLM}}(r_{n-i}) \leq B \exp((h_{\text{NP}} + \varepsilon'_{\text{err}})\tau) \left( \sum_{r \in \mathcal{P}(n-i, \tau)} f_{\text{FLM}}(r_{n-i-1}) \right) \quad (13)$$

Here,  $B$  is the maximum value the function  $b$  takes over its compact support.

Combining (8) and (12), we get the estimate we need to repeat  $n - 2\mathfrak{s}$  times.

$$\sum_{r \in \mathcal{P}(n-i+1, \tau)} f_{\text{FLM}}(r_{n-i}) \leq C' \tau \exp(-\tau) \exp((h_{\text{NP}} + \varepsilon'_{\text{err}})\tau) \left( \sum_{r \in \mathcal{P}(n-i, \tau)} f_{\text{FLM}}(r_{n-i-1}) \right) \quad (14)$$

Here, we upper bound the  $\nu_N(B_{\tau}^{\varepsilon_t}(r_{n-2}))$  with  $C'' \exp((h_{\text{NP}} + \varepsilon'_{\text{err}})\tau)$ , where  $\varepsilon'_{\text{err}}$  is a constant smaller than  $\varepsilon_{\text{err}}$ , and  $C''$  is some large constant. The term  $C'$  is equal to  $CC''c'$ .

We now iterate (13)  $2\mathfrak{s}$  times and (14)  $n - 2\mathfrak{s}$  times.

$$\begin{aligned} \sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-1}) &\leq (Be)^{2\mathfrak{s}} \cdot (C'\tau)^{n-2\mathfrak{s}} \exp((h_{\text{NP}} - 1 + \varepsilon'_{\text{err}})n\tau) f_{\text{FLM}}(r_0) \\ &= f_{\text{FLM}}(r_0) \cdot \exp \left( \left( h_{\text{NP}} - 1 + \varepsilon'_{\text{err}} + \frac{\log(C'\tau) \left( \frac{n-2\mathfrak{s}}{n} \right) + \log(Be) \left( \frac{2\mathfrak{s}}{n} \right)}{\tau} \right) n\tau \right) \end{aligned}$$

By picking  $\tau$  large enough so that  $\mathfrak{s} \leq 4$  and  $\varepsilon'_{\text{err}} + \frac{\log(C'\tau) \left( \frac{n-2\mathfrak{s}}{n} \right) + \log(Be) \left( \frac{2\mathfrak{s}}{n} \right)}{\tau} < \varepsilon_{\text{err}}$ , we get the claimed result.  $\square$

Proposition 4.11 tells us that a random walk on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  is biased away from the thin part of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ . It does so by proving strong upper bounds on the probability that a

random walk trajectory with  $n$  steps stays in the thin part is less than  $\exp((-1 + \varepsilon_{\text{err}})n\tau)$ : in other words, a random walk returns to  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$  with high probability.

**4.3.2. Why the random walk approach fails for  $\mathcal{T}(\mathcal{N}_g)$ .** If we wanted to make Proposition 4.11 work on  $\mathcal{T}(\mathcal{N}_g)$ , we would need to similarly show the random walk on  $\mathcal{T}(\mathcal{N}_g)$  is recurrent in a similarly strong sense: i.e. the probability of a length  $n$  trajectory staying in the thin part decays exponentially in  $n$ . A consequence of this requirement is that the expected return time to the thick part is finite.

Unlike  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ ,  $\mathcal{T}(\mathcal{N}_g)$  has two kinds of thin regions.

- Thin region where only two-sided curves get short.
- Thin region where some one-sided curve also gets short.

It is the second kind of thin region that poses a problem for  $\mathcal{T}(\mathcal{N}_g)$ . Minsky's product region theorem (Theorem A.3) tells us that up to additive error, the metric on these thin regions looks like a product of metrics on some copies of  $\mathbb{R}$  (corresponding to the one-sided short curves), some copies of  $\mathbb{H}$  (corresponding to the two-sided short curves), and a Teichmüller space of lower complexity. Since the random walk is controlled by the metric, the random walk on this product metric space is a product of random walks on each of the components.

In particular, the random walk on the  $\mathbb{R}$  component is a symmetric random walk on a net in  $\mathbb{R}$ : i.e. a symmetric random walk on  $\mathbb{Z}$ . Symmetric random walks on  $\mathbb{Z}$  are known to be recurrent, but only in a weak sense: they recur to compact subsets infinitely often, but the expected return time is unbounded.

This means we cannot hope to prove exponentially decaying upper bounds on the probability that a long random walk trajectory stays in the thin part, since that would lead to finite expected return times. This is why the random walk approach fails for  $\mathcal{T}(\mathcal{N}_g)$ .

**4.3.3. Recurrence for geodesic segments.** In this section, we reduce the problem of counting geodesic segments that travel in the thin part to counting trajectories of random walks that do the same.

**Proposition 4.12.** *For any  $\varepsilon_{\text{err}} > 0$ , there exists a constant  $C'$ , and a large enough  $R$ , such that the following estimate holds for the counting function  $M_{r_0}(R)$ .*

$$M_{r_0}(R) \leq C' \exp((h_{\text{NP}} - 1 + \varepsilon_{\text{err}})R)$$

Here,  $M_{r_0}(R)$  is the number of concave lattice points in a ball of radius  $R$  centered at  $r_0$ , where  $r_0$  is a point in  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$  at which we start our random walk, and  $h_{\text{NP}} = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ .

*Proof.* We first check if  $\tau$  we picked in the proof of Proposition 4.11 satisfies  $\frac{2\varepsilon_n}{\tau} < \frac{\varepsilon_{\text{err}}}{2}$ : if not, we pick a larger  $\tau$ .

Let  $\gamma$  be a mapping class such that  $\gamma p$  is a concave lattice point. We consider now the concavity detecting path for  $\gamma p$ : recall that this is a path that starts at  $p$ , and ends at  $\gamma p$ , and the middle segment obtained by deleting a prefix of length  $2 \cdot \text{diam}(\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$  and a suffix of length  $2 \cdot \text{diam}(\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$  stays outside  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ . We turn this path into an  $\mathfrak{s}$ -concave random walk trajectory by marking off points at distance  $\tau(1 - \frac{2\varepsilon_n}{\tau})$  on the segment, and then replacing those points with the nearest net point. All but the first  $\mathfrak{s}$  and the last  $\mathfrak{s}$  points in the trajectory lie outside  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ . Furthermore, the distance

between the adjacent points on the trajectory are at most  $\tau$ . The number of steps in this trajectory is  $n := \left\lceil \frac{R}{\tau} \right\rceil$ .

Let  $\mathcal{P}$  denote the collection of trajectories obtained via this construction. We apply Proposition 4.11 to count the number of such trajectories.

$$\#\mathcal{P} \leq C \exp \left( \left( h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2} \right) n\tau \right) \quad (15)$$

$$\leq C \exp \left( \left( h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2} \right) (R + \tau) \right) \quad (16)$$

We now determine how many different geodesic segments can map to the same random walk trajectory. If two geodesic segments  $[r_0, \gamma_1 r_0]$  and  $[r_0, \gamma_2 r_0]$  map to the same random walk trajectory, we must have that they fellow travel for most of their length, and as a result,  $d_{\varepsilon_t}(r_0, \gamma_2^{-1} \gamma_1 r_0)$  is bounded above by a constant value that only depends on  $\tau$ . Combining the above fact with (16) gives us a constant  $C'$  such that the following bound on  $M_{r_0}(R)$  holds.

$$\begin{aligned} M_{r_0}(R) &\leq C' \exp \left( \left( h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2} \right) (R + \tau) \right) \\ &= C' \exp \left( \left( h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2} \right) \left( 1 + \frac{\tau}{R} \right) (R) \right) \end{aligned}$$

Picking a value of  $R$  large enough gives us the result.  $\square$

We can now tie all of these calculations together to state our results on statistical convexity of  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ . Proposition 4.12 gives us an upper bound on  $h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$  (by applying the result for smaller and smaller values of  $\varepsilon_{\text{err}}$ ).

$$h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \leq h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1 \quad (17)$$

To prove Theorem 4.4, it will suffice to prove the following equality relating the lattice point entropy and net point entropy.

$$h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1 < h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \quad (18)$$

For convenience, we also define the undistorted versions of these entropy terms, using the Teichmüller metric  $d$  rather than the induced metric  $d_{\varepsilon_t}$ .

**Definition 4.13** ((Undistorted) lattice point entropy for  $\mathcal{T}(\mathcal{N}_g)$ ). Let  $p$  be a point in  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ , and let  $N_p(R)$  be the lattice point counting function.

$$N_p(R) := \# (\gamma \in \text{MCG}(\mathcal{N}_g) \mid d(p, \gamma p) \leq R)$$

The lattice point entropy  $h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g))$  is the following quantity.

$$h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g)) := \lim_{R \rightarrow \infty} \frac{\log N_p(R)}{R}$$

**Definition 4.14** ((Undistorted) net point entropy). Let  $K_p(R, \varepsilon_t)$  be the counting function for net points, where  $p \in \text{core}(\mathcal{T}(\mathcal{N}_g))$ .

$$K_p(R, \varepsilon_t) := \# (y \in \mathfrak{N} \mid d(p, y) \leq R)$$

The net point entropy  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)))$  is the following function defined in terms of  $K_p$ .

$$h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) := \lim_{R \rightarrow \infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

Note that the net point entropy does not depend on the precise value of  $\varepsilon_t$ , even though it is counting net-points in  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ , since the different values of  $\varepsilon_t$  change the counting function by a multiplicative term that does not depend on  $R$ .

Recall now Theorem 3.7, which for any  $\varepsilon_d > 0$ , provides a  $\varepsilon_t > 0$  such that the ratio of  $d_{\varepsilon_t}$  and  $d$  is bounded above by  $1 + \varepsilon_d$ . A consequence of this is that the distorted and the undistorted versions of the entropy terms differ by at most  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) \cdot \varepsilon_d$  and  $h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g)) \cdot \varepsilon_d$ .

In particular, if we show  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g))$ , (18) will follow (for small enough  $\varepsilon_t$ ), and so will Theorem 4.4. We package up this result as a theorem, which we will use in subsequent sections.

**Theorem 4.15.** *If  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g))$ , then  $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$  is statistically convex, and the action of  $\text{MCG}(\mathcal{N}_g)$  on  $\text{core}(\mathcal{T}(\mathcal{N}_g))$  is statistically convex-cocompact.*

## 5. EQUALITY OF LATTICE POINT ENTROPY AND NET POINT ENTROPY

In this, and the following section, we will prove that  $h_{\text{LP}} = h_{\text{NP}}$ , which will let us apply Theorem 4.15 to conclude that the  $\text{MCG}(\mathcal{N}_g)$  action on  $\text{core}(\mathcal{T}(S))$  is SCC for surfaces  $S$  of finite type.

**Theorem 5.1** (Entropy equality). *For any surface  $S$  of finite type, the following relationship holds between the  $h_{\text{NP}}$  and  $h_{\text{LP}}$ .*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(S), \varepsilon_t)$$

*Remark.* In the case where  $S$  is an orientable surface, the theorem is a corollary of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12, Theorem 1.2]. However, the proof of the stronger theorem in the orientable setting uses facts about the dynamics of the geodesic flow on the moduli space, which we don't have in the non-orientable setting. The proof of the weaker theorem only uses coarse geometric methods, and works equally well for orientable and non-orientable surfaces.

**5.1. Base case.** We will prove this theorem by inducting on the Euler characteristic of the surface  $S$ . The 4 base cases we need to check are the 3 non-orientable surfaces, and one orientable surface with Euler characteristic  $-1$ .

- $\mathcal{S}_{1,1,0}$ : This is the torus with 1 boundary component and 0 crosscaps attached.
- $\mathcal{S}_{1,0,1}$ : This is a torus with 0 boundary components, and 1 crosscap attached.
- $\mathcal{S}_{0,2,1}$ : This is a sphere with 2 boundary components, and 1 crosscap attached.
- $\mathcal{S}_{0,1,2}$ : This is a sphere with 1 boundary component, and 2 crosscaps attached.

**Lemma 5.2** (Entropy equality: base case). *For a surface  $S$  in  $\{\mathcal{S}_{1,1,0}, \mathcal{S}_{1,0,1}, \mathcal{S}_{0,2,1}, \mathcal{S}_{0,1,2}\}$ , the following relationship holds between the  $h_{\text{NP}}$  and  $h_{\text{LP}}$ .*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(S), \varepsilon_t)$$

*Proof.* For  $S = \mathcal{S}_{1,1,0}$ , we will directly prove the lemma, and for the remaining three non-orientable surfaces, we will use a description of their Teichmüller spaces and mapping class groups from Gendulphe [Gen17] to reduce to the first case, or show that the result follows trivially.

- $\mathcal{S}_{1,1,0}$ : Since  $\mathcal{S}_{1,1,0}$  is orientable, we have that  $\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})) = \mathcal{T}(\mathcal{S}_{1,1,0})$ , so it suffices to look at the full Teichmüller space. The Teichmüller space of  $\mathcal{S}_{1,1,0}$  is the upper

half plane  $\mathbb{H}^2$ , and the mapping class group is  $\mathrm{SL}(2, \mathbb{Z})$ . In this case, the number of lattice points in a ball of radius  $R$  grows like  $\exp(R)$ . More precisely, we have the following inequality for some constants  $c$  and  $c'$ .

$$c \leq \frac{\#(B_R(p) \cap p \cdot \mathrm{SL}(2, \mathbb{Z}))}{\exp(R)} \leq c' \quad (19)$$

Here,  $p$  is a lattice point, and  $B_R(p)$  is the ball of radius  $R$  centered at  $p$ .

To count the net points in the ball of radius, we parameterize the net points by how far from the orbit of  $p$  they lie. Since we're looking for net points in a ball of radius  $R$ , the furthest away they can be from the orbit is  $R$ . We have the following sum decomposition (for an arbitrary choice of  $\varepsilon_b > 0$ ) for the cardinality of the net points.

$$\#(B_R(p) \cap \mathfrak{N}) = \#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}) + \#(B_R(p) \cap \mathfrak{N}_{> \varepsilon_b R}) \quad (20)$$

Here,  $\mathfrak{N}_{\leq \varepsilon_b R}$  denotes the net points that lie within distance  $\varepsilon_b R$  of the orbit of  $p$ , and  $\mathfrak{N}_{> \varepsilon_b R}$  denotes the net points that lie more than distance  $\varepsilon_b R$  of the orbit of  $p$ .

We will show that the first term is at most  $p(R) \exp(R(1 + \varepsilon_b))$ , for some polynomial  $p(R)$ , and that the second term grows slower than the first term. Since the choice of  $\varepsilon_b$  was arbitrary, this will prove the equality of the two entropy terms.

Let  $\mathfrak{N}_{\leq \varepsilon_b R}(\gamma)$  denote the subset of  $\mathfrak{N}_{\leq \varepsilon_b R}$  whose closest lattice point is  $\gamma p$ . Observe that  $d(p, \gamma p)$  is at most  $R + \varepsilon_b R$ , by the triangle inequality. We also have the following inequality for any  $\gamma$ , and for some polynomial  $p$ , by Lemma 5.6.

$$\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}(\gamma)) \leq p(R)$$

Using the two facts we stated, we get the following upper bound for  $\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R})$ .

$$\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}) \leq p(R) \cdot (\exp(R(1 + \varepsilon_b)))$$

This is precisely the bound we needed for the first term in (20).

Now we show that the second term of (20) grows slower than  $\exp(R(1 - \frac{\varepsilon_b}{2}))$ . For any point  $x$  in  $\mathfrak{N}_{> \varepsilon_b R}$ , we can replace the geodesic  $[p, x]$  with two shorter segments,  $[p, x_0]$  and  $[x_0, x]$ , where  $x_0$  is the net point closest to the last point on the  $[p, x]$  which stays within some bounded distance of a lattice point. We also have that  $d(p, x_0) \leq R(1 - \varepsilon_b)$ , by our assumption, which lets us count the number of such points  $x_0$ . There are at most  $\exp(R(1 - \varepsilon_b))$  such points. Now we fix an  $x_0$ , and we need to estimate the number of possibilities for  $x$ , given that  $[x_0, x]$  stays entirely within the thin part of  $\mathrm{SL}(2, \mathbb{R})/\mathrm{SL}(2, \mathbb{Z})$ . Note that this reduces to estimating the volume of the intersection of a ball  $B_{\varepsilon_b R}(x_0)$  with a horoball  $H$  which has  $x_0$  in its boundary. Working in the upper half plane model for  $\mathbb{H}$ , where  $x_0 = i$ , and the region  $H$  is the set of points whose imaginary component is greater than 1, we get that the region of integration is contained in a rectangle, bounded by  $-C \exp(\frac{\varepsilon_b R}{2}) \leq \mathrm{Re}(z) \leq C \exp(\frac{\varepsilon_b R}{2})$  and  $1 \leq \mathrm{Im}(z) \leq \exp(\varepsilon_b R)$ , where  $C$  is some fixed constant that we do not explicitly write down. The volume of this region is given by the following



integral.

$$\text{Vol}(B_R(x_0) \cap H) \leq \int_1^{\exp(\varepsilon_b R)} \int_{-C \exp(\frac{\varepsilon_b R}{2})}^{C \exp(\frac{\varepsilon_b R}{2})} \frac{1}{y^2} dx dy \quad (21)$$

$$\leq C' \exp\left(\frac{\varepsilon_b R}{2}\right) \quad (22)$$

We thus have the following upper bound on  $\#(B_R(p) \cap \mathfrak{N}_{>\varepsilon_b R})$  for large enough values of  $R$ .

$$\begin{aligned} \#(B_R(p) \cap \mathfrak{N}_{>\varepsilon_b R}) &\leq \exp\left(\frac{\varepsilon_b R}{2}\right) \cdot \exp(R(1 - \varepsilon_b)) \\ &\leq \exp\left(R\left(1 - \frac{\varepsilon_b}{2}\right)\right) \end{aligned}$$

This finishes proving the two claims we made about the terms of (20), and proves the result for  $\mathcal{S}_{1,1,0}$ .

