THE DIAMETER OF THE THICK PART OF MODULI SPACE AND SIMULTANEOUS WHITEHEAD MOVES

KASRA RAFI and JING TAO

Abstract
Let S be a surface of genus g with p punctures with negative Euler characteristic. We study the diameter of the ϵ-thick part of moduli space of S equipped with the Teichmüller or Thurston's Lipschitz metric. We show that the asymptotic behaviors in both metrics are of order $\log(\frac{g+p}{\epsilon})$. The same result also holds for the ϵ-thick part of the moduli space of metric graphs of rank n equipped with the Lipschitz metric. The proof involves a sorting algorithm that sorts an arbitrarily labeled tree with n labels using simultaneous Whitehead moves, where the number of steps is of order $\log(n)$. As a related combinatorial problem, we also compute, in the appendix of this paper, the asymptotic diameter of the moduli space of pants decompositions on S in the metric of elementary moves.

1. Introduction

Let $\mathcal{M}_{g,p}$ be the moduli space of complete finite-volume hyperbolic surfaces of genus g with p *labeled* punctures. We equip $\mathcal{M}_{g,p}$ with the Teichmüller metric d_T. Let ϵ_M be the Margulis constant (two curves of length less than ϵ_M on a hyperbolic surface do not intersect), and let $\epsilon \leq \epsilon_M$. Let $\mathcal{M}_{g,p}^{\epsilon}$ be the ϵ-thick part of $\mathcal{M}_{g,p}$, that is, the space of surfaces where the length of every essential closed curve is at least ϵ. By a theorem of Mumford, $\mathcal{M}_{g,p}^{\epsilon}$ is compact. We are interested in a better understanding of the "shape" of $\mathcal{M}_{g,p}^{\epsilon}$. As a first step, we study the asymptotic behavior of the Teichmüller diameter of $\mathcal{M}_{g,p}^{\epsilon}$ as g and p go to infinity. In this paper, we prove the following.

THEOREM A
There exists K such that, for every g and p with $2g - 2 + p > 0$ and every $\epsilon \leq \epsilon_M$,

$$\frac{1}{K} \log\left(\frac{g+p}{\epsilon}\right) \leq \operatorname{diam}_T(\mathcal{M}_{g,p}^{\epsilon}) \leq K \log\left(\frac{g+p}{\epsilon}\right). \tag{1}$$

We will adopt a shorthand notation and rewrite equation (1) as $\mathrm{diam}_T(\mathcal{M}_{g,p}^\epsilon) \asymp \log(\frac{g+p}{\epsilon})$.

There are several variations of this theorem. First, there are two natural possible metrics on $\mathcal{M}_{g,p}^\epsilon$. For any two points $X, Y \in \mathcal{M}_{g,p}^\epsilon$, we can consider the Teichmüller distance between X and Y or the induced path metric distance in $\mathcal{M}_{g,p}^\epsilon$ between them. Also, the moduli space has an alternative definition where the punctures are not marked. This is equivalent to considering the quotient space $\mathcal{M}_{g,p}/\mathrm{Sym}_p$ where the symmetric group Sym_p acts on $\mathcal{M}_{g,p}$ by permuting the labeling of the punctures. Our theorem holds for both spaces and in both senses of diameter.

As we shall see, an essential component of the proof of our theorem is that the Teichmüller metric is an L^∞-metric. Hence, another variation is to consider another L^∞-metric on $\mathcal{M}_{g,p}$: the asymmetric Lipschitz metric d_L as defined by Thurston [22]. There is a simple inequality relating the two metrics (see Section 2). For any $X, Y \in \mathcal{M}_{g,p}$,

$$\frac{1}{2} d_L(X, Y) \le d_T(X, Y).$$

For the proof of Theorem A, we will in fact use the Lipschitz metric to obtain the lower bound and the Teichmüller metric to obtain the upper bound. In view of the above equation, this will simultaneously establish the same asymptotics for the diameter of $\mathcal{M}_{g,p}^\epsilon$ in both metrics. Henceforth, when we say *distance* or $d(\cdot, \cdot)$, *diameter* or $\mathrm{diam}(\cdot)$, without reference to a metric, we will mean either one of the two metrics.

Another metric to consider would be the Weil–Petersson metric on Teichmüller space. Heuristically, the Weil–Petersson metric is an L^2-metric: the norm of a vector is an L^2-average of the amount of deformation throughout the surface. For this reason, two points that are distance 1 apart in the Teichmüller metric have Weil–Petersson distance at most $\sqrt{\mathrm{Area}}$. Hence, our theorem provides an upper bound of order $\sqrt{g+p}\log(g+p)$ for the Weil–Petersson diameter of moduli space. This implies Theorem 1.2 in [9] where the growth rate of the Weil–Petersson diameter of moduli space is studied. In this case a matching lower bound seems to be more difficult and remains open.

Width and height

As a general philosophy, one can study the geometry of a surface by decomposing it into pants along the shortest possible curves. If the curves are sufficiently short, then the geometry of the surface is essentially determined by the combinatorics of the pants decomposition, which is encoded by the dual graph of the pants decomposition. When a surface does not admit a short pants decomposition, then the lengths of the curves in the shortest pants decomposition and the twisting information along these curves are also relevant information. For these reasons, the proof of Theorem A naturally

breaks down into two parts. One part considers the subset $\mathcal{B}_{g,p} \subset \mathcal{M}_{g,p}^{\epsilon}$ consisting of surfaces that can be decomposed into pants by curves of length ϵ_M. We will refer to the diameter of $\mathcal{B}_{g,p}$ as the *width* of $\mathcal{M}_{g,p}^{\epsilon}$. The other part considers the Hausdorff distance between $\mathcal{M}_{g,p}^{\epsilon}$ and $\mathcal{B}_{g,p}$. We represent this quantity by $\mathrm{HD}(\mathcal{M}_{g,p}^{\epsilon}, \mathcal{B}_{g,p})$ and refer to it as the *height* of $\mathcal{M}_{g,p}^{\epsilon}$. We prove the following.

THEOREM B (Width of $\mathcal{M}_{g,p}^{\epsilon}$)
We have

$$\mathrm{diam}(\mathcal{B}_{g,p}) \asymp \log(g + p).$$

THEOREM C (Height of $\mathcal{M}_{g,p}^{\epsilon}$)
We have

$$\mathrm{HD}(\mathcal{M}_{g,p}^{\epsilon}, \mathcal{B}_{g,p}) \asymp \log\left(\frac{g + p}{\epsilon}\right).$$

This pair of theorems can be viewed as a refinement of Theorem A. Using the triangle inequality, the upper bounds for the width and height of $\mathcal{M}_{g,p}^{\epsilon}$ provide the upper bound for the diameter of $\mathcal{M}_{g,p}^{\epsilon}$. The lower bound for the height is a lower bound for the diameter. However, since our main interest is a better understanding of the shape of $\mathcal{M}_{g,p}^{\epsilon}$, we include the lower bound arguments for both the width and the height in this paper.

The diameter of space of graphs
The argument for the width involves solving a combinatorial problem of independent interest. By considering the dual graph of the pants decomposition, we can associate to every element $X \in \mathcal{B}_{g,p}$ a graph of rank g with p *marked* valence 1 vertices and $(2g - 2 + p)$ valence 3 vertices (see Section 2). Let $\mathrm{Graph}(g, p)$ be the space of all such graphs. We consider the metric of *simultaneous Whitehead moves* d_S on $\mathrm{Graph}(g, p)$: a simultaneous Whitehead move on a graph is a composition of an arbitrary number of commuting Whitehead moves. This is a suitable metric for our purposes since we are considering L^{∞}-metrics on moduli space. We show that the spaces $\mathrm{Graph}(g, p)$ (equipped with the d_S-metric) and $\mathcal{B}_{g,p}$ are quasi-isometric and hence their diameters are of the same order.

THEOREM D (Diameter of $\mathrm{Graph}(g, p)$)
We have

$$\mathrm{diam}_S\big(\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big) \asymp \mathrm{diam}_S\big(\mathrm{Graph}(g, p)\big) \asymp \log(g + p).$$

The main argument for the upper bound of Theorem D boils down to an efficient sorting algorithm for labeled trees using simultaneous Whitehead moves (see Section 3). Let Tree(n) be the space of rooted binary trees with labels $0, \ldots, n$, equipped with the metric simultaneous Whitehead moves. There is a distance-increasing embedding of Graph(g, p) into Tree(n), where $n = 2g - 2 + p$ (see Section 2). We introduce a sorting algorithm using simultaneous Whitehead moves and show that the algorithm can sort any tree in Tree(n) in $\log(n)$ steps. This bound also gives the desired upper bound for the diameter of Graph(g, p).

The lower bound for Theorem D can be obtained by a simple example (see Section 5). In fact, more can be said for $\mathcal{B}_g$ and Graph(g). By the work of Pinsker [20], a generic point in Graph(g) is an expander graph (see Definition 5.5). We show that the surfaces associated to expander graphs have the following property: By a *dividing curve* on a surface of genus g we will mean a separating curve which divides the surface into two pieces, both of genus of order g.

THEOREM E
For any surface in $\mathcal{B}_g$ associated to an expander graph, the length of any dividing curve is at least of order g.

If we fix a surface $X \in \mathcal{B}_g$ with a dividing curve of length of order 1, then the above theorem implies that the distance between X and any surface in $\mathcal{B}_g$ associated to an expander graph is at least of order $\log(g)$. This gives a lower bound of order $\log(g)$ for the diameter of $\mathcal{B}_g$. However, noting that $\mathcal{B}_g$ is quasi-isometric to Graph(g), this also shows that expanders are not equidistributed in Graph(g), nor are they coarsely dense. This answers a question of Mirzakhani in the negative.

COROLLARY F
Expander graphs are not coarsely equidistributed in the space Graph(g) equipped with the metric d_S. In fact, Graph(g) contains a ball of radius r, where r is on the order of the diameter of Graph(g), that contains no expander graphs.

Theorem E also follows from Buser's work in [6].

Related combinatorial problems
Similar combinatorial problems have been considered previously in the literature. In [4], an algorithm was established, using $\log(n)$ simultaneous flips, to transform any triangulation of an n-gon into any other. This can be rephrased in terms of a simultaneous-type metric on the space of *unlabeled planar* trees. However, the result in [4] and Theorem D do not imply each other. The trees in [4] do not have labels,

and they require their elementary moves to preserve a given embedding of a tree in the plane.

Another related problem of interest is computing the diameter of the space of pants decompositions on a surface $S_{g,p}$ or $S_{g,p}/\mathrm{Sym}_p$ up to homeomorphisms, equipped with the metric of elementary moves. This is equivalent to computing the diameter of $\mathrm{Graph}(g,p)$ or $\mathrm{Graph}(g,p)/\mathrm{Sym}_p$ in the metric of Whitehead moves: two graphs have distance 1 if they differ by a single Whitehead move. One can compute these diameters using the existing works of [3] and [21], though neither the details nor the statements are contained in any existing literature as far as the authors know. (In the case of $\mathrm{Graph}(g)$, an alternative proof is presented in [8].) We have included in this paper an appendix estimating the diameters of $\mathrm{Graph}(g,p)$ and $\mathrm{Graph}(g,p)/\mathrm{Sym}_p$ in the metric of Whitehead moves (see Theorem A.1 and Theorem A.2). What is worth noting is that in this metric the diameters of $\mathrm{Graph}(g,p)$ and $\mathrm{Graph}(g,p)/\mathrm{Sym}_p$ are not the same.

Theorem 2.3 in [21] is very general and gives a uniform upper bound for the growth rate of the number of elements in a ball of radius r in any space of shapes when shapes are allowed to evolve through locally supported elementary moves. Whitehead moves certainly fit that description. Another example of interest is the mapping class group equipped with the word *metric* coming from the Lickorish generators (see [17]). It is an immediate consequence of [21, Theorem 2.3] that the growth rate of the mapping class group with an appropriate set of generators is independent of the complexity of the surface.

Outer space

Our results can be extended to the setting of metric graphs, which is of interest to the study of outer automorphisms of free groups. Let R_n be a wedge of n circles. The *moduli space of metric graphs* $\mathcal{X}_n$ is the set of nondegenerate metrics graphs of volume n with homotopy type R_n (see Section 4). (This is the same as the quotient of outer space by the outer automorphism group of $\mathbb{F}_n$; see [10].) A well-studied metric on $\mathcal{X}_n$ is the Lipschitz metric d_L, defined in the same way as for surfaces. The ϵ-thick part $\mathcal{X}_n^\epsilon$ of $\mathcal{X}_n$ consists of graphs with a lower bound ϵ for the length of the shortest loop. We prove the following.

THEOREM G (Diameter of $\mathcal{X}_n^\epsilon$)
We have

$$\mathrm{diam}_L(\mathcal{X}_n^\epsilon) \asymp \log\left(\frac{n}{\epsilon}\right).$$

Outline of the paper

The organization of the paper is as follows:

Section 2 contains the background material for the paper. In this section, we will also introduce the metric of simultaneous Whitehead moves on graphs and explain how to associate to any surface in $\mathcal{B}_{g,p}$ a rooted binary tree with labels.

Section 3 contains the main algorithm, which provides the upper bound for the diameter of Tree(n) in the metric of simultaneous Whitehead moves. As applications, we obtain the upper bound of Theorem D and Theorem B.

Section 4 applies the work on trees to obtain the upper bound of Theorem G.

Section 5 is devoted to constructing some interesting examples of surfaces, which includes an example for Theorem E. These examples also provide the lower bounds for Theorem B, Theorem C, and Theorem D. We also complete the proof of Theorem G in this section.

Section 6, the final section of the paper, contains the argument for the upper bound of Theorem C. We also collect our results together to obtain Theorem A.

In the appendix, we compute the asymptotic diameter of Graph(g, p) and Graph(g, p)/ Sym$_p$ in the metric of Whitehead moves. The results and proofs of this section are independent of the rest of the paper.

