THE GEOMETRY OF THE DISK COMPLEX

HOWARD MASUR AND SAUL SCHLEIMER

Contents

1. Introduction
2. Background on complexes
3. Background on coarse geometry
4. Natural maps
5. Holes in general and the lower bound on distance
6. Holes for the non-orientable surface
7. Holes for the arc complex
8. Background on three-manifolds
9. Holes for the disk complex
10. Holes for the disk complex – annuli
11. Holes for the disk complex – compressible
12. Holes for the disk complex – incompressible
13. Axioms for combinatorial complexes
14. Partition and the upper bound on distance
15. Background on Teichmüller space
16. Paths for the non-orientable surface
17. Paths for the arc complex
18. Background on train tracks
19. Paths for the disk complex
20. Hyperbolicity
21. Coarsely computing Hempel distance
Acknowledgments
References

1. Introduction

Suppose $M$ is a compact, connected, orientable, irreducible three-manifold. Suppose $S$ is a compact, connected subsurface of $\partial M$ so that no component of $\partial S$ bounds a disk in $M$. In this paper we study the intrinsic geometry of the disk complex $D(M,S)$. The disk complex has a natural simplicial inclusion into the curve complex $\mathcal{C}(S)$. Surprisingly, this inclusion is generally not a quasi-isometric embedding; there are disks which are close in the curve complex yet far apart in the disk complex. As we show, any obstruction to joining such disks via a short path is a subsurface $X \subset S$ that is topologically meaningful for $M$. We call such subsurfaces holes. A path in the disk complex must travel into and then out of
these holes; paths in the curve complex may skip over a hole by using the vertex representing the boundary of the subsurface. We classify holes for \( D(M, S) \).

**Theorem 1.1.** Suppose \( X \) is a hole for the disk complex \( D(M, S) \) of diameter at least 57.

- \( X \) is not an annulus.
- If \( X \) is compressible, then there are disks \( D, E \) with boundaries contained in and filling \( X \).
- If \( X \) is incompressible, then there is an \( I \)-bundle \( \rho_F : T \to F \) embedded in \( M \) so that \( X \subset \partial_h T \subset S \).

See Theorems 10.1, 11.10, and 12.1 for more precise statements. The \( I \)-bundles appearing in the classification lead us to study the arc complex \( A(F) \) of the base surface \( F \). Since the bundle \( T \) may be twisted the surface \( F \) may be non-orientable.

Thus, as a necessary warm-up to the difficult case of the disk complex, we analyze the holes for the curve complex of a non-orientable surface, as well as the holes for the arc complex.

**Topological application.** It is a long-standing open problem to decide, given a Heegaard diagram, whether the underlying splitting surface is reducible. For example, see [19, Section 2B], [48, Section 4], [19, Problem 1.11(c)], and [17, page 462]. This problem has deep connections to the geometry, topology, and algebra of three-manifolds; its resolution would give new solutions to both the three-sphere recognition problem and the triviality problem for three-manifold groups. The difficulty of deciding reducibility is underlined by its connection to the Poincaré conjecture: several approaches to the Poincaré Conjecture fell at essentially this point. See [11] for a survey of the literature on this topic.

One generalization of deciding reducibility is to find an algorithm that, given a Heegaard diagram, computes the Hempel distance of the Heegaard splitting [25]. See [1, Section 2]. The classification of holes for the disk complex leads to a coarse answer to this question.

**Theorem 21.1.** In every genus \( g \) there is a constant \( K = K(g) \) and an algorithm that, given a Heegaard diagram, computes the distance of the Heegaard splitting with error of at most \( K \).

In addition to the classification of holes, the algorithm relies on the Gromov hyperbolicity of the curve complex [30] and the quasi-convexity of the disk set inside of the curve complex [32]. However, the algorithm does not depend on our geometric application of Theorem 1.1 which we now discuss.

**Geometric application.** The hyperbolicity of the curve complex and the classification of holes are needed in the proof of the following.

**Theorem 20.4.** The disk complex is Gromov hyperbolic.

Again, as a warm-up to the proof of Theorem 20.4 we prove, for a non-orientable surface \( F \) and for an orientable surface \( S \), that \( C(F) \) and \( A(S) \) are hyperbolic. See Corollary 6.3 and Theorem 20.3. Note that Bestvina and Fujiwara [3] have previously dealt with the curve complex of a non-orientable surface, following Bowditch [7].

These results cannot be deduced from knowing that \( C(F) \), \( A(S) \), and \( D(M, S) \) can be realized as quasi-convex subsets of \( C(S) \). This is because the curve complex
is locally infinite. For a very similar example to these, consider the inclusion of the three-valent tree $T_3$ into the dual of the Farey triangulation. Thus $T_3$ is quasi-convex inside of a Gromov hyperbolic space; also $T_3$ is Gromov hyperbolic. However, the second fact cannot be deduced from the first. To see this take the Cayley graph of $\mathbb{Z}^2$ with the standard generating set. Then the cone $C(\mathbb{Z}^2)$ of height one-half is a Gromov hyperbolic space and $\mathbb{Z}^2$ is a quasi-convex subset.

The proof of Theorem 20.4 requires a distance estimate (Theorem 19.9): the distance in any one of $C(F)$, $A(S)$, or $D(M,S)$ is coarsely equal to the sum of subsurface projection distances in holes. Our theorem is modelled on the estimates for the marking graph and pants graph [31] Theorem 6.12 and Section 8] obtained by the first author and Minsky. However, we cannot use that paper's hierarchy machine; this is because hierarchies are too floppy to respect a symmetry and, at the same time, too rigid to deal with disks. For $C(F)$ we use the extremely rigid Teichmüller geodesic machine, due to Rafi [41]. For $D(M,S)$ we use the highly flexible train-track machine, developed by ourselves with Mosher [33].

Theorems 19.9 and 20.4 are part of a more general framework. Given a combinatorial complex $\mathcal{G}$ we classify the holes: the geometric obstructions lying between $\mathcal{G}$ and the curve complex. In Sections 13 and 14 we give axioms for $\mathcal{G}$ that imply a distance estimate. Hyperbolicity also follows from the axioms; this is proven in Section 20.

The axioms are stated in terms of a path of markings, a sequence in the combinatorial complex, and their relationship. For the disk complex the combinatorial sequence is a surgery sequence of essential disks while the marking path is provided by a train-track splitting sequence; both constructions are due to the first author and Minsky [32] (Section 18). The verification of the axioms (Section 19) relies on our work with Mosher: analyzing train-track splitting sequences in terms of subsurface projections [33].

We do not study non-orientable surfaces directly; instead we focus on symmetric multicurves in the double cover. This time the marking path is provided by a Teichmüller geodesic, using the fact that the symmetric Riemann surfaces form a totally geodesic subset of Teichmüller space. The combinatorial sequence is given by the systole map. We use results of Rafi [41] to verify the axioms for the complex of symmetric curves. (See Section 15) Section 17 verifies the axioms for the arc complex, also using Teichmüller geodesics and the systole map. Interestingly, for the arc complex our axioms can be verified using any one of Teichmüller geodesics, hierarchies, or train-track sequences.

The distance estimates for the marking graph and the pants graph [31] partly inspired this paper but do not fit our framework. Indeed, neither the marking graph nor the pants graph are Gromov hyperbolic. It is crucial here that all holes interfere; this leads to hyperbolicity. When there are non-interfering holes, it is unclear how to partition the marking path to obtain the distance estimate.

2. Background on complexes

We use $S_{g,b,c}$ to denote the compact, connected surface of genus $g$ with $b$ boundary components and $c$ cross-caps. If the surface is orientable we omit the subscript $c$ and write $S_{g,b}$. The complexity of $S = S_{g,b}$ is $\xi(S) = 3g - 3 + b$. If the surface is closed and orientable we simply write $S_g$. 

2.1. Arcs and curves. Suppose \( \alpha \subset S \) is a properly embedded simple closed curve or simple arc; thus \( \alpha \cap \partial S = \partial \alpha \). Then \( \alpha \) is inessential if \( \alpha \) cuts a disk off of \( S \). When \( \alpha \) is an essential curve, then \( \alpha \) is peripheral if it cuts an annulus off of \( S \).