- $\mathcal{S}_{1,0,1}$ : This surface is very similar to the previous case: it's obtained by gluing together the boundary component of  $\mathcal{S}_{1,1,0}$  via the antipodal map. It's a theorem of Scharlemann [Sch82] and also Gendulphé [Gen17] that there is a unique one-sided curve  $\kappa$  in  $\mathcal{S}_{1,0,1}$  whose complement is  $\mathcal{S}_{1,1,0}$ . As a consequence,  $\text{MCG}(\mathcal{S}_{1,0,1}) \cong \text{MCG}(\mathcal{S}_{1,1,0})$ , and  $\mathcal{T}(\mathcal{S}_{1,1,0}) \hookrightarrow \mathcal{T}(\mathcal{S}_{1,0,1})$ , where the inclusion map is given by considering a point in  $\mathcal{T}(\mathcal{S}_{1,1,0})$ , where the boundary component has length  $\varepsilon_t$ , and gluing it via the antipodal map to get a point in  $\mathcal{T}(\mathcal{S}_{1,0,1})$ . The inclusion map is also equivariant with respect to the action of  $\text{MCG}(\mathcal{S}_{1,1,0})$  and  $\text{MCG}(\mathcal{S}_{1,0,1})$ .

We consider now  $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ : the curve  $\kappa$  cannot get shorter than the threshold specified by the core. We now show that  $\kappa$  cannot be arbitrary long either. At any point  $z \in \text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ , let  $\kappa'$  be the shortest curve that intersects  $\kappa$  exactly once (see Figure 1). We have an upper bound for the length of  $\kappa'$ : namely the length of the orthogeodesic arc on  $\mathcal{S}_{1,0,1} \setminus \kappa$  that starts and ends at  $\kappa$ . It follows from hyperbolic trigonometry that if the length of  $\kappa$  goes to  $\infty$ , then the length of the orthogeodesic, hence the length of  $\kappa'$  will approach 0, the point in  $\mathcal{T}(\mathcal{S}_{1,0,1})$  will leave  $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ .

If we consider the pants decomposition of the surface along  $\kappa$ , and any two sided curve, we see that the length coordinates of  $\kappa$  in  $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$  are contained in a compact interval  $[t_1, t_2]$ , where  $t_1 > 0$ . This means that  $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$  is a bounded neighbourhood of the image of  $\mathcal{T}(\mathcal{S}_{1,1,0})$ .

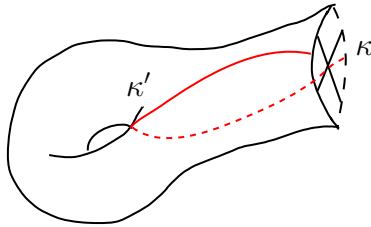


FIGURE 1. The curves  $\kappa$  and  $\kappa'$  on  $\mathcal{S}_{1,0,1}$ .

From the previous case, we already have  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0}), \varepsilon_t)$ , and since their mapping class groups are isomorphic, we also have  $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0}), \varepsilon_t) =$

$h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,0,1}), \varepsilon_t)$ . We now need to prove that  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})), \varepsilon_t) = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1})), \varepsilon_t)$  to prove the result for this case. We have that the net for  $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$  lies in a bounded neighbourhood of the net for  $\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))$ : this implies that the cardinalities of the net points in a ball of radius  $r$  differ by at most a multiplicative constant.

$$\#(B_R(p) \cap \mathfrak{N}_{\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))}) \leq c \cdot \#(B_R(p) \cap \mathfrak{N}_{\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))})$$

Since the two cardinalities differ by at most a multiplicative constant, they have the same exponential growth rate.

- $\mathcal{S}_{0,2,1}$ : The mapping class group of this surface is finite: in fact, it is isomorphic to  $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$  (see Gendulphie [Gen17]). This means  $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{0,2,1}), \varepsilon_t) = 0$ . This surface has exactly two simple geodesics  $\kappa$  and  $\kappa'$ , which intersect each other exactly once, such that deleting either one of them results in a pair of pants (see Figure 2).

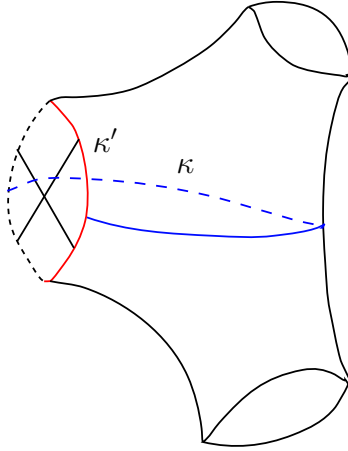


FIGURE 2. The curves  $\kappa$  and  $\kappa'$  on  $\mathcal{S}_{0,2,1}$ .

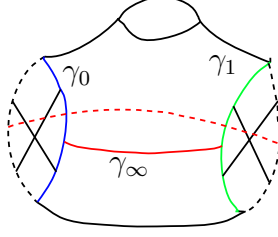
Picking a pants decomposition along either  $\kappa$  or  $\kappa'$ , we see that  $\mathcal{T}(\mathcal{S}_{0,2,1})$  is homeomorphic to  $\mathbb{R}_{>0}$ , where the homeomorphism is given by the length coordinate.

If we now consider  $\text{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$ , the lengths of  $\kappa$  and  $\kappa'$  are bounded below by the threshold. But they are also bounded above, by an argument similar to the previous case, namely if either  $\kappa$  or  $\kappa'$  are very long, the other one sided curve must be very short. This proves that  $\text{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$  is compact, and as a result  $h_{\text{NP}}(\text{core}(\mathcal{S}_{0,2,1}), \varepsilon_t) = 0$ . This proves the lemma for  $\mathcal{S}_{0,2,1}$ .

- $\mathcal{S}_{0,1,2}$ : This surface has a unique two-sided element, which we denote by  $\gamma_\infty$ . The one sided curves on this surface are indexed by  $\mathbb{Z}$ , where  $\gamma_n = D_n \gamma_0$ , and  $D_n$  is the Dehn twist about  $\gamma_\infty$  (see Figure 3).

The mapping class group of this surface is also virtually generated by  $D_n$ . If we consider the pants decomposition along  $\gamma_\infty$ , we get a Fenchel-Nielsen map from  $\mathcal{T}(\mathcal{S}_{0,1,2})$  to the upper half plane  $\mathbb{H}^2$ , where the  $y$ -coordinate is  $\frac{1}{\ell(\gamma_\infty)}$ , and the  $x$ -coordinate is the twist around  $\gamma_\infty$ . Furthermore, this map is also an isometry, and with respect to these coordinates,  $D_n$  is the action of  $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  on  $\mathbb{H}^2$ .

If we now consider  $\text{core}(\mathcal{T}(\mathcal{S}_{0,1,2}))$ , that consists of the points in  $\mathbb{H}^2$  whose  $y$ -coordinate is greater than some threshold value, i.e. a horoball in  $\mathbb{H}^2$ . We showed in (21) that for the action of  $D_n$  on a horocycle, the volume growth entropy and the

FIGURE 3. The curves  $\gamma_\infty$ ,  $\gamma_0$  and  $\gamma_1$  on  $\mathcal{S}_{0,1,2}$ .

lattice point growth entropy are both equal to  $\frac{1}{2}$ . The former entropy is precisely  $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{0,1,2})), \varepsilon_t)$ , and the latter entropy is  $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{0,1,2}), \varepsilon_t)$ .

This concludes the proof of the lemma for the 4 surfaces with  $\chi(S) = -1$ .  $\square$

**5.2. Good points and bad points.** The proof of Theorem 5.1 will split up into counting two kinds of net points, which we will call *good points* and *bad points*.

**Definition 5.3** (Good points). A point in  $B_R(p) \cap \mathfrak{N}$  is good if it is at most distance  $\varepsilon_b R$  away from a lattice point  $\gamma p$ . The set of good points is denoted by  $\mathfrak{N}_g(p, R, \varepsilon_b)$ .

**Definition 5.4** (Bad points). A point in  $B_R(p) \cap \mathfrak{N}$  is bad if it is more than distance  $\varepsilon_b R$  away from the nearest lattice point. The set of bad points is denoted by  $\mathfrak{N}_b(p, R, \varepsilon_b)$ .

Observe that the classification of a point as good or bad depends on the choice of  $R$ ,  $p$ , and an additional parameter  $\varepsilon_b > 0$ .

We also further subdivide  $\mathfrak{N}_g(p, R, \varepsilon_b)$  based on what the closest lattice point is.

**Definition 5.5** (Good point in the domain of  $\gamma$ ). For  $\gamma \in \text{MCG}(S)$ , the set  $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$  denotes the subset of  $\mathfrak{N}_g(p, R, \varepsilon_b)$  whose closest lattice point is  $\gamma p$ .

We will now prove a lemma that provides an upper bound on the number of good points when restricted to a fundamental domain.

**Lemma 5.6.** *There exists a polynomial function  $q$ , whose degree only depends on the topological type of  $S$ , such that for any  $\gamma \in \text{MCG}(S)$ , the following inequality holds for the cardinality of points in  $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$ .*

$$\#(\mathfrak{N}_g(\gamma, p, R, \varepsilon_b) \cap B_R(\gamma p)) \leq q(R)$$

*Remark.* Eskin and Mirzakhani [EM11, Lemma 3.2] prove this lemma for Teichmüller spaces of orientable surfaces, by comparing the extremal lengths of various curves on the underlying surfaces. We adapt the same proof for non-orientable surfaces, replacing extremal length for hyperbolic lengths instead.

Before we prove Lemma 5.6, we will need the following lemma on packing an  $\varepsilon_n$ -separated set into a ball of fixed radius in Teichmüller space (where  $\varepsilon_n$  is the parameter associated to our net  $\mathfrak{N}$ ).

**Lemma 5.7** (Packing bound). *For an constants  $C > 0$  and  $\varepsilon_n > 0$ , there exists a constant  $D(C, \varepsilon_n, S)$  depending on the constants  $C$ ,  $\varepsilon_n$ , and the topological type of the surface  $S$  such that any ball  $B_C(p)$  (independent from the choice of  $p$ ) in  $\mathcal{T}(S)$  cannot contain more than  $D$  points that are pairwise distance at least  $\varepsilon_n$  apart.*

*Proof.* First of all, note that the above lemma holds for  $S = \mathcal{S}_{1,1,0}$ , since  $\mathcal{T}(\mathcal{S}_{1,1,0})$  is  $\mathbb{H}^2$ , which is homogeneous.

Next, note that the lemma also holds for compact metric spaces, because we can express  $D$  as a upper semi-continuous function of the point  $p$ , which will achieve a maximum on a compact metric space.

Next, note that if the lemma holds for metric spaces  $X$  and  $Y$ , it also holds for  $X \times Y$ , where the metric on  $X \times Y$  is the sup product without an additive error. To see this, we consider the minimal number of balls  $E(C, \varepsilon_n, X \times Y)$  of radius  $\frac{\varepsilon_n}{2}$  needed to cover  $X \times Y$ . The covering number and the packing number  $D$  are related by the following standard inequality.

$$D(C, \varepsilon_n, X \times Y) \leq E(C, \varepsilon_n, X \times Y) \leq D(C, 2\varepsilon_n, X \times Y)$$

Furthermore, one can easily see that for a sup-product  $X \times Y$ , we can bound the covering number of  $X \times Y$  by a product of the covering number for  $X$  and  $Y$ , by taking a product of coverings for  $X$  and  $Y$ .

$$E(C, \varepsilon_n, X \times Y) \leq E(C, \varepsilon_n, X) \times E(C, \varepsilon_n, Y)$$

Combining the two inequalities, get a bound for  $D(C, \varepsilon_n, X \times Y)$  in terms of  $D(C, 2\varepsilon_n, X)$  and  $D(C, 2\varepsilon_n, Y)$ .

$$D(C, \varepsilon_n, X \times Y) \leq D(C, 2\varepsilon_n, X) \times D(C, 2\varepsilon_n, Y)$$

We now show that the lemma also holds, with a worse constant, for metric spaces that are sup-products with an additive error  $c$ , where  $c$  is a constant smaller than  $\varepsilon_n$ . Let  $Z$  be a metric space for which the lemma holds, and  $Z'$  be a metric space with the same underlying points, but whose metric differs from  $Z$  by an additive error  $c$ .

$$|d_Z(x, y) - d_{Z'}(x, y)| \leq c$$

Let  $\mathcal{Z}$  be a set of points in a ball of radius  $C$  in  $Z'$  that are pairwise distance at least  $\varepsilon_n$  apart. Consider  $\mathcal{Z}$  as a subset of  $Z$  instead, we have that they are contained in a ball of radius at most  $C + c$ , and are pairwise distance at least  $\varepsilon_n - c$  apart. We thus get the inequality relating the packing numbers for  $Z$  and  $Z'$ .

$$D(C, \varepsilon_n, Z') \leq D(C + c, \varepsilon_n - c, Z)$$

Suppose now that we have the lemma for the Teichmüller spaces of all surfaces with Euler characteristic at least  $-n$ . To show the lemma for a Teichmüller space of a surface  $S$  with Euler characteristic  $-n - 1$ , we break up the Teichmüller space into the thick part, where all curves are at least  $\varepsilon$  or longer, and the thin part. We pick  $\varepsilon$  such that the metric thin part on the thin part is equal to the sup-product metric, up to an additive error  $c$ , where  $c < \varepsilon_n$ , by Minsky's product region theorem [Min96, Theorem 6.1] (or Theorem A.3). The mapping class group acts co-compactly on the thick, so the first reduction applies, and on the thin part, the metric is the sup product up to an additive error of  $c$ , so the second reduction applies.  $\square$

*Proof of Lemma 5.6.* We begin by making three simplifying reductions. First, it will suffice to prove the following stronger claim instead.

$$\#(\mathfrak{N}(\gamma) \cap B_R(\gamma p)) \leq q(R) \tag{23}$$

Here,  $\mathfrak{N}(\gamma)$  denotes the set of net points whose closest lattice point is  $\gamma p$ :  $\mathfrak{N}(\gamma)$  is therefore a superset of  $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$ .

Next, note that it suffices to prove (23) for  $\gamma = 1$ , since our choice of basepoint  $p$  was arbitrary.

And finally, it will suffice to prove the following claim.

*Claim.* There exists a set  $\mathcal{Z} \subset \mathcal{T}(S)$  such that  $\#\mathcal{Z} \leq R^{f(S)}$ , and for  $y \in B_R(p)$ , there exists a  $z \in \mathcal{Z}$  and  $\kappa \in \text{MCG}(S)$  such that  $d(y, \kappa z) \leq C$ , for some value  $f(S)$  that only depends on the topological type of  $S$ , and some fixed constant  $C$ .

To see why this suffices, suppose we have such a  $\mathcal{Z}$ . Without loss of generality, we can assume that for all points  $z \in \mathcal{Z}$ , the closest lattice point is  $p$ : otherwise we could replace such a point  $z$  by  $\kappa z$  for an appropriate choice of  $\kappa$ . We then have that for any  $n \in \mathfrak{N}(1) \cap B_p(R)$ , there exists some  $z \in \mathcal{Z}$ , such that  $d(z, n) \leq 2C$ . Since  $\#\mathcal{Z} \leq R^{f(\chi(S))}$ , we have that  $\#\mathfrak{N}(1) \cap B_p(R) \leq C' R^{f(\chi(S))}$ , for some other constant  $C'$ , by Lemma 5.7.

*Proof of claim:* We consider short markings on the point  $p \in \text{teich}(S)$ . Given a pants decomposition  $\{\alpha_1, \dots, \alpha_k\}$  of a surface, a set of short transverse curves is a collection of curves  $\{\beta_1, \dots, \beta_k\}$ , such that  $\beta_i$  only intersects  $\alpha_i$ , and is the shortest such curve amongst all the curves intersecting only  $\alpha_i$ . The set  $\{\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k\}$  is called a marking. A marking is said to be short if the total length of the pants curves  $\alpha_i$  is minimized amongst the mapping class group orbit of  $\{\alpha_1, \dots, \alpha_k\}$ .

Note that there are only many short markings at point  $p$ . We know that for each of these short markings, the lengths of the pants curves are bounded above by some constant  $T$ . Each of these pants multicurves have  $N = -3\chi(S) - b$  pants curves on them, where  $b$  is the number of boundary components of  $S$ . Let  $\{M_1, \dots, M_J\}$  denote the set of short markings.

We construct the points  $z \in \mathcal{Z}$  by just varying the lengths of these pants curves: the set of lengths we will allow are the following.

$$\text{Acceptable lengths} = \{T, T \exp(-1), T \exp(-2), \dots, T(\exp(-\lceil R \rceil) - \log(s))\}$$

Here  $s$  is the length of the shortest curve on the point  $p$  in  $\mathcal{T}(S)$ . We define the point  $z_{j, i_1, i_2, \dots, i_N}$  to be the point in  $\mathcal{T}(S)$  obtained by considering the marking  $M_j$  at  $p$ , and setting the length of the  $k^{\text{th}}$  pants curve to be  $i_k$ , where the  $i_k$  is one of the acceptable lengths. It's clear that the cardinality of  $\mathcal{Z}$  is at most  $J \cdot R^N$ , which is a polynomial only depending on the topological type of the surface  $S$  and the basepoint  $p$ .