2. Preliminaries

Moduli spaces

Let $S_{g,p}$ be a connected, oriented surface of genus g with p labeled punctures. We require the Euler characteristic $\chi(S_{g,p}) = 2 - 2g - p$ to be negative. Let $\mathcal{M}_{g,p}$ be the moduli space of complete, finite-volume, hyperbolic surfaces of homeomorphism type $S_{g,p}$, up to label-preserving isometries. The quotient $\mathcal{M}_{g,p}/ \mathrm{Sym}_p$ of $\mathcal{M}_{g,p}$ by the symmetric group Sym_p on p letters is the moduli space of unlabeled punctured surfaces. One can also think of $(\mathcal{M}_{g,p})$ $\mathcal{M}_{g,p}/ \mathrm{Sym}_p$ as the quotient of the Teichmüller space of $S_{g,p}$ by the (pure) mapping class group of $S_{g,p}$. We refer to [12], [13], and [11] for more details.

By a *curve* on $S_{g,p}$, we will always mean a free homotopy class of a simple closed curve which is not homotopic to a point or to a puncture. For a hyperbolic surface X, any curve has a unique geodesic representative which is the shortest in its homotopy class. Given a curve α on X, let $\ell_X(\alpha)$ be the length of the geodesic representative of α on X. A curve α is called a *systole* on X if $\ell_X(\alpha)$ is minimal among all curves on X. We will let $\ell(X)$ be the length of a systole on X. Given $\epsilon > 0$, the ϵ-*thick part* of $\mathcal{M}_{g,p}$ is

$$\mathcal{M}_{g,p}^{\epsilon} = \left\{ X \in \mathcal{M}_{g,p} : \ell(X) \geq \epsilon \right\}.$$

To make $\mathcal{M}_{g,p}^{\epsilon}$ nonempty and connected, we consider only $\epsilon \leq \epsilon_M$, where ϵ_M is a fixed constant such that for any $X \in \mathcal{M}_{g,p}$, if two distinct curves on X have lengths

less than ϵ_M, then they are disjoint. The constant ϵ_M is called the *Margulis constant* and is independent of g and p.

Two L^∞-metrics

We consider two L^∞-metrics on $\mathcal{M}_{g,p}$. Let $X, Y \in \mathcal{M}_{g,p}$.

- *Teichmüller metric*:

$$d_T(X,Y) = \frac{1}{2} \inf_f \{\log K(f) : f : X \to Y \text{ is } K(f)\text{-quasiconformal}\}.$$

- *Lipschitz metric*:

$$d_L(X,Y) = \inf_f \{\log L(f) : f : X \to Y \text{ is } L(f)\text{-Lipschitz}\}.$$

The Lipschitz metric was introduced by Thurston in [22]. Unlike the Teichmüller metric, the Lipschitz metric is not symmetric, and one needs to be careful when choosing the order of the two points when computing the distance. Both metrics induce the same topology on $\mathcal{M}_{g,p}$. We have the following inequality:

$$\frac{1}{2} d_L(X,Y) \le d_T(X,Y). \tag{2}$$

Equation (2) follows from two facts. The first fact, due to Wolpert, asserts that, under any K-quasiconformal map, the hyperbolic length of any curve changes by at most a factor of K (see [23, Lemma 3.1]). The second fact, stated below, is due to Thurston [22].

THEOREM 2.1 (Thurston)
For any $X, Y \in \mathcal{M}_{g,p}$,

$$d_L(X,Y) = \sup_\alpha \inf_f \log \frac{\ell_Y(f(\alpha))}{\ell_X(\alpha)},$$

where the sup *is taken over all curves on X and the* inf *is taken over all Lipschitz maps from X to Y.*

To compute distances in the Teichmüller metric, Kerckhoff has a similar formula using extremal lengths of curves (see [14]). For any $X \in \mathcal{M}_{g,p}$, the *extremal length* of a curve α on X is defined to be

$$\mathrm{Ext}_X(\alpha) := \sup_\rho \frac{\ell_\rho(\alpha)^2}{\mathrm{Area}(\rho)}$$

(see [1]). Here, ρ is any metric in the conformal class of X, $\ell_\rho(\alpha)$ is the ρ-length of the shortest curve in the homotopy class of α, and $\mathrm{Area}(\rho)$ is the area of the surface X equipped with the metric ρ.

THEOREM 2.2 (Kerckhoff)
For any $X, Y \in \mathcal{M}_g$,

$$d_T(X,Y) = \frac{1}{2} \sup_{\alpha} \inf_f \log \frac{\mathrm{Ext}_Y(f(\alpha))}{\mathrm{Ext}_X(\alpha)},$$

where the sup *is taken over all curves on X and the* inf *is taken over quasiconformal maps from X to Y.*

Notations
Throughout this paper, we will use the following set of notations. Given two quantities a and b, we will write $a = O(b)$ to mean $a \leq Kb$, for some uniform constant K. Similarly, $a = \Omega(b)$ if $a \geq Kb$. We will say a is *on the order* of b and write $a \asymp b$ if $a = O(b)$ and $a = \Omega(b)$. To control notations in a string of inequalities, it is sometimes convenient to replace $a = O(b)$ by $a \prec b$, and $a = \Omega(b)$ by $a \succ b$.

Pants decomposition and Bers' constant
Two isotopy classes of curves on $S_{g,p}$ will be called *disjoint* if they have disjoint representatives. A *multicurve* on $S_{g,p}$ is a (nonempty) collection of distinct curves on $S_{g,p}$ which are pairwise disjoint. A *pants decomposition* P of $S_{g,p}$ is a multicurve such that each component of $S_{g,p} \setminus P$ is a three-holed sphere, also called a *pair of pants*. The number of curves in P is equal to the *complexity* $\xi(S_{g,p}) = 3g - 3 + p$, and the number of pants in a decomposition is equal to $|\chi(S_{g,p})| = 2g - 2 + p$.

Let $B(X)$ be the minimal number such that X admits a pants decomposition P with $\ell_X(\alpha) \leq B(X)$ for all $\alpha \in P$. Let

$$B_{g,p} = \sup_{X \in \mathcal{M}_{g,p}} B(X)$$

be the *Bers' constant* for $S_{g,p}$. It was originally proved by Bers that $B_{g,p}$ is finite for all g and p. For closed surfaces, Buser gave explicit upper and lower bounds for B_g in [7, Section 5.2].

THEOREM 2.3 (Buser)
We have

$$\sqrt{6g} - 2 \leq B_g \leq 21(g-1).$$

One can extend the proof of Theorem 2.3 to obtain the upper and lower bounds for the Bers' constant in the punctured case as well: $\Omega(\sqrt{g+p}) = B_{g,p} = O(g+p)$. The lower bound for $B_{g,p}$ is obtained by Buser's hairy torus construction. His construction does not lie in the thick part of moduli space. We give another construction

in the thick part that gives the same lower bound (see Lemma 5.9). (In [2], Parlier and Balacheff improved the upper bound of $B_{0,p}$ to match the lower bound, but we will not need that here.)

Width and height of $\mathcal{M}_{g,p}^{\epsilon}$

Let $\mathcal{B}_{g,p} \subset \mathcal{M}_{g,p}^{\epsilon}$ be the set of surfaces $X \in \mathcal{B}_{g,p}$ such that X admits a pants decomposition where the length of every curve is exactly ϵ_M. By our choice of ϵ_M, intersecting curves cannot have lengths ϵ_M, and therefore such a pants decomposition on X is unique.

Let $\mathrm{diam}_L(\mathcal{B}_{g,p})$ be the maximal Lipschitz distance between any two points in $\mathcal{B}_{g,p}$. We will call this quantity the Lipschitz *width* of $\mathcal{M}_{g,p}^{\epsilon}$. Let $\mathrm{HD}_L(\mathcal{M}_{g,p}^{\epsilon}, \mathcal{B}_{g,p})$ be the Lipschitz Hausdorff distance between $\mathcal{M}_{g,p}^{\epsilon}$ and $\mathcal{B}_{g,p}$, defined to be

$$\mathrm{HD}_L(\mathcal{M}_{g,p}^{\epsilon}, \mathcal{B}_{g,p}) = \sup_{Y \in \mathcal{M}_{g,p}^{\epsilon}} \inf_{X \in \mathcal{B}_{g,p}} \max\{d_L(X,Y), d_L(Y,X)\}.$$

This quantity will be called the Lipschitz *height* of $\mathcal{M}_{g,p}^{\epsilon}$. The Teichmüller width and height of $\mathcal{M}_{g,p}^{\epsilon}$ are defined similarly.

Dual graphs to pants decompositions

To compare geometries of various surfaces in $\mathcal{B}_{g,p}$, we can look at the dual graph of the pants decompositions. Given $X \in \mathcal{B}_{g,p}$, let P be the associated pants decomposition on X. The *dual graph* Γ_P of P has a vertex for each pair of pants in $X \setminus P$ or for each puncture of X. Two (not necessarily distinct) vertices are connected by an edge if either they represent two (not necessarily distinct) pairs of pants glued along some curve in P, or if one vertex is a puncture contained in the pants represented by the other vertex. See Figure 1 for an example in genus 8.

For closed surfaces of genus g, the dual graph to a pants decomposition is always a trivalent graph of rank g. For punctured surfaces, the dual graph is a graph of rank g with p *marked* valence 1 vertices and $2g - 2 + p$ valence 3 vertices. Let $\mathrm{Graph}(g, p)$ be the set of all such graphs. Let

$$\psi : \mathcal{B}_{g,p} \to \mathrm{Graph}(g, p)$$

be the map defined by the dual graph construction. The map ψ is surjective. Shearing along each pants curve does not change the dual graph; thus each fiber is a $(3g - 3 + p)$-dimensional torus, and since each pants curve has length ϵ_M, each fiber has uniformly bounded diameter.

Whitehead moves on graphs

Given $\Gamma \in \mathrm{Graph}(g, p)$, we will call an edge e of Γ *interior* if both vertices of e have valence 3; otherwise, e is *exterior*. From any Γ, there is a way of deriving a new graph
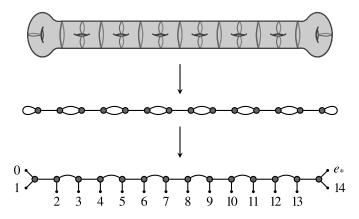
Figure 1. From pants decomposition to trivalent graph; from trivalent graph to labeled tree.

by modifying the local gluing structure about an interior edge e, called a *Whitehead move* on e. A Whitehead move on e is a process of collapsing e and reopening in a different direction. We allow two ways to reopen, as illustrated in Figure 2 on the left.

One may endow $\text{Graph}(g, p)$ with the metric d_W of Whitehead moves: $d_W(\Gamma_1, \Gamma_2) = 1$ if and only if Γ_1 and Γ_2 differ by a Whitehead move. Since our graph Γ corresponds to a pants decomposition P on a surface, there is a natural interpretation of Whitehead moves as *elementary moves* on pants decompositions. Each edge e in Γ corresponds to two pairs of pants glued along a common curve α. An elementary move on α changes P by fixing all curves in $P \setminus \alpha$, and replacing α by a
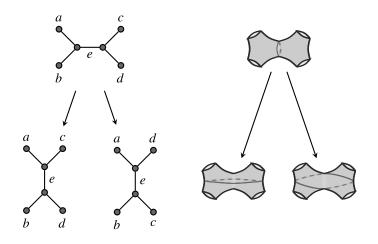

Figure 2. Whitehead move on e on left; corresponding elementary move on right.

transverse curve that intersects it minimally. Since only Whitehead moves associated to interior edges with distinct vertices are nontrivial, we only consider the case where α lies in a four-holed sphere. The two different directions of reopening represent the two minimally intersecting transverse curves up to homeomorphisms of the surface. (See the right-hand side of Figure 2.) Therefore, $\mathrm{Graph}(g, p)$ equipped with the metric of Whitehead moves is isometric to the space of homeomorphism types of pants decompositions on $S_{g,p}$ equipped with the metric of elementary moves.

Simultaneous Whitehead moves

Since we are considering L^∞-metrics on moduli spaces, the metric of Whitehead moves on $\mathrm{Graph}(g, p)$ is not the correct model metric for $\mathcal{B}_{g,p}$. Our goal is to equip $\mathrm{Graph}(g, p)$ with an appropriate metric so that its diameter is of the same order as the diameter of $\mathcal{B}_{g,p}$. Making one Whitehead move corresponds to modifying a surface $X \in \mathcal{B}_{g,p}$ by an elementary move in a four-holed sphere. But modifying X in several disjoint four-holed spheres at the same time contributes the same amount of distortion after taking the sup. This observation leads to the definition of *simultaneous Whitehead moves*.

For any graph $\Gamma \in \mathrm{Graph}(g, p)$, we will say two edges of Γ are *disjoint* if they do not share any vertices. Note that Whitehead moves on disjoint edges commute. Disjoint edges in Γ correspond to four-holed spheres which have disjoint interiors. Hence the corresponding elementary moves also commute with each other. A *simultaneous Whitehead move* on Γ is a composition of Whitehead moves on an arbitrary number of pairwise disjoint edges in Γ.

We equip $\mathrm{Graph}(g, p)$ with the metric d_S of simultaneous Whitehead moves: $d_S(\Gamma_1, \Gamma_2) = 1$ if and only if Γ_1 and Γ_2 differ by a simultaneous Whitehead move. The following lemma allows us to bound distances in $\mathcal{B}_{g,p}$ by distances in $\mathrm{Graph}(g, p)$. Recall the dual graph map $\psi : \mathcal{B}_{g,p} \to \mathrm{Graph}(g, p)$.

LEMMA 2.4
There exists a uniform constant K such that, for any g and p, if $X, Y \in \mathcal{B}_{g,p}$, then

$$d_T(X, Y) \leq K d_S\big(\psi(X), \psi(Y)\big).$$

Proof
Let S be a hyperbolic surface of genus 0 with four geodesic boundary components $\gamma_1, \ldots, \gamma_4$ such that each γ_i has length ϵ_M and, furthermore, S contains a nonperipheral curve α of length ϵ_M. The moduli space of all such surfaces is compact, since there are only finitely many ways α can separate the curves γ_i into two groups, and the amount of shearing along α is bounded by its length. Thus, there exists a uniform constant K_0 such that, for any other such surface S' with boundaries γ_i' and an essen-

tial curve α' of length ϵ_M, there is a K_0-quasiconformal map $S \to S'$ taking each γ_i to γ_i' and α to α'.