Define \( \mathcal{C}(S) \) to be the set of ambient isotopy classes of essential, non-peripheral curves in \( S \). Define \( \mathcal{A}(S) \) to be the set of ambient isotopy classes of essential arcs. When \( S = S_{0,2} \) is an annulus define \( \mathcal{A}(S) \) to be the set of essential arcs, up to ambient isotopy fixing the boundary pointwise. For any surface define \( \mathcal{AC}(S) = \mathcal{A}(S) \cup \mathcal{C}(S) \).

For \( \alpha, \beta \in \mathcal{AC}(S) \) the geometric intersection number \( \iota(\alpha, \beta) \) is the minimum intersection possible between ambient isotopy representatives of \( \alpha \) and \( \beta \). When \( S = S_{0,2} \) we do not count intersection points occurring on the boundary. When two representatives \( \alpha \) and \( \beta \) realize their geometric intersection number we say that \( \alpha \) is tight with respect to \( \beta \). If they do not realize their geometric intersection, then we may tighten \( \alpha \) until they do. In the rest of the paper we use the same notation for isotopy classes and for their representatives.

A subset \( \Delta \subset \mathcal{AC}(S) \) is a multicurve if for all \( \alpha, \beta \in \Delta \) we have \( \iota(\alpha, \beta) = 0 \). Following Harvey [22] we may impose the structure of a simplicial complex on \( \mathcal{AC}(S) \): the simplices are exactly the multicurves. Also, \( \mathcal{C}(S) \) and \( \mathcal{A}(S) \) naturally span subcomplexes.

Note that the curve complexes \( \mathcal{C}(S_{0,4}), \mathcal{C}(S_{1}), \) and \( \mathcal{C}(S_{1,1}) \) have no edges. In these cases it is useful to alter the definition. Place edges between all vertices with geometric intersection exactly two, if \( S = S_{0,4} \), or exactly one, if \( S = S_{1} \) or \( S_{1,1} \). The result is the Farey graph \( \mathcal{F} = \mathcal{F}(S) \). Two vertices \( \alpha, \beta \) spanning an edge of \( \mathcal{F} \) are called Farey neighbors.

With the current definition, \( \mathcal{C}(S) \) is empty if \( S = S_{0,2} \). Thus for the annulus we alter the definition, taking \( \mathcal{AC}(S) = \mathcal{C}(S) = \mathcal{A}(S) \).

**Definition 2.2.** For vertices \( \alpha, \beta \in \mathcal{C}(S) \) define the distance \( d_{S}(\alpha, \beta) \) to be the minimum possible number of edges of a path in the one-skeleton \( \mathcal{C}^{1}(S) \) that starts at \( \alpha \) and ends at \( \beta \).

Note that if \( d_{S}(\alpha, \beta) \geq 3 \), then \( \alpha \) and \( \beta \) fill the surface \( S \). We denote distance in the one-skeleton of \( \mathcal{A}(S) \) and of \( \mathcal{AC}(S) \) by \( d_{A} \) and \( d_{AC} \), respectively. Recall that the geometric intersection of two curves gives an upper bound for their distance.

**Lemma 2.3.** Suppose \( S \) is a compact, connected surface which is not an annulus. For any \( \alpha, \beta \in \mathcal{C}^{0}(S) \) with \( \iota(\alpha, \beta) > 0 \) we have \( d_{S}(\alpha, \beta) \leq 2 \log_{2}(\iota(\alpha, \beta)) + 2 \). □

For closed orientable surfaces a proof of Lemma 2.3 is given in [25, Lemma 2.1]. A proof in the bounded orientable case is given in [45, Lemma 1.21]. The non-orientable case is an exercise. When \( S = S_{0,2} \) an induction proves that

\[
d_{S}(\alpha, \beta) = 1 + \iota(\alpha, \beta)
\]

for distinct vertices \( \alpha, \beta \in \mathcal{C}(S) \). See [31, Equation 2.3].

**Lemma 2.5.** Suppose \( S \) is a connected compact surface. The following are equivalent.

- \( S \) admits a pseudo-Anosov map or \( S \in \{ S_{0,2}, S_{1} \} \).
- \( S \) admits an ending lamination or \( S \in \{ S_{0,2}, S_{1} \} \).
- \( \mathcal{AC}(S) \) has infinite diameter.
Definition 2.7. A component of \( \mathcal{A} \) is disjoint, then we say that this component of \( \mathcal{A} \) is essential in \( \partial X \). For every component of \( \partial S \) also is cleanly embedded in \( \mathcal{A} \). The difficult cases, \( S_{102} \), and \( S_{003} \), are discussed by Scharlemann [44, Section 2].) To finish, all surfaces with \( \chi(S) < -1 \), and also \( S_{11} \), admit pseudo-Anosov maps. The orientable cases follow from Thurston’s construction [47, Theorem 7]. Penner’s generalization [39, Theorem 4.1] covers the non-orientable cases.

We call a surface \( S \) non-simple if it satisfies any one of, hence all of, the conditions in Lemma 2.5.

2.6. Subsurfaces. Suppose \( X \) is a connected compact subsurface of \( S \). If \( X \) is an annulus with peripheral core curve, then we call \( X \) a peripheral annulus. If \( X \) is not a peripheral annulus, and if every element of \( \partial X \) is essential in \( S \), then we call \( X \) essential.

Definition 2.7. An essential subsurface \( X \subset S \) is cleanly embedded if the following property holds. For every component \( \delta \) of \( \partial X \), if \( \delta \) is peripheral in \( S \), then \( \delta \) is a component of \( \partial S \).

We say that \( \alpha \in \mathcal{A}(S) \) cuts \( X \) if all representatives of \( \alpha \) intersect \( X \). If some representative is disjoint, then we say that \( \alpha \) misses \( X \).

Definition 2.8 (II page 26). Suppose \( X \) and \( Y \) are essential subsurfaces of \( S \). If \( X \) is cleanly embedded in \( Y \), then we say that \( X \) is nested in \( Y \). If \( \partial X \) cuts \( Y \) and also \( \partial Y \) cuts \( X \), then we say that \( X \) and \( Y \) overlap.

2.9. Markings. A finite set of vertices \( \mu \subset \mathcal{A}(S) \) is called a marking. A marking \( \mu \) fills \( S \) if for all \( \beta \in \mathcal{C}(S) \) there is some \( \alpha \in \mu \) so that \( \iota(\alpha, \beta) > 0 \). For a marking \( \mu \subset \mathcal{A}(S) \) define

\[
\iota(\mu) = \frac{1}{2} \sum_{\alpha, \beta \in \mu} \iota(\alpha, \beta).
\]

A marking \( \mu \) is a \( K \)-marking if \( \iota(\mu) \leq K \). Two markings \( \mu, \nu \) are \( L \)-close if \( \iota(\mu, \nu) := \iota(\mu \cup \nu) \leq L \). For any \( K, L \) we define the marking graph \( \mathcal{M}_{K,L}(S) \) to be the graph where vertices are filling \( K \)-markings and edges are given by \( L \)-closeness. As an example, consider \( \mathcal{M} = \mathcal{M}_{1,\delta}(S_1) \). Vertices of \( \mathcal{M} \) are \( 1 \)-markings and correspond to Farey neighbors. Two vertices of \( \mathcal{M} \) are \( 3 \)-close if and only if their union is a Farey triangle. It follows that \( \mathcal{M} \) is quasi-isometric to a Cayley graph for \( \text{SL}(2, \mathbb{Z}) = \text{MCG}(S_1) \). This generalizes.
Definition 2.10 ([31 Section 2.5]). A complete clean marking $\mu = \{(\alpha_i, \beta_i)\}$ consists of

- base curves $\text{base}(\mu) = \{\alpha_i\}$: a maximal simplex in $C(S)$ and
- transversals $\{\beta_i\}$: for each $i$ define $X_i = S - (\cup_{j \neq i} \alpha_j)$ and let $\beta_i \in C(X_i)$ be a Farey neighbor of $\alpha_i$.