Suppose now that  $y$  is some other point in  $B_R(p)$ . We pick a  $\kappa \in \text{MCG}(S)$  such that the shortest marking on  $\kappa y$  is one of the markings  $M_j$  for  $1 \leq j \leq J$ . We now need to show that one of the  $z \in \mathcal{Z}$  is close to  $\kappa y$ . Pick the  $z$  such that the corresponding lengths of the pants curves are closest to the lengths of the pants curves on  $\kappa y$ . We can now invoke the combinatorial distance formula for Teichmüller metric (proved by Rafi [Raf07] for the orientable setting, and Theorem A.6 for the non-orientable case).

$$d(z, \kappa y) \doteq \sum_Y [d_Y(z, \kappa y)]_k + \sum_{\alpha \notin \Gamma} \log [d_\alpha(z, \kappa)]_k + \max_{\alpha \in \Gamma} d_{\mathbb{H}_\alpha}(z, \kappa y)$$

In the above formula, the first term is the distance between the short markings when projected to non-annular subsurfaces, the second term is the distance between the short markings when projected to annular subsurfaces whose core curves are not the pants curves in the marking, and the third term corresponds to the length and twist parameters of the short curves.

Since both  $z$  and  $\kappa y$  have the same short markings, the first two terms in the above sum become 0. Also, since we picked  $z$  to be the element of  $\mathcal{Z}$  such that the lengths were closest to those on  $\kappa y$ , the third term is bounded by some constant, which proves the result.  $\square$

**5.3. Using complexity length to count bad points.** In this section we introduce an alternative to the Teichmüller metric, called the *complexity length* (see Definition 6.23). Complexity length (denoted by  $\mathfrak{L}$ ) was constructed by Dowdall and Masur [DM23], in order to get better estimates on net points contained in the thin part of Teichmüller space (for orientable surfaces). We adapt the construction of complexity to the Teichmüller space of non-orientable surfaces in Section 6. In this section, we state the main results about complexity length we need in order to prove Theorem 5.1.

For this section, we will state the results with a rescaled version of complexity length, in order to compare it with Teichmüller length.

**Definition 5.8** (Rescaled complexity length). Let  $S$  be a surface of finite type. The rescaled complexity length  $d_{\text{comp}}$  on  $\text{core}(\mathcal{T}(S))$  is given by the following formula.

$$d_{\text{comp}}(x, y) = \frac{\mathfrak{L}(x, y)}{h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t)}$$

Here,  $\mathfrak{L}(x, y)$  is the complexity length between points  $x$  and  $y$ .

The first result we will need is a count of the net points with respect to the rescaled complexity length.

**Theorem 5.9** (Theorem 12.1 of [DM23], Theorem 6.24). *There exists a polynomial function  $p(R)$  that depends on the net  $\mathfrak{N}$ , and a parameter  $\varepsilon_{\text{err}} > 0$ , such that the following inequality holds for any  $\varepsilon_{\text{err}} > 0$ .*

$$\#(y \in \mathfrak{N} \mid d_{\text{comp}}(p, x) \leq R) \leq p(R) \exp((h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t) + \varepsilon_{\text{err}}) \cdot R)$$

The next result, which is the main theorem of Section 6, is that if  $y$  is a bad point that is Teichmüller distance  $R$  away from  $p$ , then its rescaled complexity distance to  $p$  is smaller than  $R$  by a definite amount. We state this theorem with an additional hypothesis on the net point entropy of subsurfaces. We will establish that this hypothesis holds inductively in Section 5.5.

**Theorem 5.10** (Linear gap in complexity length, Theorem 6.25). *Suppose that for all proper subsurfaces  $V$  of  $\mathcal{S}$ , the following inequality holds.*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}), \varepsilon_t))$$

*Then for any  $\varepsilon_b > 0$ , there exists  $c > 0$ , such that for all  $R > 0$ , and for any bad point  $y$ , i.e. a point in  $\mathfrak{N}_b(p, R, \varepsilon_b)$ , the following upper bound on the complexity distance between  $p$  and  $y$  holds.*

$$d_{\text{comp}}(p, y) \leq R(1 - c)$$

*Remark.* We give a brief outline of why the above result should hold. Complexity length can be thought of as a weighted version of Teichmüller length, where a specific segment is assigned a weight based on whether it's traveling in the thick part of Teichmüller space, or the thin part. In the latter case, it is assigned a smaller weight that is proportional to the net point entropy of the product region it is traveling in. The Teichmüller geodesics associated to bad points spend a significant fraction traveling in the thin part, by the very definition of



bad points. It stands to reason then that the complexity length assigned to them is smaller by a definite amount, due to the time they spend in the thin part.

Combining Theorems 5.9 and 5.10, it follows that as  $R$  goes to  $\infty$ , the proportion of bad points goes to 0, which is what we need for Theorem 5.1.

**5.4. Entropy gap.** In the previous section, we saw that the key hypothesis we need for the complexity length estimate was that for any proper subsurface  $V$  of  $\mathcal{S}$ , the following strict inequality held.

$$h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) \quad (24)$$

In this section, we will prove that (24) holds by proving a similar inequality for the lattice point entropy, and using the fact that Theorem 5.1 holds for all proper subsurfaces  $V$ , by the inductive hypothesis.

**Lemma 5.11** (Lattice point entropy gap). *Let  $\mathcal{S}$  be a surface, and  $\chi(\mathcal{S}) \leq -2$ . If  $V$  is a proper subsurface and Theorem 5.1 holds for  $V$ , then we have the following strict inequality between their lattice point entropy.*

$$h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) < h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t)$$

*Remark.* We do actually need the hypothesis  $\chi(\mathcal{S}) \leq -2$  in the statement of the lemma for two reasons. The first reason is that the lemma is actually false for  $\mathcal{S}_{1,0,1}$ . Recall that this surface has the torus with one boundary component as a subsurface, but their mapping class groups are isomorphic, and have the same lattice point growth entropy. Another reason why we need the hypothesis is that the proof of the lemma proceeds via a construction of pseudo-Anosov elements on  $\mathcal{S}$ , and  $\mathcal{S}_{1,0,1}$  does not admit any pseudo-Anosov mapping classes.

*Proof of Lemma 5.11.* Observe that  $\text{MCG}(V)$  is a subgroup of  $\text{MCG}(\mathcal{S})$ . We will first construct an intermediate subgroup  $H = \mathbb{Z} * \text{MCG}(V)$ , which is the free product of a pseudo-Anosov element in  $\text{MCG}(\mathcal{S})$  with  $\text{MCG}(V)$ , and show that  $h_{\text{LP}}(H, \varepsilon_t) > h_{\text{LP}}(\text{MCG}(V), \varepsilon_t)$ . This is enough to prove the result, since  $H$  is a subgroup of  $\text{MCG}(\mathcal{S})$ , we have that  $h_{\text{LP}}(\text{MCG}(\mathcal{S}), \varepsilon_t) \geq h_{\text{LP}}(H, \varepsilon_t)$ .

We now need to show that  $\text{MCG}(\mathcal{S})$  contains a pseudo-Anosov element. We can invoke Penner's construction of pseudo-Anosov mapping classes ([Pen88, Theorem 4.1]), as long as we can construct a filling collection of *two-sided* curves in  $\mathcal{S}$ . This may not be always possible for  $\mathcal{S}$  where  $\chi(\mathcal{S}) = -1$ , but for  $\mathcal{S}$  with  $\chi(\mathcal{S}) \leq -2$ , this is always possible (see [LS18] and [KPW23] for explicit constructions). Let  $\kappa$  denote the pseudo-Anosov mapping class we construct.

By [Yan18, Proposition 6.6], there exists a large enough  $n$  such that  $\kappa^n$  and  $\text{MCG}(V)$  generate their free product in  $\text{MCG}(\mathcal{S})$ : call this subgroup  $H$ .

We now need to show that the lattice point entropy for  $H$  is strictly larger than the  $\text{MCG}(V)$ . To see this, we recall an equivalent definition of the lattice point entropy. The lattice point entropy is the infimum of the set of exponents  $h$  such that the following Poincaré series transitions converges for any  $x \in \mathcal{T}(\mathcal{S})$ .

$$\sum_{\gamma \in H} \exp(-h \cdot d_{\varepsilon_t}(x, \gamma x)) \quad (25)$$

Since  $H = \mathbb{Z} * \text{MCG}(V)$ , we can represent  $\gamma \in H$  as  $a_1 \cdot b_1 \cdot a_2 \cdots a_k \cdot b_k$ , where  $a_i$  belong in  $\mathbb{Z}$  and  $b_i$  belong in  $\text{MCG}(V)$ . We use this along with the triangle inequality to get an upper

bound for  $d(x, \gamma x)$ .

$$d_{\varepsilon_t}(x, \gamma x) \leq \sum_{i=1}^k d_{\varepsilon_t}(x, a_i x) + d_{\varepsilon_t}(x, b_i x) \quad (26)$$

We plug inequality (26) into (25) to get a lower bound.

$$\sum_{\gamma \in H} \exp(-h \cdot d_{\varepsilon_t}(x, \gamma x)) = \sum_{k=1}^{\infty} \left( \sum_{a_1} \cdots \sum_{a_k} \sum_{b_1} \cdots \sum_{b_k} \exp(-h \cdot d_{\varepsilon_t}(x, a_1 \cdot b_1 \cdots a_k \cdot b_k x)) \right) \quad (27)$$

$$\geq \sum_{k=1}^{\infty} \left( \sum_{a \in \mathbb{Z}} \exp(-h \cdot d_{\varepsilon_t}(x, ax)) \right)^k \left( \sum_{b \in \text{MCG}(V)} \exp(-h \cdot d_{\varepsilon_t}(x, bx)) \right)^k \quad (28)$$

We have that Theorem 5.1 holds for  $V$ , which means that  $\text{core}(\mathcal{T}(V))$  is SCC. Corollary 5.4 of [Yan18] states that group actions that are SCC have Poincaré series that diverge at the critical exponent. This means there's small enough  $\varepsilon > 0$  such that for  $h = h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) + \varepsilon$ , the series converges to a value greater than 1. But that means the Poincaré series for  $H$  diverges at  $h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) + \varepsilon$ , since we have a lower bound by a geometric series whose ratio is greater than 1. This proves that the critical exponent for  $H$  is strictly greater than the critical exponent for  $\text{MCG}(V)$ .  $\square$

We can now prove the entropy gap result for  $h_{\text{NP}}$ .

**Lemma 5.12** (Net point entropy gap). *Let  $\mathcal{S}$  be a surface, and  $\chi(\mathcal{S}) \leq -2$ . If  $V$  is a proper subsurface and Theorem 5.1 holds for  $V$ , then we have the following strict inequality between their net point entropy.*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$$

*Proof.* We have the following inequality, which follows trivially from the definition of  $h_{\text{LP}}$  and  $h_{\text{NP}}$ .

$$h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \leq h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) \quad (29)$$

From Lemma 5.11, we get the following inequality.

$$h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) < h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \quad (30)$$

Finally, since we have Theorem 5.1 for  $V$ , we have the following equality.

$$h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) = h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) \quad (31)$$

Chaining together (29), (30), and (31) gives us the result.  $\square$

**5.5. Proof of Theorem 5.1.** We now have all the lemmas we need in order to prove Theorem 5.1.

*Proof of Theorem 5.1.* We will prove this lemma by inducting on the complexity of the surface  $\mathcal{S}$ . Lemma 5.2 proves the result for surfaces with Euler characteristic equal to  $-1$ , which serves as the base case of the theorem.

We now assume that Theorem 5.1 already holds for all proper subsurfaces  $V$  of  $\mathcal{S}$ : it will suffice to show that the result holds for  $\mathcal{S}$ .

We will establish that for any  $\varepsilon_b > 0$ , there exists a polynomial  $q(R)$ , and  $R$  large enough, such that the following bound holds.

$$\#(B_R(p) \cap \mathfrak{N}) \leq q(R) \cdot \exp(h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \cdot R \cdot (1 + 2\varepsilon_b))$$

We first count the good points in  $B_R(p)$ , by partitioning them according to the nearest lattice point.

$$\mathfrak{N}_g(p, R, \varepsilon_b) = \bigsqcup_{\gamma \in \text{MCG}(\mathcal{S})} \mathfrak{N}_g(\gamma, p, R, \varepsilon_b) \quad (32)$$

Observe that if  $y \in \mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$ , then  $d(p, \gamma p) \leq R(1 + \varepsilon_b)$ , since  $d(p, y) \leq R$  and  $d(y, \gamma p) \leq \varepsilon_b R$ . This observation leads to the following upper bound on  $\#(\mathfrak{N}_g(p, R, \varepsilon_b))$ .

$$\#(\mathfrak{N}_g(p, R, \varepsilon_b)) \leq \sum_{\substack{\gamma \in \text{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1 + \varepsilon_b)}} \#(B_{\varepsilon_b R}(\gamma p) \cap \mathfrak{N}_g(\gamma, p, R, \varepsilon_b)) \quad (33)$$

By Lemma 5.6, there exists a polynomial  $q(R)$  such that each term in the above sum is at most  $q(R)$ .

$$\#(\mathfrak{N}_g(p, R, \varepsilon_b)) \leq \sum_{\substack{\gamma \in \text{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1 + \varepsilon_b)}} q(R) \quad (34)$$

$$\leq q(R) \cdot \exp(h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \cdot R \cdot (1 + 2\varepsilon_b)) \quad (35)$$

Here, we estimated the cardinality of  $\gamma$  such that  $d(p, \gamma p) \leq R(1 + \varepsilon_b)$  as at most  $\exp(h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \cdot R \cdot (1 + 2\varepsilon_b))$ , for large enough  $R$ . We have the desired upper bound on the cardinality for the good points. Now we show that the number of bad points is much smaller than the total number of points in the ball, which will then prove the result.

From the inductive hypothesis, we have that Theorem 5.1 holds for all proper subsurfaces  $V$ . By Lemma 5.12, we have that  $h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$ : this is precisely the hypothesis we need to apply Theorem 5.10. Applying the theorem, we see that if  $y$  is a bad point,  $d_{\text{comp}}(p, y) \leq R(1 - c)$ . We then apply Theorem 5.9 to get an upper bound on the number of bad points.

$$\#(\mathfrak{N}_b(p, R, \varepsilon_b)) \leq kR^k \cdot \exp((h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) + \varepsilon_{\text{err}}) \cdot R \cdot (1 - c))$$

We pick  $\varepsilon_{\text{err}}$  small enough such that the above term satisfies the following inequality for large enough  $R$ .

$$\exp((h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) + \varepsilon_{\text{err}}) \cdot R \cdot (1 - c)) < \exp\left(h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) \cdot R \cdot \left(1 - \frac{2c}{3}\right)\right)$$

On the other hand, we have that for large enough  $R$ , the total number of net points is at least  $\exp(h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t) \cdot R \cdot (1 - \frac{c}{2}))$ . Combining these two facts, we see that the proportion of bad points goes to 0 as  $R$  goes to  $\infty$ , which proves the result.  $\square$

## 6. LINEAR GAP IN COMPLEXITY LENGTH

**6.1. An example of counting in product regions.** Before we define complexity length, we will look at an example that illustrates why we need complexity length. Theorem 1.3 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12] proves an estimate on the volume of balls in Teichmüller space. From this volume estimate, we can obtain an estimate on the cardinality of net points of an  $(\varepsilon_n, 2\varepsilon_n)$ -net  $\mathfrak{N}$ .

**Theorem 6.1** (Theorem 1.3 of [ABEM12]). *For a point  $p$  in  $\mathcal{T}(S)$  (where  $S$  is a genus  $g$  surface with  $b$  boundary components), the number of net points in a ball of radius  $R$  centered at the origin satisfies the following asymptotic as  $R$  goes to  $\infty$ .*

$$\#(\mathfrak{N} \cap B_R(p)) \asymp \exp((6g - 6 + 2b)R)$$

Here, the multiplicative and additive constants showing up in  $\asymp$  only depend on  $p$  and  $\varepsilon_n$ .

Suppose now that we want to use the above theorem to count net points in a product region. More concretely, let  $p$  be a point in  $\mathcal{T}(S)$  such that a non-separating curve  $\gamma$  is very short:  $\ell_\gamma(p) \leq \delta \cdot \exp(-R_0)$ , for some  $\delta > 0$ , and some large  $R_0$ , and we want to estimate the cardinality of  $\mathfrak{N} \cap B_R(p)$  for  $R < R_0$ . Note that since  $R < R_0$ , the ball  $B_R(p)$  is still contained in the product region of Teichmüller space where  $\ell_\gamma \leq \delta$ .

Since the entire ball  $B_R(p)$  is in a product region, we have by Minsky's product region theorem (see Theorem A.3 for a precise statement) that the ball decomposes (up to an additive error) as the product of a ball in  $\mathcal{T}(S \setminus \gamma)$  and ball in  $\mathbb{H}$  (which corresponds to the length and twist around  $\gamma$ ). This gives us an alternative estimate for  $\#(\mathfrak{N} \cap B_R(p))$ .

$$\#(\mathfrak{N} \cap B_R(p)) \leq C \cdot (\mathfrak{N}_1 \cap B_R(p, S \setminus \gamma)) \cdot (\mathfrak{N}_2 \cap B_R(p, \mathbb{H})) \quad (36)$$

Here  $\mathfrak{N}_1$  and  $\mathfrak{N}_2$  are  $(\varepsilon_n, 2\varepsilon_n)$  nets for  $\mathcal{T}(S \setminus \gamma)$  and  $\mathbb{H}$ , and  $B_R(p, S \setminus \gamma)$  and  $B_R(p, \mathbb{H})$  are projections on the ball  $B_R(p)$  to the two components. Applying Theorem 6.1 to the right hand side of (36), we get a better estimate than we would have gotten with a direct application of Theorem 6.1.

$$\begin{aligned} \#(\mathfrak{N} \cap B_R(p)) &\leq C \cdot (\mathfrak{N}_1 \cap B_R(p, S \setminus \gamma)) \cdot (\mathfrak{N}_2 \cap B_R(p, \mathbb{H})) \\ &\asymp \exp((6(g-1) - 6 + 2(b+2))R) \cdot \exp(R) \\ &= \exp((6g - 6 + 2b - 1)R) \end{aligned}$$

This example illustrates that in order to count net points accurately, it's not sufficient to just estimate the distance between the base point  $p$  and the net point  $n$ : if the geodesic segment  $[p, n]$  travels in a product region, the count will be lower than what Theorem 6.1 predicts. In fact, the count will also depend on the type of the product region. In the above example, the product region had just one curve  $\gamma$  becoming short, but in general, a product region can have multiple curves getting short, in which case, the net point count will be even smaller.