Similarly, we can consider the space of hyperbolic surfaces homeomorphic to a four-holed sphere, with a punctures and b geodesic boundaries of length ϵ_M so that $a + b = 4$. Let K_a be a uniform constant depending only on a such that there is a K_a-quasiconformal map between any two such surfaces fixing the boundaries and the punctures. Let K' be the maximum of the K_a's.

Now suppose that $X, Y \in \mathcal{B}_{g,p}$ have $d_S(\psi(X), \psi(Y)) = 1$. This means that there is a set of disjoint four-holed spheres in X on which we need to make a modification as above. We can construct a K'-quasiconformal map from X to some surface Y', locally using maps as above, where Y and Y' have the same dual graph. That is, $d_T(X, Y) \leq \frac{1}{2} \log K'$ and $\psi(Y) = \psi(Y')$. But, as mentioned before, the preimage of a point under ψ is a compact set with uniform diameter. Hence, $d_T(Y, Y') = O(1)$. By the triangle inequality, we have $d_T(X, Y) \leq K$, for some uniform K.

In the general situation where $d_S(\psi(X), \psi(Y)) = n$. Let $X = X_0, X_1, \ldots, X_n = Y$ be a sequence of elements in $\mathcal{B}_{g,p}$ with $d_S(\psi(X_{i-1}), \psi(X_i)) = 1$ for all $i = 1, \ldots, n$. From above, $d_T(X_{i-1}, X_i) \leq K$. Thus, by the triangle inequality, we obtain the desired statement of the lemma:

$$d_T(X, Y) \leq \sum_i^n d_T(X_{i-1}, X_i) \leq Kn = K d_S(\psi(X), \psi(Y)). \qquad \square$$

We conclude with the following corollary.

COROLLARY 2.5
We have

$$\mathrm{diam}_T(\mathcal{B}_{g,p} / \mathrm{Sym}_p) \leq \mathrm{diam}_T(\mathcal{B}_{g,p}) \prec \mathrm{diam}_S(\mathrm{Graph}(g, p)).$$

From graphs to labeled trees
By Corollary 2.5, the problem of bounding the width of $\mathcal{M}_{g,p}^\epsilon$ from above can be replaced by the problem of bounding $\mathrm{diam}_S(\mathrm{Graph}(g, p))$ from above. For the latter problem, it will be more convenient to cut each graph into a binary tree and label the ends in pairs to remember the gluing. We explain how to do this after some definitions.

By a *binary tree* (or a *tree* for short) we mean a connected graph with no loops, so that the valence at each vertex is either 1 or 3. A *rooted* tree has a distinguished exterior edge e_*. A *labeled* tree is a rooted tree where all ends (exterior edges except for e_*) are labeled with numbers $\{0, 1, 2, \ldots, n\}$, where n will be called the *complexity* of the tree. Note that a tree of complexity n has $n + 2$ exterior edges and $n - 1$ interior edges. Two labeled trees are said to be equal if there is a homeomorphism

between them taking root to root preserving the labels. After this identification, there are finitely many labeled trees for each fixed complexity n. Let Tree(n) be the space of labeled trees of complexity n. Whitehead moves or simultaneous Whitehead moves for trees are defined as before. We retain the notations d_W and d_S for the metrics of Whitehead moves and simultaneous Whitehead moves on Tree(n), respectively.

We now construct a map Graph(g, p) $\rightarrow$ Tree(n), $n = 2g - 2 + p$. Let $\Gamma \in$ Graph(g, p). The graph Γ has p exterior edges which are labeled. (They inherit the labeling from the marked vertices.) We may identify the labeling as $0, \dots, p - 1$. Now arbitrarily pick a spanning tree in Γ. The complement of the spanning tree in Γ contains exactly g edges. We cut each such edge in half, resulting in a tree T with $2g$ unlabeled exterior edges. We label these edges arbitrarily from $p, \dots, 2g + p - 1$ with the only restriction being that $p + 2k$ is glued to $p + 2k + 1$ in Γ, for $k = 0, \dots, g - 1$. Finally, erase the edge with the highest label, $2g + p - 1$, and make that the root of T. The resulting tree T is an element in Tree(n) associated to Γ. See Figure 1 for an example.

If the two trees associated to two graphs differ by a simultaneous Whitehead move, then the two graphs also differ by one simultaneous Whitehead move, and hence

$$\text{diam}_S\big(\text{Graph}(g, p)/\,\text{Sym}_p\big) \leq \text{diam}_S\big(\text{Graph}(g, p)\big) \leq \text{diam}_S\big(\text{Tree}(n)\big). \quad (3)$$

By Corollary 2.5, we have

$$\text{diam}_T\big(\mathcal{B}_{g,p}/\,\text{Sym}_p\big) \leq \text{diam}_T\big(\mathcal{B}_{g,p}\big) \prec \text{diam}_S\big(\text{Tree}(n)\big). \quad (4)$$

3. Trees and upper bound on width

In this section, we describe two algorithms for transforming a binary tree into a desired shape efficiently using simultaneous Whitehead moves. We prove the following.

THEOREM 3.1
For any n,

$$\text{diam}_S\big(\text{Tree}(n)\big) = O\big(\log(n)\big).$$

In view of (3) and (4), we obtain the upper bound of Theorem D and the upper bound for the width of $\mathcal{M}^{\epsilon}_{g,p}$. We remark that the lower bound for Theorem D is easy, but we will postpone a proof to Section 5, where we discuss the lower bound for the Lipschitz width of $\mathcal{M}^{\epsilon}_{g,p}/\,\text{Sym}_p$.

This section is organized as follows. In Section 3.1, we will introduce an algorithm which makes any tree more compact, by reducing its height to be of order $\log(n)$

in $O(\log(n))$ simultaneous Whitehead moves. Then in Section 3.2, we will introduce a distinguished element $T_n \in \text{Tree}(n)$, called the *fully sorted tree*. In Section 3.3, we will describe how to sort the labels of a compact tree to be fully sorted in $O(\log(n))$ simultaneous Whitehead moves.

3.1. Reducing the height

Given $T \in \text{Tree}(n)$, the root e_* defines a partial order "$<$" on the set of edges of T where e_* is the minimal element: Given edges e_1 and e_2, we say e_2 is a *descendant* of e_1, and we write $e_1 < e_2$, if the path from e_2 to e_* contains e_1. (We note that, in our figures, "descendants" are drawn as "ascendants.") If e_1 and e_2 are adjacent and $e_1 < e_2$, then e_1 and e_2 are in a *parent–child* relationship. The maximal elements of this relation are called the *ends* of T. Given any edge e in T, let T_e be the subtree of T consisting of e and its descendants. We will say T_e is *rooted* at e or e is the root of T_e. The *size* of T_e will be the number of edges of T_e. The edge e defines a partial ordering on the edges of T_e which is the one inherited from T. The maximal elements of T_e will be called the *ends* of T_e, and the *labels* of T_e will be the labels of the ends of T_e. We will say the root e of T_e has *height* 1, its children have height 2, and inductively define the height of all edges of T_e. The maximal possible height will be called the *height* of T_e. The height of T will be the height of $T_{e_*} = T$. Given a subtree T_e of height h, we will say T_e is *full* if T_e has 2^h ends.

For each interior edge e, we will label its children e_l (the left) and e_r (the right) and we call the trees T_{e_r} and T_{e_l} the *children* subtrees of T_e. Similarly, we label the left and right edges of e_l and e_r by e_{ll}, e_{lr}, e_{rl}, and e_{rr} and refer to the associated subtrees as the *grandchildren* subtrees.

PROPOSITION 3.2
Any tree T can be transformed to have a height $6 \log_2(n)$ after $O(\log(n))$ simultaneous moves.

Proof
To prove the proposition, it suffices to show that, if the height of T is larger than $6 \log_2(n)$, one can apply one simultaneous move to reduce the height by a definite multiplicative factor.

Let e be any edge. We define a special Whitehead move, called the *balance move at e*. Compare the sizes of the subtrees $T_{e_{ll}}$, $T_{e_{lr}}$, $T_{e_{rl}}$, and $T_{e_{rr}}$. If there is an absolute maximum among them (say, $T_{e_{ll}}$), we apply one Whitehead move to the edge e_l which reduces the height of e_{ll}, does not change the height of e_{lr}, and increases the heights of both e_{rl} and e_{rr} (see Figure 3). If there is no absolute maximum, then we apply no move.
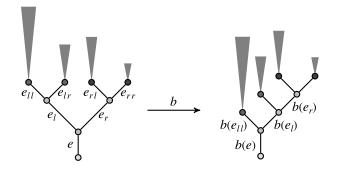
Figure 3. The map $\xrightarrow{b}$ is a balance move at e.

Now consider the following simultaneous Whitehead move consisting of a balance move at each edge of odd height. Since the supports of these balance moves are disjoint, their composition defines a simultaneous Whitehead move. We show that after one such simultaneous Whitehead move, any edge whose height was larger than $6\log_2(n)$ will have its height reduced by at least a multiplicative factor $5/6$. That is, if the height of T was larger than $6\log_2(n)$, then it will be reduced by a factor $5/6$. The algorithm stops when the height of T is less than $6\log_2(n)$.

Let e be any edge of height $h_e \geq 6\log(n)$. Consider the path P of length h_e in T connecting e to the root e_*. The height of e is affected by the balance moves at edges of odd height along this path. Some will decrease the height of e, some will increase it, and some will leave it unchanged. Moving down from e to e_*, let $e_1, e_2, \ldots, e_s$ be the set of edges in P of odd height and such that, for $j = 1, \ldots, s$, the balance move at e_j does not decrease the height of e. We bound the number of such edges, that is, we show that most edges along P will decrease the height of e.

By assumption, $T_e \subset T_{e_1}$ has at least one edge. For the edge e_1, the grandchild tree containing e is not of maximal size. Hence, T_{e_1} contains at least two subtrees of size T_e plus the edge e_1 and its children. That is,

$$\mathrm{size}(T_{e_1}) \geq 2\,\mathrm{size}(T_e) + 3 \geq 5.$$

Similarly, at each e_j, the grandchild tree containing e is not of maximal size. Hence, T_{e_j} contains two grandchildren subtrees of size at least $T_{e_{j-1}}$. That is,

$$\mathrm{size}(T_{e_j}) \geq 2\,\mathrm{size}(T_{e_{j-1}}) + 3.$$

By induction, $\mathrm{size}(T_{e_j}) \geq 2^{j+1} + 1$. The total number of edges in T, $2n + 1$, is larger than $\mathrm{size}(T_{e_s})$. Therefore,

$$s \leq \log_2(n).$$

A balance move at e_j, for all $j = 1, \ldots, s$, may increase the height of e, but a balance move at every other edge along P of odd height will reduce the height of e. The number of edges that decrease the height is at least $h_e/2 - \log_2(n)$, and the number of edges that increase the height is at most $\log_2(n)$. Hence, after the simultaneous move, the height of e is no more than

$$h_e - \frac{h_e}{2} + 2\log_2(n) \leq \frac{5}{6} h_e.$$

Since the maximum height of any tree is $n + 1$, the number of simultaneous moves required to reduce the height of T to a height less than $6 \log_2(n)$ is at most $\log_{6/5}(n + 1)$. This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.2. Fully sorted tree

In this section, we inductively construct a tree $T_n \in \text{Tree}(n)$, for each n. Considering T_n as a base point of $\text{Tree}(n)$, an upper bound for the distance between any tree and T_n will provide an upper bound for $\text{diam}_S(\text{Tree}(n))$.

For $n = 0$, T_0 is just the root edge e_* with one end labeled 0. Now assume that we have already constructed a tree $T_k \in \text{Tree}(k)$ for all $k < n$. Let m be the largest number so that $2^m \leq n$, and let $k = n - 2^m \geq 0$. Take the root edge e_*. On the left, we attach T_{2^m-1} (a full tree of height m) and on the right we attach a copy of T_k. We then change the labels of ends of T_k by adding 2^m to their values. The tree T_n is the tree giving the binary expansion of numbers 0 to n. See Figure 4 for some examples of T_n.

We can also think of T_n as a *fully sorted* tree. This is the description of T_n used in Proposition 3.8.
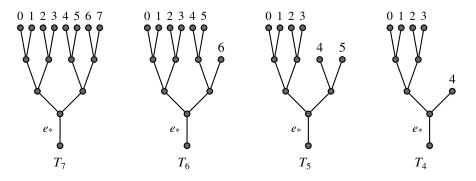

Figure 4. Examples of fully sorted trees.

Definition 3.3
Let d be the number of digits needed for the binary expansion of n. Take $T \in \mathrm{Tree}(n)$. Write every label of T as a d-digit number, possibly starting with several zeros. For an edge $e \in T$, let n_e be the number of ends of T_e, and let d_e be the number of digits in the binary expansion of n_e. By the kth digit of a label, we always mean the kth digit from the right. We say an edge $e \in T$ is *k-sorted* if the digits $(k+1)$ to d of all labels of T_e are the same and either

(a1)   all labels of T_e have the same kth digit as well, or

(a2)   all labels of T_e whose kth digit is 0 appear as ends of T_{e_l} and all labels of T_e whose kth digit is 1 appear as ends of T_{e_r}.

We say T is *fully sorted* if every edge e is d_e-sorted. Note that if an edge e is k-sorted, then it is also j-sorted for all $j \geq k$. All descendants of e are also at least k-sorted. On the other hand, if $k < d_e$, then e cannot be k-sorted. (There would not be enough free digits to represent n_e different numbers.)

Here is the second characterization of the base tree T_n.

LEMMA 3.4
A tree $T \in \mathrm{Tree}(n)$ is fully sorted if and only if $T = T_n$.