If $\mu$ is a complete clean marking, then $\iota(\mu) \leq 2\xi(S) + 6|\chi(S)|$. As discussed in [31] any two complete clean markings are connected by a sequence of elementary moves. Twisting about $\alpha_i$ replaces the transversal $\beta_i$ by a new transversal $\beta_i'$ which is a Farey neighbor of both $\alpha_i$ and $\beta_i$. Flipping swaps the roles of $\alpha_i$ and $\beta_i$. (After flipping, some of the other transversals must be cleaned.)

Following [31] Section 7.1 for any surface $S$ there are choices of $K, L$ so that $\mathcal{M}(S)$ is non-empty, connected, and quasi-isometric to the word metric on $\text{MCG}(S)$. We use $d_M$ to denote distance in the marking graph.

2.11. Three-manifolds and disks. Suppose $M$ is a compact, connected, orientable three-manifold. Recall that $M$ is irreducible if every embedded two-sphere in $M$ bounds a three-ball. Suppose $S$ is a compact, connected subsurface of $\partial M$. We make the following standing assumption.

Definition 2.12. The pair $(M, S)$ is spotless. That is, $M$ is irreducible and no component of $\partial S$ bounds a disk in $M$.

A properly embedded disk $(D, \partial D) \subset (M, S)$ is essential if $\partial D$ is essential in $S$. Let $\mathcal{D}(M, S)$ be the set of essential disks up to ambient isotopy preserving $S$. A subset $\Delta \subset \mathcal{D}(M, S)$ is a multidisk if for all $D, E \in \Delta$ we have $\iota(\partial D, \partial E) = 0$.

Following McCullough [31] Section 5] we place a simplicial structure on $\mathcal{D}(M, S)$ by taking multidisks to be simplices. As with the curve complex, define $d_D$ to be the distance in the one-skeleton of $\mathcal{D}(M, S)$. When $S = \partial M$ we simply write $\mathcal{D}(M)$.

3. Background on coarse geometry

We review a few ideas from coarse geometry. See [9, 14, 18] for a fuller discussion.

3.1. Quasi-isometry. Suppose $r, s, A$ are non-negative real numbers, with $A \geq 1$. If $s \leq A \cdot r + A$, then we write $s \leq_A r$. If $s \leq_A r$ and $r \leq_A s$, then we write $s =_A r$ and call $r$ and $s$ quasi-equal with constant $A$. Define the cut-off function $[r]_c$ by $[r]_c = 0$ if $r < c$ and $[r]_c = r$ if $r \geq c$.

Suppose $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ are metric spaces. For subsets $U, V \subset \mathcal{X}$ define

$$d_\mathcal{X}(U, V) = \text{diam}_\mathcal{X}(U \cup V).$$

Suppose $f \subset \mathcal{X} \times \mathcal{Y}$ is a relation. In a slight abuse of notation, we write $f: \mathcal{X} \to \mathcal{Y}$ and, for $x \in \mathcal{X}$, we write $f(x) = \{y \in \mathcal{Y} \mid xf\}$. For examples of usage, see Definitions 3.6, 3.2 and 1.3.

Fix $A \geq 1$. A relation $f: \mathcal{X} \to \mathcal{Y}$ is an $A$–quasi-isometric embedding if for every $x, y \in \mathcal{X}$ we have $f(x) \neq \emptyset$ and

$$d_\mathcal{X}(x, y) =_A d_\mathcal{Y}(f(x), f(y)).$$

The relation $f$ is a quasi-isometry, and $\mathcal{X}$ is quasi-isometric to $\mathcal{Y}$, if $f$ is an $A$–quasi-isometric embedding and the image of $f$ is $A$–dense: the $A$–neighborhood of $f(\mathcal{X})$ equals $\mathcal{Y}$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
3.2. Geodesics. Fix an interval \([u, v] \subset \mathbb{R}\). A geodesic, connecting \(x\) to \(y\) in \(\mathcal{X}\), is an isometric embedding \(f: [u, v] \to \mathcal{X}\) with \(f(u) = x\) and \(f(v) = y\). Denote the image of \(f\) by \([x, y] \subset \mathcal{X}\).

Fix intervals \([m, n], [p, q] \subset \mathbb{Z}\). An \(A\)-quasi-isometric embedding \(g: [m, n] \to \mathcal{X}\) is called an \(A\)-quasi-geodesic. A function \(g: [m, n] \to \mathcal{X}\) is an \(A\)-unparameterized quasi-geodesic if

- there is an increasing function \(\rho: [p, q] \to [m, n]\) so that \(g \circ \rho: [p, q] \to \mathcal{X}\) is an \(A\)-quasi-geodesic in \(\mathcal{X}\) and
- for all \(i \in [p, q - 1]\), \(\text{diam}_{\mathcal{X}} g[\rho(i), \rho(i + 1)] \leq A\).

This is an adaptation the \((K, \delta, s)\)-quasi-geodesics defined in [30, Section 7].

A subset \(\mathcal{Y} \subset \mathcal{X}\) is \(Q\)-quasi-convex if every \(\mathcal{X}\)-geodesic connecting two points of \(\mathcal{Y}\) lies within a \(Q\)-neighborhood of \(\mathcal{Y}\).

3.3. Hyperbolicity. We now assume that \(\mathcal{X}\) is a connected metric graph where all edges have length one.

Definition 3.4. The graph \(\mathcal{X}\) is \(\delta\)-hyperbolic if, for any three points \(x, y, z\) in \(\mathcal{X}\), any geodesic triangle connecting them is \(\delta\)-thin. That is, the \(\delta\)-neighborhood of any two edges contains the third. We say that \(\mathcal{X}\) is Gromov hyperbolic if \(\mathcal{X}\) is \(\delta\)-hyperbolic for some \(\delta \geq 0\).

An important tool for this paper is the following theorem of the first author and Minsky.

Theorem 3.5 ([30, Theorem 1.1]). The curve complex of an orientable surface is Gromov hyperbolic.

For the remainder of this section assume \(\mathcal{X}\) is \(\delta\)-hyperbolic, assume \(x, y, z \in \mathcal{X}\) are points, and fix geodesics \(k = [x, y], g = [y, z]\), and \(h = [z, x]\).

Definition 3.6. Define \(\rho_k: \mathcal{X} \to k\) to be the closest points relation where

\[\rho_k(z) = \{w \in k \mid d_{\mathcal{X}}(z, w) \leq d_{\mathcal{X}}(z, v) \text{ for all } v \in k\}.\]

The next several lemmas are used in Section 20. The proofs are left as exercises.

Lemma 3.7. For any \(w \in \rho_k(z)\) there is a point of \(g\) within distance \(2\delta\) of \(w\). □

Lemma 3.8. The diameter of \(\rho_k(z)\) is at most \(4\delta\). □

Lemma 3.9. The diameter of \(\rho_g(x) \cup \rho_h(y) \cup \rho_k(z)\) is at most \(6\delta\). □

Lemma 3.10. Suppose \(z'\) is another point in \(\mathcal{X}\) with \(d_{\mathcal{X}}(z, z') \leq R\). Then \(d_{\mathcal{X}}(\rho_k(z), \rho_k(z')) \leq R + 6\delta\). □

Lemma 3.11. Suppose \(k'\) is another geodesic in \(X\) where the endpoints of \(k'\) are within distance \(R\) of the points \(x\) and \(y\). Then \(d_{\mathcal{X}}(\rho_k(z), \rho_{k'}(z)) \leq R + 11\delta\). □

Here is a consequence of Morse stability of quasi-geodesics in Gromov hyperbolic graphs, used in Section 13.15.