We now consider a geodesic  $[x, y]$  (where  $x$  and  $y$  are net points) that travels through several product regions  $\pi_i$ , and possibly the thick part, which we will also consider a product region, albeit a trivial one. Let  $h_i$  be the exponent associated to the product region  $\pi_i$ : this is the exponent that will appear when we invoke Theorem 6.1 to count net points in the product region  $\pi_i$ . The order in which  $[x, y]$  travels through the product region is specified in Figure 4.

Let  $z_i$  denote the points on the geodesic segment that correspond to the times when the geodesic enters or exits a product region: in Figure 4, we have labeled  $z_i$  for  $1 \leq i \leq 8$ . Let  $\mathcal{J}_i$  denote the interval  $[x, z_1]$  for  $i = 0$ , the interval  $[z_i, z_{i+1}]$  for  $1 \leq i \leq 7$ , and  $[z_8, y]$  for  $i = 8$ . Let  $\ell_i$  be the length of  $\mathcal{J}_i$ , and  $e_i$  be the sum of the  $h_i$  for each of the product regions that the interval  $\mathcal{J}_i$  is in. Also, for each  $z_i$ , let  $z'_i$  denote the nearest net point.

Keeping  $x$  fixed, we can try to count the number of net points  $y$  that satisfy the configuration we have described. We have about  $O(\exp(e_0 \ell_0))$  possibilities for  $z'_1$ , and then keeping

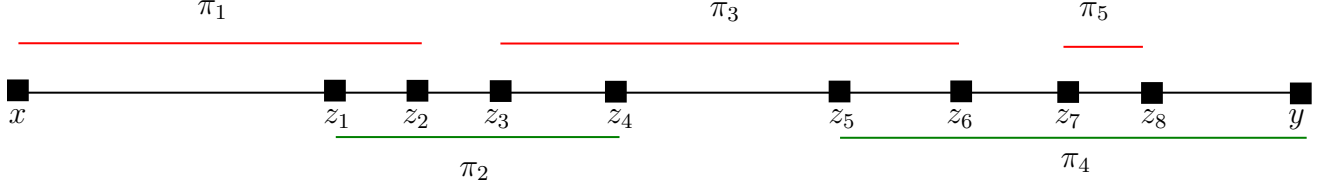


FIGURE 4. A schematic of the geodesic  $[x, y]$  traveling through several product regions.

a  $z'_1$  fixed,  $O(\exp(e_i \ell_1))$  possibilities for  $z'_2$  and so on. Multiplying all these estimates, we have the following upper bound for cardinality of  $y$ .

$$\#(y) \leq \exp \left( \sum_{i=0}^8 e_i \ell_i \right)$$

The quantity  $\sum e_i \ell_i$  serves as a re-weighted version of length of the geodesic in manner that works well with the counting function. This is a primitive version of the *complexity length* of  $[x, y]$ , and motivates the actual definition.

Before we define complexity length, we note two ways in which the above estimate overestimates the actual number of net points: that happens when a point in the geodesic is simultaneously in two or more product regions, which can happen in two ways.

- (i) The active subsurfaces associated to product regions are disjoint: In this case, we are accounting the length of the geodesic segment multiple times: once for each product region we are in. However, this overcounting is still better than directly invoking Theorem 6.1, since the sums of the exponents  $h_i$  associated to each of the disjoint product regions are smaller than the exponent associated to the entire surface.
- (ii) The active subsurfaces associated to product regions are nested: In this case as well, we are accounting for the length of a geodesic multiple times, once for each product region we are in. Unlike in the previous case, in this case, the exponents associated to each product region can add up to a quantity larger than the exponent associated to the entire surface, which means the presence of nested product regions can give a worse estimate than Theorem 6.1. We will get around this problem by looking only at product regions associated to special subsurfaces which are called *witnesses*.

**6.2. An overview of complexity length.** Now that we have motivated the need for complexity length, as well as considering special subsurfaces called witnesses, we formally define them in this section. This section is a summary for Sections 7 through 12 of Dowdall and Masur [DM23], so we refer the reader to those sections for details we elide. One difference in our presentation is that we care about these constructions for both orientable and non-orientable surfaces, while the original authors only work with orientable surfaces. However, their constructions and proofs go through for non-orientable surfaces, as long as we provide a proof of the non-orientable versions of some of the foundational results they use. We list those theorems here, and link to the proof of the non-orientable version that appears in Section A.

- (i) Minsky's product region theorem (see Theorem A.3).
- (ii) Distance formula for Teichmüller space (see Theorem A.6).
- (iii) Active intervals for subsurfaces (see Proposition A.7).

(iv) Consistency and realization (see Theorem A.9).

Let  $\mathcal{S}$  be a surface (not necessarily orientable), and  $\mathbf{C}$  some large arbitrary constant, and  $\varepsilon_t > 0$  a small constant we pick later. We also pick constants  $N_V$ , for each  $V \sqsubset \mathcal{S}$ , such that  $N_V$  only depends on the topological type of  $V$ . The precise values of the  $N_V$ 's is specified via Proposition 10.13 of [DM23]. We will also abuse notation slightly and use  $h_{\text{NP}}(V)$  to refer to  $h_{\text{NP}}(\text{core}(\mathcal{T}(V)))$  whenever  $V$  is a non-orientable surface: when  $V$  is orientable,  $h_{\text{NP}}(V)$  will refer to  $h_{\text{NP}}(\mathcal{T}(V))$ .

Let  $[x, y]$  be a geodesic segment in  $\mathcal{T}(\mathcal{S})$ : we describe the set  $\Upsilon(x, y)$  of subsurfaces along which  $[x, y]$  has large projections.

**Definition 6.2** (Active subsurfaces). A subsurface  $V \sqsubset \mathcal{S}$  is an active subsurface, i.e. in  $\Upsilon(x, y)$ , if one of the following two conditions hold.

- (i) The projection to  $\mathcal{C}(V)$  has diameter at least  $N_V$ .
- (ii) If  $V$  is annular with core curve  $\gamma$ , then

$$\min(\ell_\gamma(x), \ell_\gamma(y)) < \varepsilon_t$$

Associated to each active subsurface  $V$ , there is a non-empty connected sub-interval of  $[x, y]$ , which we call an active interval, and denote  $\mathcal{I}_V^{\varepsilon_t}$ , which we obtain via an application of Proposition A.7. The active intervals associated to active subsurfaces enjoy the following properties.

- (i)  $\ell_\alpha(z) < \varepsilon_t$  for  $z \in \mathcal{I}_V^{\varepsilon_t}$  and  $\alpha \in \partial V$ .
- (ii) For  $z \notin \mathcal{I}_V^{\varepsilon_t}$ ,  $\ell_\alpha(z) > \varepsilon_t'$  for some  $z \in \partial V$ , and some  $\varepsilon_t' < \varepsilon_t$  that only depends on  $\varepsilon_t$ .
- (iii) For  $[w, z] \subset [x, y]$  with  $[w, z] \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$ ,  $d_V(w, z) \leq M_{\varepsilon_t}$  for some  $M_{\varepsilon_t}$  that only depends on  $\varepsilon_t$ .
- (iv) For  $U \pitchfork V$ ,  $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$ .

For pairs of transverse subsurfaces  $U \pitchfork V$ , since  $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$  we can also determine which of the subsurfaces are active first.

**Definition 6.3** (Behrstock partial order). If  $U$  and  $V$  are a pair of transverse subsurfaces in  $\Upsilon(x, y)$ , we say  $U \leq V$  if  $\mathcal{I}_U^{\varepsilon_t}$  appears to the left of  $\mathcal{I}_V^{\varepsilon_t}$  in  $[x, y]$ .

Observe that when restricted to  $\mathcal{I}_V^{\varepsilon_t}$ , the geodesic is traveling in a product region, one of whose components is  $\mathcal{T}(V)$ , but trying to apply the technique from the previous subsection leads to the problem of overcounting, namely overcounting arising from subsurfaces either nested in  $V$ , or subsurfaces  $V$  is nested in.

To deal with this issue, we will consider a subset of  $\Upsilon(x, y)$ , called a *witness family*. However, to avoid overcounting, some additional properties are required of the witness families. Rather than defining all of those properties without context, we introduce them one at a time, after motivating the need for the property.

**Definition 6.4** (Witness family). A witness family  $\Omega(x, y)$  associated to the geodesic  $[x, y]$  is a subset of  $\Upsilon(x, y)$  satisfying the following properties.

- (i) For any  $Z \in \Upsilon(x, y)$ ,  $Z \sqsubset W$  for some  $W \in \Omega(x, y)$ .
- (ii) If  $Z \sqsubset W$ , and  $Z \in \Omega(x, y)$  and  $W \in \Upsilon(x, y)$ , then  $W$  must also either be a witness, or must be transverse to a witness  $V \in \Omega(x, y)$  such that  $Z \sqsubset V$ .

The first condition of the definition ensures that when we restrict our attention from all active subsurfaces to witnesses, we do not lose information, i.e. every active subsurface



contributes to whichever witness it is contained in. The second condition is a more technical requirement that is required to ensure that the other properties we define later work nicely.

We now make the notion of an active subsurface *contributing to a witness* more precise.

**Definition 6.5** (Complete witness family). For an active subsurface  $V$ , a witness  $W$  is said to be the  $\Omega$ -completion of  $V$ , denoted  $\bar{V}^\Omega$  if  $W$  is the minimal (by inclusion) witness containing  $V$ . If  $W = \bar{V}^\Omega$ , we say  $V$  *contributes* to  $W$ . Furthermore, a witness family is *complete* if every active subsurface has a unique  $\Omega$ -completion.

By partitioning off the collection active subsurfaces into classes, where each class is represented by a witness, and only considering the product regions associated to the witnesses, rather than all the active subsurfaces, we can cut down on the overcount we obtain by considering all active subsurfaces.

We now look at an extreme example of a complete witness families to motivate further properties that we will need from the witness families in order to count well.

**Example 6.6** (Trivial witness family). Let  $\gamma$  be a pseudo-Anosov mapping class on  $\mathcal{S}$ , such that  $\gamma$  has large translation distance on  $\mathcal{C}(\mathcal{S})$ , and  $\delta$  a reducible mapping class, acting on a subsurface  $V$  such that the action of  $\delta$  on  $\mathcal{C}(V)$  has large translation distance as well.

Let  $x$  be a point in  $\mathcal{T}(\mathcal{S})$ , and  $y = \delta\gamma\delta^{-1}x$ . The active subsurfaces for  $[x, y]$  contain the surfaces  $\mathcal{S}$ ,  $V$ , and  $\gamma V$ : however, we can pick  $\Omega(x, y) = \{\mathcal{S}\}$ , and check that this is a complete witness family.

In the above example, since we only have one subsurface in our witness family, we certainly do not overcount via overlapping product regions, but we do end up ignoring the fact that the geodesic  $[x, y]$  travels in a smaller product region near the beginning of the segment, as well as the end. For the initial and the final segment of the geodesic, the witness  $\mathcal{S}$  is too big for the subsurface the geodesic is actually traveling in. This suggest that a better choice of a witness family would be to include both  $V$  and  $\gamma V$  as witnesses too. We can take this approach further, and include every subsurface in  $\Upsilon(x, y)$  as a witness: this will still form a complete witness family. However, this approach also leads to multiple witnesses nested within one another, which is something we want to avoid as much as possible.

The drawback of Example 6.6 motivates the next property we will require from witness families, which is the notion of being *insulated*. Informally, a witness family  $\Omega(x, y)$  is insulated if all the maximal active subsurfaces that are active near the beginning or end of  $[x, y]$  are also witnesses.

**Definition 6.7** (Insulated witness family). A witness family  $\Omega(x, y)$  is insulated if for every  $E \in \Omega(x, y)$ , all subsurfaces  $V \sqsubset E$  satisfying the following properties are also witnesses.

- (i)  $V \in \Upsilon(x, y)$ .
- (ii)  $d_E(\mathcal{C}(V), x) \leq 9C$ , or  $d_E(\mathcal{C}(V), y) \leq 9C$ , where we consider  $\mathcal{C}(V)$  to be a subset of  $\mathcal{C}(E)$ .
- (iii)  $V$  is topologically maximal among the subsurfaces that satisfy (i) and (ii).

Once we have an insulated witness family, we can order a nested pair of witnesses  $W \sqsubset V$  based on whether  $W$  is active near the beginning or the end of the geodesic  $[x, y]$  projected to  $\mathcal{C}(V)$ .

**Definition 6.8** (Subordering). Let  $[x, y]$  be a geodesic in  $\mathcal{T}(\mathcal{S})$  and  $\Omega(x, y)$  a complete insulated witness family associated to  $[x, y]$ . Then for each nested pair of witnesses  $W \sqsubset V$ , a subordering is an assignment of exactly one of the following two possibilities:

- (i)  $W \swarrow V$
- (ii)  $V \searrow W$

The orderings  $\swarrow$  and  $\searrow$  satisfy the following properties.

- (i) If  $Z$ ,  $V$ , and  $W$  are witnesses such that  $Z \sqsubset V \sqsubset W$ , then  $Z \swarrow W$  iff  $V \swarrow W$  (equivalently,  $W \searrow Z$  iff  $W \searrow V$ ).
- (ii) If  $Z$ ,  $V$  and  $W$  are witnesses such that  $Z \swarrow V \searrow W$ , then  $Z \pitchfork_V W$  and  $Z \leq W$ .
- (iii) If  $Z$  and  $V$  are witnesses, and  $W$  an active subsurface such that  $Z \swarrow V \leq W$ , or  $W \leq V \searrow Z$ , then  $Z \pitchfork_V W$ .
- (iv) If  $Z$  and  $V$  are witnesses, such that  $Z \swarrow V$  (or  $Z \searrow V$ ), then there does not exist any active subsurface  $W$  such that the  $\Omega$ -closure of  $W$  is  $V$  and  $W \leq Z$  (or  $Z \leq W$ ).

Here  $Z \pitchfork_V W$  refers to notion of two subsurfaces cutting each other relative to  $V$ .

**Definition 6.9** (Relative cutting). Given a subsurface  $V$  of  $S$ , we say two subsurfaces  $Z$  and  $W$  of  $S$  *cut relative to  $V$*  if for any subsurfaces  $Z' \sqsubset Z$  and  $W' \sqsubset W$  that intersect  $V$ ,  $Z' \pitchfork W'$ .

We now provide some motivation for the various conditions that appear in the above definition. First of all, when we see  $Z \swarrow W$ , we are to read that as *the geodesic  $[x, y]$  makes progress in the nested subsurface  $Z$ , before making progress in the supersurface  $W$* . Similarly, when we see  $W \searrow Z$ , we are to read that as *the geodesic  $[x, y]$  make progress in the supersurface  $W$  before making progress in the nested subsurface  $Z$* . With this description of the subordering, conditions (i) and (iv) of the definition are easy to understand. The conditions (ii) and (iii) let us upgrade  $\swarrow$  and  $\searrow$  to transversality and time-ordering. A more intuitive reading of condition (ii) for instance would be, if  $Z \swarrow V \searrow W$ , that means the geodesic makes progress in  $Z$  before  $V$ , and then makes progress in  $W$ . That means if we just look at  $V$  and  $W$ , it makes progress in  $V$  and then  $W$ . And since neither of them are nested in the other, the only way they can be time-ordered is by cutting relative to  $V$ .

We can also see how the subordering on a witness family interacts with the witness family being insulated: recall the pair of witnesses  $V \sqsubset E$  from Definition 6.7.

- If  $d_E(\mathcal{C}(V), x) \leq 9\mathbf{C}$ , then  $V \swarrow E$ , since the geodesic makes progress in  $V$  before  $E$ .
- If  $d_E(\mathcal{C}(V), y) \leq 9\mathbf{C}$ , then  $E \searrow V$ , since the geodesic makes progress in  $E$  before  $V$ .

However, the above example does not capture all the ways in which we can have  $V \swarrow E$  or  $E \searrow V$ . Consider a decomposition of a Teichmüller geodesic by the active intervals corresponding to witnesses illustrated in Figure 5.

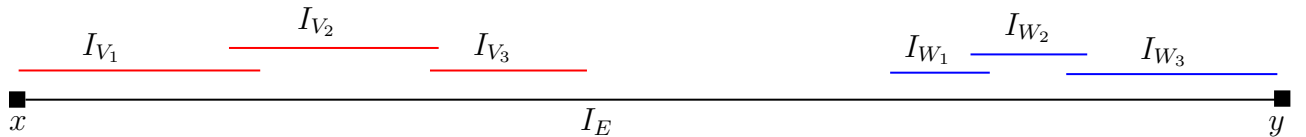


FIGURE 5. Examples of  $V_i \swarrow E$  and  $E \searrow W_i$ .

In this example, all of the witnesses  $V_i$  satisfy  $V_i \swarrow E$ , and all of the witnesses  $W_i$  satisfy  $E \searrow W_i$ , but  $d_E(\mathcal{C}(V_i), x)$  need not be less than  $9\mathbf{C}$  for  $i = 2$  or  $i = 3$ , and similarly,  $d_E(\mathcal{C}(W_i), y)$  need not be less than  $9\mathbf{C}$  for  $i = 2$  or  $i = 3$ .

In fact, the above example illustrates that for witnesses  $V$  that are not within distance  $9\mathbf{C}$  from one of the endpoints, the choice between assigning  $V \swarrow E$  and  $E \searrow V$  is ambiguous, which is what the property of being *wide* tries to fix. The property of being *wide* also tells when an active subsurface  $V$  nested in a witness  $E$  contributes to  $E$ : this happens when  $V$  appears in the “middle” of the segment  $[x, y]$ .

**Definition 6.10** (Wide witness families). An insulated complete subordered witness family is wide if for each  $V$  in the witness family, both of the following quantities are at most  $\frac{N_V}{3}$ .

