

## ISOPERIMETRIC INEQUALITIES FOR THE HANDLEBODY GROUPS

URSULA HAMENSTÄDT AND SEBASTIAN HENSEL

ABSTRACT. We show that the mapping class group of a handlebody  $V$  of genus at least 2 has a Dehn function of at most exponential growth type.

### 1. INTRODUCTION

A handlebody of genus  $g \geq 2$  is a compact orientable 3-manifold  $V$  whose boundary  $\partial V$  is a closed surface of genus  $g$ . The *handlebody group*  $\text{Map}(V)$  is the group of isotopy classes of orientation preserving homeomorphisms of  $V$ . It can be identified with a subgroup of the mapping class group  $\text{Map}(\partial V)$  of  $\partial V$  via the restriction homomorphism. This subgroup is of infinite index, and it surjects onto the outer automorphism group of the fundamental group of  $V$ , which is the free group with  $g$  generators.

The handlebody group is finitely presented [Wa98]. Thus it can be equipped with a word norm which is unique up to quasi-isometry. The large-scale geometry induced by such a metric is not compatible with the large-scale geometry of the ambient group  $\text{Map}(\partial V)$ . Namely, we showed in [HH11] that for every  $g \geq 2$  the handlebody group is an exponentially distorted subgroup of the mapping class group of  $\partial V$ . Here, a finitely generated subgroup  $H < G$  of a finitely generated group  $G$  is called *exponentially distorted* if the following holds. First, the word norm in  $H$  of every element  $h \in H$  can be bounded from above by an exponential function in the word norm of  $h$  in  $G$ . On the other hand, there is no such bound with sub-exponential growth rate.

As a consequence, geometric properties of the mapping class group may be very different from geometric properties of the handlebody group.

A particularly useful geometric invariant of a finitely presented group  $G$  is its *Dehn function*, which can be defined as the isoperimetric function of a presentation complex for  $G$  (see Section 4 for a complete definition). Although the Dehn function itself depends on the choice of a finite presentation of  $G$ , its growth type does not. In fact, the growth type of the Dehn function is a quasi-isometry invariant of  $G$ .

The mapping class group  $\text{Map}(\partial V)$  of  $\partial V$  is automatic [Mo95] and hence has a quadratic Dehn function. On the other hand, for  $g \geq 3$  the Dehn function of the outer automorphism group  $\text{Out}(F_g)$  of a free group  $F_g$  on  $g$  generators is exponential [HV96, BV95, BV10, HM10]. Since the kernel of the projection from the handlebody

---

Received by the editors February 7, 2012.

AMS subject classification: 20F65, 57M07.

Both authors are partially supported by the Hausdorff Center Bonn and the Hausdorff Institut Bonn. The second author is supported by the Max-Planck Institut für Mathematik Bonn.

group to  $\text{Out}(F_g)$  is infinite (and even infinitely generated [McC85]), this fact does not restrict the Dehn function of  $\text{Map}(V)$ .

The goal of this note is to give an upper bound for the Dehn function of  $\text{Map}(V)$ . We show

**Theorem.** *The handlebody group  $\text{Map}(V)$  satisfies an exponential isoperimetric inequality, i.e. the growth of its Dehn function is at most exponential.*

The strategy of proof for the theorem is similar to the strategy employed in [HV96] to show an exponential upper bound for the Dehn function of outer automorphism groups of free groups. We construct a graph which is a geometric model for the handlebody group. A similar graph is used in [HH11] to show that the distortion of the handlebody group in the mapping class group is at most exponential. Vertices of this graph correspond to isotopy classes of special cell decompositions of  $\partial V$  containing the boundary of a *simple disk system* in their one-skeleton. A simple disk system is a collection of pairwise disjoint, pairwise non-homotopic embedded disks in  $V$  which decompose  $V$  into simply connected regions. We then use a surgery procedure for disk systems to define a distinguished class of paths in the handlebody group. The example at the end of this note shows that these paths are in general not uniform quasi-geodesics. However, they are sufficiently well-behaved so that they can be used to fill a cycle with area bounded by an exponential function in the length of the cycle.

The organization of this note is as follows. In Section 2 we introduce disk systems and construct special paths in the disk system graph. Section 3 discusses a geometric model for the handlebody group. This model is used in Section 4 for the proof of the main theorem.

**Acknowledgement.** We would like to thank the anonymous referee for providing us with helpful comments and suggestions.

## 2. DISK EXCHANGE PATHS

In this section we collect some facts about properly embedded disks in a handlebody  $V$  of genus  $g \geq 2$ . In particular, we describe a surgery procedure that is central to the construction of paths in the handlebody group used in later sections.

A disk  $D$  in  $V$  is called *essential* if it is properly embedded and if  $\partial D$  is an essential simple closed curve on  $\partial V$ . A *disk system* for  $V$  is a set of pairwise disjoint essential disks in  $V$  no two of which are homotopic. Here, and in the sequel, all homotopies are required to be proper. A disk system is called *simple* if all of its complementary components are simply connected. It is called *reduced* if in addition it has a single complementary component.

We usually consider disks and disk systems only up to proper isotopy. Furthermore, we will always assume that disks and disk systems are in minimal position if they intersect. Here we say that two disk systems  $\mathcal{D}_1, \mathcal{D}_2$  are in *minimal position* if their boundary multicurves intersect in the minimal number of points in their respective isotopy classes and if every component of  $\mathcal{D}_1 \cap \mathcal{D}_2$  is an embedded arc in  $\mathcal{D}_1 \cap \mathcal{D}_2$  with endpoints in  $\partial \mathcal{D}_1 \cap \partial \mathcal{D}_2$ . Note that although the minimal intersection pattern of the boundary curves  $\partial \mathcal{D}_1$  and  $\partial \mathcal{D}_2$  is determined by the isotopy classes of  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , this is not true for the minimal position of the disks  $\mathcal{D}_1$  and  $\mathcal{D}_2$  themselves. In particular the intersection pattern  $\mathcal{D}_1 \cap \mathcal{D}_2$  between disks in minimal position is not determined by the isotopy classes of  $\mathcal{D}_1$  and  $\mathcal{D}_2$ .

The following easy fact will be used frequently throughout the article.

**Lemma 2.1.** *The handlebody group acts transitively on the set of isotopy classes of reduced disk systems. Every mapping class of  $\partial V$  which fixes the isotopy class of a simple disk system is contained in the handlebody group.*

*Proof.* The first claim follows from the fact that the complement of a reduced disk system in  $V$  is a ball with  $2g$  spots and any two such manifolds are homeomorphic. The second claim is immediate since every homeomorphism of the boundary of a spotted ball extends to a homeomorphism of the interior.  $\square$

Let  $\mathcal{D}$  be a disk system. An *arc relative to  $\mathcal{D}$*  is a continuous embedding  $\rho : [0, 1] \rightarrow \partial V$  whose endpoints  $\rho(0)$  and  $\rho(1)$  are contained in  $\partial \mathcal{D}$ . Such an arc  $\rho$  is called *essential* if it cannot be homotoped into  $\partial \mathcal{D}$  with fixed endpoints as an arc on  $\partial V$ . In the sequel we always assume that arcs are essential and that the number of intersections of  $\rho$  with  $\partial \mathcal{D}$  is minimal in its isotopy class.

Choose an orientation of the curves in  $\partial \mathcal{D}$ . Since  $\partial V$  is oriented, this choice determines a left and a right side of a component  $\alpha$  of  $\partial \mathcal{D}$  in a small annular neighborhood of  $\alpha$  in  $\partial V$ . We then say that an endpoint  $\rho(0)$  (or  $\rho(1)$ ) of an arc  $\rho$  *lies to the right (or to the left) of  $\alpha$* , if a small neighborhood  $\rho([0, \epsilon])$  (or  $\rho([1 - \epsilon, 1])$ ) of this endpoint is contained in the right (or left) side of  $\alpha$  in a small annulus around  $\alpha$ . A *wave relative to  $\mathcal{D}$*  is an arc both of whose endpoints lie on the same side of the boundary  $\partial D$  of a disk  $D$  in  $\mathcal{D}$ , and whose interior is disjoint from  $\partial \mathcal{D}$ .