Proof
The statement is clear for $n = 0$. Assume that T is a fully sorted tree and that $n \geq 1$. The edge e_* is d-sorted, since $d_{e*} = d$. In this case, condition (a2) must hold, because the labels of T cannot all start with the same number. That is, all the labels at the ends of $T_{(e_*)_l}$ start with the digit 0 and all the labels at the ends of $T_{(e_*)_r}$ start with the digit 1.

We now cut e_* out, obtaining the two children trees. We modify the labels of the left tree by removing the first digit 0 from all labels. The labels on the right may have several digits in common. We modify the labels by removing all these digits. That is, the number of digits are the minimum needed to represent the labels. Denote these modified trees simply by T_l and T_r.

We now check that these two trees are still fully sorted. Consider the tree T_r, and assume that s digits have been removed in the modifications of the labels. For $e \in T_r$, we need to show that e is d_e-sorted in T_r. Since the number of ends n_e of T_e does not change by cutting out the root e_* from T, the edge e was d_e-sorted in T. This means that all digits from $d_e + 1$ to d of the labels of T_e are the same in T, and either condition (a1) or (a2) held true for e in T. Removing s unnecessary digits from T_r means that, viewing T_e as a subtree of T_r, the labels must agree on all digits from $d_e + 1$ to $d - s$. Furthermore, if condition (a1) or (a2) held true for e in T, then it

would still hold for e in T_r. Therefore, every edge e of T_r is d_e-sorted. The proof for T_l is the same ($s = 1$).

Note that T_l has 2^m ends where m is the largest number with $n \geq 2^m$. Since T_l is fully sorted, by induction $T_l = T_{2^m-1}$. Similarly, $T_r = T_k$, where $k = n - 2^m$, because T_r has $k + 1$ ends. That is, $T = T_n$. □

3.3. Sorting

We now present the sorting algorithm which will transform any tree T of height less than $6 \log_2(n)$ to a fully sorted tree (which we know has to equal T_n) in $O(\log(n))$ simultaneous Whitehead moves.

Note that the ends of a tree are always 1-sorted. Essentially our algorithm sorts the tree at different digits wherever possible by applying a *simultaneous sort move* which we describe below.

We say an edge e is k-*presorted*, if the following conditions hold:

(b1)   the children e_r and e_l are k-sorted;

(b2)   the digits $(k + 1)$ to d of all the labels at the ends of T_e are the same; and

(b3)   the edge e is not k-sorted.

We say an edge is *presorted* if it is presorted for some value of k.

To make this well defined for $k = d$ we assume that all edges are always $(d + 1)$-sorted. (One can think of the digit $(d + 1)$ as always being 0.) Given any k-presorted edge e, one can apply a *sort move* at e to make e j-sorted, for some $j \leq k$. There are essentially three types of such moves depicted in Figure 5. The first type requires three Whitehead moves supported at e_l and e_r. The second and third type require only one Whitehead move supported at one of e_l or e_r.

We claim the following statement.

LEMMA 3.5

Let E be the collection of presorted edges of T. Then the sort moves associated to the edges in E have disjoint support.

Proof

The supports of the sort moves at two edges are always disjoint if the edges are not in a parent–child relationship. Hence, it is enough to show that if e and e_l are presorted, then the support of the sort move at e is at e_r. This is equivalent to showing that, in this case, the sort move at e is of Type 2. The support of the sort move at e_l is a subset of $\{e_{ll}, e_{lr}\}$, which are disjoint from e_r. Hence, these two moves do not interfere with one another.

Suppose that e is k-presorted and that e_l is j-presorted. Condition (b1) applied to e implies that e_l is k-sorted. But e_l is j-presorted. Thus $j < k$. By condition (b2),
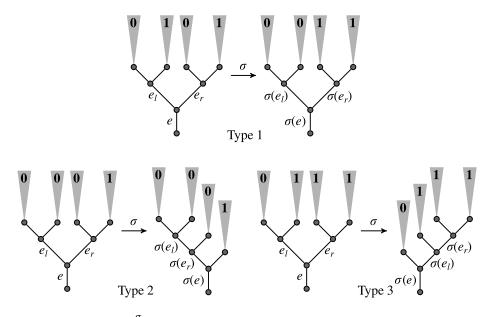
Figure 5. The map $\xrightarrow{\sigma}$ represents a sort move at e. There are three types of sort moves.

all digits $j + 1 \leq k < d$ of the labels of T_{e_l} must be the same. That is, the sort move at e is of type 2. $\qquad\square$

Definition 3.6

We will say a tree T is *ripe* if when any edge $e \in T$ is k-sorted, then each child of e is either $(k-1)$-presorted or $(k-1)$-sorted. Note that if T is ripe, then any subtree of T is ripe. We will say an edge $e \in T$ is *ripe* if T_e is ripe.

Now let $\overline{E} \subset E$ be the set of presorted edges e so that T_e is ripe. A *simultaneous sort move* is a composition of the sort moves associated to edges in $\overline{E}$. A simultaneous sort move is a composition of 3 simultaneous Whitehead moves.

We use $T \xrightarrow{\sigma} T'$ to mean that T' is obtained from T from one simultaneous sort move. There is a natural identification of the edges in T with the edges in T' via the map σ. We denote the edge in T' associated to an edge $e \in T$ by $\sigma(e)$.

LEMMA 3.7

Suppose $T \xrightarrow{\sigma} T'$, and let $\sigma(e)$ be an edge in T'. If the children of e are ripe in T, then $\sigma(e)$ is ripe in T'.

Proof

Let k be the minimal number such that $\sigma(e)$ is k-sorted. We need to show that each child of $\sigma(e)$ is either $(k-1)$-presorted or $(k-1)$-sorted. We argue in five cases depending on how the local picture around e changes under σ. In each case, the lemma essentially follows from the definition.

Case 1: e is presorted, and the sort move is of type 1.   In this case, the children of $\sigma(e)$ are images of the children of e and the grandchildren of e are mapped to the grandchildren of $\sigma(e)$. Since $\sigma(e)$ is k-sorted, the digits $(k+1)$ to d of the ends of $T'_{\sigma(e)}$ match and the kth digits are as depicted in Figure 5. This means e_r and e_l are k-sorted. But e_r and e_l are ripe; hence the grandchildren of e are $(k-1)$-sorted or $(k-1)$-presorted. Thus, the grandchildren of $\sigma(e)$ are $(k-1)$-sorted. Hence, the children of $\sigma(e)$ are either $(k-1)$-sorted or $(k-1)$-presorted. (The conditions (b1) and (b2) hold, but (b3) may or may not hold.) That is, $\sigma(e)$ is ripe in T'.

Case 2: e is presorted, and the sort move is of type 2 *or* 3.   By symmetry, we may assume type 2. In this case, a child of $\sigma(e)$ is an image of either e_r or e_{rr}. First consider $\sigma(e_r) = \sigma(e)_l$. As before, e_l is k-sorted and its children are at least $(k-1)$-presorted. But since the kth digit of labels at the ends of T_{e_l} match, e_{ll} and e_{lr} are in fact at least $(k-1)$-sorted. Hence, e_l is either $(k-1)$-sorted or $(k-1)$-presorted and $\sigma(e_l)$ is at least $(k-1)$-sorted (see Figure 5).

   Note also that $\sigma(e)_{lr}$ is an image of a grandchild of e and, as argued in the previous case, it is at least $(k-1)$-sorted. Thus, the children of $\sigma(e)_l$ are both at least $(k-1)$-sorted and hence $\sigma(e)_l$ is either $(k-1)$-presorted or $(k-1)$-sorted.

   The argument is easier for $\sigma(e_{rr}) = \sigma(e)_r$; since $\sigma(e)_r$ is an image of a grandchild of e, it is $(k-1)$-sorted and therefore $\sigma(e)$ is ripe.

Case 3: e is not presorted, and the children of e are mapped to the children of $\sigma(e)$. In this case, e is as sorted as $\sigma(e)$. Hence, e_r and e_l are at least k-sorted and, since they are ripe, the grandchildren are $(k-1)$-sorted or $(k-1)$-presorted. That is, the children of $\sigma(e)$ are either $(k-1)$-presorted or $(k-1)$-sorted. This implies that $\sigma(e)$ is ripe.

Case 4: e is not presorted, but σ contains a sort move of type 1 *at the parent of e.*   Let f be the parent of e. The sort move swaps the grandchildren of f. Since σ contains a sort move at f, all descendants of f are ripe.

   Since $\sigma(e)$ is k-sorted, the digits $(k+1)$ to d of ends of $T'_{\sigma(e)}$ match. That means the preimage of the children of $\sigma(e)$ are k-sorted. As before, using ripeness, we have that the children of $\sigma(e)$ are $(k-1)$-sorted or $(k-1)$-presorted. Hence $\sigma(e)$ is ripe.

Case 5: e is not presorted, and σ contains a sort move of type 2 or 3 at the parent of e.   Again, by symmetry, we may assume type 2. Let f be the parent of e. The case $e = f_l$ is already covered in case 3. Assume $e = f_r$. From the figure, we have that all the children of $\sigma(e)$ have the same kth digits. The proof now follows identically to case 1.   □

PROPOSITION 3.8
Let $T_n \in \text{Tree}(n)$ be the fully sorted tree, and let $T \in \text{Tree}(n)$ be any tree of height at most $6\log_2(n)$. Then

$$d_S(T, T_n) = O\big(\log(n)\big).$$

Proof
We will show that T can be transformed to T_n in $O(\log(n))$ simultaneous sort moves.

Let h be the height of T. First, we show that e_* will be d-sorted and that every edge is ripe after $(h - 3)$ steps. At the beginning, every edge is $(d + 1)$-sorted, and edges at the ends are d-sorted. In fact, the edges whose children are ends are also d-sorted after relabeling left and right edges. Hence, every edge at height $(h - 2)$ is either d-sorted or d-presorted and ripe. After the first step, every edge at height $(h - 2)$ or higher is d-sorted and every edge at height $(h - 3)$ is either d-sorted or d-presorted and, by Lemma 3.7, ripe. Note that, if T_e is not ripe, there are no sort moves at edges in the path connecting e to e_*. That is, this path is preserved identically under σ and, in particular, the height of e does not change. Hence, the maximum height of an edge e where T_e is not ripe goes down by at least 1 after every simultaneous sort move.

Continuing in this way, we get that after $(h - 3)$ steps, every edge at height $(h - 2 - (h - 3)) = 1$ or higher is d-sorted and ripe. That is, e_* is d-sorted and ripe. Let T^1 be the resulting tree. We have shown that $d_s(T, T^1) \leq h - 3$.

We now claim that if a tree $T^1 \in \text{Tree}(n)$ has the property that its root e_* is $d_{e_*} = d$-sorted and T^1 is ripe, then T^1 will be fully sorted after at most $d = \lceil \log_2(n) \rceil$ simultaneous sort moves. We will prove this by induction on n. When $n = 1$ there is nothing to prove. Now suppose that $n > 1$. By assumption, T^1 is ripe and e_* is d-sorted. Therefore, any future sort move will preserve the children subtrees of e_*. Let e be a child of e_*. The subtree T_e^1 is also ripe, since it is a subtree of T_1. After removing all digits which are common to all labels of T_e^1, e is either d_e-sorted or d_e-presorted. After applying one sort move $T_e^1 \xrightarrow{\sigma} T_e^{1'}$, the root e becomes d_e-sorted and $T_e^{1'}$ remains ripe. By the induction hypothesis, $T_e^{1'}$ can be transformed to T_{d_e} in $d_e \leq d - 1$ simultaneous sort moves. Therefore, after at most d simultaneous sort moves, both subtrees attached to e_* are fully sorted. This exactly means that T^1 is fully sorted after at most d sort moves.

We conclude that

$$d_S(T, T_n) \leq d_S(T, T^1) + d_S(T^1, T_n) \leq (h - 3) + d = O(\log(n)). \qquad \square$$

This completes the proof of Theorem 3.1. Using (3) and (4), we deduce the following respective corollaries.

COROLLARY 3.9
We have

$$\text{diam}_S(\text{Graph}(g, p)/\text{Sym}_p) \leq \text{diam}_S(\text{Graph}(g, p)) = O(\log(g + p)).$$

COROLLARY 3.10
We have

$$\text{diam}_T(\mathcal{B}_{g,p}/\text{Sym}_p) \leq \text{diam}_T(\mathcal{B}_g) = O(\log(g + p)).$$

4. Application to the moduli space of metric graphs

We now give an application of the algorithms in Section 3 to the moduli space of metric graphs. Our goal is to prove the upper bound of Theorem G of the introduction. The lower bound is worked out in the next section.

4.1. Lipschitz diameter of moduli space of metric graphs

Let R_n be a wedge of n circles. Let $\mathcal{X}_n$ be the space of isometry classes of metric graphs G with the following properties:

- G is homotopy equivalent to R_n.
- The valence of each vertex of G is at least 3.
- The sum of the lengths of edges or the *volume* of G is n.

We call $\mathcal{X}_n$ the *moduli space of metric graphs*. It is also naturally the quotient of *outer space* by the group of outer automorphisms of $\mathbb{F}_n$ (see [10]). We equip $\mathcal{X}_n$ with the Lipschitz metric: for any two graphs G and H, define

$$d_L(G, H) = \min_f \{\log L(f)\},$$

where f is an $L(f)$-Lipschitz map from G to H. The thick part of $\mathcal{X}_n$ is the subset $\mathcal{X}_n^\epsilon$ containing those graphs with no loop shorter than ϵ. We will show the following.

THEOREM 4.1
We have

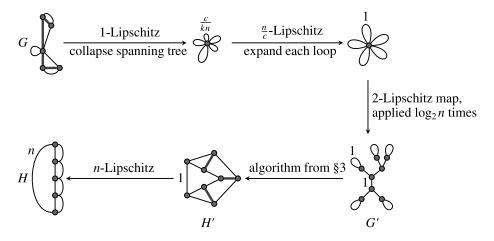$$\text{diam}_L(\mathcal{X}_n^\epsilon) = O\left(\log\left(\frac{n}{\epsilon}\right)\right).$$

Figure 6. An example in $\mathcal{X}_5^\epsilon$.