- For  $W$  a witness such that  $W \swarrow V$ , the quantity  $\text{diam}_V(x, \mathcal{C}(W))$ .
- For  $W$  a witness such that  $V \searrow W$ , the quantity  $\text{diam}_V(y, \mathcal{C}(W))$ .

The idea behind a wide witness family is to create a buffer zone of length at least  $\frac{N_V}{3}$  in the middle of the projection of the geodesic to  $\mathcal{C}(V)$  for any witness  $V$  such that:

- If any subsurface  $W$  is active to the left of the buffer zone, it contributes to a subsurface  $Z$  such that  $Z \swarrow V$ .
- If any subsurface  $W$  is active to the right of the buffer zone, it contributes to a subsurface  $Z$  such that  $V \searrow Z$ .
- If a subsurface  $W$  is active within the buffer zone, it contributes to  $V$ .

The upshot of defining wide, insulated, subordered, and complete witness families (which will be abbreviated to WISC witness families) is that it gives us a better idea the order in which progress is made in various active subsurfaces. If we were working with just the collection of active subsurfaces, the only time we can tell if a geodesic makes progress in a subsurface  $V$  followed by the subsurface  $W$  is when  $V \cap W$ . When working with a WISC witness family, we can do that, but we can also make similar statements about pairs of nested witnesses  $V \sqsubset W$ , namely we can have either  $V \swarrow W$  or  $W \searrow V$ .

The following lemma asserts that WISC witness families exist, and their cardinality can be uniformly bounded.

**Lemma 6.11** (Lemmas 7.29 and 7.30 from [DM23]). *Let  $\mathcal{S}$  be surface, and  $[x, y]$  a geodesic segment in  $\mathcal{T}(\mathcal{S})$ . Then there exists a WISC witness family  $\Omega(x, y)$  for  $[x, y]$ . Furthermore, the cardinality of  $\Omega(x, y)$  depends only on  $\mathcal{S}$ , and not the points  $x$  and  $y$ .*

*Remark.* While the statements of Lemmas 7.29 and 7.30 in Dowdall and Masur [DM23] are for orientable surfaces, they go through without any changes for non-orientable surfaces as well.

We now get to the *raison d'être* of witness families: turning points on the geodesic segment  $[x, y]$  in  $\mathcal{T}(\mathcal{S})$  into points in  $\mathcal{T}(V)$ , where  $V$  is a witness in  $\Omega(x, y)$ . We will do so by assigning to each point  $w$  in a neighbourhood of  $[x, y]$  a point  $\tilde{w}_Z$  in  $\mathcal{C}(Z)$  for all subsurfaces  $Z$  contained in  $V$ , and then showing this assignment is consistent. Then the realization theorem (Theorem A.9) will give us a point  $\hat{w}_V^\Omega$  in  $\mathcal{T}(V)$  which has the same projections in  $\mathcal{C}(Z)$  as the original point  $W$ .

**Definition 6.12** (Projection tuple). Let  $\Omega(x, y)$  be a WISC witness family for a Teichmüller geodesic  $[x, y]$  in  $\mathcal{T}(\mathcal{S})$ . Let  $w$  be a point in  $\mathcal{T}(\mathcal{S})$  satisfying the following bound for every subsurface  $V$ .

$$d_V(x, w) + d_V(w, y) \leq d_V(x, y) + 9\mathbf{C}$$

Then for any  $U \in \Omega(x, y)$ , the projection tuple  $\tilde{w}$  of  $w$  is the point in  $\prod_{Z \sqsubseteq U} \mathcal{C}(Z)$  given by the following formula (where  $\pi_Z$  is the usual projection map from  $\mathcal{T}(\mathcal{S})$  to  $\mathcal{C}(Z)$ ).

$$\tilde{w}_Z = \begin{cases} \pi_Z(y), & \text{if } Z \in \Upsilon(x, y) \text{ and } \overline{Z}^\Omega \not\searrow U \\ \pi_Z(x), & \text{if } Z \in \Upsilon(x, y) \text{ and } U \searrow \overline{Z}^\Omega \\ \pi_Z(w), & \text{otherwise} \end{cases}$$

Observe that this is different from the usual projection map from  $\mathcal{T}(\mathcal{S})$  to  $\mathcal{C}(Z)$ : for subsurfaces  $Z$  that contribute to a witness nested in  $U$ , and consequently,  $\overline{Z}^\Omega \not\searrow U$  or  $U \searrow \overline{Z}^\Omega$ , we change the projection from  $\pi_Z(w)$  to  $\pi_Z(y)$  or  $\pi_Z(x)$  respectively.

This new projection map, despite being a modification of the usual projection map, is still consistent.

**Proposition 6.13** (Proposition 8.4 of [DM23]). *The projection tuple  $\tilde{w}_Z$  is  $k$ -consistent for some  $k$  depending only on  $\mathbf{C}$ .*

Using the above proposition, and the realization theorem for non-orientable surfaces (Theorem A.9), we can turn a projection tuple into a point in  $\mathcal{T}(U)$ , which Dowdall and Masur [DM23] refer to as *resolving a point  $w$  in  $\mathcal{T}(U)$* .

**Definition 6.14** (Resolution point). Let  $[x, y]$  be a geodesic segment in  $\mathcal{T}(\mathcal{S})$ , and  $\Omega(x, y)$  an associated WISC witness family. For  $w \in \{x, y\}$ , and  $U \in \Omega(x, y)$ , we define  $\hat{w}_U^\Omega$  as follows.

- If  $U$  is non-annular, then  $\hat{w}_U^\Omega \in \mathcal{T}(U)$  is the thick point whose projections to  $\mathcal{C}(V)$  for  $V \sqsubseteq U$  are coarsely equal to the projection tuple  $\hat{w}_U$  (which exists due to the Realization theorem (Theorem A.9)).
- If  $U$  is annular, then  $\tilde{w}_U$  is an element of  $\mathbb{Z}$ , and we set  $\hat{w}_U^\Omega$  to be the point in  $\mathbb{H}$  whose twist coordinate is  $\tilde{w}_U$ , and whose length coordinate is  $\frac{1}{\min(\varepsilon_t, \ell_{\partial U}(w))}$ .

We can now define the complexity length associated to a witness family  $\Omega$ .

**Definition 6.15** (Complexity of witness family). Let  $[x, y]$  be a geodesic segment in  $\mathcal{T}(\mathcal{S})$ , and  $\Omega$  an associated WISC witness family. The complexity  $\mathfrak{L}_\Omega(x, y)$  of  $\Omega$  is the following quantity.

$$\mathfrak{L}_\Omega(x, y) := \sum_{U \in \Omega} h_{\text{NP}}^*(U) \cdot d_{\mathcal{T}(U)}(\hat{x}_U^\Omega, \hat{y}_U^\Omega)$$

Here,  $h_{\text{NP}}^*(U)$  is the net point growth entropy for  $\mathcal{T}(U)$  when  $U$  is non-annular, and when  $U$  is annular,  $h_{\text{NP}}^*(U)$  is 1 when both  $\hat{x}_U^\Omega$  and  $\hat{y}_U^\Omega$  are  $\varepsilon_t$ -thick, and 2 if not.

We now address why we used a modified version of the projection map in Definition 6.12 instead of the usual projection map to curve complexes, by revisiting Example 6.6.

**Example 6.16.** Let  $\gamma$  be a pseudo-Anosov mapping class on  $\mathcal{S}$ , with large translation distance on  $\mathcal{C}(\mathcal{S})$  and small projections elsewhere. Let  $\delta$  be a reducible mapping class, which is pseudo-Anosov on a subsurface  $V$ , with large translation distance on  $\mathcal{C}(V)$  and small translation distance everywhere. Let  $x$  be a point in  $\mathcal{T}(\mathcal{S})$  such that  $\partial V$  is a component of the short marking on  $x$ , and  $y = \delta\gamma\delta^{-1}x$ . We first verify that  $\Upsilon(x, y) = \{\mathcal{S}, V, \delta\gamma V\}$ . To see this, we consider the following tuple of points  $(x, \delta x, \delta\gamma x, \delta\gamma\delta^{-1}x)$ . We claim that each

of the points in the tuple lies coarsely on the geodesic  $[x, y]$ . To see this, we compute the projections of adjacent pairs of points in tuple to various curve complexes.

- $[x, \delta x]$  has large projections on  $\mathcal{C}(V)$  and small projections on other curve complexes.
- $[\delta x, \delta \gamma x]$  has large projections on the  $\mathcal{C}(\delta \mathcal{S})$ , which is the same as  $\mathcal{C}(\mathcal{S})$ , and small projections elsewhere.
- $[\delta \gamma x, \delta \gamma \delta^{-1} x]$  has large projections on  $\mathcal{C}(\delta \gamma V)$ , and small projections elsewhere.

Consider the path  $\kappa$  obtained by concatenating Teichmüller geodesics between  $x$  and  $\delta x$ ,  $\delta x$  and  $\delta \gamma x$ , and  $\delta \gamma x$  and  $\delta \gamma \delta^{-1} x$ . The curve complex calculations above show that the projection of this path to any curve complex is a quasi-geodesic. That indicates this path is a *hierarchy path*, in the language of hierarchically hyperbolic spaces. Furthermore,  $\kappa$  does not have large projections on disjoint subsurfaces: this indicates the hierarchy path  $\kappa$  follow travels the geodesic  $[x, y]$ .

Let  $\Omega(x, y) = \Upsilon(x, y) = \{\mathcal{S}, V, \delta \gamma V\}$ . One can verify that this is a WISC witness family for  $[x, y]$ . Furthermore, we have that  $V \swarrow \mathcal{S}$  and  $\mathcal{S} \searrow \gamma V$ .

We now compute the resolution of the points  $x$  and  $y$  in  $\mathcal{T}(V)$ ,  $\mathcal{T}(\gamma V)$ , and  $\mathcal{T}(\mathcal{S})$ . Observe that the geodesic  $[x, y]$  almost immediately moves into a product region associated to  $V$  at the beginning, leaves that product region at some point  $w$  along the geodesic, and then enters the product region associated to  $\gamma V$  at some point  $z$ , and then stays in that product region almost all the way up to the end. We will abuse notation slightly, and refer to  $x$  and  $w$  as points in  $\mathcal{T}(V)$ , when we mean their projection via the product region map, and  $z$  and  $y$  will refer to points in  $\mathcal{T}(\gamma V)$ . Resolving points in  $\mathcal{T}(V)$  and  $\mathcal{T}(\gamma V)$  is easy, since there's no other witnesses nested in them, which means the projection tuple for those subsurfaces is the usual projection map.

$$\begin{aligned}\widehat{x}_V^\Omega &= x \\ \widehat{y}_V^\Omega &= w \\ \widehat{x}_{\gamma V}^\Omega &= z \\ \widehat{y}_{\gamma V}^\Omega &= y\end{aligned}$$

To resolve points in  $\mathcal{T}(\mathcal{S})$ , we have to use our modified projection map, instead of the usual one. Doing so, the points  $x$  and  $y$  resolve in the following manner.

$$\begin{aligned}\widehat{x}_\mathcal{S}^\Omega &= w \\ \widehat{y}_\mathcal{S}^\Omega &= z\end{aligned}$$

With these resolutions, we get the following estimate for complexity in terms of Teichmüller distance.

$$\begin{aligned}\mathfrak{L}_\Omega(x, y) &= h_{\text{NP}}^*(V) \cdot d_{\mathcal{T}(V)}(x, w) + h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(w, z) + h_{\text{NP}}^*(\gamma V) \cdot d_{\mathcal{T}(\gamma V)}(z, y) \\ &< h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)\end{aligned}$$

Compare this to the complexity  $\mathfrak{L}'_\Omega$  we would have gotten if we used the usual projection map instead of the modified projection map.

$$\begin{aligned}\mathfrak{L}'_\Omega(x, y) &= h_{\text{NP}}^*(V) \cdot d_{\mathcal{T}(V)}(x, w) + h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y) + h_{\text{NP}}^*(\gamma V) \cdot d_{\mathcal{T}(\gamma V)}(z, y) \\ &> h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)\end{aligned}$$

If we use Rafi's distance formula to estimate the distance between the resolution points  $\hat{x}_V^\Omega$  and  $\hat{x}_V^\Omega$  we do not get very good bounds for  $d_{\mathcal{T}(V)}(\hat{x}_V^\Omega, \hat{x}_V^\Omega)$ : at best, we accrue multiplicative and additive errors. For our applications however, the most we can tolerate is additive error. To do this, we will need to refine to notion of active interval for a subsurface to something more useful for the estimate: the *contribution set*  $\mathcal{A}_V^\Omega$  of a witness  $V$ . We first define two intermediate collections of subintervals of a geodesic segment  $[x, y]$ .

$$M(V) := \bigcup \{ \mathcal{I}_W^{\varepsilon_t} \mid W \in \Omega \text{ with } W \sqsubset V \}$$

$$C(V) := \bigcup \{ \mathcal{I}_Z^{\varepsilon_t} \mid Z \text{ contributes to } V \}$$

**Definition 6.17** (Contribution set). For a witness  $V \in \Omega$ , the contribution set  $\mathcal{A}_V^\Omega$  is a subset of the geodesic segment  $[x, y]$  defined in the following manner.

$$\mathcal{A}_V^\Omega := (\mathcal{I}_V^{\varepsilon_t} \setminus M(V)) \cup C(V)$$

*Remark.* The reason we remove  $M(V)$  and then later add  $C(V)$  again is because it is possible for several different subsurfaces to be active at the same time: orthogonal subsurfaces for instance. One can have  $V$  and  $W$  as witnesses, with  $W \sqsubset V$ , and  $Z \sqsubset V$  an active subsurface but not a witness, such that  $W \perp Z$  with  $\mathcal{I}_W^{\varepsilon_t}$  and  $\mathcal{I}_Z^{\varepsilon_t}$  overlapping. In that case, removing  $M(V)$  would also remove part of  $\mathcal{I}_Z^{\varepsilon_t}$ , and adding back  $C(V)$  would add back the deleted portion.

The following theorem estimates  $d_{\mathcal{T}(V)}(\hat{x}_V^\Omega, \hat{y}_V^\Omega)$  using  $\mathcal{A}_V^\Omega$ .

**Theorem 6.18** (Theorem 9.4 of [DM23]). *There exists a uniform constant  $C$  such that the following bound holds for any  $x, y$ , and witness  $V$ .*

$$d_{\mathcal{T}(V)}(\hat{x}_V^\Omega, \hat{y}_V^\Omega) \leq \int_x^y \mathbb{1}_{\mathcal{A}_V^\Omega} + C$$

Contribution sets help us make precise the notion of “overcounting” when multiple product regions are active at the same time. More precisely, when a segment of  $[x, y]$  is a part of two or more contribution sets, that segment shows up multiple times when computing  $\mathfrak{L}_\Omega(x, y)$ , thanks to Theorem 6.18. If the overlapping segment is sufficiently long, one could even end up having  $\mathfrak{L}_\Omega(x, y) > h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)$ , which as we will see, leads to a worse count for net points than the usual methods. This phenomenon of contribution sets overlapping is called *badness*, and while we will not be able to eliminate it entirely, we will be able to minimize it.

**Definition 6.19** (Bad set). We say a point  $p$  in  $\mathcal{A}_V^\Omega$  is bad if there exists some other witness  $W$  such that  $p$  also belongs in  $\mathcal{A}_W^\Omega$ . The bad set  $\mathcal{B}_V^\Omega$  denotes the set of all bad points in  $\mathcal{A}_V^\Omega$ , and  $|\mathcal{B}_V^\Omega|$  denotes the total length of this set, when we think of  $\mathcal{B}_V^\Omega$  as a subset of the geodesic segment  $[x, y]$ .

For our applications, we won't need to eliminate badness entirely, or even bound the length of the bad set uniformly: it will suffice to show that the length of the bad set is a very small multiple of  $d_{\mathcal{T}(\mathcal{S})}(x, y)$ .

**Definition 6.20** (Admissible and limited). A witness family  $\Omega$  associated to a geodesic segment  $[x, y]$  is said to be:



- *admissible* if  $|\mathcal{B}_V^\Omega| \leq \frac{d_{\mathcal{T}(\mathcal{S})}(x, y)}{K_V \mathbf{C}}$ , for all  $V \in \Omega$ , and some constants  $K_V$  that only depend on the topological type of  $V$ .
- *limited* if  $|\Omega|$  is uniformly bounded, independent of  $x$  and  $y$ .

*Remark.* Our definition of limited is a weaker version of Definition 10.7 from Dowdall and Masur [DM23], but since we don't need the stronger version, we present this version instead.

Dowdall and Masur [DM23] prove that WISC witness families that are admissible and limited exist. They call these witness families WISCAL witness families.

**Proposition 6.21** (Section 10.3 of [DM23]). *For all  $[x, y]$ , there exists an associated WISC witness family that is also admissible and limited.*

For WISCAL witness families, the following result relating complexity and Teichmüller distance follows easily from Theorem 6.18 and the definition of admissible.

**Proposition 6.22.** *If  $\Omega$  is a WISCAL witness family associated to  $[x, y]$ , then the following inequality holds.*

$$\mathfrak{L}_\Omega(x, y) \leq \left( h_{\text{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}} \right) d_{\mathcal{T}(\mathcal{S})}(x, y) + K\mathbf{C}$$

Here,  $K$  is some uniform constant depending only on  $\mathcal{S}$ .

We now define complexity length, which follows from the definition of the complexity of a witness family.

**Definition 6.23** (Complexity length). For a pair of points  $x$  and  $y$  in  $\mathcal{T}(\mathcal{S})$ , the complexity length  $\mathfrak{L}(x, y)$  is defined to be the following.

$$\mathfrak{L}(x, y) := \inf_{\Omega} \mathfrak{L}_\Omega(x, y)$$

Here, we take the infimum over all WISCAL witness families for  $[x, y]$ .

With the machinery of complexity length set up, it is now possible to count net points with respect to complexity length.