For every disk  $E$  which is not disjoint from  $\mathcal{D}$  there are at least two distinct waves of  $\partial E$  relative to  $\mathcal{D}$ . Namely, consider the intersection pattern of  $E$  with the disks in  $\mathcal{D}$ . Then there are at least two components  $E', E''$  of  $E \setminus \mathcal{D}$  whose boundary consists of a single arc in  $\partial E$  and a single arc in  $\mathcal{D}$ . The boundary arc in  $\partial E$  of such an outermost component is a wave relative to  $\mathcal{D}$ .

Let now  $\mathcal{D}$  be a simple disk system and let  $\rho$  be a wave whose endpoints are contained in the boundary of some disk  $D \in \mathcal{D}$ . Then  $\partial D \setminus \{\rho(0), \rho(1)\}$  is the union of two (open) intervals  $\gamma_1$  and  $\gamma_2$ . Put  $\alpha_i = \gamma_i \cup \rho$ . Up to isotopy,  $\alpha_1$  and  $\alpha_2$  are simple closed curves in  $\partial V$  which are disjoint from  $\partial \mathcal{D}$  (compare [St99] for this construction). Therefore both  $\alpha_1$  and  $\alpha_2$  bound disks in the handlebody which we denote by  $Q_1$  and  $Q_2$ . We say that  $Q_1$  and  $Q_2$  are obtained from  $D$  by *simple surgery along the wave  $\rho$* .

The following observation is well known (compare [HH11], [M86, Lemma 3.2], or [St99]).

**Lemma 2.2.** *If  $\Sigma$  is a reduced disk system and  $\rho$  is a wave with endpoints on  $D \in \Sigma$ , then for exactly one choice of the disks  $Q_1, Q_2$  defined as above, say the disk  $Q_1$ , the disk system obtained from  $\Sigma$  by replacing  $D$  by  $Q_1$  is reduced.*

The disk  $Q_1$  is characterized by the requirement that the two spots in the boundary of  $V \setminus \Sigma$  corresponding to the two copies of  $D$  are contained in distinct connected components of  $V \setminus (\Sigma \cup Q_1)$ . It only depends on  $\Sigma$  and the wave  $\rho$ . We call the interval  $\gamma_1$  used in the construction of the disk  $Q_1$  the *preferred interval defined by the wave*.

**Definition 2.3.** Let  $\Sigma$  be a reduced disk system. A *disk exchange move* is the replacement of a disk  $D \in \Sigma$  by a disk  $D'$  which is disjoint from  $\Sigma$  and such that  $(\Sigma \setminus D) \cup D'$  is a reduced disk system. If  $D'$  is determined as in Lemma 2.2 by a

wave of a disk in a disk system  $\mathcal{D}$  then the modification is called a *disk exchange move of  $\Sigma$  in direction of  $\mathcal{D}$*  or simply a *directed disk exchange move*.

A sequence  $(\Sigma_i)$  of reduced disk systems is called a *disk exchange sequence in direction of  $\mathcal{D}$*  (or *directed disk exchange sequence*) if each  $\Sigma_{i+1}$  is obtained from  $\Sigma_i$  by a disk exchange move in direction of  $\mathcal{D}$ .

The following lemma is an easy consequence of the fact that simple surgery reduces the geometric intersection number (see e.g [HH11, Lemma 5.4] for a proof).

**Lemma 2.4.** *Let  $\Sigma_1$  be a reduced disk system and let  $\mathcal{D}$  be any other disk system. Then there is a disk exchange sequence  $\Sigma_1, \dots, \Sigma_n$  in direction of  $\mathcal{D}$  such that  $\Sigma_n$  is disjoint from  $\mathcal{D}$ .*

To estimate the growth rate of the Dehn function of the handlebody group we will need to compare disk exchange sequences starting in disjoint reduced disk systems. This is made possible by considering a different type of surgery sequence for disk systems which we describe in the remainder of this section.

To this end, let  $\mathcal{D}$  be any simple disk system and let  $\rho$  be a wave. A *full disk replacement defined by  $\rho$*  modifies  $\mathcal{D}$  to a simple disk system  $\mathcal{D}'$  as follows. Let  $D \in \mathcal{D}$  be the disk containing the endpoints of the wave  $\rho$ . Replace  $D$  by both disks  $Q_1, Q_2$  obtained from  $D$  by the simple surgery defined by  $\rho$ . The disks  $Q_1, Q_2$  are disjoint from each other and from  $\mathcal{D}$ . If one (or both) of these disks is isotopic to a disk  $Q$  from  $\mathcal{D} \setminus D$  then this disk will be discarded (i.e. we retain a single copy of  $Q$ ; compare [Ha95] for a similar construction). We say that a sequence  $(\mathcal{D}_i)$  is a *full disk replacement sequence in direction of  $\mathcal{D}$*  (or *directed full disk replacement sequence*) if each  $\mathcal{D}_{i+1}$  is obtained from  $\mathcal{D}_i$  by a full disk replacement along a wave contained in  $\partial\mathcal{D}$ .

The following two lemmas relate full disk replacement sequences to disk exchange sequences. Informally, these lemmas state that every directed disk exchange sequence may be extended to a full disk replacement sequence, and conversely every full disk replacement sequence contains a disk exchange sequence. To make this idea precise, we use the following

**Definition 2.5.** Let  $\mathcal{D}$  be an arbitrary disk system. Suppose that  $\mathcal{D}_1, \dots, \mathcal{D}_n$  is a full disk replacement sequence in direction of  $\mathcal{D}$  and that  $\Sigma_1, \dots, \Sigma_k$  is a disk exchange sequence in direction of  $\mathcal{D}$ .

We say that the sequences  $(\mathcal{D}_i)$  and  $(\Sigma_i)$  are *compatible*, if there is a non-decreasing surjective map  $r : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$  such that  $\Sigma_{r(i)} \subset \mathcal{D}_i$  for all  $i$ .

**Lemma 2.6.** *Let  $\Sigma$  be a reduced disk system, let  $\mathcal{D}$  be a simple disk system containing  $\Sigma$  and let  $\mathcal{D} = \mathcal{D}_1, \dots, \mathcal{D}_n$  be a full disk replacement sequence in direction of a disk system  $\mathcal{D}'$ . Then there is a disk exchange sequence  $\Sigma = \Sigma_1, \dots, \Sigma_k$  in direction of  $\mathcal{D}'$  which is compatible with  $(\mathcal{D}_i)$ .*

*Proof.* We proceed by induction on the length of the full disk replacement sequence  $(\mathcal{D}_i)$ . If this length equals zero there is nothing to show. Assume that the claim holds true whenever this length does not exceed  $n - 1$  for some  $n > 0$ .

Let  $\mathcal{D}_1, \dots, \mathcal{D}_n$  be a full disk replacement sequence of length  $n$  and let  $\Sigma \subset \mathcal{D}_1$  be a reduced disk system. Let  $D \in \mathcal{D}_1$  be the disk replaced in the full disk replacement move connecting  $\mathcal{D}_1$  to  $\mathcal{D}_2$ .

If  $D \in \Sigma$  then for one of the two disks obtained from  $D$  by simple surgery, say the disk  $D'$ , the disk system  $\Sigma_2 = (\Sigma \setminus D) \cup D'$  is reduced. However,  $\Sigma_2 \subset \mathcal{D}_2$  and the claim now follows from the induction hypothesis.