Lemma 3.12. For every \(\delta\) and \(A\) there is a constant \(C\) with the following property. If \(\mathcal{X}\) is \(\delta\)-hyperbolic and \(g: [0, N] \to \mathcal{X}\) is an \(A\)-unparameterized quasi-geodesic, then for any \(m < n < p\) in \([0, N]\) we have

\[d_{\mathcal{X}}(x, y) + d_{\mathcal{X}}(y, z) < d_{\mathcal{X}}(x, z) + C,\]

where \(x, y, z = g(m), g(n), g(p)\). □
3.13. **A hyperbolicity criterion.** Here we give a hyperbolicity criterion tailored to our requirements. We thank Brian Bowditch for pointing out an error in our first proof of Theorem 3.15 and for informing us of Gilman’s work [15, 16].

Suppose $\mathcal{X}$ is a graph where all edges have length one. Suppose $\gamma: [0, N] \to \mathcal{X}$ is a loop in $\mathcal{X}$ with unit speed. Any pair of points $a, b \in [0, N]$ gives a chord of $\gamma$. If $N/4 \leq |b - a| \leq 3N/4$, then the chord is $1/4$--separated. The length of the chord is $d_\mathcal{X}(\gamma(a), \gamma(b))$.

Following Gilman [15] Theorem B] we have the following.

**Theorem 3.14.** Suppose $\mathcal{X}$ is a graph where all edges have length one. Then $\mathcal{X}$ is Gromov hyperbolic if and only if there is a constant $K$ so that every loop $\gamma: [0, N] \to \mathcal{X}$ has a $1/4$–separated chord of length at most $N/7 + K$. □

Gilman’s proof goes via the subquadratic isoperimetric inequality [18, Criterion 6.8.M]. See also [6]. We now give our criterion, noting that it is closely related to another paper of Gilman’s [16].

**Theorem 3.15.** Suppose $\mathcal{X}$ is a graph where all edges have length one. Then $\mathcal{X}$ is Gromov hyperbolic if and only if there is a constant $M \geq 0$ and, for all unordered pairs $x, y \in \mathcal{X}^0$ there is a connected subgraph $g_{x,y} \subset \mathcal{X}$ containing $x$ and $y$, with the following properties.

- *(Local)* If $d_\mathcal{X}(x, y) \leq 1$, then $g_{x,y}$ has diameter at most $M$.
- *(Slim triangles)* For all $x, y, z \in \mathcal{X}^0$ the subgraph $g_{x,y}$ is contained in an $M$–neighborhood of $g_{y,z} \cup g_{z,x}$.

**Proof.** To prove the forward direction suppose that $\mathcal{X}$ is $\delta$–hyperbolic. For every $x, y \in \mathcal{X}^0$ take $g_{x,y}$ to be any geodesic connecting $x$ to $y$. Setting $M = \max\{\delta, 1\}$ gives the two properties.

For the backwards direction suppose that $\gamma: [0, N] \to \mathcal{X}$ is a loop. Let $e$ be the empty string and define $I_e = [0, N]$. For any binary string $\omega$ let $I_{\omega 0}$ and $I_{\omega 1}$ be the first and second half of $I_\omega$. Note that if $|\omega| \geq \lceil \log_2 N \rceil$, then $|I_\omega| \leq 1$.

Fix a string $\omega$ and let $[r, s] = I_\omega$. Let $g_\omega = g_{\gamma(r), \gamma(s)}$ be the given connected subgraph of $\mathcal{X}$ containing $\gamma(r)$ and $\gamma(s)$. Note that $g_{0} = g_{1}$ because $\gamma(0) = \gamma(N)$ and because we use unordered pairs as subscripts. Also, for any binary string $\omega$ the subgraphs $g_{\omega 0}, g_{\omega 0}, g_{\omega 1}$ form an $M$–slim triangle. If $|\omega| \leq \lceil \log_2 N \rceil$, then every $x \in g_\omega$ has some point $b \in I_\omega$ so that

$$d_\mathcal{X}(x, \gamma(b)) \leq M(\lceil \log_2 N \rceil - |\omega|) + 2M.$$ 

Since $g_0$ is connected there is a point $x \in g_0$ that lies within the $M$–neighborhoods both of $g_{00}$ and of $g_{01}$. Pick some $b \in J_0$ so that $d_\mathcal{X}(x, \gamma(b))$ is bounded as in the previous paragraph. It follows that there is a point $a \in J_0$ so that $a, b$ are $1/4$–separated and so that

$$d_\mathcal{X}(\gamma(a), \gamma(b)) \leq 2M \lceil \log_2 N \rceil + 2M.$$ 

Thus there is an additive error $K$ large enough so that $\mathcal{X}$ satisfies the criterion of Theorem 3.14 and we are done. □

4. **Natural maps**

There are several natural maps between the complexes and graphs defined in Section 2. Here we review what is known about their geometric properties and give examples relevant to the rest of the paper.
4.1. **Lifting, surgery, and subsurface projection.** Suppose $S$ is not simple. Choose a hyperbolic metric on the interior of $S$ so that every end has infinite area. Fix $X$, a compact essential subsurface of $S$, that is not a peripheral annulus. Let \( \rho_X : S^X \to S \) be the covering map where $X$ lifts homeomorphically and $S^X \cong \text{interior}(X)$. For any $\alpha \in AC(S)$ let $\alpha^X = \rho_X^{-1}(\alpha)$ be the full preimage.

The induced homeomorphism between $X$ and the Gromov compactification of $S^X$ identifies $AC(X)$ with $AC(S^X)$.

**Definition 4.2.** The cutting relation $\kappa_X : AC(S) \to AC(X)$ is defined as follows. A vertex $\beta$ lies in $\kappa_X(\alpha)$ if and only if $\beta$ is an essential non-peripheral component of $\alpha^X$.

We also use the notation $\alpha|X = \kappa_X(\alpha)$. Note that $\alpha$ cuts $X$ if and only if $\alpha|X$ is non-empty.

**Definition 4.3.** Suppose $S$ is not an annulus. The surgery relation $\sigma_X : AC(S) \to C(S)$ is defined as follows. A vertex $\beta$ lies in $\sigma_S(\alpha)$ if and only if $\beta \in C(S)$ is a boundary component of a regular neighborhood of $\alpha \cup \partial S$.

**Definition 4.4.** The subsurface projection relation $\pi_X : AC(S) \to C(X)$ is defined by $\pi_X = \sigma_X \circ \kappa_X$. When $X$ is an annulus define $\pi_X = \kappa_X$.

If $\alpha, \beta \in AC(S)$ both cut $X$ define
\[
d_X(\alpha, \beta) = \text{diam}_X(\pi_X(\alpha) \cup \pi_X(\beta)).
\]
This is the subsurface projection distance between $\alpha$ and $\beta$ in $X$. If $\alpha, \beta \subset S$ are disjoint there is a bound on their subsurface projection distance following [31, Lemma 2.3] and the remarks in the section Projection Bounds of [35, page 29].

**Lemma 4.5.** Suppose $\alpha$ and $\beta$ span an edge in $AC(S)$ and both cut $X$. Then $\text{diam}_X(\pi_X(\alpha))$ is at most two. Also, $d_X(\alpha, \beta)$ is at most two (unless $S = S_{1,1}$ and $X = S_{0,2}$; the bound is then replaced by three).

**Corollary 4.6.** Fix $X \subset S$. Suppose $\{\beta_i\}_{i=0}^N$ is a path in $AC(S)$, with $N \geq 1$. Suppose $\beta_i$ cuts $X$, for all $i$. Then $d_X(\beta_0, \beta_N) \leq 2N$ (unless $S = S_{1,1}$ and $X = S_{0,2}$; the bound is then replaced by $3N$).

It is crucial to note that if some vertex of $\{\beta_i\}$ misses $X$, then the projection distance $d_X(\beta_0, \beta_N)$ may be arbitrarily large compared to $N$. Corollary 4.6 can be greatly strengthened when the path is a geodesic.

**Theorem 4.7** ([31, Theorem 3.1] (Bounded geodesic image)). There is a constant $M_0 = M_0(S)$ with the following property. Fix $X \subset S$. Suppose $\{\beta_i\}_{i=0}^n$ is a geodesic in $C(S)$. Suppose $\beta_i$ cuts $X$, for all $i$. Then $d_X(\beta_0, \beta_n) \leq M_0$.