**Theorem 6.24** (Theorem 12.1 of [DM23]). *For any large enough  $\mathbf{C} > 0$ , and any  $\varepsilon_{\text{err}} > 0$ , there exists a polynomial function  $p(r)$ , and  $r > 0$  large enough such that the following bound holds for net points in  $\mathcal{T}(\mathcal{S})$ .*

$$\#(y \in \mathfrak{N} \mid \mathfrak{L}(x, y) \leq r) \leq p(r) \cdot \exp((1 + \varepsilon_{\text{err}})r)$$

*Remark.* The above theorem is a weaker version of the theorem that appears in Dowdall and Masur [DM23]: their version does not have the  $\varepsilon_{\text{err}}$ . The reason we have the weaker version is that in the proof of their theorem, they count the number of net points in  $\mathcal{T}(V)$  in a ball of radius  $R$ , where  $V$  is a witness, using Theorem 6.1, which gives them that the number of net points is equal, up to multiplicative error, to  $\exp(h_{\text{NP}}(U)R)$ . Since Theorem 6.1 only holds orientable surfaces, and we want to state our results for non-orientable surfaces as well, we will need to use a weaker counting result to count net points in  $\mathcal{T}(V)$ , namely the following bound, which holds for any  $\varepsilon_{\text{ent}} > 0$  and large enough  $R$ .

$$\#(y \in \mathfrak{N} \mid d_{\mathcal{T}(V)}(x, y) \leq R) \leq \exp((h_{\text{NP}}(V) + \varepsilon_{\text{ent}})R)$$

We sketch out a proof of Theorem 6.24 below: the proof proceeds identically to the proof in Dowdall and Masur [DM23], except at one point, where we plug in our weaker bound for net points in  $\mathcal{T}(V)$  for witnesses  $V$ .

*Sketch of proof for Theorem 6.24.* For each  $y \in \mathfrak{N}$  such that  $\mathfrak{L}(x, y) \leq r$ , we have a WISCAL witness family  $\Omega$  such that  $\mathfrak{L}_\Omega(x, y) \leq r$ . We can turn that witness family into a graph in the following manner.

- Add a vertex for every witness  $V \in \Omega$ .
- Label the vertex associated with  $V$  with the tuple  $(h_{\text{NP}}^*(V), \lfloor d_{\mathcal{T}(V)}(\hat{x}_V^\Omega, \hat{y}_V^\Omega) \rfloor)$ .
- If we have a pair of witnesses  $V \swarrow W$ , we join the vertices associated to them with a directed edge labeled “SW”:  $V \xrightarrow{\text{SW}} W$ .
- If we have a pair of witnesses  $W \searrow V$ , we join the vertices associated to them with a directed edge labeled “SE”:  $W \xrightarrow{\text{SE}} V$ .
- If we have a pair of witnesses  $W \pitchfork V$ , with  $W \triangleleft V$ , we join the vertices associated to them with a directed edge labelled “P”:  $W \xrightarrow{P} V$ .

We first count how many distinct possibilities are there for such labeled graphs that correspond to  $y$  for which  $\mathfrak{L}(x, y) \leq r$ . Since the cardinality of a WISCAL family is uniformly bounded, there are at most  $k$  many vertices, for some constant  $k$ . As for the labels on the vertices, there are at most  $\frac{r}{h_{\text{NP}}^*(V)}$  possibilities for a label on vertex which corresponds to a subsurface which is homeomorphic to  $V$ . From this, we conclude that there are at most  $p(r)$  possibilities for the combinatorial type of the graph, where  $p(r)$  is a polynomial in  $r$ .

It will suffice to compute how many distinct net points give rise to witness families whose graph is of a given type. To do so, we consider *initial subsets* of the graph, i.e. a subset  $\mathcal{W}$  of the vertices  $\mathcal{V}$  of the graph such that there is no directed edge from  $\mathcal{V} \setminus \mathcal{W}$  to  $\mathcal{W}$ .

Given an initial subset  $\mathcal{W}$  of the graph, we construct points  $y$  such that the witness family associated to  $[x, y]$  has the combinatorial type  $\mathcal{W}$ . We then consider an enlargement of  $\mathcal{W}$  by one-additional vertex  $v$ , such that the enlargement is still an initial subset.

*Claim.* The entire graph  $\mathcal{V}$  can be built up from such one-step enlargements.

We then count the number of net points whose associated witness families have the combinatorial type  $\mathcal{W} \cup \{v\}$ , after we fix one witness family associated to  $\mathcal{W}$ . More concretely, let  $w$  be a point such that the combinatorial type of the witness family associated to  $[x, w]$  is  $\mathcal{W}$ . Suppose now that we add a vertex  $(h, r_0)$  to the graph  $\mathcal{W}$ . To extend  $\Omega(x, w)$  so that its combinatorial type is  $\mathcal{W} \cup \{(h, r_0)\}$ , we need to add a witness  $U$  whose net point entropy is  $h$ , and a point  $y \in \mathcal{T}(U)$  such that the following holds.

$$d_{\mathcal{T}(U)}(\hat{x}_U^\Omega, y) \leq r_0$$

There are only finitely many choices for such subsurfaces  $U$  (because their boundary curves must get short near  $w$ ), and once we’ve made a choice of  $U$ , we have a choice  $\exp((h_{\text{NP}}(V) + \varepsilon_{\text{ent}})r_0)$  points for  $y$ .

Multiplying out the counts for each vertex added, we get the following estimate for the cardinality associated to each combinatorial type.

$$\begin{aligned} \#(y \mid \Omega(x, y) \text{ has combinatorial type } \mathcal{V}) &= \sum_{(h, s) \in \mathcal{V}} \exp((h + \varepsilon_{\text{ent}})s) \\ &\leq \exp((1 + \varepsilon_{\text{err}})r) \end{aligned}$$

We get the second inequality by picking  $\varepsilon_{\text{ent}}$  small enough, and observing that  $\sum hs \leq r$ .  $\square$

**6.3. Linear gap for bad points.** In this subsection, we will prove our main result involving complexity length: on the complexity length of bad points.

**Theorem 6.25.** *Suppose that for all proper subsurfaces  $V$  of  $\mathcal{S}$ , the following inequality holds.*

$$h_{\text{NP}}(V) < h_{\text{NP}}(\mathcal{S})$$

*Then for any  $\varepsilon_b > 0$ , there exists  $c > 0$ , and  $R$  large enough, such that for any bad point  $y$ , i.e. a point in  $\mathfrak{N}_b(p, R, \varepsilon_b)$ , the following upper bound holds for the complexity length between  $p$  and  $y$ .*

$$\mathfrak{L}(p, y) \leq h_{\text{NP}}(\mathcal{S})(1 - c)R$$

*Proof.* Let  $\Omega$  be a WISCAL witness family for  $[p, y]$ : the proof of Theorem 6.25 splits into two cases depending on whether the surface  $\mathcal{S}$  is a witness in  $\Omega$  or not.

The case where  $\mathcal{S}$  is a witness is harder, so we deal with that first.

We consider the triple of points  $(p, \hat{y}_S^\Omega, y)$ , and first estimate  $\mathfrak{L}(p, \hat{y}_S^\Omega)$ .

By applying Proposition 6.22, we get a bound for  $\mathfrak{L}(p, \hat{y}_S^\Omega)$ .

$$\mathfrak{L}(p, \hat{y}_S^\Omega) \leq \left( h_{\text{NP}}(\mathcal{S}) + \frac{K}{C} \right) (d_{\mathcal{T}(\mathcal{S})}(p, \hat{y}_S^\Omega)) + KC \quad (37)$$

We next estimate  $\mathfrak{L}(\hat{y}_S^\Omega, y)$ : we claim that there exists a WISCAL witness family  $\Omega'$  for  $[\hat{y}_S^\Omega, y]$  that does not have  $\mathcal{S}$  as a witness. The first thing we need in order to get such a witness family is verify that on the Teichmüller geodesic  $[\hat{y}_S^\Omega, y]$ ,  $\mathcal{S}$  is not an active subsurface. This follows from the fact that the  $\mathcal{S}$ -coordinate in the projection tuple of  $y$  is equal to the  $\mathcal{S}$ -coordinate of the projection of  $\hat{y}_S^\Omega$  to  $\mathcal{C}(\mathcal{S})$ , by construction of  $\hat{y}_S^\Omega$ . We now use Proposition 6.21 to get a witness family which does not contain  $\mathcal{S}$ , since  $\mathcal{S}$  is not an active subsurface.

Since the witness family  $\Omega'$  does not have  $\mathcal{S}$  as a witness, we can do better than Proposition 6.22 when estimating  $\mathfrak{L}(\hat{y}_S^\Omega, y)$ . We have from our hypothesis that  $h_{\text{NP}}(\mathcal{S}) > h_{\text{NP}}(V)$ , so there exists a constant  $h$  such that  $h < h_{\text{NP}}(\mathcal{S})$  but  $h > h_{\text{NP}}(V)$ . Using Theorem 6.18, we get the following estimate for  $\mathfrak{L}(\hat{y}_S^\Omega, y)$ .

$$\mathfrak{L}(\hat{y}_S^\Omega, y) \leq \left( h + \frac{K}{C} \right) d_{\mathcal{T}(\mathcal{S})}(\hat{y}_S^\Omega, y) + KC \quad (38)$$

From the triangle inequality for complexity length, we also have the following

$$\mathfrak{L}(p, y) \leq \mathfrak{L}(p, \hat{y}_S^\Omega) + \mathfrak{L}(\hat{y}_S^\Omega, y) \quad (39)$$

We plug in (37) and (38) into (39).

$$\begin{aligned} \mathfrak{L}(p, y) &\leq \left( h_{\text{NP}}(\mathcal{S}) + \frac{K}{C} \right) (d_{\mathcal{T}(\mathcal{S})}(p, \hat{y}_S^\Omega)) + \left( h + \frac{K}{C} \right) d_{\mathcal{T}(\mathcal{S})}(\hat{y}_S^\Omega, y) + 2KC \\ &= \left( h_{\text{NP}}(\mathcal{S}) + \frac{K}{C} \right) (d_{\mathcal{T}(\mathcal{S})}(p, \hat{y}_S^\Omega) + d_{\mathcal{T}(\mathcal{S})}(\hat{y}_S^\Omega, y)) \\ &\quad - (h_{\text{NP}}(\mathcal{S}) - h) d_{\mathcal{T}(\mathcal{S})}(\hat{y}_S^\Omega, y) \\ &\quad + 2KC \end{aligned}$$

We now claim that  $\hat{y}_S^\Omega$  is within a bounded distance of a point  $q$  on  $[p, y]$ . Consider the triple  $(x, \hat{y}_S^\Omega, y)$ : this is a strongly aligned tuple, i.e. its projections onto all curve complexes

satisfy a coarse reverse triangle inequality (see Definition 3.21 of [DM23]). Lemma 9.10 of [DM23] asserts the existence of  $q$  on  $[x, y]$  that satisfies the following properties.

- For any subsurface  $V$  such that  $\bar{V}^\Omega = \mathcal{S}$ ,  $d_V(\hat{y}_\mathcal{S}^\Omega, q) \leq M$  for some fixed constant  $M$ .
- For any subsurfaces  $V$  such that  $\bar{V}^\Omega \not\prec \mathcal{S}$  or  $\mathcal{S} \searrow \bar{V}^\Omega$ ,  $q$  lies outside the active interval for  $V$ .

In the case where  $\bar{V}^\Omega \not\prec \mathcal{S}$ , we have that  $d_V(q, y) \leq M$ , since  $q$  lies outside the active interval for  $V$ . But note that by construction of the projection tuple,  $\pi_V(\hat{y}_\mathcal{S}^\Omega) = y$ , thus  $d_V(\hat{y}_\mathcal{S}^\Omega, q) \leq M$ . Similarly, we have  $d_V(\hat{y}_\mathcal{S}^\Omega, q) \leq M$  for  $\mathcal{S} \searrow \bar{V}^\Omega$  as well. Thus, by Rafi's distance formula, we have that  $\hat{y}_\mathcal{S}^\Omega$  is within a bounded distance of  $q$  on  $[p, y]$ : we have  $d_{\mathcal{T}(\mathcal{S})}(p, \hat{y}_\mathcal{S}^\Omega) + d_{\mathcal{T}(\mathcal{S})}(\hat{y}_\mathcal{S}^\Omega, y) \leq R + J$ , for some constant  $J$ . Furthermore,  $d_{\mathcal{T}(\mathcal{S})}(\hat{y}_\mathcal{S}^\Omega, y) \geq \varepsilon_b R$ , by the hypothesis of  $y$  being a bad point, since  $\hat{y}_\mathcal{S}^\Omega$  is in the thick part of Teichmüller space. This simplifies the expression for  $\mathfrak{L}(x, y)$ .

$$\begin{aligned} \mathfrak{L}(x, y) &\leq \left( h_{\text{NP}}(\mathcal{S}) + \frac{K}{C} \right) R - (h_{\text{NP}}(\mathcal{S}) - h)(\varepsilon_b)R + 2KC \\ &= R \left( h_{\text{NP}}(\mathcal{S}) - d + \frac{K}{C} \right) + 2KC \end{aligned}$$

Here,  $d = \varepsilon_b(h_{\text{NP}}(\mathcal{S}) - h)$ , which is a positive constant, since  $h_{\text{NP}}(\mathcal{S}) > h$ . By picking  $C$  and  $R$  large enough, we get the statement of the theorem, which proves the result in the first case of  $\mathcal{S}$  being in the witness family.

When  $\mathcal{S}$  is not in the witness family, we set  $\hat{y}_\mathcal{S}^\Omega = p$ , and the rest of the proof follows identically.  $\square$

## APPENDIX A. GEOMETRY OF $\mathcal{T}(\mathcal{N}_g)$

In this section, we prove some standard results about the geometry of Teichmüller spaces of non-orientable surfaces that we use in Section 6. We do so by lifting the hyperbolic structures and markings on the non-orientable surfaces to their double covers, which give us points in the Teichmüller space and curve complex of the double cover.

The fact that these lifts are well-defined and respect the metric properties are encapsulated in the following two theorems.

**Theorem A.1** (Isometric embedding of Teichmüller spaces (Theorem 2.1 of [Kha23])). *The map  $i : \mathcal{T}(\mathcal{N}_g) \rightarrow \mathcal{T}(\mathcal{S}_{g-1})$  given by lifting the hyperbolic structure and marking from  $\mathcal{N}_g$  to  $\mathcal{S}_{g-1}$  is an isometric embedding. Furthermore, the image of  $\mathcal{T}(\mathcal{N}_g)$  in  $\mathcal{T}(\mathcal{S}_{g-1})$  is the subset of  $\mathcal{T}(\mathcal{S}_{g-1})$  is fixed by  $\iota^*$ , where  $\iota^*$  is the map induced by the orientation reversing deck transformation  $\iota$  on  $\mathcal{S}_{g-1}$ .*

**Theorem A.2** (Quasi isometric embedding of curve complexes (Lemma 6.3 from [MS13])). *The map  $\mathcal{C}(\mathcal{N}_g) \rightarrow \mathcal{C}(\mathcal{S}_{g-1})$  obtained by lifting curves in  $\mathcal{N}_g$  to  $\mathcal{S}_{g-1}$  is a quasi-isometric embedding.*

We will use the above two theorems, along with Lemma A.5, to reduce statements about the geometry of  $\mathcal{T}(\mathcal{N}_g)$  to statements about the geometry of  $\mathcal{T}(\mathcal{S}_{g-1})$ . However, we postpone the statement and the proof of Lemma A.5 until Section A.3, since it's not required for Section A.1.

We set up some notation for this section.

- $d(x, y)$  and  $d(\tilde{x}, \tilde{y})$ : Given points  $x$  and  $y$  in  $\mathcal{T}(\mathcal{N}_g)$ ,  $d(x, y)$  is the distance in Teichmüller metric between them, and  $d(\tilde{x}, \tilde{y})$  is the distance in  $\mathcal{T}(\mathcal{S}_{g-1})$  between their images,  $\tilde{x}$  and  $\tilde{y}$ .
- $\pi_V(\mu_x)$  and  $\pi_V(x)$ : If  $\mu_x$  is a marking/curve on a surface, the  $\pi_V(\mu_x)$  denotes the subsurface projection to the subsurface  $V$ . If  $x$  is a point in the Teichmüller space, the  $\pi_V(x) = \pi_V(\mu_x)$ , where  $\mu_x$  is the Bers marking on  $x$ .
- $d_V(\mu_x, \mu_y)$  and  $d_V(x, y)$ : If  $\mu_x$  and  $\mu_y$  are markings/curves on a surface, and  $V$  is a subsurface, then  $d_V(\mu_x, \mu_y)$  refers to the curve complex distance between the subsurface projections of  $\mu_x$  and  $\mu_y$  in  $\mathcal{C}(V)$ . When  $x$  and  $y$  are points in Teichmüller space,  $d_V(x, y)$  refers to  $d_V(\mu_x, \mu_y)$ , where  $\mu_x$  and  $\mu_y$  are the Bers marking on  $x$  and  $y$ .

**A.1. Minsky's product region theorem.** In this section, we prove a version of Minsky's product region theorem [Min96, Theorem 6.1] for non-orientable surfaces.

We recall the following objects that were defined in Section 3.3.

- (i) The multicurve  $\gamma$  on  $\mathcal{N}_g$ .
- (ii) The metric space  $X_\gamma$ , and the projection map  $\Pi$ .
- (iii) The thin region  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ .

**Theorem A.3** (Product region theorem for non-orientable surfaces). *For any  $c > 0$ , there exists a small enough  $\varepsilon_t > 0$ , such that the restriction of  $\Pi$  to  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$  is an isometry with additive error at most  $c$ , i.e. the following holds for any  $x$  and  $y$  in  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ .*

$$|d(x, y) - d_{X_\gamma}(\Pi(x), \Pi(y))| \leq c$$

*Proof.* We will prove this result by reducing the distance calculation in  $\mathcal{T}(\mathcal{N}_g)$  to a distance calculation in  $\mathcal{T}(\mathcal{S}_{g-1})$ , where  $\mathcal{S}_{g-1}$  is the orientation double cover, and invoking the classical product region theorem in that setting.