If  $D \notin \Sigma$  then  $\Sigma \subset \mathcal{D}_2$  by definition and once again, the claim follows from the induction hypothesis.  $\square$

**Lemma 2.7.** *Let  $\Sigma_1, \dots, \Sigma_k$  be a disk exchange sequence of reduced disk systems in direction of a disk system  $\mathcal{D}'$ . Then for every simple disk system  $\mathcal{D}_1 \supset \Sigma_1$  there is a full disk replacement sequence  $\mathcal{D}_1, \dots, \mathcal{D}_n$  in direction of  $\mathcal{D}'$  which is compatible with  $(\Sigma_i)$ .*

*Proof.* We proceed by induction on the length  $k$  of the directed disk exchange sequence.

The case  $k = 1$  is trivial, so assume that the lemma holds true for directed disk exchange sequences of length at most  $k - 1$  for some  $k \geq 2$ . Let  $\Sigma_1, \dots, \Sigma_k$  be a directed disk exchange sequence of length  $k$ . Suppose  $\Sigma_2$  is obtained from  $\Sigma_1$  by replacing a disk  $D \in \Sigma_1$ . Let  $\rho$  be the wave with endpoints on  $D$  defining the disk replacement, and let  $D_2$  be the disk in  $\Sigma_2$  which is the result of the simple surgery.

We distinguish two cases. In the first case,  $\rho \cap \mathcal{D}_1 = \rho \cap D$ . Then  $\rho$  is a wave relative to  $\mathcal{D}_1$ . Let  $\mathcal{D}_2$  be the disk system obtained from  $\mathcal{D}_1$  by the full disk replacement defined by  $\rho$ . One of the two disks obtained by simple surgery along  $\rho$  is the disk  $D_2$  and hence  $D_2 \in \mathcal{D}_2$ . The claim now follows from the induction hypothesis, applied to the disk exchange sequence  $\Sigma_2, \dots, \Sigma_k$  of length  $k - 1$  and the simple disk system  $\mathcal{D}_2$  containing  $\Sigma_2$ .

In the second case, the wave  $\rho$  intersects  $\mathcal{D}_1 \setminus D$ . Then  $\rho \setminus (\mathcal{D}_1 \setminus D)$  contains a component  $\rho'$  which is a wave with endpoints on a disk  $Q \in \mathcal{D}_1 \setminus \{D\}$ . A replacement of the disk  $Q$  by both disks obtained from  $Q$  by simple surgery using the wave  $\rho'$  reduces the number of intersection components of  $\rho \cap (\mathcal{D}_1 \setminus D)$ . Moreover, the resulting disk system contains  $D$ . In finitely many surgery steps, say  $s \geq 1$  steps, we obtain a simple disk system  $\mathcal{D}_s$  with the following properties.

- (1)  $\mathcal{D}_s$  contains  $D$  and is obtained from  $\mathcal{D}_1$  by a full disk replacement sequence.
- (2)  $\rho \cap (\mathcal{D}_s \setminus D) = \emptyset$ .

Define  $r(i) = 1$  for  $i = 1, \dots, s$ , where  $r$  is the function required in the definition of compatibility. We now can use the procedure from the first case above, applied to  $\Sigma_1, \mathcal{D}_s$  and  $\rho$  to carry out the induction step. This completes the proof of the lemma.  $\square$

Lemma 2.6 and Lemma 2.7 show the following. Let  $\Sigma, \Sigma'$  be reduced disk systems which are disjoint. Let  $\mathcal{D}$  be any disk system. Then for every disk exchange sequence connecting  $\Sigma$  to a disk system disjoint from  $\mathcal{D}$  there is a disk exchange sequence connecting  $\Sigma'$  to a disk system disjoint from  $\mathcal{D}$  so that these sequences are contained in the same full disk replacement sequence.

### 3. THE GRAPH OF RIGID RACKS

The goal of this section is to describe a construction of paths in the handlebody group whose geometry is easy to control. A version of these paths was already used in [HH11] to establish an upper bound for the distortion of the handlebody group in the mapping class group. The main objects are given by the following

**Definition 3.1.** A rack  $R$  in  $V$  is given by a reduced disk system  $\Sigma(R)$ , called the *support system* of the rack  $R$ , and a collection of pairwise disjoint essential embedded arcs in  $\partial V \setminus \partial\Sigma(R)$  with endpoints on  $\partial\Sigma(R)$ , called *ropes*, which are pairwise non-homotopic relative to  $\partial\Sigma(R)$ . At each side of a support disk  $D \in \Sigma(R)$ , there is at least one rope which ends at the disk and approaches the disk from this side. A rack is called *large* if the set of ropes decomposes  $\partial V \setminus \partial\Sigma(R)$  into simply connected regions.

We will consider racks up to an equivalence relation called “rigid isotopy” which is defined as follows.

**Definition 3.2.** i) Let  $R$  be a large rack. The union of the boundary of the support system and the system of ropes of  $R$  defines the 1–skeleton of a cell decomposition of the surface  $\partial V$  which we call the *cell decomposition induced by  $R$* .  
ii) Let  $R$  and  $R'$  be racks. We say that  $R$  and  $R'$  are *rigidly isotopic* if there is an isotopy of  $\partial V$  which maps the support system of  $R$  to the support system of  $R'$  and defines an isotopy of the cell decompositions induced by  $R$  and  $R'$ .

Since stabilizers of cell-decompositions in the mapping class group are finite, and the stabilizer in the mapping class group of a reduced disk system is contained in the handlebody group we have the following

**Lemma 3.3.** *The handlebody group acts on the set of rigid isotopy classes of large racks with finite quotient and finite point stabilizers.*

For simplicity of notation, we call a rigid isotopy class of a large rack simply a *rigid rack*. Lemma 3.3 allows us to use rigid racks as the vertex set of a  $\text{Map}(V)$ –graph. More precisely, we make the following

**Definition 3.4.** The *graph of rigid racks*  $\mathcal{RR}_K(V)$  is the graph whose vertex set is the set of large rigid racks. Two such vertices corresponding to racks  $R$  and  $R'$  are joined by an edge if the support systems of  $R$  and  $R'$  are disjoint, and furthermore up to isotopy, the 1–skeleta of the cell decompositions induced by  $R$  and  $R'$  intersect in at most  $K$  points.

In [HH11] it is shown that the constant  $K$  may be chosen in such a way that  $\mathcal{RR}_K(V)$  is connected (see the discussion after Lemma 7.3 in [HH11]). Furthermore, we may choose  $K$  large enough so that any two racks with the same support system  $\Sigma$  can be joined by an edge-path all of whose vertices correspond to racks with support system  $\Sigma$ . In the sequel, we will always use this choice of  $K$  and suppress the mention of  $K$  from our notation. It then follows from Lemma 3.3 and the Svarc-Milnor lemma that the graph  $\mathcal{RR}(V)$  is quasi-isometric to  $\text{Map}(V)$ .

For a reduced disk system  $\Sigma$  let  $\mathcal{RR}(V, \Sigma)$  be the complete subgraph of  $\mathcal{RR}(V)$  whose vertices are rigid racks with support system  $\Sigma$ . We have

**Lemma 3.5.** *Let  $\Sigma$  be a reduced disk system for  $V$ .*

- i)  $\mathcal{RR}(V, \Sigma)$  is a connected subgraph of  $\mathcal{RR}(V)$  which is equivariantly quasi-isometric to the stabilizer of  $\partial\Sigma$  in the mapping class group of  $\partial V$ .
- ii)  $\mathcal{RR}(V, \Sigma)$  is quasi-isometrically embedded in  $\mathcal{RR}(V)$ .

*Proof.*  $\mathcal{RR}(V, \Sigma)$  is connected by choice of the constant  $K$  in the definition of the graph of rigid racks.