Here is a kind of converse to Lemma 4.5.

**Lemma 4.8.** For every $K \in \mathbb{N}$ there is a number $L \in \mathbb{N}$ with the following property. For any $\alpha, \beta \in AC(S)$, if $d_X(\alpha, \beta) \leq K$ for all $X \subset S$, then $\iota(\alpha, \beta) \leq L$.

Corollary D of [12] gives a more precise relation between projection distance and intersection number.

**Proof of Lemma 4.8.** We only sketch the contrapositive. Suppose we are given a sequence of curves $\alpha_n, \beta_n$ so that $\iota(\alpha_n, \beta_n)$ tends to infinity. Passing to subsequences and applying elements of the mapping class group we may assume that $\alpha_n = \alpha_0$
for all \(n\). Setting \(c_n = \iota(\alpha_0, \beta_n)\) and passing to subsequences again we may assume that \(\beta_n/c_n\) converges to \(\lambda \in \mathcal{PML}(S)\), the projectivization of Thurston’s space of measured laminations. Let \(Y\) be any connected component of the subsurface filled by \(\lambda\), chosen so that \(\alpha_0\) cuts \(Y\). Note that \(\pi_Y(\beta_n)\) converges to \(\lambda|Y\). Again applying Kobayashi’s argument [27, Proposition 2.2], the distance \(d_Y(\alpha_0, \beta_n)\) tends to infinity. \(\Box\)

4.9. Inclusions. We now record a well-known fact.

**Lemma 4.10.** The inclusion \(\nu: \mathcal{C}(S) \to \mathcal{AC}(S)\) is a quasi-isometry. The surgery map \(\sigma_S: \mathcal{AC}(S) \to \mathcal{C}(S)\) is a quasi-inverse for \(\nu\).

**Proof.** Fix \(\alpha, \beta \in \mathcal{C}(S)\). Since \(\nu\) is simplicial we have \(d_{\mathcal{AC}}(\alpha, \beta) \leq d_S(\alpha, \beta)\). In the other direction, let \(\{\alpha_i\}_{i=0}^N\) be a geodesic in \(\mathcal{AC}(S)\) connecting \(\alpha\) to \(\beta\). Since every \(\alpha_i\) cuts \(S\) we apply Corollary 4.6 and deduce that \(d_S(\alpha, \beta) \leq 2N + 2\).

Note that the composition \(\sigma_S \circ \nu\) is the identity on \(\mathcal{C}(S)\). Also, for any arc \(\alpha \in \mathcal{A}(S)\) we have \(d_{\mathcal{AC}}(\alpha, \nu(\sigma_S(\alpha))) = 1\). Finally, \(\mathcal{C}(S)\) is \(1\)-dense in \(\mathcal{AC}(S)\), as any arc \(\gamma \subset S\) is disjoint from the one or two curves of \(\sigma_S(\gamma)\). \(\Box\)

Brian Bowditch raised the question, at the Newton Institute in August 2003, of the geometric properties of the inclusion \(\nu: \mathcal{A}(S) \to \mathcal{AC}(S)\). The natural assumption, that this inclusion is again a quasi-isometric embedding, is false. In this paper we exactly characterize how the inclusion distorts distance.

We now move up a dimension. Suppose \((M, S)\) is a spotless pair. The natural map \(\nu: \mathcal{D}(M, S) \to \mathcal{C}(S)\) takes an essential disk \(D\) to its boundary \(\partial D\). Since \((M, S)\) is spotless, the curve \(\partial D\) determines the disk \(D\), up to isotopy; also, \(\partial D\) cannot be peripheral in \(S\). Thus \(\nu\) is injective and well-defined. We call the image \(\nu(D)\) the disk set.

The first author and Minsky have shown the following.

**Theorem 4.11** ([32, Theorem 1.1]). The disk set is a quasi-convex subset of the curve complex. \(\Box\)

It is natural to ask if \(\nu: \mathcal{D}(M, S) \to \mathcal{C}(S)\) is a quasi-isometric embedding, as that would directly imply the hyperbolicity of \(\mathcal{D}(M, S)\). In fact, \(\nu\) again badly distorts distance; this paper gives an exact characterization.

4.12. Markings and the mapping class group. As discussed in Section 2.9 the marking graph \(\mathcal{M}(S)\) is quasi-isometric to the word metric on the mapping class group. Using subsurface projections, the first author and Minsky obtained a distance estimate for the marking graph and thus for the mapping class group.

**Theorem 4.13** ([31, Theorem 6.12] (Distance estimate)). There is a constant \(C_0 = C_0(S)\) so that, for any \(c \geq C_0\) there is a constant \(A\) with

\[
d_{\mathcal{M}}(\mu, \nu) = A \sum [d_X(\mu, \nu)]_c
\]

independent of the choice of \(\mu\) and \(\nu\). Here the sum ranges over all essential subsurfaces \(X \subset S\). \(\Box\)

This, and their similar estimate for the pants graph, is a model for the distance estimates given below. Notice that a filling marking \(\mu \in \mathcal{M}(S)\) cuts all essential subsurfaces of \(S\). It is not an accident that the sum ranges over the same set.
5. Holes in general and the lower bound on distance

Suppose that $S$ is a compact, connected surface. In this paper the vertices of any combinatorial complex $G = G(S)$ will be multicurves in $S$. We typically assume that vertices of $G$ are connected by edges if and only if there are representatives that are disjoint. This assumption is made only to simplify the proofs; all arguments work in the case where, with the marking graph, we replace multicurves by markings and where adjacent vertices are allowed to have uniformly bounded intersection. In all cases $G$ will be connected. There is a natural map $\nu: G \to \mathcal{AC}(S)$ taking a vertex of $G$ to the corresponding multicurve. Examples in the literature include the marking graph \cite{31}, the pants graph \cite{10, 2}, the Hatcher-Thurston graph \cite{23}, the complex of separating curves \cite{8}, the arc complex, and the curve complex.

For any combinatorial complex $G$ defined in this paper other than the curve complex we denote distance in the one-skeleton of $G$ by $d_G$. Distance in the one-skeleton of $\mathcal{AC}(S)$ will always be denoted by $d_S$.

5.1. Holes, defined. Suppose $S$ is not simple. Suppose $G = G(S)$ is a combinatorial complex. Suppose $X$ is a cleanly embedded subsurface of $S$. A vertex $\alpha \in G$ cuts $X$ if some component of $\alpha$ cuts $X$.

**Definition 5.2.** We say that $X$ is a hole for $G$ if every vertex of $G$ cuts $X$.

Almost equivalently, if $X$ is a hole, then the subsurface projection $\pi_X: G \to \mathcal{C}(X)$ never takes the empty set as a value. Note that the entire surface $S$ is always a hole, regardless of our choice of $G$. A peripheral annulus cannot be cleanly embedded (unless $S$ is also an annulus), so generally cannot be a hole. A hole $X \subset S$ is strict if $X$ is not homeomorphic to $S$.

**Definition 5.3.** If $X$ is a hole for $G(S)$ and if $\pi_X(G) \subset \mathcal{C}(X)$ has diameter at least $R$ we say that the hole $X$ has diameter at least $R$.

We now classify holes for the arc complex.

**Example 5.4.** Suppose $S = S_{g,b}$ with $b > 0$ and consider the arc complex $\mathcal{A}(S)$. The holes, up to isotopy, are exactly the cleanly embedded surfaces which contain $\partial S$. So, for example, if $S$ is planar, then only $S$ is a hole for $\mathcal{A}(S)$. The same holds for $S = S_{1,1}$. In these cases it is an exercise to show that $\mathcal{C}(S)$ and $\mathcal{A}(S)$ are quasi-isometric. In all other cases the arc complex admits infinitely many holes. Since the mapping class group acts on the arc complex, all non-simple holes for $\mathcal{A}(S)$ have infinite diameter.