We begin the proof by constructing some points in  $\mathcal{T}(\mathcal{S}_{g-1})$  and a multicurve on  $\mathcal{S}_{g-1}$ . Recall that  $\mathcal{T}(\mathcal{N}_g)$  isometrically embeds inside  $\mathcal{T}(\mathcal{S}_{g-1})$ : let  $\tilde{x}$  and  $\tilde{y}$  denote the points in  $\mathcal{T}(\mathcal{S}_{g-1})$  that are the images of  $x$  and  $y$  under the embedding. Let  $\tilde{\gamma}$  denote the lift of the multicurve  $\gamma$ : if  $\gamma_i$  is a two-sided curve, it will have two disjoint lifts in the cover, and if  $\gamma_i$  is a one-sided curve, it will have single lift in the double cover. We have that the region  $\mathcal{T}_{\tilde{\gamma} \leq \varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$  intersects the image of  $\mathcal{T}(\mathcal{N}_g)$  at the image of  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g) \subset \mathcal{T}(\mathcal{N}_g)$ . Let  $\iota$  denote the orientation reversing deck transformation on  $\mathcal{S}_{g-1}$  which corresponds to the covering map.

*Claim.* Let  $\Pi_k$  denote the projection map from  $\mathcal{T}(\mathcal{N}_g)$  to the  $k^{\text{th}}$  component of  $X_\gamma$ , and  $\widetilde{\Pi}_k$  denote the projection map from  $\mathcal{T}(\mathcal{S}_{g-1})$  to the lift of the  $k^{\text{th}}$  component of  $\gamma$  to  $\mathcal{S}_{g-1}$ . This map is an isometric embedding.

$$d(\Pi_k(x), \Pi_k(y)) = d(\widetilde{\Pi}_k(\tilde{x}), \widetilde{\Pi}_k(\tilde{y}))$$

*Proof of claim.* We need to verify the claim on the three kinds of factors of  $X_\gamma$ .

- (i)  $\mathcal{N}_g \setminus \gamma$ : The lift of  $\mathcal{N}_g \setminus \gamma$  to  $\mathcal{S}_{g-1}$  will have two components if  $\mathcal{N}_g \setminus \gamma$  is orientable, which we call  $S_1$  and  $S_2$ . Both  $S_1$  and  $S_2$  are homeomorphic to  $\mathcal{N}_g \setminus \gamma$ . If  $\mathcal{N}_g \setminus \gamma$  is non-orientable, then its lift in  $\mathcal{S}_{g-1}$  is the orientation double cover.

In the first case,  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$  maps to the diagonal subspace in  $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$ , and the metric on  $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$  is the sup metric. The space  $\mathcal{T}(\mathcal{N}_g)$  maps to the diagonal subspace because its image must be invariant under the map  $\iota$ , which isometrically

swaps  $S_1$  and  $S_2$ . This map is an isometric embedding, and thus for any points  $x$  and  $y$  in  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ , the distance between their images in  $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$  is the same as the distance in  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ .

In the second case, we have that  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$  also isometrically embeds inside the Teichmüller space of its double cover, by Theorem A.1, so the claim follows.

- (ii)  $\gamma_i$  (for  $\gamma_i$  two-sided): The lift of  $\gamma_i$  in this case are two disjoint curves on  $\mathcal{S}_{g-1}$ , which are swapped by the deck transformation  $\iota$ . This means the  $\mathbb{H}$ -coordinate given by length and twist of  $\gamma_i$  maps to the diagonal in  $\mathbb{H} \times \mathbb{H}$ , which correspond the length and twist around the two lifts. Since  $\mathbb{H}$  mapped to the diagonal in  $\mathbb{H} \times \mathbb{H}$  is an isometric embedding with sup metric, the claim follows in this case.
- (iii)  $\gamma_i$  (for  $\gamma_i$  one-sided): The lift of  $\gamma_i$  in this case is a single curve  $\tilde{\gamma}_i$  on  $\mathcal{S}_{g-1}$  which is left invariant by the deck transformation  $\iota$ . We will show that the twist coordinate around  $\tilde{\gamma}_i$  cannot be changed without leaving the image of  $\mathcal{T}(\mathcal{N}_g)$  in  $\mathcal{T}(\mathcal{S}_{g-1})$ , i.e. any  $\tilde{x}$  and  $\tilde{y}$  have the same twist coordinate around  $\tilde{\gamma}_i$ . Once we have established that, the claim will follow, since only the length coordinate of  $\gamma_i$  can be changed, which corresponds to  $\mathbb{R}_{>0}$ .

Suppose now that  $x$  is a point in  $\mathcal{T}(\mathcal{N}_g)$  and  $\tilde{x}$  the corresponding point in  $\mathcal{T}(\mathcal{S}_{g-1})$ . Consider a pants decomposition on  $\mathcal{N}_g$  that contains  $\gamma_i$  as one of the curves. There is a unique one-sided curve  $\kappa$  that intersects  $\gamma_i$  and does not intersect any of the other pants curves. Let  $\tilde{\kappa}$  be the lift of  $\kappa$  to  $\mathcal{S}_{g-1}$ : we will use this curve to measure twisting around  $\tilde{\gamma}_i$ . Let  $x'$  be another point in  $\mathcal{T}(\mathcal{S}_{g-1})$  obtained by taking  $\tilde{x}$ , and twisting by some amount around  $\tilde{\gamma}_i$ , without changing the length of  $\tilde{\gamma}_i$ . On  $x'$ , the length of  $\tilde{\kappa}$  will be different from the length on  $\tilde{x}$ . However, this means that  $x'$  is not contained in the image of  $\mathcal{T}(\mathcal{N}_g)$ , since if it were, the length of  $\tilde{\kappa}$  would have to be the same, since that's the lift of the curve  $\kappa$ , whose length only depends on the length of  $\gamma_i$ .

□

The following equality follows from the claim.

$$d_{X_{\tilde{\gamma}}}(\Pi(\tilde{x}), \Pi(\tilde{y})) = d_{X_{\gamma}}(\Pi(x), \Pi(y))$$

We also have that  $\mathcal{T}(\mathcal{N}_g)$  isometrically embeds into  $\mathcal{T}(\mathcal{S}_{g-1})$ .

$$d(\tilde{x}, \tilde{y}) = d(x, y)$$

And finally, have that the region  $\mathcal{T}_{\tilde{\gamma} \leq \varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$  intersects the image of  $\mathcal{T}(\mathcal{N}_g)$  at the image of  $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g) \subset \mathcal{T}(\mathcal{N}_g)$ . Combining these three facts, and applying Minsky's product region theorem for orientable surfaces, the result follows. □

**A.2. Uniform bounds for the volume of a ball.** In this section, , we show that the for balls of fixed radius in  $\text{core}(\mathcal{T}(\mathcal{N}_g))$ , the  $\nu_N$  volume of the ball is bounded above and below by constants that are independent of the center of the ball.

Let  $\mathcal{P}$  be a pants decomposition for  $\mathcal{N}_g$ : recall the formula for  $\nu_N$ .

$$\nu_N = \left( \bigwedge_{\gamma_i \text{ one-sided}} \coth(\ell(\gamma_i)) d\ell(\gamma_i) \right) \wedge \left( \bigwedge_{\gamma_i \text{ two-sided}} d\tau(\gamma_i) \wedge d\ell(\gamma_i) \right)$$

Here  $\ell(\gamma_i)$  denotes the length of the curve  $\gamma_i$ , and  $\tau(\gamma_i)$  denotes the twist, when  $\gamma_i$  is two-sided.



**Proposition A.4.** *For any  $\kappa > 0$ , and  $\varepsilon_t > 0$  small enough, there exist positive constants<sup>2</sup>  $c_1$  and  $c_2$  (depending only on  $\kappa$  and  $\varepsilon_t$ ) such the  $\nu_N$  volume of a ball  $B_\kappa^{\varepsilon_t}(x)$  of radius  $\kappa$  centered at  $x \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  are bounded below and above by  $c_1$  and  $c_2$ .*

$$c_1 \leq \nu_N(B_\kappa^{\varepsilon_t}(x)) \leq c_2$$

*Proof.* Note that since the points we are considering lie in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ , we have the following upper bound and lower bound for  $\coth(\ell(\gamma_i))$ , where  $\gamma_i$  is a one sided curve.

$$1 \leq \coth(\ell(\gamma_i)) \leq \coth(\varepsilon_t) \quad (40)$$

In particular, the  $\nu_N$  volume of a ball can be bounded above and below by  $\coth(\varepsilon_t)\nu'_N$  and  $\nu'_N$ , where  $\nu'_N = \left(\bigwedge_{\gamma_i \text{ one-sided}} d\ell(\gamma_i)\right) \wedge \left(\bigwedge_{\gamma_i \text{ two-sided}} d\kappa(\gamma_i) \wedge d\ell(\gamma_i)\right)$ .

We now split up  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$  into two regions:  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ , and the complementary region. Since  $\text{MCG}(\mathcal{N}_g)$  acts cocompactly on  $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ , and  $\nu_N(B_\kappa^{\varepsilon_t}(x))$  is continuous in  $x$ , the desired bounds hold in this region. It will therefore suffice to prove the bounds in the complementary region.

Note that for any  $x$  in the complementary region, there is some two-sided curve  $\gamma$  that is short. By Theorem A.3, the ball  $B_\kappa^{\varepsilon_t}(x)$  is contained in a product of balls, one in  $\mathbb{H}$ , and one in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g \setminus \gamma)$ . We pick  $\gamma$  to be part of a pants decomposition  $\mathcal{P}$ , and write  $\nu_N$  as follows.

$$\nu_N = (d\kappa(\gamma) \wedge d\ell(\gamma)) \wedge \nu_N^{\mathcal{N}_g \setminus \gamma} \quad (41)$$

Here,  $\nu_N^{\mathcal{N}_g \setminus \gamma}$  denotes the volume form on  $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ . As a result, we have that the  $\nu_N$  measure of a product of the two balls is the product of the corresponding measures of those balls.

The measure of any ball of a fixed radius in  $\mathbb{H}$  is constant, since  $\mathbb{H}$  is homogeneous. The  $\nu_N^{\mathcal{N}_g \setminus \gamma}$  measure of a ball in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g \setminus \gamma)$  is again bounded above and below by fixed constants, by inducting on a surface of lower complexity.

Since we have uniform bounds for both the terms in the product, we get uniform bounds for the measure of a ball in  $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ .  $\square$

**A.3. Teichmüller geodesics and geodesics in the curve complex.** In this section, we will deduce some standard results about Teichmüller geodesics and the corresponding curve complex geodesics for non-orientable surfaces by reducing to the orientable case. The following lemma will be the main tool for the reduction to the orientable case.

**Lemma A.5.** *Let  $[x, y]$  be a Teichmüller geodesic segment in  $\mathcal{T}(\mathcal{N})$ , where  $\mathcal{N}$  is a non-orientable surface, and  $[\tilde{x}, \tilde{y}]$  be its image in  $\mathcal{T}(\mathcal{S})$ , where  $\mathcal{S}$  is the orientable double cover of  $\mathcal{N}$ . Let  $V$  be a subsurface of  $\mathcal{S}$ : then the following statements hold for  $d_V(\tilde{x}, \tilde{y})$ .*

- (i) *If  $V$  is the lift of an orientable subsurface  $W$  in  $\mathcal{N}$ , then  $d_V(\tilde{x}, \tilde{y}) = d_{\iota(V)}(\tilde{x}, \tilde{y}) = d_W(x, y)$ .*
- (ii) *If  $V$  is the lift of a non-orientable subsurface  $W$  in  $\mathcal{N}$ , then  $d_V(\tilde{x}, \tilde{y}) \dot{\asymp} d_W(x, y)$ .*
- (iii) *If  $V$  is not a lift of a subsurface in  $\mathcal{N}$ , then there exists a uniform constant  $k_0$ , independent of  $x, y$ , and  $V$ , such that  $d_V(\tilde{x}, \tilde{y}) \leq k_0$ .*

*Proof.* We deal with the proof in cases.

- (i) If  $V$  is the lift of an orientable surface, we have that the covering map restricted to  $V$  is a homeomorphism, and the same holds for  $\iota(V)$ , so the result follows in this case as well.

---

<sup>2</sup>We will also consider  $c_1$  and  $c_2$  as functions of  $\kappa$  elsewhere in the paper.

- (ii) If  $V$  is the lift of a non-orientable subsurface  $W$ , then by Theorem A.2, we have that  $\mathcal{C}(W)$  quasi-isometrically embeds into  $\mathcal{C}(V)$ , and the result follows.
- (iii) If  $V$  is not a lift at all, that means  $V$  and  $\iota(V)$  are transverse subsurfaces. By the Behrstock inequality, there exists a  $k_0$  such that the following holds.

$$\min(d_V(\tilde{x}, \partial\iota(V)), d_{\iota(V)}(\tilde{x}, \partial V)) \leq \frac{k_0}{2} \quad (42)$$

But we also have that  $\tilde{x}$  is fixed by  $\iota$ , which gives us the following equality of curve complex distances.

$$d_V(\tilde{x}, \partial\iota(V)) = d_{\iota(V)}(\tilde{x}, \partial V) \quad (43)$$

Combining (42) and (43), we get the following bound on the  $\mathcal{C}(V)$  distance between  $\tilde{x}$  and  $\partial\iota(V)$ .

$$d_V(\tilde{x}, \partial\iota(V)) \leq \frac{k_0}{2}$$

We have that the same bounds also hold for  $\tilde{y}$ , so the result follows from the above inequality and the triangle inequality.

This shows the result for all the cases and concludes the proof.  $\square$

We begin by proving the distance formula for points in Teichmüller space. Let  $x$  and  $y$  be a pair of points in  $\mathcal{T}(\mathcal{N}_g)$ , and let  $\Gamma$  be the set of curves that are short on both  $x$  and  $y$ ,  $\Gamma_x$  the set of curves that are only short on  $x$ , and  $\Gamma_y$  the set of curves that are only short on  $y$ . Let  $\mu_x$  and  $\mu_y$  be short markings on  $x$  and  $y$  respectively. Let  $\mathcal{C}^+$  and  $\mathcal{C}^-$  denote the set of two-sided and one-sided curves on  $\mathcal{N}_g$ . Finally, let  $[x]_k$  be the function which is 0 for  $x \leq k$ , and identity for  $x > k$ .

**Theorem A.6** (Distance formula). *The distance between  $x$  and  $y$  in  $\mathcal{T}(\mathcal{N}_g)$  is given by the following formula.*

$$\begin{aligned} d(x, y) &\doteq \sum_Y [d_Y(\mu_x, \mu_y)]_k + \sum_{\alpha \in \Gamma^c \cap \mathcal{C}^+} [\log(d_\alpha(\mu_x, \mu_y))]_k \\ &+ \max_{\alpha \in \Gamma \cap \mathcal{C}^+} d_{\mathbb{H}_\alpha}(x, y) + \max_{\alpha \in \Gamma \cap \mathcal{C}^-} d_{(\mathbb{R}_{>0})_\alpha}(x, y) \\ &+ \max_{\alpha \in \Gamma_x} \log \frac{1}{\ell_x(\alpha)} + \max_{\alpha \in \Gamma_y} \log \frac{1}{\ell_y(\alpha)} \end{aligned} \quad (44)$$

*Proof.* Let  $\tilde{x}$  and  $\tilde{y}$  be the images of  $x$  and  $y$  in  $\mathcal{T}(\mathcal{S}_{g-1})$  under the isometric embedding map. Since  $d(x, y) = d(\tilde{x}, \tilde{y})$ , it will suffice to estimate  $d(\tilde{x}, \tilde{y})$  using distances in the curve complexes. Let  $\tilde{\mu}_x$  and  $\tilde{\mu}_y$  be the lifts of  $\mu_x$  and  $\mu_y$ . Both  $\tilde{\mu}_x$  and  $\tilde{\mu}_y$  are short markings on  $\tilde{x}$  and  $\tilde{y}$  respectively. We have by Rafi's distance formula [Raf07, Theorem 6.1], the following estimate on  $d(\tilde{x}, \tilde{y})$ .

$$\begin{aligned} d(\tilde{x}, \tilde{y}) &\doteq \sum_Y [d_Y(\tilde{\mu}_x, \tilde{\mu}_y)]_k + \sum_{\alpha \in \tilde{\Gamma}^c} [\log(d_\alpha(\tilde{\mu}_x, \tilde{\mu}_y))]_k \\ &+ \max_{\alpha \in \tilde{\Gamma}} d_{\mathbb{H}_\alpha}(\tilde{x}, \tilde{y}) \\ &+ \max_{\alpha \in \tilde{\Gamma}_{\tilde{x}}} \log \frac{1}{\ell_{\tilde{x}}(\alpha)} + \max_{\alpha \in \tilde{\Gamma}_{\tilde{y}}} \log \frac{1}{\ell_{\tilde{y}}(\alpha)} \end{aligned} \quad (45)$$

Here,  $\tilde{\Gamma}$ ,  $\tilde{\Gamma}_{\tilde{x}}$ , and  $\tilde{\Gamma}_{\tilde{y}}$  are curves on  $\tilde{x}$  and  $\tilde{y}$  that are simultaneously short, short on  $\tilde{x}$  and not on  $\tilde{y}$ , and short on  $\tilde{y}$  and not on  $\tilde{x}$  respectively.

It will suffice to show that for a large enough choice of  $k$ , the right hand side of (44) is equal to the right hand side of (45), up to an additive and multiplicative constant. We consider the first term in the right hand side of (45), namely the sum over the non-annular subsurfaces  $Y$ . There are three possibilities for  $Y$  in  $\mathcal{S}_{g-1}$ , which we deal with using Lemma A.5.