Let G and H be two graphs in $\mathcal{X}_n^\epsilon$. We will construct a map $G \to H$ in four steps. The idea is to interpolate G and H by two trivalent graphs, G' and H', on which we can apply the algorithms of the previous section. The reader may wish to look at the example in Figure 6.

We equip R_n with a metric by assigning length 1 to each circle.

LEMMA 4.2
The graph G can be mapped to R_n by an $O(n/\epsilon)$-Lipschitz maps.

Proof
Let T be the shortest spanning tree of G. Since $G \in \mathcal{X}_n^\epsilon$, every edge e in the complement of T has length at least ϵ/kn for some universal k. To see this, consider the unique loop consisting of e and an embedded path in T. Each edge in the loop cannot be longer than e since T is the shortest spanning tree. Since there are at most $O(n)$ edges in the loop and the total length of the loop is at least ϵ, this gives the lower bound on the length of e. Now map T to the vertex of R_n and each edge in the complement of T to a circle of R_n via a linear map. This map is at most (kn/ϵ)-Lipschitz. $\square$

LEMMA 4.3
R_n can be mapped to a trivalent graph G' with all edge lengths 1 via a composition of $\lceil \log_2 n \rceil$ 2-Lipschitz maps.

Proof
Divide the circles of R_n into two sets with roughly $n/2$ circles each. For each circle,

mark off two segments of length $1/4$ starting from the vertex of R_n. For each set, fold all the circles together along the marked segments. Note that folding is a 1-Lipschitz map. The resulting graph has an edge e of length 1, and each end point of e is attached to roughly $n/2$ loops of length $1/2$. Now stretch each loop to have length 1 by a 2-Lipschitz map and proceed inductively. At each end point of e, divide the circles into two sets of roughly $n/4$ circles each. Then fold and stretch. After $\lceil \log_2 n \rceil$ steps, we obtain a trivalent graph G' with all edge lengths 1. The composition map $R_n \to G'$ has Lipschitz constant at most $2^{\lceil \log_2 n \rceil} = O(n)$.   □

LEMMA 4.4
There is a trivalent graph H' with all edge lengths 1 that can be mapped to H via an n-Lipschitz map.

Proof

For each constant b, choose a binary tree t_b with b exterior edges. Let v be a vertex of H of valence $b > 3$. Remove a small neighborhood of v in H which does not contain any other vertex of H, and glue the end points to the end points of t_b in an arbitrary way. Now erase the vertices of valence 2 to obtain a trivalent graph H'. Equip H' with the metric so that each edge has length 1. There is a natural map from H' to H obtained by collapsing the edges of H' that are the image of the interior edges of t_b and then rescaling the remaining edges of H'. Collapsing is a 1-Lipschitz map, and since edges in H' have length 1 and edges in H cannot be longer than n, this map is n-Lipschitz.   □

LEMMA 4.5
G' can be sent to H' via a composition of $O(\log(n))$ L-Lipschitz maps, where L is a uniform constant.

Proof

There is a uniform constant L such that, for any two trivalent graphs with edge lengths 1, if they differ by one simultaneous Whitehead move, then they differ by a Lipschitz map with Lipschitz constant at most L. The graphs G' and H' can be cut into binary trees of complexity $2n - 2$. By (3) and Theorem 3.1, G' can be transformed into H' by $O(\log(n))$ simultaneous Whitehead moves, and hence the statement.   □

Proof of Theorem 4.1
We construct a Lipschitz map $G \to H$ as a composition of the maps coming from the four lemmas above:

$$G \xrightarrow{O(\frac{n}{\epsilon})} R_n \xrightarrow{O(n)} G' \xrightarrow{L^{O(\log n)}} H' \xrightarrow{n} H.$$

The Lipschitz constant of the composition $G \to H$ is bounded by the product of the four Lipschitz constants, which is bounded by n^d/ϵ for some uniform constant d. Thus, $d_L(G, H) = O(\log(n/\epsilon))$. $\qquad\square$

5. Examples of surfaces

In this section, we construct some examples of surfaces in the thick part of moduli space. These examples will provide the required lower bound for the width and the height, and hence the diameter, of the thick part. They also showcase some interesting behaviors which are of independent interest.

Let us for the moment restrict our attention to closed surfaces.

To bound the width from below, we construct three surfaces in $\mathcal{B}_g$ which are pairwise $\Omega(\log(g))$ apart in the Lipschitz metric. These surfaces are constructed using graphs. The first two surfaces, the line surface X and the bouquet surface Y, are constructed from two graphs which have a large ratio between their diameters. This ratio computes a lower bound on the Lipschitz constant from Y to X. The third surface, called the *expander surface Z*, is constructed using an *expander graph*, or a graph with high connectivity. We will show that Theorem E holds for Z: every separating curve on Z is of length $\Omega(g)$. This will contrast with X and Y, where both contain separating curves of length ϵ_M. Then the length ratio of separating curves will provide a lower bound on the Lipschitz distance from X or Y to Z.

To bound the height from below, we will construct a surface $H \in \mathcal{M}_g^\epsilon$ that cannot be decomposed into pants by curves shorter than $\sqrt{g}$. Buser already has such a construction, called the *hairy torus*, but it does not lie in $\mathcal{M}_g^\epsilon$. Our construction essentially takes two copies of Buser's hairy torus and glues them along the hairs. The resulting surface H has $B(H) = \Omega(\sqrt{g})$. Using length ratios we obtain a lower bound on the Lipschitz constant from H to any surface in $\mathcal{B}_g$.

Most of our constructions generalize easily punctured surfaces. The only exception is the expander surface, as the notion of expanders does not exist for graphs in $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$, $p > 0$, so this example is skipped. For the height, we give a construction that works for all $p \geq 0$ and all $g \geq 1$, by combining a double hairy torus and a punctured torus. For the remaining case of genus 0, we will refer to the construction in [2, Section 8] of a hairy sphere.

Since $\mathcal{M}_{g,p}$ covers $\mathcal{M}_{g,p}/\mathrm{Sym}_p$, the height and width of $\mathcal{M}_{g,p}$ are bounded below by the height and width of $\mathcal{M}_{g,p}/\mathrm{Sym}_p$.

5.1. Shadow map

Let $\psi : \mathcal{B}_{g,p} \to \mathrm{Graph}(g, p)$ be the dual graph map. We will regard elements in $\mathrm{Graph}(g, p)$ as metric graphs by assigning length 1 to each edge. Let $X \in \mathcal{B}_{g,p}$. Outside of the cusps, X is quasi-isometric to $\psi(X)$. To make this precise, we introduce the *shadow map* $\Upsilon : X \to \psi(X)$.

For each puncture p of X, let N_p be the horocyclic neighborhood of p such that the horocyclic boundary of N_p has length equal to ϵ_M. Let $\overline{X}$ be the closure of $X \setminus \bigcup N_p$, where p ranges over all punctures of X. We call $\overline{X}$ the *truncated surface* obtained from X.

Let P be the associated pants decomposition on X. We may assume that ϵ_M is small enough so that P is contained in $\overline{X}$ and that the distance from every curve in P to the boundaries of $\overline{X}$ is of order 1.

Given a constant A, for any $\alpha \in P$, let

$$N_\alpha = \big\{x \in X : d_X(x, \alpha) \leq A\big\},$$

and for each boundary component $\gamma \in \partial \overline{X}$, let

$$N_\gamma = \big\{x \in \overline{X} : d_X(x, \gamma) \leq A\big\}.$$

Choose A so that $N = \bigcup_{\alpha \in P} N_\alpha \cup \bigcup_{\gamma \in \partial \overline{X}} N_\gamma$ is a disjoint union of embedded annuli. Each component of $\overline{X} \setminus N$ is a pair of pants with diameter bounded uniformly by a constant D. Foliate each N_α and N_γ by closed loops equidistant from α and γ, respectively. The shadow map

$$\Upsilon : X \to \psi(X)$$

sends each component in $X \setminus N$ to a vertex and each N_α or N_γ to an edge by collapsing leaves and then mapping linearly onto the edge. The map Υ is essentially distance decreasing. For any $x, y \in \overline{X}$,

$$d_{\psi(X)}\big(\Upsilon(x), \Upsilon(y)\big) \prec d_X(x, y).$$

On the other hand,

$$d_X(x, y) \leq (A + D)\big(d_{\psi(X)}\big(\Upsilon(x), \Upsilon(y)\big) + 2\big).$$

Thus, Υ is a quasi-isometry from $\overline{X}$ to $\psi(X)$.

5.2. Line and bouquet surfaces

We construct two graphs in $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$, one of which has diameter $g + p$ and the other has diameter $\log(g + p)$.

Consider the tree T with $g + p$ exterior edges as in Figure 7. We can make T into an element Γ in $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$ by attaching g loops to g of those edges. The diameter of Γ is at least $g + p$.

Now consider any tree T' of height $\log_2(g + p)$ with $g + p$ exterior edges. For instance, one can pick the fully sorted tree (see Figure 4). Again, T' can be made into
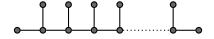
Figure 7. The tree T giving rise to the line surface.

a graph $\Gamma' \in \text{Graph}(g, p)/\text{Sym}_p$ by attaching g loops. The graph Γ' has diameter at most $\log_2(g + p) + 2$.

Let $X = X_{g,p}$ and $Y = Y_{g,p}$ be elements of $\mathcal{B}_{g,p}/\text{Sym}_p$ such that $\psi(X) = \Gamma$ and $\psi(Y) = \Gamma'$. We will refer to X and Y as a *line surface* and *bouquet surface*, respectively.

LEMMA 5.1
We have

$$\text{diam}_L(\mathcal{B}_{g,p}/\text{Sym}_p) = \Omega\big(\log(g + p)\big).$$

Proof
We will use $d_L(Y, X)$ to achieve this lower bound, where X is the line surface and Y is the bouquet surface in $\mathcal{B}_{g,p}/\text{Sym}_p$.

The section on shadow map implies that

$$\text{diam}(\overline{X}) \succ g + p \qquad \text{and} \qquad \text{diam}(\overline{Y}) \prec \log(g + p),$$

where $\overline{X}$ and $\overline{Y}$ are truncated surfaces obtained from X and Y, respectively.

Consider any L-Lipschitz map $f : Y \to X$. We can choose two points x_1 and x_2 in $\overline{X}$ a distance $\Omega(g + p)$ apart. Let $y_i = f^{-1}(x_i)$. Since x_i has injectivity radius at least ϵ_0, the injectivity radius at y_i is at least ϵ_0/L. So y_i has distance at most $\log L$ from $\overline{Y}$. We can connect y_1 to y_2 by an arc ω in Y with

$$\ell_Y(\omega) \prec \log(g + p) + 2\log L.$$

The image $f(\omega)$ is an arc connecting x_1 to x_2, so $\ell_X(f(\omega)) \succ (g + p)$. We have

$$L \geq \frac{\ell_X(f(\omega))}{\ell_Y(\omega)} \succ \frac{g + p}{\log(g + p) + 2\log L}.$$

If $L \geq g + p$, then $d_L(Y, X) = \Omega(\log(g + p))$. In the case that $L \leq g + p$, then the above becomes

$$L \succ \frac{g + p}{3\log(g + p)}.$$

Thus, we also obtain

$$d_L(Y, X) \succ \log\left(\frac{g + p}{3 \log(g + p)}\right) = \Omega\big(\log(g + p)\big). \qquad \square$$

Combining Lemma 5.1 with Corollary 3.10, we obtain the following.

COROLLARY 5.2
The Teichmüller and the Lipschitz width of $\mathcal{M}_{g,p}^{\epsilon}$ and $\mathcal{M}_{g,p}^{\epsilon}/\operatorname{Sym}_p$ are all of order $\log(g + p)$.

Combining Lemma 5.1 with Corollary 2.5 and Corollary 3.9, we also obtain the following.

COROLLARY 5.3
We have

$$\operatorname{diam}_S\big(\operatorname{Tree}(g + p)\big) \asymp \operatorname{diam}_S\big(\operatorname{Graph}(g, p)\big)$$
$$\asymp \operatorname{diam}_S\big(\operatorname{Graph}(g, p)/\operatorname{Sym}_p\big) \asymp \log(g + p).$$

Finally, we can derive the lower bound to the diameter of $\mathcal{X}_n^{\epsilon}$. Together with the upper bound coming from Theorem 4.1, we obtain the following.

LEMMA 5.4
We have

$$\operatorname{diam}_L(\mathcal{X}_n^{\epsilon}) \asymp \log\left(\frac{n}{\epsilon}\right).$$

Proof
We show the lower bound.

Let $\Gamma \in \mathcal{X}_n^{\epsilon}$ be a graph of diameter of order n. For instance, pick Γ to be the graph inducing the line surface X_n, but renormalized to have volume n. The wedge R_n of n circles (with edge lengths 1) has diameter 1. Thus $d_L(R_n, G) \succ \log(n)$. On the other hand, let $H \in \mathcal{X}_n^{\epsilon}$ be any graph which has a loop of length ϵ. Then $d_L(H, R_n) \geq \log(1/\epsilon)$. It follows that

$$\operatorname{diam}_L(\mathcal{X}_n^{\epsilon}) \succ \frac{1}{2}\left(\log(n) + \log\left(\frac{1}{\epsilon}\right)\right) = \frac{1}{2}\log\left(\frac{n}{\epsilon}\right). \qquad \square$$

5.3. Expander surfaces
In this section, we consider only closed surfaces. In this case, the dual graph to a pants decomposition is a trivalent graph. We will use trivalent graphs with "high"

connectivity to construct surfaces in $\mathcal{B}_g$ with long separating curves. These surfaces will be $\Omega(\log(g))$ away from the examples of the previous section, providing another proof of Lemma 5.1. To formalize the notion of connectivity, we define the Cheeger constant of a graph.