The classification of holes for the disk complex is more difficult and is postponed until Sections \[122\]. We here content ourselves with the first non-trivial example.

**Example 5.5.** Suppose $F$ is a non-simple, orientable surface with boundary. Let $M = F \times I$. Let $X = F \times \{0\}$ and $Y = F \times \{1\}$. Since $Y$ is incompressible (Definition \[5.2\]) in $M$ we deduce that $X$ is a hole for $\mathcal{D}(M)$; similarly $Y$ is a hole.

This example was the genesis of our program to understand the intrinsic geometry of $\mathcal{D}(M,S)$. The occurrence or non-occurrence of disjoint holes $X, X' \subset S$ is highly important for the intrinsic geometry of $G(S)$. In the presence of symmetry there can be a relationship between $\pi_X|G$ and $\pi_{X'}|G$, as follows.
Definition 5.6. Suppose \(X, X'\) are holes for \(G\), both of infinite diameter. Then \(X\) and \(X'\) are paired if there is a homeomorphism \(\tau: X \to X'\) and a constant \(L_3\) so that
\[
d_{X'}(\pi_{X'}(\gamma), \tau(\pi_X(\gamma))) \leq L_3
\]
for every \(\gamma \in G\). Furthermore, if \(Y \subset X\) is a hole, then \(\tau\) pairs \(Y\) with \(Y' = \tau(Y)\). Lastly, pairing is symmetric; if \(\tau\) pairs \(X\) with \(X'\), then \(\tau^{-1}\) pairs \(X'\) with \(X\).

Definition 5.7. Two holes \(X\) and \(Y\) interfere if either
- \(X \cap Y \neq \emptyset\),
- \(X\) is paired with \(X'\) and \(X' \cap Y \neq \emptyset\), or
- \(Y\) is paired with \(Y'\) and \(X \cap Y' \neq \emptyset\).

Examples arise in the symmetric arc complex and, as in Example 5.5, in the discussion of \(I\)-bundles inside of a three-manifold.

5.8. Projection to holes is coarsely Lipschitz. The following “coarse Lipschitz projection” lemma is used repeatedly throughout the paper.

Lemma 5.9. Suppose \(G(S)\) is a combinatorial complex. Suppose \(X\) is a hole for \(G\). Then for any \(\alpha, \beta \in G\) we have
\[
d_X(\alpha, \beta) \leq 2 + 2 \cdot d_G(\alpha, \beta).
\]
The additive error is required only when \(\alpha = \beta\).

Proof. This follows from Corollary 4.6 and our assumption that vertices of \(G\) connected by an edge represent disjoint multicurves. \(\square\)

5.10. Infinite diameter holes. We may now state a first answer to Bowditch’s question.

Lemma 5.11. Suppose \(G(S)\) is a combinatorial complex. Suppose there is a strict hole \(X\) having infinite diameter. Then \(\nu: \mathcal{G} \to \mathcal{AC}(S)\) is not a quasi-isometric embedding. \(\square\)

This lemma and Example 5.4 completely determine when the inclusion of \(\mathcal{A}(S)\) into \(\mathcal{AC}(S)\) is a quasi-isometric embedding. It quickly becomes clear that the set of holes tightly constrains the intrinsic geometry of a combinatorial complex.

Lemma 5.12. Suppose \(G(S)\) is a combinatorial complex invariant under the natural action of \(\text{MCG}(S)\) on \(\mathcal{AC}(S)\). Then every non-simple hole for \(G\) has infinite diameter. Furthermore, if \(X\) and \(Y\) are disjoint non-simple holes for \(G\), then there is a quasi-isometric embedding of \(\mathbb{Z}^2\) into \(\mathcal{G}\). \(\square\)

We do not use Lemmas 5.11 or 5.12 and so omit the proofs. Instead our interest lies in proving the far more powerful distance estimate (Theorems 5.14 and 13.1) for \(G(S)\).

5.13. A lower bound on distance. The sum of projection distances in holes gives a lower bound for distance.

Theorem 5.14. Fix \(S\), a compact, connected, non-simple surface. Suppose \(G(S)\) is a combinatorial complex. Then there is a constant \(C_0\) so that for all \(c \geq C_0\) there is a constant \(A\) satisfying
\[
\sum [d_X(\alpha, \beta)]_c \leq A \cdot d_G(\alpha, \beta).
\]
Here \(\alpha, \beta \in G\) and the sum is taken over all holes \(X\) for the complex \(G\).
Proof. We follow the proofs of Theorems 6.10 and 6.12 of [31], practically word for word. The only changes necessary are as follows. We replace the sum over all subsurfaces by the sum over all holes. We replace Lemma 2.5 of [31], which records how markings differing by an elementary move project to an essential subsurface, by Lemma 5.9 of this paper, which records how \( \mathcal{G} \) projects to a hole.

One major goal of this paper is to give criteria sufficient to obtain the reverse inequality: Theorem 13.1.

6. Holes for the non-orientable surface

Fix \( F \) a compact, connected, non-orientable surface. We take \( \mathcal{C}_2(F) \subset \mathcal{C}(F) \) to be the subcomplex spanned by vertices representing two-sided curves. Note that \( \mathcal{C}_2(F) \) is isometrically embedded in and is \( 1 \)-dense in \( \mathcal{C}(F) \). Thus these two complexes are quasi-isometric.

Let \( S \) be the orientation double cover with covering map \( \rho_F: S \to F \). Let \( \tau: S \to S \) be the associated involution; so for all \( x \in S \), \( \rho_F(x) = \rho_F(\tau(x)) \).

Definition 6.1. A multicurve \( \gamma \subset \mathcal{AC}(S) \) is symmetric if \( \tau(\gamma) \cap \gamma = \emptyset \) or \( \tau(\gamma) = \gamma \). A multicurve \( \gamma \) is invariant if there is a curve or arc \( \delta \subset F \) so that \( \gamma = \rho_F^{-1}(\delta) \). The same definitions hold for subsurfaces \( X \subset S \).

Definition 6.2. The invariant complex \( \mathcal{C}(S) \) is a simplicial complex with vertex set consisting of invariant multicurves (without arcs). There is a \( k \)-simplex for every collection of \( k+1 \) distinct isotopy classes having pairwise disjoint representatives.

Notice that \( \mathcal{C}(S) \) is simplicially isomorphic to \( \mathcal{C}_2(F) \). There is also a natural relation \( \nu: \mathcal{C}(S) \to \mathcal{C}(S) \).

Lemma 6.3. \( \nu: \mathcal{C}(S) \to \mathcal{C}(S) \) is a quasi-isometric embedding.

Before giving the proof, note the following.

Corollary 6.4 ([3, Section 5]). \( \mathcal{C}(F) \) is Gromov hyperbolic.

Proof of Lemma 6.3. Since \( \nu \) sends adjacent vertices in \( \mathcal{C}(S) \) to adjacent simplices in \( \mathcal{C}(S) \) we have

\[
ds_S(\alpha, \beta) \leq d_{\mathcal{C}(S)}(\alpha, \beta),
\]

as long as \( \alpha \) and \( \beta \) are distinct in \( \mathcal{C}(S) \).

The other half of the proof of Lemma 6.3 consists of showing that \( S \) is the only hole for \( \mathcal{C}(S) \) with large diameter. After a discussion of Teichmüller geodesics we prove the following.

Lemma [16.8] There is a constant \( K \) with the following property. Suppose \( \alpha, \beta \) are invariant multicurves in \( S \). Suppose \( X \subset S \) is an essential subsurface where \( d_X(\alpha, \beta) > K \). Then \( X \) is symmetric.

Corollary 6.6. With \( K \) as in Lemma 16.8. If \( X \) is a hole for \( \mathcal{C}(S) \) with diameter greater than \( K \), then \( X = S \).

Proof. Suppose \( X \) is a cleanly embedded strict subsurface of \( S \). Suppose \( \text{diam}_X(\mathcal{C}(S)) \geq K \). Thus \( X \) is symmetric. It follows that \( \partial X - \partial S \) is also symmetric. Since \( \partial X \) does not cut \( X \), we deduce that \( X \) is not a hole for \( \mathcal{C}(S) \). This proves the corollary.