Let  $G$  be the stabilizer of  $\partial\Sigma$  in the mapping class group of  $\partial V$ . The group  $G$  is contained in the handlebody group since every homeomorphism of the boundary of a spotted ball extends to the interior. The group  $G$  acts on  $\mathcal{RR}(V, \Sigma)$  with finite quotient and finite point stabilizers. To show this, note that there are only finitely many isotopy classes of cell decompositions of a bordered sphere with uniformly few cells up to the action of the mapping class group. Thus, by the Svarc-Milnor lemma,  $\mathcal{RR}(V, \Sigma)$  is equivariantly quasi-isometric to  $G$ . This proves part *i*).

The stabilizer  $G$  of  $\partial\Sigma$  is quasi-isometrically embedded in the full mapping class group of  $\partial V$  (see [MM00] or [H09b, Theorem 2]). Hence  $G$  is also quasi-isometrically embedded in the handlebody group. Together with *i*) this shows *ii*).  $\square$

Next we recall a family of distinguished paths in the graph of rigid racks from [HH11]. The paths are inspired by splitting sequences of train tracks on surfaces. Since we do not use the details of the construction in this note, we only summarize the important properties of the paths in the following theorem, which is an immediate consequence of the proof of Theorem 7.9 from [HH11].

**Theorem 3.6.** *Let  $R, R'$  be two rigid racks. Then there is a disk system  $\mathcal{D}$  depending only on the support system of  $R'$  with the following property. Let  $\Sigma(R) = \Sigma_1, \dots, \Sigma_n$  be a disk exchange sequence in direction of  $\mathcal{D}$  such that  $\Sigma_n$  is disjoint from  $\mathcal{D}$ .*

*Then there is an edge-path  $(R_j)$  in the graph of rigid racks with the following properties.*

- i) There is a sequence of numbers  $1 = r_1 < \dots < r_n$  such that the support system of  $R_j$  is  $\Sigma_i$  for all  $r_i \leq j \leq r_{i+1} - 1$ .*
- ii) For  $r_i \leq j \leq r_{i+1} - 1$ , the sequence  $(R_j)$  is a uniform quasi-geodesic in the graph  $\mathcal{RR}(V, \Sigma_i)$ .*
- iii) The length of the sequence  $(R_j)$  is bounded by  $e^{k_0 \cdot d}$ , where  $d$  is the distance between  $R$  and  $R'$  in the graph of rigid racks and  $k_0 > 0$  is a uniform constant.*

We call a sequence  $(R_i)$  given by the theorem a  $\mathcal{D}$ -splitting sequence, and call the sequence  $(\Sigma_i)$  the disk exchange sequence associated to  $(R_i)$ .

Here and in the sequel, we say that a path is a *uniform* quasigeodesic if the quasigeodesic constants of the path depend only on the genus of the handlebody. Similarly, we say that a number is *uniformly bounded*, if there is a bound depending only on the genus of  $V$ .

In Section 2 we demonstrated that  $\mathcal{D}$ -disk exchange sequences starting in disjoint reduced disk systems can be compared using full disk replacement sequences. In the rest of this section we develop a slight generalization of racks, which will allow to similarly compare  $\mathcal{D}$ -splitting sequences starting in adjacent vertices of  $\mathcal{RR}(V)$ .

Namely, define an *extended rack*  $R$  in the same way as a rack except that the support system  $\mathcal{D}(R)$  of an extended rack  $R$  may be any simple disk system instead of a reduced disk system. The cell decomposition induced by an extended rack is defined in the obvious way, and similarly we can define rigid isotopies between extended racks. The rigid isotopy class of an extended rack is called a *rigid extended rack*.

Rigid extended racks can be used in the same way as racks to define a geometric model for the handlebody group.

**Definition 3.7.** The graph of rigid extended racks  $\mathcal{RE}\mathcal{R}_K(V)$  is the graph whose vertex set is the set of large rigid extended racks. Two such vertices corresponding to rigid extended racks  $R$  and  $R'$  are connected by an edge of length one if the following holds. The support systems of  $R$  and  $R'$  are disjoint, and in addition up to isotopy the 1-skeleta of the cell decompositions induced by  $R$  and  $R'$  intersect in at most  $K$  points.

Again, the constant  $K$  may be chosen in such a way that the graph of rigid extended racks is connected. For a simple disk system  $\mathcal{D}$  let  $\mathcal{RE}\mathcal{R}(V, \mathcal{D})$  be the complete subgraph of  $\mathcal{RE}\mathcal{R}(V)$  whose vertices are rigid extended racks with support system  $\mathcal{D}$ . We may choose  $K$  large enough such that for any simple disk system  $\mathcal{D}$  the subgraph  $\mathcal{RE}\mathcal{R}(V, \mathcal{D})$  is connected. We denote the resulting graph by  $\mathcal{RE}\mathcal{R}(V)$ . The analog of Lemma 3.5 holds for the graphs  $\mathcal{RE}\mathcal{R}(V, \mathcal{D})$ , with the same proof.

An analog of Lemma 3.3 holds for rigid extended racks as well, and implies that the handlebody group acts on the graph of rigid extended racks with finite quotient. Thus, the graph of rigid extended racks is quasi-isometric to the handlebody group. Note also that every large rack is a large extended rack. Thus the graph of rigid racks embeds as a subgraph in the graph of rigid extended racks. This inclusion is a quasi-isometry.

The proof of Theorem 7.9 of [HH11] also implies the following theorem which allows to connect two rigid racks with a distinguished type of path.

**Theorem 3.8.** *There is a number  $k_1 > 0$  with the following property. Let  $\hat{R}$  be a rigid rack, and let  $\mathcal{D}$  be the disk system given by Theorem 3.6. Let  $R$  be another rigid rack and let  $\mathcal{D}(R) = \mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  be a full disk replacement sequence in direction of  $\mathcal{D}$  such that  $\mathcal{D}_n$  is disjoint from  $\mathcal{D}$ .*

*Then there is a rigid extended rack  $R'$  which is at distance at most  $k_1$  to  $\hat{R}$  in  $\mathcal{RE}\mathcal{R}(V)$ , and there is an edge-path  $(R_j)$  in the graph of rigid extended racks with the following properties.*

- i) There is a sequence of numbers  $1 = r_1 < \dots < r_k$  such that the support system of  $R_j$  is  $\mathcal{D}_i$  for all  $r_i \leq j \leq r_{i+1} - 1$ .*
- ii) For  $r_i \leq j \leq r_{i+1} - 1$ , the sequence  $(R_j)$  is a uniform quasi-geodesic in the graph  $\mathcal{RE}\mathcal{R}(V, \mathcal{D}_i)$ .*
- iii) The length of the sequence  $(R_j)$  is bounded by  $e^{k_1 d}$ , where  $d$  is the distance between  $R$  and  $\hat{R}$  in the graph of rigid extended racks.*

We call a sequence  $(R_j)$  given by this theorem a *full  $\mathcal{D}$ -splitting sequence* and say that the sequence  $(\mathcal{D}_i)$  is the *full disk exchange sequence* associated to  $(R_i)$ .

Combining Lemmas 2.6 and 2.7 with Theorem 3.8 above, we obtain the following.