Let Γ be any graph with n edges. For any subgraph Δ in Γ, let $|\Delta|$ be the number of edges in Δ. We will let $\partial\Delta \subset \Delta$ be the subset of edges in Δ which share a vertex with an edge outside of Δ. The *Cheeger constant* of Γ is defined to be

$$\mathrm{ch}(\Gamma) = \min_{1 \le |\Delta| \le n/2} \frac{|\partial\Delta|}{|\Delta|},$$

where the minimum is taken over all subgraphs Δ with at most $n/2$ edges.

Definition 5.5
An infinite family $\mathcal{E}$ of d-regular graphs is a δ-expander family if $\mathrm{ch}(E) \ge \delta$ for every $E \in \mathcal{E}$.

THEOREM 5.6 ([18], [20])
For every d, there exists a δ_d-expander family of d-regular graphs.

Fix δ, and let $\mathcal{E}$ be a δ-expander family of trivalent graphs. We will call a surface $Z_g \in \mathcal{B}_g$ an *expander surface* if its dual graph E_g is an element $\mathcal{E}$. By a *dividing curve* on a closed surface S_g, we will mean a separating curve on S_g which divides S_g into two pieces, each of which has genus on the order of g. We will prove the following fact about expander surfaces, which was known to Buser [6].

THEOREM 5.7
If $Z_g \in \mathcal{B}_g$ is an expander surface, then the shortest dividing curve on Z_g has length $\Omega(g)$.

Proof
Let α be any dividing curve on Z_g, and let U be the closure of one of the components of $Z_g \setminus \alpha$. By assumption, the genus of U is of order g.

Let $\Upsilon : Z_g \to \Gamma_g$ be the shadow map. We claim that

$$|\Upsilon(U)| = \Omega(g) \qquad \text{and} \qquad 4|\partial\Upsilon(U)| \le |\Upsilon(\alpha)|. \tag{5}$$

To see the first statement, let U' be the largest subsurface of Z_g with the same shadow as U (i.e., the preimage of $\Upsilon(U)$). Since U is a subsurface of U', we have $|\chi(U')| \ge |\chi(U)| = \Omega(g)$. But U' is a union of pairs of pants, exactly one associated to a vertex in $\Upsilon(U)$. Hence $|\chi(U')| \asymp |\Upsilon(U)|$. That is, $|\Upsilon(U)| = \Omega(g)$. So the first statement follows.

For the second statement, it is sufficient to construct a map from $\partial \Upsilon(U)$ to $\Upsilon(\alpha)$ so that the preimage of every edge has a uniformly bounded size. Let $e \in \partial \Upsilon(U)$. Then e shares a vertex v with an edge outside of $\Upsilon(U)$. The vertex v corresponds to a pair of pants and α has to intersect this pair of pants nontrivially. Hence, there is an edge connected to v that is in $\Upsilon(\alpha)$. We send e to this edge. The preimage of an edge in $\Upsilon(\alpha)$ under this map has size at most 4.

Using (5) and the fact that $\mathcal{E}$ is an expander family, we obtain

$$\big|\Upsilon(\alpha)\big| \geq 4\big|\partial \Upsilon(U)\big| \geq 4\delta\big|\Upsilon(U)\big| = \Omega(g).$$

This bounds the length of $\Upsilon(\alpha)$ from below. Since the shadow map Υ is essentially distance decreasing, we obtain $\ell_{Z_g}(\alpha) = \Omega(g)$. $\square$

LEMMA 5.8
Let X_g, Y_g, and Z_g be a line surface, bouquet surface, and expander surface, respectively. We have

$$d_L(X_g, Z_g) = \Omega\big(\log(g)\big), \qquad d_L(Y_g, Z_g) = \Omega\big(\log(g)\big).$$

Proof
To see $d_L(X_g, Z_g) = \Omega(\log(g))$, let Γ_g be the graph from which we constructed X_g. One sees that Γ_g can be divided into two roughly equal-sized pieces by one edge. The associated pants decomposition of X_g contains a dividing curve of length ϵ_M.

However, from Theorem 5.7, any dividing curve on Z_g has length $\Omega(g)$. Any homeomorphism $X_g \to Z_g$ must take dividing curves to dividing curves. Using the length ratio, we obtain a lower bound for the Lipschitz constant:

$$d_L(X_g, Z_g) \succ \log\Big(\frac{g}{\epsilon_M}\Big) = \Omega\big(\log(g)\big).$$

Similarly, Y_g also admits a dividing curve of length ϵ_M. So the same argument also shows that $d_L(Y_g, Z_g) = \Omega(\log(g))$. $\square$

5.4. Hairy torus example and a lower bound for height

Recall that surfaces in $\mathcal{B}_{g,p}/\operatorname{Sym}_p$ can be decomposed into pants by curves of length ϵ_M. In the following, we will construct a surface in $\mathcal{M}_{g,p}^\epsilon/\operatorname{Sym}_p$ which cannot be decomposed into pants by curves all shorter than $\sqrt{g+p}$.

LEMMA 5.9
There exists a surface $H \in \mathcal{M}_{g,p}^\epsilon/\operatorname{Sym}_p$ such that $B(H) = \Omega(\sqrt{g+p})$.

Assuming Lemma 5.9, we derive the lower bound for the height of the thick part of moduli space.

COROLLARY 5.10
We have

$$\mathrm{HD}_L(\mathcal{M}^\epsilon_{g,p}, \mathcal{B}_{g,p}) \geq \mathrm{HD}_L(\mathcal{M}^\epsilon_{g,p}/\mathrm{Sym}_p, \mathcal{B}_{g,p}/\mathrm{Sym}_p) = \Omega\left(\log\left(\frac{g+p}{\epsilon}\right)\right).$$

Proof
Let H be the surface of Lemma 5.9. For any $X \in \mathcal{B}_{g,p}/\mathrm{Sym}_p$, any Lipschitz map from X to H must take a pants decomposition of X to a pants decomposition of H. Thus some curve on X must get stretched by a factor $\Omega(\frac{\sqrt{g+p}}{\epsilon_M})$. This means that

$$d_L(X, H) \succ \log\left(\frac{g+p}{\epsilon_M}\right).$$

Now consider a surface $Y \in \mathcal{M}^\epsilon_{g,p}/\mathrm{Sym}_p$ which has a curve of length $\epsilon \leq \epsilon_M$. Any Lipschitz map from Y to X must stretch this curve by a factor at least ϵ_M/ϵ, so

$$d_L(Y, X) \geq \log\left(\frac{\epsilon_M}{\epsilon}\right).$$

It follows then that

$$\mathrm{HD}_L(\mathcal{M}^\epsilon_{g,p}/\mathrm{Sym}_p, \mathcal{B}_{g,p}/\mathrm{Sym}_p) \succ \frac{1}{2}\left(\log\left(\frac{g+p}{\epsilon_M}\right) + \log\left(\frac{\epsilon_M}{\epsilon}\right)\right)$$

$$= \frac{1}{2}\log\left(\frac{g+p}{\epsilon}\right). \qquad \square$$

The rest of this section is dedicated to constructing H for Lemma 5.9. In the case of genus $g = 0$, $p > 0$, this has already been done by Balacheff and Parlier in [2]. Their hairy sphere construction gives rise to a surface $H \in \mathcal{M}^\epsilon_{0,p}/\mathrm{Sym}_p$ with $B(H) = \Omega(\sqrt{p})$. For higher genus, we will use a variation of Buser's hairy torus (see [7, Section 5.3]). We will first explain the construction in the case of closed surfaces. Then we will extend the construction to punctured surfaces with at least one genus.

Start with a right-angled geodesic pentagon in the hyperbolic plane with side lengths a, b, c, d, and e. We set $a = b$, $c = e$, and $d = 0.25$. Glue four copies of such pentagons together to form a $2a \times 2a$ hyperbolic square R with an inner geodesic boundary component γ of length 1 (see Figure 8).

Given any positive integer m, we can paste together m^2 copies of R to obtain a larger square R_m of side lengths $2am$ with m^2 inner boundary components. We index these boundaries by γ_{ij}. Identifying the opposite sides of R_m yields a hyperbolic surface (a hairy torus) T_m of genus 1 with m^2 boundary components. Now take two copies of T_m and glue them along the γ_{ij}'s. The resulting closed surface H_m has genus $g = 1 + m^2$ (see Figure 9).
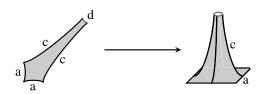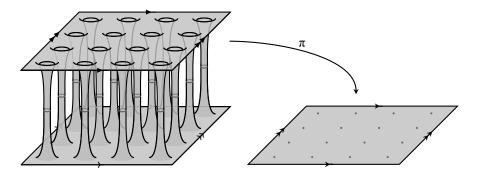
Figure 8. Gluing pentagons.

Figure 9. Hairy torus.

To show that H_m satisfies Lemma 5.9, we define a Lipschitz map $\pi : H_m \to F_m$, where F_m is a flat torus obtained from gluing the opposite sides of a $2am \times 2am$ Euclidean square. The map π is defined locally on each hyperbolic square R with an inner boundary γ.

Let F be a $2a \times 2a$ Euclidean square. Let $\pi_R : R \to F$ be any Lipschitz map that takes the sides of R to the sides of F and γ to the center of F. Let L be the Lipschitz constant of π_R. Divide F_m into m^2 subsquares. Let $\pi : H_m \to F_m$ be the map such that, on each hyperbolic square R in H_m, π restricted to R is mapped to the corresponding subsquare F in F_m via the map π_R (see Figure 9). The Lipschitz constant of π is at most the Lipschitz constant of π_R. We have shown the following.

LEMMA 5.11
The map $\pi : H_m \to F_m$ is L-Lipschitz for a uniform L.

The map π has the following property.

LEMMA 5.12
For any pants decomposition P on H_m, there exists a curve $\alpha \in P$ such that $\pi(\alpha)$ is not trivial in homology. Therefore, H_m satisfies Lemma 5.9.

Proof

For the purpose of this proof, we can consider π up to homotopy. The map $\pi : H_m \to F_m$ is surjective in homology, since π maps each hairy torus in H_m onto F_m. If a curve $\beta \in P$ maps to a trivial curve in the homology of F_m, then one can find a disk D in F_m such that $\pi(\beta)$ lies in D. We may assume ∂D avoids the singular points (the centers of the subsquares) of F_m. The complement $F_m \setminus D$ is a one-holed torus. Its preimage $Z = \pi^{-1}(F_m \setminus D)$ in H_m is an essential subsurface which is not a pair of pants (since $\pi(Z)$ maps onto the homology of F_m). Since P is a pants decomposition, some curve $\alpha \in P$ must intersect Z. The image $\pi(\alpha)$ cannot be homotoped away from $F_m \setminus D$, and thus it is nontrivial in the homology of F_m.

For such an α, we have $\ell_{F_m}(\pi(\alpha)) \geq 2am = 2a\sqrt{g-1}$. Since π is L-Lipschitz, $\ell_{H_m}(\alpha) = \Omega(m) = \Omega(\sqrt{g})$. This is true for any pants decomposition P, so $B(H_m) = \Omega(\sqrt{g})$. $\qquad\square$

The double-torus construction yields a closed surface H_m of genus $1 + m^2$. We now extend the construction to every genus g. Given an arbitrary g, let m be the largest integer such that $1 + m^2 \leq g < 1 + (m+1)^2$. Let $r = g - (1+m^2) < 2m+1$. Let H_m be the closed surface of genus $1 + m^2$ obtained from gluing two hairy tori together. Cut H_m along one of the curves of length 1 which came from the gluing to obtain a surface of genus m^2 with two boundary components. For the pair of boundary components β and β', we insert a surface with boundary as follows. Let W be a hyperbolic surface of genus r with two geodesic boundary components. We require the boundary components of W to have length 1 and all nonperipheral curves on W to have length at least ϵ. We glue a copy of W along its boundary components to β and β'. The resulting surface H is a closed surface of genus $g = m^2 + r + 1$. In the map previously defined on $H_m \to F_m$, the pair β and β' is mapped to the same point in F. Thus, this map can be extended to W by a constant map. This defines a Lipschitz map $H \to F_m$ with the same constant as the map $H_m \to F_m$. The same proof in Lemma 5.12 also works to show $B(H) = \Omega(m) = \Omega(\sqrt{g})$.

Now we explain the construction in the case of punctured surfaces of genus 1. For any a, there exists a hyperbolic quadrilateral with three right angles and one ideal vertex, such that the two sides opposite the ideal vertex have length a. Glue four copies of such quadrilaterals together to form a $2a \times 2a$ hyperbolic square R' with a puncture. The surface R' can be mapped to F minus the center by a uniformly Lipschitz map. Pasting together m^2 copies of R' and gluing opposite sides yields a surface Q_m of genus 1 with $p = m^2$ punctures. Gluing the maps on each R' equips Q_m with a Lipschitz map π to the flat torus F_m. By the same proof as in Lemma 5.12, any pants decomposition P on Q_m must contain a curve α which does not vanish in the homology of F. Hence $\ell_F(\pi(\alpha)) \geq 2am = 2a\sqrt{p}$, which implies that $\ell_{Q_m}(\alpha) = \Omega(\sqrt{p}) = B(Q_m)$.

For an arbitrary p, we do a similar modification as in the case of closed surfaces. Let $r = p - m^2 < 2m + 1$. Let Q_m be the genus 1 surface with m^2 punctures as constructed above. Now remove an R'-square of Q_m, and replace it with an R-square. The result is a surface Q'_m of genus 1 with $m^2 - 1$ punctures and one boundary component of length 1. One checks that $B(Q'_m) \asymp B(Q_m) = \Omega(m)$. To the inner boundary of R, attach a sphere with $r + 1$ punctures with a boundary component of length 1 such that all essential curves also have length at least ϵ. The result is a surface Q of genus 1 with p punctures. The map $Q \to F_m$ is defined locally in each R and R' and extended by the constant map to the attached sphere. The same reasoning as before shows that $B(Q) = \Omega(m) = \Omega(\sqrt{p})$.