□
This corollary, together with the upper bound (Theorem 13.1), completes the proof of Lemma 6.3.

7. Holes for the arc complex

We generalize the definition of the arc complex and classify its holes.

Definition 7.1. Suppose $F$ is a non-simple surface with boundary. Suppose $\Delta$ is a non-empty union of components of $\partial F$. The arc complex $\mathcal{A}(F, \Delta)$ is the subcomplex of $\mathcal{A}(F)$ spanned by essential arcs $\alpha \subset F$ with $\partial \alpha \subset \Delta$.

Note that $\mathcal{A}(F, \partial F)$ and $\mathcal{A}(F)$ are identical.

Lemma 7.2. Suppose $X$ is cleanly embedded in $F$. Then $X$ is a hole for $\mathcal{A}(F, \Delta)$ if and only if $\Delta \subset \partial X$.

This follows from the definition of a hole. We now have a straightforward observation.

Lemma 7.3. If $X$ and $Y$ are holes for $\mathcal{A}(F, \Delta)$, then $X \cap Y \neq \emptyset$.

The proof follows immediately from Lemma 7.2. Lemma 5.12 indicates that Lemma 7.3 is necessary to prove that $\mathcal{A}(F, \Delta)$ is Gromov hyperbolic.

Suppose now that $F$ is non-simple, has non-empty boundary, and is non-orientable. Let $\rho_F : S \to F$ be the orientation double cover and let $\tau : S \to S$ be the induced involution. Fix $\Delta' \subset \partial F$ and let $\Delta = \rho_F^{-1}(\Delta')$.

Definition 7.4. We define $\mathcal{A}^\tau(S, \Delta)$ to be the invariant arc complex: vertices are invariant multi-arcs and simplices arise from disjointness.

Again, $\mathcal{A}^\tau(S, \Delta)$ is simplicially isomorphic to $\mathcal{A}(F, \Delta')$. If $X \cap \tau(X) = \emptyset$ and $\Delta \subset X \cup \tau(X)$, then the subsurfaces $X$ and $\tau(X)$ are paired holes, as in Definition 5.6. Notice as well that all non-simple symmetric holes $X$ for $\mathcal{A}^\tau(S, \Delta)$ have infinite diameter.

Unlike $\mathcal{A}(F, \Delta')$ the complex $\mathcal{A}^\tau(S, \Delta)$ may have disjoint holes. Nonetheless, we have the following.

Lemma 7.5. Any two non-simple holes for $\mathcal{A}^\tau(S, \Delta)$ interfere.

Proof. Suppose $X$ and $Y$ are non-simple holes for the $\tau$–invariant arc complex $\mathcal{A}^\tau(S, \Delta)$. It follows from Lemma 16.8 that $X$ is symmetric and $\Delta \subset X \cup \tau(X)$. The same holds for $Y$. Thus $Y$ must cut $X$, $\tau(X)$, or both.

8. Background on three-manifolds

We review the necessary material regarding three-manifolds. See [24, 43] for detailed presentations. Throughout $M$ is a compact, connected, orientable three-manifold. Recall that $(M, S)$ is assumed to be spotless (Definition 2.12).

If $N$ is a compact submanifold of $M$, then fr$(N)$, the frontier of $N$ in $M$, is the closure of $\partial N - \partial M$.

8.1. Surgery. Suppose $F$ is a surface embedded in $M$. We consider two cases. Either $F \subset \partial M$ is a subsurface of the boundary or $(F, \partial F) \subset (M, \partial M)$ is properly embedded.
Suppose \((D, \partial D) \subset (M, F)\) is an embedded disk. We call \(D\) a \textit{surgery disk} for \(F\) if

- \(D \cap \partial F = \emptyset,\)
- \(D \cap F = \partial D,\) and
- \(D \cap \partial M = \partial D \cap \partial M.\)

We may \textit{surge} \(F\) along \(D\) to obtain \(F_D\), as follows. Let \(N\) be a closed regular neighborhood of \(D\). Remove from \(F\) the annulus \(N \cap F\). Form \(F_D\) by gluing on both disk components of \(\text{fr}(N) - F\) and taking the closure. When \(F \subset \partial M\) we must also isotope \(\text{int}(F_D)\) into \(\text{int}(M)\) to ensure that \(F_D\) is properly embedded.

\textbf{Definition 8.2.} A surgery disk \(D\) for \(F \subset M\) is a \textit{compressing disk} if \(\partial D \subset F\) is an essential curve. If \(F\) admits no compressing disk, then \(F\) is \textit{incompressible} in \(M\).

The triple \((B, \alpha, \beta)\) is a \textit{bigon} exactly when \(B\) is a disk and \(\alpha, \beta\) are arcs in \(\partial B\) so that \(\alpha \cup \beta = \partial B\) and \(\alpha \cap \beta = \partial \alpha = \partial \beta\). Suppose \((B, \alpha, \beta) \subset (M, F, \partial M)\) is an embedded bigon. We call \(B\) a \textit{surgery bigon} for \(F\) if

- \(B \cap \partial F = \partial \alpha = \partial \beta,\)
- \(B \cap F = \alpha,\) and
- \(B \cap \partial M = \partial B\) if \(F \subset \partial M\) while
- \(B \cap \partial M = \beta\) if \(F\) is properly embedded.

Again, we may surge \(F\) along \(B\). Let \(N\) be a closed regular neighborhood of \(B\). Remove the rectangle \(N \cap F\) from \(F\). Form \(F_B\) by gluing on the two bigon components of \(\text{fr}(N) - F\) and taking the closure. If \(F\) lies in \(\partial M\) we isotope \(\text{int}(F_B)\) into \(\text{int}(M)\) to ensure that \(F_B\) is properly embedded.

\textbf{Definition 8.3.} A surgery bigon \((B, \alpha, \beta)\) for \(F \subset M\) is a \textit{boundary compression} if \(\beta\) is an essential arc in \(\partial M - \partial F\). A boundary compression is \textit{essential in} \(F\) if \(\alpha\) is an essential arc in \(F\).

In other work, boundary compressions \(B\) for \(F\) are required to have the latter property. We divide the definition in two because, for us, \(F\) will typically be a properly embedded disk in \(M\).

Suppose now that \((F, \partial F)\) is properly embedded in \((M, S)\). Suppose \(X\) is a subsurface of \(S\). Properly isotope \(F\), fixing \(\partial S\) pointwise, to make \(\partial F\) and \(\partial X\) tight. If \((B, \alpha, \beta)\) is a boundary compression of \(F\) so that \(\beta \subset X\), then we say that \(F\) is boundary compressible \textit{into} \(X\).

\textbf{8.4. Boundary compression.} We now begin our study of boundary compressions of disks.

\textbf{Remark 8.5.} Suppose \(\Gamma\) is a multicurve in \(S\). Suppose \((D, \partial D) \subset (M, S)\) is an essential disk. Suppose that \(\Gamma\) and \(\partial D\) are tight. Suppose \(B\) is a boundary compression of \(D\) into \(S - n(\Gamma)\), where \(n(\Gamma)\) is an open product neighborhood. Then \(\iota(D_B, \Gamma) \leq \iota(D, \Gamma)\).

In a slight abuse of terminology, in the above definition we allow the multicurve \(\Gamma\) to have parallel components.

\textbf{Lemma 8.6.} Suppose \(X\), a connected essential subsurface of \(S\), compresses in \(M\). Suppose \((D, \partial D) \subset (M, S)\) is an essential disk. Suppose \(\partial X\) and \(\partial D\) are tight.
Then
\begin{itemize}
\item $D$ is boundary compressible into $X$ or
\item $D$ is disjoint from some compressing disk $(E, \partial E) \subset (M, X)$.
\end{itemize}

Proof. Suppose $(E, \partial E) \subset (M, X)$ is a compressing disk. If $\iota(\partial D, \partial E) = 0$, then we are done, using the irreducibility of $M$. Suppose instead that the geometric intersection number is positive. Ambiently isotope $E$, fixing $\partial X$ pointwise, to make $\partial E$ tight with respect to $\partial D$. This can be done because $\partial D$ and $\partial X$ are tight. Since $M$ is irreducible we may further isotope $E$, fixing $\partial E$ pointwise, to remove all simple closed curves of $D \cap E$. Since $D \cap E$ remains non-empty let $B$ be an outermost bigon of $E - D$. So $(B, \alpha, \beta) \subset (M, D, X)$ is a surgery bigon.