**Corollary 3.9.** *There is a number  $k_2 > 0$  with the following property.*

- i) Let  $R_i, i = 1, \dots, N$  be a  $\mathcal{D}$ -splitting sequence of racks with associated disk exchange sequence  $(\Sigma_j)$ . Let  $(\mathcal{D}_j)$  be a full disk replacement sequence compatible with  $(\Sigma_j)$ . Then there is a full  $\mathcal{D}$ -splitting sequence  $\tilde{R}_k, k = 1, \dots, K$  such that the following holds. The full disk replacement sequence associated to  $(\tilde{R}_k)$  is  $(\mathcal{D}_j)$ . Furthermore,  $\tilde{R}_1 = R_1$  and the distance between  $\tilde{R}_K$  and  $R_N$  is at most  $k_2$ . The length  $K$  of any such sequence is at most  $e^{k_2 d}$ , where  $d$  is the distance between  $R_1$  and  $R_N$  in the graph of rigid racks.*
- ii) Conversely, suppose that  $\tilde{R}_k, k = 1, \dots, K$  is a full  $\mathcal{D}$ -splitting sequence with associated full disk replacement sequence  $(\mathcal{D}_j)$ . Suppose further that  $(\Sigma_j)$  is a*



disk exchange sequence compatible with  $(\mathcal{D}_i)$ . If  $\tilde{R}_1$  is a large rack, then there is a  $\mathcal{D}$ -splitting sequence  $R_1, \dots, R_N$  whose associated disk exchange sequence is  $(\Sigma_j)$  such that  $R_N$  is of distance at most  $k_1$  to  $\tilde{R}_K$ . The length  $N$  of any such sequence is at most  $e^{k_2 d}$ , where  $d$  is the distance between  $R_1$  and  $R_N$  in the graph of rigid racks.

#### 4. ESTIMATING THE DEHN FUNCTION OF THE HANDLEBODY GROUP

In this section we prove the main result of this note.

**Theorem 4.1.** *The Dehn function of the handlebody group has at most exponential growth rate.*

To begin, we recall the definitions of the Dehn function and growth rate. Let  $G$  be a finitely presented group, and choose a finite presentation  $G = \langle \mathcal{S} | \mathcal{R} \rangle$ . Let  $F(\mathcal{S})$  be the free group with generating set  $\mathcal{S}$ , and let  $R$  be the subgroup of  $F(\mathcal{S})$  normally generated by  $\mathcal{R}$ . Thus  $G$  is isomorphic to the quotient  $F(\mathcal{S})/R$ .

Every  $r \in R < F(\mathcal{S})$  can be written as a product of conjugates of elements in  $\mathcal{R}$ :

$$r = \prod_{i=1}^n r_i^{\gamma_i}, \quad r_i \in \mathcal{R}, \gamma_i \in G.$$

We call the minimal length  $n$  of such a product *the area*  $\text{Area}(r)$  *needed to fill the relation*  $r$ . On the other hand,  $r$  can be viewed as an element of  $F(\mathcal{S})$  and hence it can be written as a word in the elements of  $\mathcal{S}$ . We call the minimal length of such a word *the length*  $l(r)$  *of the loop*  $r$ .

The *Dehn function* of  $G$  is then defined by

$$\delta(n) = \max\{\text{Area}(r) \mid r \in R \text{ with } l(r) \leq n\}.$$

The function  $\delta$  depends on the choice of finite presentation  $\langle \mathcal{S} | \mathcal{R} \rangle$ . However, the Dehn functions obtained from different finite presentations are of the same growth type. To say that two functions  $f, g : \mathbb{N}_{\geq 0} \rightarrow \mathbb{N}_{\geq 0}$  are of the same growth type means that there are numbers  $K, L > 0$  with

$$f(x) \leq L \cdot g(K \cdot x + K) + L \cdot x + L$$

for all  $x \in \mathbb{N}_{\geq 0}$ , and that there is an analogous inequality bounding  $g$  in terms of  $f$ .

To estimate the Dehn function of the handlebody group we will consider loops in the graph  $\mathcal{RER}(V)$  of rigid extended racks. Let  $\gamma$  be a loop of length  $R$  in  $\mathcal{RER}(V)$ . We will show that there is a number  $k > 0$  only depending on the genus of the handlebody and that there are  $m \leq e^{kR}$  loops  $\zeta_1, \dots, \zeta_m$  of length at most  $k$  each such that  $\gamma$  can be contracted to a point in  $m$  steps consisting each of replacing a subsegment of  $\zeta_i$  by another subsegment of  $\zeta_i$ . This suffices, since each loop  $\zeta_i$  as above corresponds to a cycle in the handlebody group which can be filled with uniformly small area. By slight abuse of notation we will then say that  $\gamma$  can be filled with area  $e^{kR}$  if the above condition holds.

Recall from Section 3 the definition of the graph  $\mathcal{RER}(V, \mathcal{D})$ . The following lemma allows to control the isoperimetric function of these subgraphs.

**Lemma 4.2.** *Let  $\mathcal{D}$  be a simple disk system for  $V$ . Any loop in  $\mathcal{RER}(V, \mathcal{D})$  can be filled with area coarsely bounded quadratically in its length.*

*Proof.* The stabilizer  $G$  of  $\partial\mathcal{D}$  in the mapping class group of  $\partial V$  is a Lipschitz retract of the mapping class group of  $\partial V$  (see [HM10] for a detailed discussion of this fact which is a consequence of the work of Masur and Minsky [MM00]). Mapping class groups are automatic [Mo95] and hence have quadratic Dehn function. Then the same holds true for  $G$  (compare again [HM10]). The lemma now follows from the fact that by Lemma 3.5, the graph  $\mathcal{RE}\mathcal{R}(V, \Sigma)$  is quasi-isometric to  $G$  (the proof of Lemma 3.5 works verbatim also for rigid extended racks).  $\square$

As the next step, we use Corollary 3.9 to control splitting sequences starting at adjacent points in the graph of rigid racks. We show that these paths can be constructed in such a way that the resulting loop can be filled with controlled area. Together with the length estimate for splitting sequences from Theorem 3.8 this will imply the exponential bound for the Dehn function.

The main technical tool in this approach is given by the following

**Lemma 4.3.** *For each  $k > 0$  there is a number  $k_3 = k_3(k) > 0$  with the following property.*

*Let  $\mathcal{D}$  be a simple disk system. Let  $R_i, i = 1, \dots, N$  be a  $\mathcal{D}$ -splitting sequence of rigid racks and let  $\tilde{R}_j, j = 1, \dots, M$  be a full  $\mathcal{D}$ -splitting sequence of rigid extended racks such that the following holds.*

- i) The rigid extended racks  $R_1$  and  $\tilde{R}_1$  (respectively  $R_N$  and  $\tilde{R}_M$ ) have distance at most  $k$  in the graph of rigid extended racks.*
- ii) The disk exchange sequence associated to  $(R_i)$  is compatible with the full disk replacement sequence associated to  $(\tilde{R}_j)$ .*

*Then the loop  $\gamma$  in  $\mathcal{RE}\mathcal{R}(V)$  formed by the sequences  $(R_i)$ ,  $(\tilde{R}_j)$  and geodesics between  $R_1$  and  $\tilde{R}_1$  and  $R_N$  and  $\tilde{R}_M$  can be filled with area  $k_3(N + M)^3$ .*

*Proof.* The idea of the proof is to inductively decompose the loop  $\gamma$  into smaller loops, each of which can be filled with area at most  $k_3(N + M)^2$  for a suitable  $k_3$ .

Denote the disk exchange sequence associated to  $(R_i)$  by  $\Sigma_i$  ( $i = 1, \dots, m$ ) and the full disk replacement sequence associated to  $(\tilde{R}_j)$  by  $\mathcal{D}_j$  ( $j = 1, \dots, n$ ). Let  $r : \{1, \dots, n\} \rightarrow \{1, \dots, m\}$  be the non-decreasing surjective function given by the definition of compatibility, i.e.  $\Sigma_{r(j)} \subset \mathcal{D}_j$  for all  $j = 1, \dots, n$ .

We define

$$I(\ell) = \{k \mid \Sigma(R_k) = \Sigma_\ell\}$$

and

$$J(\ell) = \{k \mid \mathcal{D}(\tilde{R}_k) = \mathcal{D}_j \text{ and } r(j) = \ell\}.$$

Put  $i_k = \max I(k)$  and  $j_k = \max J(k)$ . We will inductively choose paths  $d_k$  connecting  $R_{i_k}$  to  $\tilde{R}_{j_k}$  and paths  $c_k$  connecting  $R_{i_{k+1}}$  to  $\tilde{R}_{j_{k+1}}$  with the following properties.