Finally, we combine the two constructions. Let $g > 1$, and let p be arbitrary. Let H be a closed surface of genus $g - 1$ obtained from the double hairy torus construction. Cut H along one of the curves of length 1 to obtain a surface H' of genus $g - 2$ with two boundary components. From above, we can construct a surface Q' with genus 1, p punctures, and one boundary component of length 1 such that $B(Q') = \Omega(\sqrt{p})$. Now glue Q' to H' via an intermediate pair of pants with three boundary components of lengths 1. The resulting surface X has the right topology. The Lipschitz map from H to F_m, where $m \asymp \sqrt{g}$, sends the boundary components of H' to the same point. Therefore, this map extends to all of X by a constant map on Q' and the intermediate pants. Likewise, the Lipschitz map from Q' to $F_{m'}$, $m' \asymp \sqrt{p}$, can be extended to X by a constant map on H' and the intermediate pants. In both cases, for any pants decomposition P on X, the image of P in either F_m or $F_{m'}$ is nontrivial in homology. Therefore, P contains a curve α and a curve α' such that $\ell_X(\alpha) = \Omega(\sqrt{g})$ and $\ell_X(\alpha') = \Omega(\sqrt{p})$. It follows then that

$$B(X) = \Omega(\sqrt{g} + \sqrt{p}) = \Omega(\sqrt{g + p}).$$

This concludes the proof of Lemma 5.9 and this section.

6. Upper bound for height

In this section, we give an upper bound for the height of $\mathcal{M}_{g,p}^{\epsilon}$. For any $X \in \mathcal{M}_{g,p}^{\epsilon}$, we will show that there exists a surface $Y \in \mathcal{B}_{g,p}$ and a map from Y to X such that, for any curve γ on X, the ratio of $\text{Ext}_X(\gamma)$ to $\text{Ext}_Y(\gamma)$ is bounded above by a polynomial function in $\frac{g+p}{\epsilon}$. Then Theorem 2.2 would provide an upper bound of $\log(\frac{g+p}{\epsilon})$ for the Teichmüller distance between X and Y.

Let $X \in \mathcal{M}_{g,p}^{\epsilon}$ be given, and let P be the shortest pants decomposition of X. By Theorem 2.3, there is an upper bound of order $(g + p)$ for the lengths of the curve in P. (The bound does not depend on ϵ.) Let $Y \in \mathcal{B}_{g,p}$ be a surface where there is a pants decomposition P' in the homeomorphism class of P where the curves have lengths ϵ_M. For $\alpha \in P'$, choose the shortest transverse curve $\overline{\alpha}$ to α which intersects

α minimally and is disjoint from the other curves in P'. All transverse curves have length of order 1. The curves in P' and their duals form a set of curves μ, which is usually referred to as a *clean marking* on Y.

There is map from Y to X that sends the curves in P' to curves in P. In fact, the homotopy class of this map is unique up to Dehn twist around curves in P. We choose such a map f so that, for any $\alpha \in P'$, the length of $f(\overline{\alpha})$ in X has an upper bound of order $M = \max\{g + p, \log(1/\epsilon)\}$. To see that such a map exists, let β be any curve in P. We will find a transverse curve $\overline{\beta}$ to β of length $O(M)$. Cutting X along all curves in P, except β, leaves a subsurface containing β which is either a torus with one boundary component or a sphere with four boundary components. The length of each boundary component has an upper bound of order $g + p$ and a lower bound of ϵ. Consider the case in which β is contained in a torus with a boundary component β'. There is a pair of arcs ω and ω' in the torus that are perpendicular to β and β'. One can use elementary hyperbolic geometry to show that the lengths of ω and ω' are $O(M)$. By a surgery using ω and ω' and arcs in β' and β, we can construct a curve that intersects β exactly once. The length of the geodesic representative $\overline{\beta}$ of the curve is bounded above by the sum of the lengths of β, β', ω, and ω', and thus $\ell_X(\overline{\beta}) = O(M)$. A similar construction also works in the case in which β is contained in a sphere with four boundary components.

Hence, we do not distinguish between P and P' and denote them both by P, and we can consider μ as a marking on X.

Let γ be a curve in Y. First we compare the extremal length and the hyperbolic length of γ in Y. Recall the definition of extremal length from the background section. If we pick ρ to be the hyperbolic metric on Y, we have the following inequality just from the definition:

$$\text{Ext}_Y(\gamma) \geq \frac{\ell_\rho(\gamma)^2}{\text{Area}(\rho)} = \frac{\ell_Y(\gamma)^2}{2\pi\chi(Y)} \succ \frac{\ell_Y(\gamma)^2}{g + p}. \tag{6}$$

We now estimate the $\ell_Y(\gamma)$ using its intersection pattern with curves in μ. Recall that curves in μ are either pants curve coming from P, usually denoted by α, or dual curves, denoted by $\overline{\alpha}$. For simplicity, when we write $\alpha \in \mu$, we allow α to be both a pants curve or a dual curve. In the case where α is a pants curve, $\overline{\alpha}$ would be the pants curve that α is dual to. (Every curve is the dual of the dual.)

LEMMA 6.1
We have

$$\ell_Y(\gamma) \succ \sum_{\alpha \in \mu} i(\gamma, \alpha)\ell_Y(\overline{\alpha}). \tag{7}$$

Proof

This is well known. Essentially, every time γ intersects α, it has to cross an annulus with thickness comparable to the length of $\overline{\alpha}$. $\qquad\square$

Note that a curve $\alpha \in \mu$ has length of order 1 in Y. Also,

$$\ell_X(\alpha) \prec M.$$

Hence, we can control how the sum on the right-hand side of (7) changes when we replace Y with X. That is,

$$\sum_{\alpha \in \mu} i(\gamma, \alpha)\ell_Y(\overline{\alpha}) \succ \frac{1}{M} \sum_{\alpha \in \mu} i(\gamma, \alpha)\ell_X(\overline{\alpha}). \tag{8}$$

This sum, in turn, provides an upper bound for the length of γ in X. The following formula was proved in [16, Proposition 3.2].

LEMMA 6.2

For any curve γ on X,

$$2\sum_{\alpha \in \mu} i(\gamma, \alpha)\ell_X(\overline{\alpha}) \geq \ell_X(\gamma). \tag{9}$$

The final step is to compare the hyperbolic length and the extremal length of γ in X.

LEMMA 6.3

For any $X \in \mathcal{M}_{g,p}^{\epsilon}$ and any curve γ,

$$\ell_X(\gamma)^2 \succ \frac{\epsilon^2}{g+p} \operatorname{Ext}_X(\gamma). \tag{10}$$

To prove this, we use the following lemma, which essentially follows from the definition of the extremal length and is a special case of [19, Lemma 4.1].

LEMMA 6.4

Given a metric σ on S and a representative $\hat{\gamma}$ of a simple closed curve γ, let $v = v(\sigma, \hat{\gamma})$ be a number so that the v-neighborhood of $\hat{\gamma}$ is homeomorphic to a standard product $\hat{\gamma} \times [0, 1]$. Then

$$\operatorname{Ext}_\sigma(\gamma) \leq \frac{\operatorname{Area}(\sigma)}{4v^2}.$$

Proof of Lemma 6.3

Let X be a point in $\mathcal{M}_g^\epsilon$. We would like to apply the above lemma. However, when $\ell_X(\gamma)$ is large, the geodesic representative of γ gets exponentially close to itself, that is, $v(\gamma, X) \asymp e^{-\ell_X(\gamma)}$. Hence, to obtain a polynomial bound, we need to perturb γ and push it away from itself as much as possible. Our approach is to triangulate the surface and spread γ locally in each triangle.

Let $\overline{X}$ be the truncated surface obtained from X (see Section 5.1). We may assume that all curves of P are contained in $\overline{X}$. Choose a triangulation T on $\overline{X}$ so that the length of any edge in T has an upper bound of $\epsilon/2$ and a lower bound of order ϵ and so that there is a uniform lower bound on the angles of every triangle in T. This T can be constructed using a Delaunay triangulation on random well-spaced points on $\overline{X}$, which has been constructed explicitly in [15]. The edge lengths of the triangles of T take values in the interval $[\epsilon/k, \epsilon/2]$ for a uniform k. The triangles in this construction have bounded circumradii on the order of ϵ. This fact provides a uniform lower bound for the angles (see [5]).

We may also perturb the triangulation so that γ does not pass through any vertex of T and is not tangent to any edge of T.

CLAIM 6.5
For any edge e in T,

$$\left| \gamma \cap e \right| \leq \frac{2\ell_X(\gamma)}{\epsilon}.$$

Proof of the claim

Let $p, q \in \gamma \cap e$ be two intersection points that appear consecutively along γ. Consider the loop formed by taking the union of the arc $\omega_\gamma \subset \gamma$ connecting p to q with the segment $\omega_e \subset e$ connecting p to q. This loop is essential in X and thus must have length at least ϵ. The length of the arc ω_e is at most $\epsilon/2$. Thus the length of arc ω_γ is at least $\epsilon/2$. Hence, the intersection number is less than $2\ell_X(\gamma)/\epsilon$. $\qquad\qquad\square$

Let $l = \ell_X(\gamma)$. We refer to Figure 10 in the following construction. For a triangle $\Delta \in T$, the restriction of γ to Δ is a collection of $O(l/\epsilon)$ arcs. For every edge e of Δ, let $i_e = |\gamma \cap e|$. When $i_e > 1$, let $\bar{e}$ be the middle third of the segment e. Mark i_e points on $\bar{e}$ subdividing it into $(i_e - 1)$ equal segments. When $i_e = 1$, $\bar{e}$ is just the middle point of e and it contains one marked point. When $i_e = 0$, $\bar{e}$ is empty. We replace the restriction of γ to Δ with a collection of straight segments that start and end with the chosen marked points. Note that the distance between segments is at least of order ϵ/l. We denote the resulting curve by $\hat{\gamma}$, which is a slight perturbation of γ and hence is homotopic to it. For every point in $\hat{\gamma}$, there is a neighborhood with radius of order ϵ that is contained in the union of at most two triangles. Thus, there is
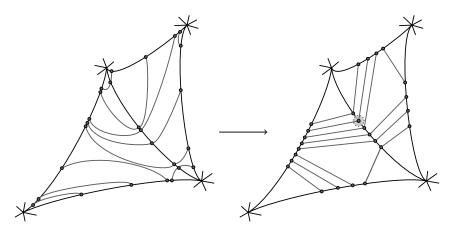
Figure 10. Perturbing arcs of P in Δ.

a neighborhood of $\hat{\gamma}$, with thickness of order ϵ/l, that is standard. That is,

$$v = v(\hat{\gamma}, X) \succ \frac{\epsilon}{l}.$$

Taking σ to be the hyperbolic metric on X and using Minsky's lemma we obtain

$$\mathrm{Ext}_X(\gamma) \prec \frac{\mathrm{Area}(X)}{\epsilon^2/l^2}.$$

The proof follows from the fact that the area of X is of order $(g + p)$. $\qquad\square$

We can now combine these results to obtain the desired upper bound for the height of the thick part.

THEOREM 6.6
We have

$$\mathrm{HD}_T(\mathcal{M}_{g,p}^{\epsilon}, \mathcal{B}_{g,p}) = O\left(\log\left(\frac{g+p}{\epsilon}\right)\right).$$

Proof
We will show that for any $X \in \mathcal{M}_{g,p}^{\epsilon}$ there is $Y \in \mathcal{B}_{g,p}$ so that

$$d_T(Y, X) = O\left(\log\left(\frac{g+p}{\epsilon}\right)\right).$$

Let γ be a curve on X. We can combine equations (6), (7), (8), (9), and (10) (multiply (6), (10), and the square of (7), (8), and (9)) to obtain

$$\mathrm{Ext}_Y(\gamma) \succ \left(\frac{\epsilon}{g+p}\right)^2 \frac{1}{M^2} \mathrm{Ext}_X(\gamma),$$

where $M = \max\{g + p, \log(1/\epsilon)\}$. We overestimate an upper bound for M for uniformity: $M \prec \frac{g+p}{\epsilon}$. Hence, after reorganizing, we have

$$\frac{\text{Ext}_X(\gamma)}{\text{Ext}_Y(\gamma)} \prec \left(\frac{g+p}{\epsilon}\right)^2 M^2 \prec \left(\frac{g+p}{\epsilon}\right)^4.$$

Since this is true for every curve γ, applying Theorem 2.2, we obtain the desired upper bound

$$d_T(X, Y) \leq \frac{1}{2} \sup_\gamma \log \frac{\text{Ext}_X(\gamma)}{\text{Ext}_Y(\gamma)} \prec 2\log\left(\frac{g+p}{\epsilon}\right). \qquad \square$$

Combining Theorem 6.6 with Corollary 5.10, we deduce the following.

COROLLARY 6.7
The Teichmüller and the Lipschitz heights of $\mathcal{M}_{g,p}^\epsilon / \text{Sym}_p$ and $\mathcal{M}_{g,p}^\epsilon$ are all of order $\log(\frac{g+p}{\epsilon})$.

Finally, combining Corollary 6.7 with Corollary 5.2 and the triangle inequality, we obtain the following.

COROLLARY 6.8
The Teichmüller and the Lipschitz diameters of $\mathcal{M}_{g,p}^\epsilon / \text{Sym}_p$ and $\mathcal{M}_{g,p}^\epsilon$ are all of order $\log(\frac{g+p}{\epsilon})$.

Proof
Note that the lower bound for the Lipschitz height of $\mathcal{M}_{g,p}^\epsilon / \text{Sym}_p$ is also a lower bound for the diameter of $\mathcal{M}_{g,p}^\epsilon / \text{Sym}_p$. Thus, $\text{diam}_L(\mathcal{M}_{g,p}^\epsilon / \text{Sym}_p) = \Omega(\log(\frac{g+p}{\epsilon}))$. What remains is to show the upper bound for the *Teichmüller* diameter of $\mathcal{M}_{g,p}^\epsilon$.