- (i)  $Y$  is one component of a lift of an orientable subsurface  $Z$  of  $\mathcal{N}_g$ : In this case we have  $d_Y(\tilde{\mu}_x, \tilde{\mu}_y) = d_Z(\mu_x, \mu_y)$  (and the same equality with  $Y$  replaced with  $\iota(Y)$ ). Thus, for every term associated to an orientable non-annular subsurface  $Z$  in (44), we get two corresponding equal terms in (45).
- (ii)  $Y$  is the lift of a non-orientable subsurface  $Z$  of  $\mathcal{N}_g$ : In this case, we have  $d_Y(\tilde{\mu}_x, \tilde{\mu}_y) \asymp d_Z(\mu_x, \mu_y)$ .
- (iii)  $Y$  is not a lift of a subsurface of  $\mathcal{N}_g$ : In this case, we have the following for some  $k_0$ .

$$d_Y(\tilde{\mu}_x, \tilde{\mu}_y) \leq k_0$$

If we pick a threshold  $k > k_0$ , the subsurfaces  $Y$  that do not arise from lifts will not contribute to the right hand side of (45).

We now do the same case analysis for annular subsurfaces: consider a curve  $\alpha$  on  $\mathcal{S}_{g-1}$  that is contained in  $\Gamma^c$ , i.e. it is not simultaneously short on  $\tilde{x}$  and  $\tilde{y}$ . There are three possibilities for  $\alpha$ .

- (i)  $\alpha$  is one component of a lift of a two-sided curve  $\gamma$  on  $\mathcal{N}_g$ : In this case,  $\alpha$  and  $\iota(\alpha)$  are disjoint, and the restriction of the covering map to these curves is a homeomorphism. We have  $d_\alpha(\tilde{\mu}_x, \tilde{\mu}_y) = d_\gamma(\mu_x, \mu_y)$ : consequently, for every term in  $\Gamma^c \cap \mathcal{C}^+$  in (44), we have two equal terms in (45).
- (ii)  $\alpha$  is the lift of a one-sided curve on  $\mathcal{N}_g$ : In this case  $\alpha = \iota(\alpha)$ , but the transformation  $\iota$  reverses orientation on the surface  $\mathcal{S}_{g-1}$ . That means  $\tilde{\mu}_x$  and  $\tilde{\mu}_y$  cannot have a relative twist between them along  $\alpha$ , because if they did,  $\iota(\tilde{\mu}_x)$  and  $\iota(\tilde{\mu}_y)$  would have the opposite twist. On the other hand  $\tilde{\mu}_i = \iota(\tilde{\mu}_i)$  for  $i = x$  and  $i = y$ , which means the relative twist must be 0. This proves that the  $\alpha$  which are lifts of one-sided curves do not contribute to the second term of (45).
- (iii)  $\alpha$  is not a lift of a curve on  $\mathcal{N}_g$ : In this case  $\alpha$  and  $\iota(\alpha)$  intersect each other, and are not equal, which means they are transverse. We deal with this the same way we dealt with transverse non-annular subsurfaces, i.e. via the Behrstock inequality.

This case analysis proves that the second terms on the right hand side of (44) and (45) are equal, up to an additive and multiplicative constant.

We now deal with the last three terms of (45). These terms deal with short curves on  $x$  or  $y$ : we claim that the short curves must be lifts of either one-sided or two-sided curves in  $\mathcal{N}_g$ . Suppose a curve  $\alpha$  is short and not a lift. Then  $\alpha$  has positive intersection number with  $\iota(\alpha)$ , but since  $\iota$  is an isometry,  $\iota(\alpha)$  must also be short. For a sufficiently small threshold for what we call short, we can't have a short curve intersecting another short curve, which proves the claim that all the short curves arise as lifts.

Since the curves in  $\tilde{\Gamma}$  are all lifts, the third term of (45) can be split up into two terms: the lifts of the two-sided and one-sided curves. For the two-sided curves, the distance calculation

involves both the length and twist coordinate, and for the one-sided curves, only the length coordinate is involved. This follows from Theorem A.3.

Finally, the last two terms in (45) are the same as the last two terms of (44), up to an additive error of  $(6g) \cdot \log(2)$ , since the lift of a short curve can double its length, and there are no more than  $6g$  short curves.

We have shown that the right hand sides of (44) and (45) are equal, up to a multiplicative and additive constant, which proves the result.  $\square$

We now verify that Teichmüller geodesics can be broken up into *active intervals associated to subsurfaces*, which are subintervals of the geodesic associated to each subsurface  $V$ , along which the projection to  $V$  is large, and outside of which, the projection is bounded. The following lemma of Dowdall and Masur [DM23, Lemma 3.26] (which itself is a generalization of Rafi [Raf07, Proposition 3.7]) describes the subsegments of  $[x, y]$  along which the geodesic makes progress in the curve complex of a subsurface.

**Proposition A.7.** *For each sufficiently small  $\varepsilon_t > 0$ , there exists  $0 < \varepsilon_t' < \varepsilon_t$  and  $M_{\varepsilon_t} \geq 0$  such that for any subsurface  $V \sqsubset S$ , there's a (possibly empty) connected interval  $\mathcal{I}_V^{\varepsilon_t} \subset [x, y]$  such that the following five conditions hold.*

- (i) *If  $d_V(x, y) \geq M_{\varepsilon_t}$ , then  $\mathcal{I}_V^{\varepsilon_t}$  is a non-empty subinterval of  $[x, y]$ .*
- (ii)  *$\ell_\alpha(z) < \varepsilon_t$  for all  $z \in \mathcal{I}_V^{\varepsilon_t}$  and  $\alpha \in \partial V$ .*
- (iii) *For all  $z \in [x, y] \setminus \mathcal{I}_V^{\varepsilon_t}$ , some component  $\alpha$  of  $\partial V$  has  $\ell_\alpha(z) > \varepsilon_t'$ .*
- (iv)  *$d_V(w, z) \leq M_{\varepsilon_t}$  for every subinterval  $[w, z] \subset [x, y]$  if  $[w, z] \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$ .*
- (v) *For a pair of traverse subsurfaces  $U$  and  $V$ ,  $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$ .*

*Proof of Theorem A.7 for non-orientable surfaces.* Let  $\mathcal{N}$  be the non-orientable surface, and  $\mathcal{S}$  its double cover. We consider the image  $[\tilde{x}, \tilde{y}]$  of the geodesic  $[x, y]$  in  $\mathcal{T}(\mathcal{S})$ . We know that the result holds for  $[\tilde{x}, \tilde{y}]$ , although with  $\varepsilon_t'$  replaced with  $\frac{\varepsilon_t'}{2}$ , since lifting can double the lengths of some curves.

The main fact we need to verify is that the only subsurfaces  $V$  that have non-empty  $\mathcal{I}_V^{\varepsilon_t}$  come from lifts. If  $V$  is a subsurface of  $\mathcal{S}$  that is not a lift, we use case (iii) of Lemma A.5 to conclude that  $d_V(x, y) \leq k_0$  for some fixed constant  $k_0$ . Picking  $M_{\varepsilon_t} > k_0$  guarantees that the only subsurfaces for which  $\mathcal{I}_V^{\varepsilon_t}$  is non-empty arise from lifts, which proves the result for non-orientable surfaces.  $\square$

Finally, we show that the consistency and the realization results (Behrstock, Kleiner, Minsky, and Mosher [BKMM12]) hold for Teichmüller spaces of non-orientable surfaces as well. We begin by recalling the definition of consistency.

**Definition A.8** (Consistency). For a connected surface  $S$ , and a parameter  $\theta \geq 1$ , we say a tuple  $(z_V) \in \prod_{V \sqsubset S} \mathcal{C}(V)$  is  $\theta$ -consistent if the following two conditions holds for all pairs of subsurfaces  $U$  and  $V$ .

- (i) If  $U \pitchfork V$ , then

$$\min(d_U(z_U, \partial V), d_V(z_V, \partial U)) \leq \theta$$

- (ii) If  $U \sqsubset V$ , then

$$\min(d_U(z_U, \pi_U(z_V)), d_V(z_V, \partial U)) \leq \theta$$

The following theorem (Behrstock, Kleiner, Minsky, and Mosher [BKMM12, Theorem 4.3]) states that the projection from Teichmüller space to the curve complexes of all the subsurfaces is coarsely surjective onto the set of consistent tuples.

**Theorem A.9** (Consistency and realization). *There is a constant  $K \geq 1$ , and function  $\mathfrak{C} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that the following holds for any surface  $S$ .*

- (Consistency) *For every  $x \in \mathcal{T}(S)$ , the projection tuple  $(\pi_V(x))_{V \sqsubset S}$  is  $K$ -consistent.*
- (Realization) *For every  $\theta$ -consistent tuple  $(z_V)_{V \sqsubset S}$ , there exists a point  $z \in \mathcal{T}(S)$  such that  $d_V(\pi_V(z), z_V) \leq \mathfrak{C}(\theta)$  for all  $V$ .*

*Proof sketch of Theorem A.9 for non-orientable surfaces.* We first show that the projection map is consistent, and then show consistent tuples lie coarsely in the image of the projection map.

- (Consistency) We map  $x$  to  $\tilde{x}$  in the Teichmüller space of the double cover  $\tilde{S}$ . By applying the theorem for orientable surfaces, we have the  $(\pi_W(\tilde{x}))_{W \sqsubset \tilde{S}}$ , and we restrict to the subsurfaces in the tuple which arise as lifts. These points lie in the image of the quasi-isometric embedding map from Theorem A.2, which means consistency also holds for the tuples in  $S$ .
- (Realization) Given a  $\theta$ -consistent tuple  $(z_V)_{V \sqsubset S}$ , we construct a  $\theta'$ -consistent tuple in the double cover  $\tilde{S}$ . For subsurfaces of  $\tilde{S}$  that arise as lifts, we use the map from Theorem A.2. For the subsurfaces  $W$  that are not lifts, we set  $z_W = \pi_W(\partial(\iota(W)))$ . The fact that this is a  $\theta'$ -consistent tuple follows from the Behrstock inequality (for some  $\theta' > \theta$ )<sup>3</sup>. We now use this point to construct  $y \in \mathcal{T}(\tilde{S})$ , and deduce that  $y$  is coarsely fixed by  $\iota^*$ , i.e. the distance between  $y$  and  $\iota^*(y)$  is bounded above by a uniform constant  $C$ . Consider the midpoint  $\tilde{x}$  of the Teichmüller geodesic segment joining  $y$  and  $\iota^*(y)$ : by Teichmüller’s uniqueness theorem,  $\tilde{x}$  is fixed under the  $\iota^*$ -action. This means there is some  $x \in \mathcal{T}(S)$  whose image is  $\tilde{x}$ , and therefore the projection maps are coarsely  $(z_V)$ .

□

## REFERENCES

- [ABEM12] J. Athreya, A. Bufetov, A. Eskin, and M. Mirzakhani. “Lattice point asymptotics and volume growth on Teichmüller space”. In: *Duke Mathematical Journal* 161.6 (2012), pp. 1055–1111. DOI: [10.1215/00127094-1548443](https://doi.org/10.1215/00127094-1548443) (cit. on pp. 18, 25, 35, 36).
- [Ara22] F. Arana-Herrera. “Counting problems from the viewpoint of ergodic theory: from primitive integer points to simple closed curves”. In: *arXiv e-prints*, arXiv:2202.04156 (Feb. 2022), arXiv:2202.04156. DOI: [10.48550/arXiv.2202.04156](https://doi.org/10.48550/arXiv.2202.04156). arXiv: [2202.04156 \[math.DS\]](https://arxiv.org/abs/2202.04156) (cit. on p. 4).
- [Bab02] M. Babillot. “On the mixing property for hyperbolic systems”. In: *Israel journal of mathematics* 129 (2002), pp. 61–76 (cit. on p. 9).
- [BHS19] J. Behrstock, M. Hagen, and A. Sisto. “Hierarchically hyperbolic spaces II: Combination theorems and the distance formula”. In: *Pacific Journal of Mathematics* 299.2 (2019), pp. 257–338 (cit. on p. 55).
- [BKMM12] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher. “Geometry and rigidity of mapping class groups”. In: *Geometry & Topology* 16.2 (2012), pp. 781–888 (cit. on p. 54).

<sup>3</sup>A longer but a more thorough way of seeing this would be to verify that the Teichmüller space of a non-orientable surface satisfies the 9 axioms for hierarchical hyperbolicity that are enumerated in Behrstock, Hagen, and Sisto [BHS19].

- [CGTY] R. Coulon, I. Gekhtman, S. Tapie, and W. Yang. *Mixing of the geodesic flow for a group with a contracting element* (cit. on pp. 2, 10, 11).
- [DM23] S. Dowdall and H. Masur. *Lattice point counting for finite-order mapping classes*. 2023. arXiv: [2305.03721 \[math.GT\]](#) (cit. on pp. 6, 32, 37, 38, 41, 42, 44–46, 48, 54).
- [EGPS23] V. Erlandsson, M. Gendulphe, I. Pasquinelli, and J. Souto. “Mapping class group orbit closures for non-orientable surfaces”. In: *Geometric and Functional Analysis* (2023), pp. 1–57 (cit. on pp. 1, 4).
- [EM11] A. Eskin and M. Mirzakhani. “Counting closed geodesics in moduli space”. In: *Journal of Modern Dynamics* 5.1 (2011), pp. 71–105 (cit. on pp. 3, 18, 29).
- [EM22] A. Eskin and S. Mozes. “10. Margulis functions and their applications”. In: *The Impact of Margulis on Modern Mathematics*. Ed. by D. Fisher, D. Kleinbock, and G. Soifer. Chicago: University of Chicago Press, 2022, pp. 342–361. DOI: [doi:10.7208/chicago/9780226804163-011](#) (cit. on pp. 18, 20).
- [FM02] B. Farb and L. Mosher. “Convex cocompact subgroups of mapping class groups”. In: *Geometry & Topology* 6.1 (2002), pp. 91–152 (cit. on p. 5).
- [Gen17] M. Gendulphe. “What’s wrong with the growth of simple closed geodesics on nonorientable hyperbolic surfaces”. In: *arXiv e-prints*, arXiv:1706.08798 (June 2017), arXiv:1706.08798. arXiv: [1706.08798 \[math.GT\]](#) (cit. on pp. 1, 4, 7, 11, 12, 25, 27, 28).
- [GM23] I. Gekhtman and B. Ma. *Dynamics of subgroups of mapping class groups*. 2023. arXiv: [2311.03779 \[math.GT\]](#) (cit. on pp. 2, 3, 11).
- [Kha23] S. Khan. “The limit set of non-orientable mapping class groups”. In: *Journal of Modern Dynamics* 19.0 (2023). DOI: [10.3934/jmd.2023007](#) (cit. on pp. 1–3, 12, 48).
- [KPW23] S. Khan, C. Partin, and R. R. Winarski. “Pseudo-Anosov homeomorphisms of punctured nonorientable surfaces with small stretch factor”. In: *Algebraic & Geometric Topology* 23.6 (2023), pp. 2823–2856 (cit. on p. 33).
- [LR06] C. J. Leininger and A. W. Reid. “A combination theorem for Veech subgroups of the mapping class group”. In: *Geometric & Functional Analysis GAFA* 16 (2006), pp. 403–436 (cit. on p. 5).
- [LS18] L. Liechti and B. Strenner. “Minimal pseudo-Anosov stretch factors on nonoriented surfaces”. In: *Algebraic & Geometric Topology* (2018) (cit. on p. 33).
- [Mar04] G. Margulis. *On some aspects of the theory of Anosov systems*. Springer, 2004 (cit. on p. 3).
- [Min96] Y. N. Minsky. “Extremal length estimates and product regions in Teichmüller space”. en. In: *Duke Math. J.* 85.1 (1996), pp. 249–286 (cit. on pp. 13, 30, 49).
- [Mir08] M. Mirzakhani. “Growth of the number of simple closed geodesies on hyperbolic surfaces”. In: *Annals of Mathematics* (2008), pp. 97–125 (cit. on pp. 3, 4).
- [MS13] H. Masur and S. Schleimer. “The geometry of the disk complex”. In: *Journal of the American Mathematical Society* 26.1 (2013), pp. 1–62 (cit. on p. 48).
- [Nor08] P. Norbury. “Lengths of geodesics on non-orientable hyperbolic surfaces”. In: *Geometriae Dedicata* 134 (2008), pp. 153–176 (cit. on pp. 1, 7).
- [Pen88] R. C. Penner. “A construction of pseudo-Anosov homeomorphisms”. In: *Transactions of the American Mathematical Society* 310.1 (1988), pp. 179–197 (cit. on p. 33).



- [Qui06] J.-F. Quint. “An overview of Patterson-Sullivan theory”. In: *Workshop The barycenter method, FIM, Zurich*. 2006 (cit. on p. 9).
- [Raf07] K. Rafi. “A combinatorial model for the Teichmüller metric”. In: *GAFAGeometric And Functional Analysis* 17 (2007), pp. 936–959 (cit. on pp. 31, 52, 54).
- [Rob03] T. Roblin. “Ergodicité et équidistribution en courbure négative”. In: (2003) (cit. on pp. 9, 13).
- [Sch82] M. Scharlemann. “The Complex of Curves on Non-Orientable Surfaces”. In: *Journal of the London Mathematical Society* 2.1 (1982), pp. 171–184 (cit. on p. 27).
- [Sul79] D. Sullivan. “The density at infinity of a discrete group of hyperbolic motions”. In: *Publications Mathématiques de l’IHÉS* 50 (1979), pp. 171–202 (cit. on p. 9).
- [Yan18] W.-y. Yang. “Statistically Convex-Cocompact Actions of Groups with Contracting Elements”. In: *International Mathematics Research Notices* 2019.23 (Feb. 2018), pp. 7259–7323. DOI: [10.1093/imrn/rny001](https://doi.org/10.1093/imrn/rny001). eprint: <https://academic.oup.com/imrn/article-pdf/2019/23/7259/31146235/rny001.pdf> (cit. on pp. 2, 10, 11, 33, 34).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI

Email address: [saykhan@umich.edu](mailto:saykhan@umich.edu)

URL: <https://www.sayantankhan.io>