- (1) The path  $c_k$  is a uniform quasigeodesic in  $\mathcal{RE}\mathcal{R}(V, \Sigma_{k+1})$ .
- (2) The path  $d_k$  is a uniform quasigeodesic in  $\mathcal{RE}\mathcal{R}(V, \Sigma_k)$ .
- (3) The paths  $c_k, d_k$  are uniform fellow-travelers.

The existence of a family of paths with these properties implies the statement of the lemma in the following way. We adopt the following notation. If  $c = (x_i)$  is a path in  $\mathcal{RE}\mathcal{R}(V)$  then we denote by  $c^{-1}$  or  $(x_i)^{-1}$  the path with the opposite orientation.

The restriction of the sequence  $R_i$  to  $I(k)$  and the restriction of  $\tilde{R}_j^{-1}$  to  $J(k)$  form together with  $c_{k-1}^{-1}$  and  $d_k$  a loop  $\gamma_k$  in  $\mathcal{RE}\mathcal{R}(V, \Sigma_k)$ . The length of  $c_{k-1}$  and  $d_k$  is coarsely bounded by  $N + M$  by the triangle inequality. Hence, the length of  $\gamma_k$  can be coarsely bounded by  $4(N + M)$ . Since  $\mathcal{RE}\mathcal{R}(V, \Sigma_k)$  admits a quadratic isoperimetric function by Lemma 4.2, this loop can be filled with area bounded by  $k_3(N + M)^2$  for some uniform constant  $k_3$ .

Similarly, the paths  $d_k^{-1}$  and  $c_{k+1}$  together with the edges connecting  $R_{i_k}$  to  $R_{i_{k+1}}$  and  $\tilde{R}_{j_{k+1}}$  to  $\tilde{R}_{j_k}$  form a loop  $\delta_k$ . The length of  $\delta_k$  can again be coarsely bounded by  $2(N + M)$  using the triangle inequality. Since the paths  $d_k$  and  $c_k$  are uniform fellow-travelers,  $\delta_k$  can be filled with area depending linearly on its length.

There are at most  $2 \max(N, M)$  loops  $\gamma_k, \delta_k$ . Hence the concatenation of all the loops  $\gamma_i$  and  $\delta_j$  can be filled with area at most  $k_3(N + M)^3$  (after possibly enlarging the constant  $k_3$ ). The paths  $c_i$  and  $d_i$  occur in the concatenation of  $\gamma_i$  and  $\delta_j$  twice, with opposite orientations, except for  $c_0$  and the last occurring arc  $d_L$ . As a consequence, the concatenation of the loops  $\gamma_i$  and  $\delta_j$  is, after erasing these opposite paths, uniformly close to  $\gamma$  in the Hausdorff metric. Thus  $\gamma$  may also be filled with area bounded by  $k_3(N + M)^3$  (again possibly increasing  $k_3$ ).

We now describe the inductive construction of the paths  $c_k$  and  $d_k$ . We set  $c_0 = d_0$  to be the constant path  $R_1$ . Suppose that the paths  $c_i, d_i$  are already constructed for  $i = 0, \dots, k - 1$ .

The support systems of  $R_{i_k}$  and  $\tilde{R}_{j_k}$  both contain  $\Sigma_k$ . We first construct the path  $d_k$  connecting  $R_{i_k}$  and  $\tilde{R}_{j_k}$ .

The reduced disk systems  $\Sigma_k$  and  $\Sigma_{k+1}$  are disjoint. The simple disk system  $\Sigma_k \cup \Sigma_{k+1}$  is disjoint from the support systems of  $R_{i_k}, R_{i_{k+1}}$  and  $\tilde{R}_{j_k}, \tilde{R}_{j_{k+1}}$  by definition of splitting sequences. Furthermore, the 1-skeleta of the cell decompositions of all four of these extended racks intersect  $\partial\Sigma_k \cup \partial\Sigma_{k+1}$  in uniformly few points. Hence, there are rigid extended racks  $U_1, U_2$  which have  $\Sigma_k \cup \Sigma_{k+1}$  as their support system and such that  $U_1$  is uniformly close to  $R_{i_k}$ , and  $U_2$  is uniformly close to  $\tilde{R}_{j_k}$  in  $\mathcal{RE}\mathcal{R}(V)$ . Let  $e$  be a geodesic path in  $\mathcal{RE}\mathcal{R}(V, \Sigma_k \cup \Sigma_{k+1})$  connecting  $U_1$  and  $U_2$ . Since  $\mathcal{RE}\mathcal{R}(V, \Sigma_k \cup \Sigma_{k+1})$  is undistorted in  $\mathcal{RE}\mathcal{R}(V)$  by Lemma 3.5, the length of  $e$  is coarsely bounded by  $N + M + 1$ . By adding uniformly short geodesic segments in  $\mathcal{RE}\mathcal{R}(V, \Sigma_k)$  at the beginning and the end of  $e$ , we obtain the path  $d_k$  with property *ii*).

By definition of  $i_k$  and  $j_k$ , we have  $i_k + 1 \in I(k + 1)$  and  $j_k + 1 \in J(k + 1)$ . Hence, both  $R_{i_{k+1}}$  and  $\tilde{R}_{j_{k+1}}$  contain  $\Sigma_{k+1}$  in their support systems. We can thus define  $c_k$  with properties *i*) and *iii*) by adding uniformly short geodesic segments in  $\mathcal{RE}\mathcal{R}(V, \Sigma_{k+1})$  to the beginning and the end of  $e$ .  $\square$

We have now collected all the tools for the proof of the main theorem.

*Proof of Theorem 4.1.* Recall that it suffices to show that every loop in the graph of rigid racks can be filled with area coarsely bounded by an exponential function of its length in the sense discussed in the beginning of this section.

Let  $(R_i)$  be a loop of length  $L$  in the graph of rigid racks based at  $R_0 = \hat{R}$ . Since the graph of rigid racks is quasi-isometric to the graph of rigid extended racks, we can consider  $R_i$  as a loop in  $\mathcal{RE}\mathcal{R}(V)$  and it suffices to show that this loop can be filled in  $\mathcal{RE}\mathcal{R}(V)$  with area bounded exponentially in its length.

Similar to the proof of Lemma 4.3, the strategy is to write the loop  $(R_i)$  as a concatenation of smaller loops whose area we can control. To this end let  $\hat{\Sigma}$  be the disk system given by Theorem 3.8 applied to  $\hat{R} = R_0$ . We will define paths  $c_i$  in  $\mathcal{RER}(V)$  with the following properties.

- (1) The path  $c_i$  connects  $R_i$  to a rack which is uniformly close to  $\hat{R}$  in  $\mathcal{RER}(V)$ .
- (2) The path  $c_i$  is a  $\hat{\Sigma}$ -splitting sequence of racks.
- (3) The loop formed by  $c_i, c_{i+1}^{-1}$ , the edge between  $R_{i+1}$  and  $R_i$ , and a geodesic connecting the other pair of endpoints of  $c_i, c_{i+1}$  can be filled with area bounded by  $e^{k_3 L}$ .

This implies that the loop  $(R_i)$  itself can be filled with area at most  $Le^{k_3 L}$ , proving the theorem.

The construction of the paths  $c_i$  again uses induction. We set  $c_0$  to be the constant path  $R_0$ . Suppose now that the path  $c_k$  has already been constructed. Since  $R_k$  and  $R_{k+1}$  are connected by an edge in the graph of rigid racks, their support systems  $\Sigma_k$  and  $\Sigma_{k+1}$  are disjoint (compare Definition 3.4). Let  $\Sigma_k^{(i)}, i = 1, \dots, n$  be the disk exchange sequence associated to the splitting sequence  $c_k$ . Put  $\mathcal{D}_1 = \Sigma_k \cup \Sigma_{k+1}$ . Using Lemma 2.7 we obtain a full disk replacement sequence  $(\mathcal{D}_i)$  compatible with  $(\Sigma_k^{(i)})$ . Corollary 3.9 part *i*) then yields a full splitting sequence  $\tilde{R}_k, k = 1, \dots, M$  with associated full disk exchange sequence  $(\mathcal{D}_i)$ . By Lemma 4.3, the loop formed by  $c_k$  and  $(\tilde{R}_k)$  can be filled with area bounded by  $k_3(N + M)^3$ , where  $N$  is the length of the path  $c_k$ .