Let X, Y be any two points in $\mathcal{M}_{g,p}^\epsilon$. By Theorem 6.6, there exist X' and Y' in $\mathcal{B}_{g,p}$ such that

$$d_T(X, X') \prec \log\left(\frac{g+p}{\epsilon}\right), \qquad d_T(Y, Y') \prec \log\left(\frac{g+p}{\epsilon}\right).$$

By Corollary 5.2, we have

$$d_T(X', Y') \prec \log\left(\frac{g+p}{\epsilon}\right).$$

Thus, by the triangle inequality (note that the Teichmüller metric is symmetric),

$$d_T(X, Y) \leq d_T(X, X') + d_T(X', Y') + d_T(Y', Y) \prec 3\log\left(\frac{g+p}{\epsilon}\right). \qquad \square$$

Appendix. The metric of Whitehead moves on the space of graphs

In this section, we compute the asymptotic diameter of $\mathrm{Graph}(g, p)$ and $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$ in the metric of Whitehead moves. Recall that this is equivalent to computing the asymptotic diameter of the moduli space of pants decompositions on $S_{g,p}$ and $S_{g,p}/\mathrm{Sym}_p$ in the metric of elementary moves. The main results and proofs are extrapolated from known results. Our purpose for writing this section is to unify what is known and to put it into our context. This section can be read independently from the rest of the paper.

Let d_W represent the metric of Whitehead moves: two graphs have distance 1 if they differ by one Whitehead move. The main results we would like to present are as follows. Note that, as a matter of convention, we have $x \log(x) = 0$, when $x = 0$.

THEOREM A.1 (Labeled punctures)
We have

$$\mathrm{diam}_W\big(\mathrm{Graph}(g, p)\big) \asymp (g + p) \log(g + p).$$

THEOREM A.2 (Unlabeled punctures)
We have

$$\mathrm{diam}_W\big(\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big) \asymp g \log(g + p) + (g + p).$$

Our arguments for the two theorems are based on the work of [3] and [21]. In the case $p = 0$, we also refer to [8] for an alternate proof of $\mathrm{diam}_W(\mathrm{Graph}(g)) \asymp g \log(g)$.

We first argue for the upper bounds. The upper bound in Theorem A.1 follows easily from Theorem D. Namely, since a simultaneous Whitehead move is a composition of at most $g + p$ Whitehead moves, we have

$$\mathrm{diam}_W\big(\mathrm{Graph}(g, p)\big) \prec (g + p) \log(g + p).$$

Similarly, if $g \geq p$, then the same argument also shows that

$$\mathrm{diam}_W\big(\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big) \prec (g + p) \log(g + p) \prec g \log(g + p).$$

This argument fails for $p > g$, so we present another one that works in general.

When $g = 0$ and $p \geq 2$, then $\mathrm{Graph}(0, p)/\mathrm{Sym}_p$ is the space of unlabeled (unrooted) trees with p ends. This case has been dealt with in [21], where it was shown that

$$\mathrm{diam}_W\big(\mathrm{Graph}(0, p)/\mathrm{Sym}_p\big) \asymp p. \tag{11}$$

(In fact, their estimate is very explicit with only a small additive error.)

Now suppose that $g \geq 1$. We use induction on g and assume Theorem A.2 for all smaller values of g. Let $\Gamma_0 \in \mathrm{Graph}(g, p)/\mathrm{Sym}_p$ be a graph which has a loop of length 1 (an edge e starting and ending at the same vertex). Let $\Gamma_1 \in \mathrm{Graph}(g-1, p+1)/\mathrm{Sym}_{p+1}$ be the graph obtained from Γ_0 after removing e. Let Γ be any other graph in $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$. It is sufficient to show that Γ is a distance $O(\log(g + p))$ from a graph Γ' with a loop e' of length 1. This is because $\Gamma' \setminus \{e'\}$ can be transformed to Γ_1 in

$$O\big((g - 1)\log\big((g - 1) + (p + 1)\big) + \big((g - 1) + (p + 1)\big)\big)$$

many steps using induction, and the theorem follows.

Let T be a spanning subtree of Γ. Then T has $(g + p)$ ends. Using (11), we can use Whitehead moves with support in T to transform Γ to a graph with a spanning tree of diameter $\log(g + p)$. We denote the new graph by Γ as well. Then Γ has a loop of length of order $\log(g + p)$. That is, for $n \prec \log(g + p)$, there is a linear map $\phi \colon [0, n] \to \Gamma$ such that the image is a nontrivial loop, $\phi(0) = \phi(n)$, and $\phi([i - 1, i])$ is an edge of Γ. One can always apply a Whitehead move at an edge on the loop to shorten the length of the loop by at least one. Hence, after $O(\log(g + p))$-many moves, there is a loop of length one. This proves the upper bound.

The lower bounds for both theorems are obtained by a counting argument. We first need an asymptotic formula for the cardinalities of $\mathrm{Graph}(g, p)$ and $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$. For the following, we introduce the notation $A \sim B$ to mean $A \asymp c^{g+p}B$, where c is a uniform constant. Since we will be applying the logarithm later, exponential factors can be ignored.

By the work of [3], we have

$$\big|\mathrm{Graph}(g, p)\big| \sim \frac{(6g + 2p)!}{g!(2g + p)!}.$$

Up to exponential factors, this simplifies to

$$\big|\mathrm{Graph}(g, p)\big| \sim \frac{(g + p)^{6g+2p}}{g^g(g + p)^{2g+p}} \sim \frac{(g + p)^{4g+p}}{g^g}. \tag{12}$$

Similarly, we have

$$\big|\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big| \sim \frac{(6g + 2p)!}{g!p!(2g + p)!} \sim \frac{(g + p)^{4g+p}}{g^g\, p^p}. \tag{13}$$

To finish the proof, we need the result in [21], which gives an upper bound of the form c^{g+p+r} for the cardinality of a ball of radius r in $\mathrm{Graph}(g, p)$ and in $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$, where c is some fixed constant (see [21, Theorem 2.3]). (In fact, their theorem is much more general and applies to any space of shapes when shapes

are allowed to evolve through locally supported elementary moves, such as Whitehead moves.)

Let r be the minimal number such that the ball of radius r contains the whole space $\mathrm{Graph}(g, p)$. By (12), we have

$$c^{g+p+r} \succ \frac{(g+p)^{4g+p}}{g^g}.$$

Taking the logarithm of both sides, we obtain

$$g + p + r \succ (g + p) \log(g + p) \implies \mathrm{diam}_W\big(\mathrm{Graph}(g, p)\big) \geq r \succ (g + p) \log(g + p).$$

This completes the proof of Theorem A.1.

For unlabeled punctures, we prove two lower bounds. Since our errors are multiplicative, their sum is also a lower bound.

Let $\Gamma, \Gamma' \in \mathrm{Graph}(g, p)/\mathrm{Sym}_p$ be, respectively, of diameters of order $(g + p)$ and $\log(g + p)$. Consider a sequence $\Gamma_1 \cdots \Gamma_n$ of Whitehead moves taking Γ to Γ'. Since

$$\big|\mathrm{diam}(\Gamma_i) - \mathrm{diam}(\Gamma_{i+1})\big| \leq 1$$

we must have

$$\mathrm{diam}_W\big(\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big) \succ (g + p) - \log(g + p) \succ (g + p). \qquad (14)$$

That is the first lower bound.

When $g \geq p$, (13) reduces to

$$\big|\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big| \sim g^g \left(\frac{g}{p}\right)^p. \qquad (15)$$

On the other hand, when $p > g$, we have

$$\big|\mathrm{Graph}(g, p)/\mathrm{Sym}_p\big| \sim p^g \left(\frac{p}{g}\right)^g. \qquad (16)$$

Let r be the minimal number such that the ball of radius r contains $\mathrm{Graph}(g, p)/\mathrm{Sym}_p$. When $g \geq p$, (15) implies

$$r \succ g \log(g) + p \log \frac{g}{p} \succ g \log(g + p).$$

Similarly, when $p > g$, (16) implies

$$r \succ g \log(p) + g \log \frac{p}{g} \succ g \log(g + p).$$

Using the above and (14) we obtain

$$\mathrm{diam}_W\big(\mathrm{Graph}(g,\,p)/\,\mathrm{Sym}_p\big) \succ g\log(g+p) + (g+p).$$

This finishes the proof of Theorem A.2.

Acknowledgments. We thank Curt McMullen for pointing out to us the example of the double hairy torus. We also thank Saul Schleimer for a helpful conversation which contributed to a simplification of the arguments contained in the last section. Finally, we thank the referee for helpful comments.

References

[1]   L. V. AHLFORS, *Conformal Invariants: Topics in Geometric Function Theory*, McGraw-Hill, New York, 1973. MR 0357743. *(1839)*

[2]   F. BALACHEFF and H. PARLIER, *Bers' constants for punctured spheres and hyperelliptic surface*, J. Topol. Anal. **4** (2012), 271–296. MR 2982444. DOI 10.1142/S179352531250015X. *(1841, 1857, 1863)*

[3]   B. BOLLOBÁS, *The asymptotic number of unlabelled regular graphs*, J. London Math. Soc. (2) **26** (1982), 201–206. MR 0675164. DOI 10.1112/jlms/s2-26.2.201. *(1837, 1872, 1873)*

[4]   P. BOSE, J. CZYZOWICZ, Z. GAO, P. MORIN, and D. R. WOOD, *Simultaneous diagonal flips in plane triangulations*, J. Graph Theory **54** (2007), 307–330. MR 2292668. DOI 10.1002/jgt.20214. *(1836)*

[5]   W. BRESLIN, *Thick triangulations of hyperbolic n-manifolds*, Pacific J. Math. **241** (2009), 215–225. MR 2507575. DOI 10.2140/pjm.2009.241.215. *(1869)*

[6]   P. BUSER, *Cubic graphs and the first eigenvalue of a Riemann surface*, Math. Z. **162** (1978), 87–99. MR 0505920. DOI 10.1007/BF01437826. *(1836, 1861)*

[7]   ———, *Geometry and Spectra of Compact Riemann Surfaces*, Progr. Math. **106**, Birkhäuser, Boston, 1992. MR 1183224. *(1840, 1863)*

[8]   W. CAVENDISH, *Growth of the diameter of the pants graph modulo the mapping class group*, preprint, https://web.math.princeton.edu/~wcavendi/PantsModMCG.pdf, 2010. *(1837, 1872)*

[9]   W. CAVENDISH and H. PARLIER, *Growth of the Weil–Petersson diameter of moduli space*, Duke Math. J. **161** (2012), 139–171. MR 2872556. DOI 10.1215/00127094-1507312. *(1834)*

[10]   M. CULLER and K. VOGTMANN, *Moduli of graphs and automorphisms of free groups*, Invent. Math. **84** (1986), 91–119. MR 0830040. DOI 10.1007/BF01388734. *(1837, 1854)*

[11]   B. FARB and D. MARGALIT, *A Primer on Mapping Class Groups*, Princeton Math. Ser. **49**, Princeton Univ. Press, Princeton, N.J., 2012. MR 2850125. *(1838)*

[12]   F. P. GARDINER and N. LAKIC, *Quasiconformal Teichmüller Theory*, Math. Surveys Monogr. **76**, Amer. Math. Soc., Providence, 2000. MR 1730906. *(1838)*

[13]   J. HUBBARD, *Teichmüller Theory and Applications to Geometry, Topology and Dynamics*, Matric Edition, Ithaca, N.Y., 2006. MR 2245223. *(1838)*

[14]   S. P. KERCKHOFF, *The asymptotic geometry of Teichmüller space*, Topology **19** (1980), 23–41. MR 0559474. DOI 10.1016/0040-9383(80)90029-4. *(1839)*

[15]   G. LEIBON and D. LETSCHER, "Delaunay triangulations and Voronoi diagrams for Riemannian manifolds" in *Proceedings of the Sixteenth Annual Symposium on Computational Geometry (Hong Kong, 2000)*, ACM, New York, 2000, 341–349. MR 1802283. DOI 10.1145/336154.336221. *(1869)*

[16]   A. LENZHEN, K. RAFI, and J. TAO, *Bounded combinatorics and the Lipschitz metric on Teichmüller space*, Geom. Dedicata **159** (2012), 353–371. MR 2944537. DOI 10.1007/s10711-011-9664-2. *(1868)*

[17]   W. B. R. LICKORISH, *A finite set of generators for the homeotopy group of a* 2-*manifold*, Proc. Cambridge Philos. Soc. **60** (1964), 769–778. MR 0171269. *(1837)*

[18]   G. A. MARGULIS, *Explicit constructions of expanders* (in Russian), Problemy Peredachi Informatsii **9** (1973), 71–80. MR 0484767. *(1861)*

[19]   Y. N. MINSKY, *Extremal length estimates and product regions in Teichmüller space*, Duke Math. J. **83** (1996), 249–286. MR 1390649. DOI 10.1215/S0012-7094-96-08310-6. *(1868)*

[20]   M. S. PINSKER, "On the complexity of a concentrator" in 7*th International Telegraffic Conference*, 1973, 318/1–318/4. *(1836, 1861)*

[21]   D. D. SLEATOR, R. E. TARJAN, and W. P. THURSTON, *Short encodings of evolving structures*, SIAM J. Discrete Math. **5** (1992), 428–450. MR 1172751. DOI 10.1137/0405034. *(1837, 1872, 1873)*

[22]   W. P. THURSTON, *Minimal stretch maps between hyperbolic surfaces*, preprint, arXiv:math.GT/9801039 [math.GT]. *(1834, 1839)*

[23]   S. A. WOLPERT, *The length spectra as moduli for compact Riemann surfaces*, Ann. of Math. (2) **109** (1979), 323–351. MR 0528966. DOI 10.2307/1971114. *(1839)*

Rafi

Department of Mathematics, University of Oklahoma, Norman, Oklahoma 73019-0315, USA; rafi@math.ok.edu

Tao

Department of Mathematics, University of Utah, Salt Lake City, Utah 84112-0090, USA; jing@math.utah.edu