Using Lemma 2.6 on the sequence  $(\mathcal{D}_i)$  and the initial reduced disk system  $(\Sigma_{k+1})$  we obtain a  $\hat{\Sigma}$ -splitting sequence  $(\Sigma_{k+1}^{(i)})$  compatible with  $(\mathcal{D}_i)$ , which starts in  $\Sigma_{k+1}$ . Corollary 3.9 part *ii*) now yields a  $\hat{\Sigma}$ -splitting sequence  $c_{k+1}$  starting in  $R_{k+1}$  and ending uniformly close to  $\hat{R}$ , thus satisfying properties (1) and (2). Applying Lemma 4.3 again, we see that the loop formed by  $c_{k+1}$  and  $\tilde{R}_k, k = 1, \dots, M$  can be filled with area bounded by  $k_3(N' + M)^3$ , where  $N'$  is the length of the path  $c_{k+1}$ .

Since both  $c_k$  and  $c_{k+1}$  are splitting sequences connecting points which are of distance at most  $L$ , their lengths can be bounded by  $k_4 e^{k_4 L}$  for a suitable  $k_4$  by Theorem 3.6. As a consequence, the paths  $c_k$  and  $c_{k+1}$  satisfy condition (3). This concludes the inductive construction of  $c_i$  and the proof of the theorem.  $\square$

The proof of the theorem would give a polynomial bound for the Dehn function provided that the length of the splitting sequences used to fill in the loops had a length which is polynomial in the distance between their endpoints. However, the following example shows that such a bound does not exist. This is similar to the behavior of paths of sphere systems used in [HV96] to show an exponential upper bound for the Dehn function of  $\text{Out}(F_n)$ ,

For simplicity of exposition, we do not construct these paths in the graph of rigid racks (or the handlebody group), but instead in a slightly simpler graph. The example given below can be extended to the full graph of rigid racks in a straightforward fashion.

We define  $\mathcal{RD}(V)$  to be the graph of reduced disk systems in  $V$ . The vertex set of  $\mathcal{RD}(V)$  is the set of isotopy classes of reduced disk systems, and two such vertices are connected by an edge of length one if the corresponding disk systems are disjoint. Every directed disk exchange sequence defines an edge-path in  $\mathcal{RD}(V)$ .

The following example shows that the length of these edge-paths may be exponential in the distance between their endpoints.

*Example 4.4.* Consider a handlebody  $V$  of genus 4. For each  $n \in \mathbb{N}$  we will construct a disk exchange sequence  $\Sigma_1^{(n)}, \dots, \Sigma_{N(n)}^{(n)}$  such that on the one hand, the length  $N(n)$  of the sequence growth exponentially in  $n$ . On the other hand, the distance between its endpoints  $\Sigma_1^{(n)}$  and  $\Sigma_{N(n)}^{(n)}$  in  $\mathcal{RD}(V)$  grows linearly in  $n$ . To simplify notation we will only construct the endpoint  $\Sigma_{N(n)}^{(n)}$  and denote it by  $\Sigma_n$ .

We choose three disjoint simple closed curves  $\alpha_1, \alpha_2, \alpha_3$  which decompose the surface  $\partial V$  into a pair of pants, two once-punctured tori and a once-punctured genus 2 surface (see Figure 1). We may choose the  $\alpha_i$  such that they bound disks in  $V$ . We denote the two solid tori in the complement of these disks by  $T_1, T_2$  and the genus 2 subhandlebody by  $V'$ .

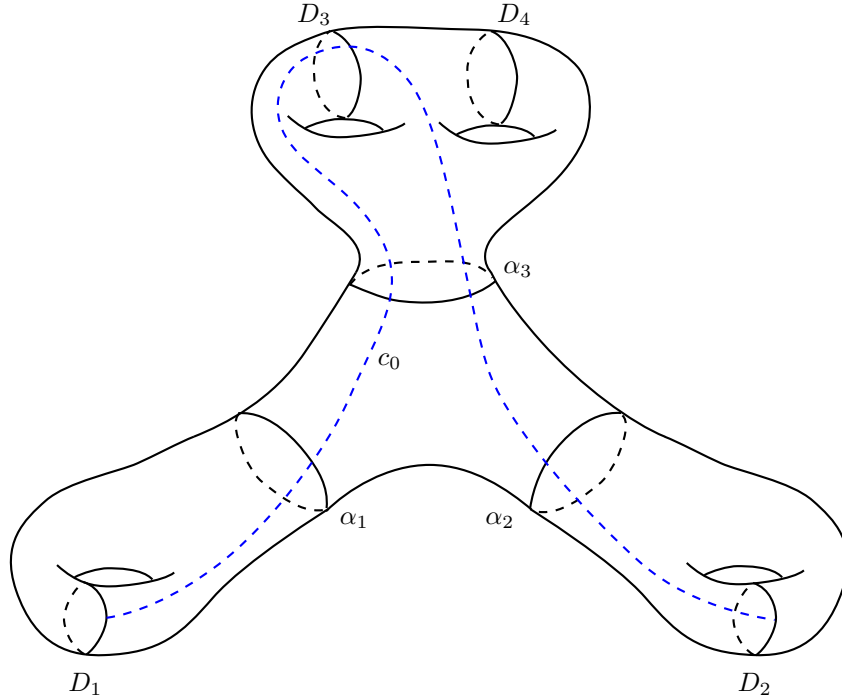


FIGURE 1. The setup for the example of a non-optimal disk exchange path. An admissible arc is drawn dashed.

Let  $\Sigma_0 = \{D_1, D_2, D_3, D_4\}$  be a reduced disk system such that  $D_1 \subset T_1, D_2 \subset T_2$  and  $D_3, D_4 \subset V'$ . Choose a base point  $p$  on  $\alpha_3$ . Let  $\gamma_1, \gamma_2$  be two disjointly embedded loops on  $\partial V \cap V'$  based at  $p$  with the following property. The loop  $\gamma_1$  intersects the disk  $D_3$  in a single point and is disjoint from  $D_4$ , while  $\gamma_2$  intersects  $D_4$  in a single point and is disjoint from  $D_3$ . Since the complement of  $D_3 \cup D_4$  in  $V'$  is simply connected, such a pair of loops generates the fundamental group of  $V'$ . Denote the projections of  $\gamma_1$  and  $\gamma_2$  to  $\pi_1(V', p)$  by  $A_1$  and  $A_2$ , respectively.

Let  $c$  be an embedded arc on  $\partial V$ . We say that  $c$  is *admissible* if the following holds. The arc  $c$  connects the disk  $D_1$  to the disk  $D_2$ . The interior of  $c$  intersects

$\alpha_1$  and  $\alpha_2$  in a single point each. Furthermore it intersects  $\alpha_3$  in two points, and its interior is disjoint from both  $D_1$  and  $D_2$ .

Let  $c$  be an admissible arc. The intersection of  $c$  with  $V'$  is an embedded arc  $c'$  connecting  $\alpha_3$  to itself. The arc  $c'$  may be turned into an embedded loop in  $V'$  based at  $p$  by connecting the two endpoints of  $c'$  to  $p$  along  $\alpha_3$ . Since the curve  $\alpha_3$  bounds a disk in  $V'$ , the image of this loop in  $\pi_1(V', p)$  is determined by the homotopy class of the arc  $c$  relative to  $\partial D_1, \partial D_2$ . We call this image the *element induced by the arc  $c$* .

Choose an admissible arc  $c_0$  in such a way that it intersects the disk  $D_3$  in a single point, and is disjoint from  $D_4$  (see Figure 1 for an example). Up to changing the orientation of  $\gamma_1$  we may assume that the element induced by  $c_0$  is  $A_1$ .

We now describe a procedure that produces essential disks from admissible arcs. To this end, let  $c$  be an admissible arc. Consider a regular neighborhood  $U$  of  $\partial D_1 \cup c \cup \partial D_2$  in  $\partial V$ . Its boundary consists of three simple closed curves. Two of them are homotopic to either  $\partial D_1$  or  $\partial D_2$ . The third one we denote by  $\beta(c)$ . Note that  $\beta(c)$  bounds a nonseparating disk in  $V$ .

Choose a fixed element  $\varphi$  of the handlebody group of  $V$  with the following properties. The mapping class  $\varphi$  fixes the isotopy classes of the curves  $\alpha_1, \alpha_2$  and  $\alpha_3$ . The restriction of  $\varphi$  to the complement of  $V'$  is isotopic to the identity. The restriction of  $\varphi$  to  $V'$  induces an automorphism of exponential growth type on  $\pi_1(V')$ . To be somewhat more precise, we may choose  $\varphi$  such that it acts on the basis  $A_i$  as the following automorphism  $\Phi$ :

$$\begin{aligned} A_1 &\mapsto A_1 A_2 \\ A_2 &\mapsto A_1^2 A_2 \end{aligned}$$

Put  $c_n = \varphi^n(c_0)$  and  $\beta_n = \beta(c_n)$ . We claim that a disk exchange sequence in direction of  $\beta_n$  that makes  $\beta_n$  disjoint from  $\Sigma_0$  has length at least  $2^n$ .

To show the claim, note that the arc  $c_n$  intersects the disks  $D_3$  and  $D_4$  in at least  $2^n$  points. Namely, the element of  $\pi_1(V', p)$  induced by  $\varphi^n(c_0)$  is equal to  $\Phi^n(A_1)$ . The cyclically reduced word describing  $\Phi^n(A_1)$  in the basis  $A_1, A_2$  has length at least  $2^n$  by construction of  $\Phi$ .

Therefore, the curve  $\beta_n$  can be described as follows. Choose a parametrization  $\beta_n : [0, 1] \rightarrow \partial V$ . Then there are numbers  $0 < t_1 < \dots < t_N < t_{N+1} < \dots < t_{2N} < 1$  such that the following holds. Each subarc  $\beta_n([t_i, t_{i+1}])$  intersects  $\Sigma_0$  only at its endpoints. The subarcs  $\beta_n([t_N, t_{N+1}])$  and  $\beta_n([0, t_1] \cup [t_{2N}, 1])$  are waves to  $\Sigma_0$ . Furthermore, the arcs  $\beta_n([t_i, t_{i+1}])$  and  $\beta_n([t_{2N-i}, t_{2N+1-i}])$  are homotopic relative to  $\Sigma_0$  for all  $i = 1, \dots, N-1$ . More generally, if there are numbers  $t_i$  with these properties for a reduced disk system  $\Sigma$  we say that  $\beta_n$  is a *long string of rectangles with respect to  $\Sigma$* . The number  $N$  is then called *the length of the string of rectangles*. By construction, the length  $N$  of the string of rectangles  $\beta_n$  defines with respect to  $\Sigma_0$  is at least  $2^n$ .

The curve  $\beta_n$  has two waves. Let  $a$  be one of them, say  $\beta_n([t_N, t_{N+1}])$  and let  $\sigma \in \Sigma_0$  denote the disk containing the endpoints of  $a$ . One of the disks obtained by simple surgery along  $a$  is isotopic to either  $D_1$  or  $D_2$  (depending on which wave we chose). The preferred interval defined by  $a$  contains every intersection point of  $\beta_n$  with  $\sigma$  except the endpoints of  $a$ .

Denote by  $\Sigma_1$  the reduced disk system obtained by simple surgery along  $a$ . By construction, the subarc  $\beta_n(t_{N-1}, t_{N+2})$  now defines a wave with respect to  $\Sigma_1$ . One



of the disks obtained by simple surgery along this wave is still properly isotopic to  $D_1$ . Furthermore, the subarcs  $\beta_n([t_i, t_{i+1}])$  and  $\beta_n([t_{2N-i}, t_{2N+1-i}])$  are still arcs with endpoints on  $\Sigma_1$  which are homotopic relative to  $\Sigma_1$  for all  $i = 1, \dots, N - 2$ . Each of these arcs cannot be homotoped into  $\partial\Sigma_1$ .

Hence the curve  $\beta_n$  has a description as a string of rectangles of length  $N - 1$  with respect to  $\Sigma_1$  and the argument can be iterated. By induction, it follows that any disk exchange sequence starting in  $\Sigma_0$  which ends in a disk system disjoint from  $\beta_n$  has length at least  $2^n$ .

On the other hand, the growth of the distance between  $\Sigma_0$  and  $\varphi^n(\Sigma_0)$  in the graph of reduced disk systems is linear in  $n$  by the triangle inequality. The curve  $\beta_n$  intersects  $\varphi^n(\Sigma_0)$  in uniformly few points, and thus the disk system  $\varphi^n(\Sigma_0)$  is uniformly close to a reduced disk system that is disjoint from  $\beta_n$ . Thus the disk systems  $\Sigma_n$  have the properties described in the beginning of the example.

## REFERENCES

- [BV95] M. Bridson, K. Vogtmann, *On the geometry of the group of automorphisms of a free group*, Bull. London Math. Soc. 27 (1995), 544–525.
- [BV10] M. Bridson, K. Vogtmann, *The Dehn function of  $\text{Aut}(F_n)$  and  $\text{Out}(F_n)$* , arXiv:1011.1506.
- [H09b] U. Hamenstädt, *Geometry of the mapping class group II: A biautomatic structure*, arXiv:0912.0137.
- [HH11] U. Hamenstädt, S. Hensel, *The geometry of the handlebody groups I: Distortion*, arXiv:1101.1838.
- [HM10] M. Handel, L. Mosher, *Lipschitz retraction and distortion for subgroups of  $\text{Out}(F_n)$* , arXiv:1009.0518.
- [Ha95] A. Hatcher, *Homological stability for automorphism groups of free groups*, Comm. Math. Helv. 70 (1995), 39–62.
- [HV96] A. Hatcher, K. Vogtmann, *Isoperimetric inequalities for automorphism groups of free groups*, Pacific J. Math. 173 (1996), no. 2, 425–441.
- [M86] H. Masur, *Measured foliations and handlebodies*, Erg. Th. & Dynam. Sys. 6 (1986), 99–116.
- [MM00] H. Masur, Y. Minsky, *Geometry of the complex of curves II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000), 902–974.
- [McC85] D. McCullough, *Twist groups of compact 3-manifolds*, Topology 24 (1985), no. 4, 461–474.
- [Mo95] L. Mosher, *Mapping class groups are automatic*, Ann. of Math. (2) 142 (1995), no. 2, 303–384.
- [St99] J. Stallings, *Whitehead graphs on handlebodies*, Geometric group theory down under (Canberra, 1996), 317–330, de Gruyter, Berlin, 1999.
- [Wa98] B. Wajnryb, *Mapping class group of a handlebody*, Fund. Math. 158 (1998), no. 3, 195–228.

URSULA HAMENSTÄDT AND SEBASTIAN HENSEL, MATHEMATISCHES INSTITUT DER UNIVERSITÄT BONN, ENDENICHER ALLEE 60, D-53115 BONN, GERMANY

*E-mail address:* ursula@math.uni-bonn.de

*E-mail address:* loplop@math.uni-bonn.de