If $\beta$ is inessential in $S - n(\partial D)$, then $\beta$ is parallel, in $S$, to an arc $\gamma \subset \partial D$. Thus there is an ambient isotopy of $D$ pushing $\gamma$ past $\beta$. This reduces $\iota(\partial D, \partial E)$, a contradiction. \hfill $\square$

8.7. Band sums. A \textit{band sum} is the inverse operation to boundary compression. Fix disjoint disks $D', D'' \subset (M, S)$. Fix a simple arc $\delta \subset S$ so that $\delta$ meets each of $D'$ and $D''$ in exactly one endpoint. Let $N \subset M$ be a closed regular neighborhood of $\delta$. Let $D$ be the disk formed by adding to $(D' \cup D'') - N$ the rectangle component of fr$(N) - (D' \cup D'')$. The disk $D$ is the result of \textit{band summing} $D'$ to $D''$ along $\delta$. Note that $D$ has a boundary compression \textit{dual} to $\delta$ yielding $D' \cup D''$: that is, there is a boundary compression $B \subset N$ for $D$ so that $B \cap \delta$ is a single point and so that $DB = D' \cup D''$.

8.8. Compression bodies. We pause to discuss a few special three-manifolds.

Definition 8.9. Suppose $F$ is a compact, connected, orientable surface. Let $V$ be a three-manifold obtained from $F \times I$ by attaching two-handles to $F \times \{0\}$ and capping off any resulting two-spheres (disjoint from $F \times \{1\}$) with three-balls. Then $V$ is a \textit{compression body with exterior boundary} $\partial_+ V$ equal to $F \times \{1\}$, with \textit{vertical boundary} $\partial_0 V$ equal to $\partial F \times I$, and with \textit{interior boundary} $\partial_- V$ equal to the closure of $\partial V - (\partial_+ V \cup \partial_0 V)$.

When $\partial_- V = \partial_0 V = \emptyset$, then $V$ is called a \textit{handlebody}. In this case the \textit{genus} of $V$ is the genus of $\partial_+ V$.

Disk components of $\partial V - \partial_+ V$ are called \textit{spots}. When all components of $\partial V - \partial_+ V$ are spots, then $V$ is homeomorphic to a handlebody (ignoring the given partition of $\partial V$).

Following [5, Theorem 2.1] for any spotless pair $(M, S)$ there is a \textit{characteristic compression body} $V \subset M$ so that $V$ has no spots, $S = \partial_+ V = V \cap \partial M$, and the inclusion induces an isomorphism $\mathcal{D}(V, S) \cong \mathcal{D}(M, S)$. If $X$ is a subsurface of $S$, then the characteristic compression body $W \subset M$ for $X$ is contained in $V$. If $X$ is not a hole for $\mathcal{D}(V, S)$, then the image of $\mathcal{D}(W, X)$ in $\mathcal{D}(V, S)$ has finite diameter. When $X$ is a hole the geometry of the inclusion $\mathcal{D}(W, X) \to \mathcal{D}(V, S)$ depends on how $W$ is contained in $V$. The inclusion need not be a quasi-isometric embedding; see Remark [19, 10] for a brief discussion.

By the above, to understand disk complexes of spotless pairs $(M, S)$ it suffices to study $\mathcal{D}(V, S)$, where $V$ is a compression body without spots. However, this does not appear to simplify any of the arguments.
8.10. Interval bundles.

**Definition 8.11.** Suppose $F$ is a compact, connected surface. Let $T$ be the orientation $I$–bundle over $F$: the unique $I$–bundle over $F$ with orientable total space. Let $\rho_F: T \to F$ be the associated bundle map. Then $T$ has vertical boundary $\partial_v T$ equal to $\rho_F^{-1}(\partial F)$ and has horizontal boundary $\partial_h T$ equal to the closure of $\partial T - \partial_v T$.

In general, if $A \subset T$ is a union of fibers of the map $\rho_F$, then $A$ is vertical. If $F$ is orientable, then $T \cong F \times I$. If $F$ is non-orientable and if $\alpha \subset F$ is an orientation-reversing simple closed curve, then $\rho_F^{-1}(\alpha) \subset T$ is a vertical one-sided Möbius band.

If $F$ is not a sphere or a projective plane, then $T$ is irreducible. Also $\partial_h T$ is always incompressible in $T$. If $F$ is not homeomorphic to a disk, then $\partial_v T$ is incompressible in $T$. If $\partial_v T \neq \emptyset$, then any proper vertical surface in $T$ admits a (vertical) surgery bigon. However, no vertical surface in $T$ may be boundary compressed into $\partial_h T$.

**Lemma 8.12.** Suppose $F$ is a compact, connected surface with $\partial F \neq \emptyset$. Let $\rho_F: T \to F$ be the orientation $I$–bundle over $F$. Let $X$ be a component of $\partial_h T$. Let $D \subset T$ be a properly embedded disk. Suppose

- $\partial D$ is essential in $\partial T$,
- $\partial D$ and $\partial X$ are tight and
- $D$ cannot be boundary compressed into $X$.

Then $D$ may be ambiently isotoped to be vertical with respect to $T$.

**Proof.** Since $\partial D$ is tight with respect to $\partial X$ we may ambiently isotope $D$ to make $D \cap \partial_v T$ vertical.

Choose $\alpha \subset F$, a multi-arc, cutting $F$ into a collection of hexagons. Let $A = \rho_F^{-1}(\alpha)$. Thus $A$ cuts $T$ into a collection of hexagonal prisms. We choose $\alpha$ so that $A$ and $D$ are in general position. Thus $A \cap \partial_v T$ is disjoint from $D \cap \partial_v T$. (If $F$ is orientable, set $Y = \partial_h T - X$. In this case choose $\alpha$ so that $D \cap Y$ is disjoint from $A \cap Y$.)

Our first goal is to ambiently isotope $D$ inside of $T$ to make $D \cap A$ vertical. Each isotopy will decrease the complexity ($|D \cap A|, |\partial D \cap \partial A|$). Also, each isotopy will preserve the following properties.

- $D \cap \partial_v T$ is vertical and disjoint from $A \cap \partial_v T$.
- $\partial D$ is tight with respect to $\partial X$.
- If $F$ is orientable, then $D \cap Y$ is disjoint from $D \cap A$.

If all components of $D \cap A$ connect opposite sides of $\partial_h A$, then we are done. Suppose instead that $\beta \subset D \cap A$ cuts off of $A$ either an innermost disk or an outermost bigon. In the former case, as $T$ is irreducible, there is an ambient isotopy of $D$ in $T$, fixing $\partial T$ pointwise, that eliminates $\beta$. This decreases $|D \cap A|$ and preserves the properties.

Suppose instead that $\beta$ cuts an outermost bigon $(B, \beta, \gamma)$ off of $A$. Note that $\gamma \subset X \cap \partial_h A$. Since $D$ cannot be boundary compressed into $X$, deduce that $\gamma$ is inessential in $\partial T - \partial D$. It follows that there is a bigon $(C, \gamma, \delta)$ lying in $\partial T$ with $\delta \subset \partial D$. Since $\gamma \cap \partial X = \emptyset$ and since $\partial D, \partial X$ are tight deduce that $C \cap \partial X = \emptyset$. Thus $C \subset X$. Thus the ambient isotopy pushing $\delta$ across $C$ reduces $|\partial D \cap \partial A|$ and preserves the properties.

Our next goal is to make $D$ vertical in the complement of $A$. Let $H$ be the closure (in the path metric) of a component of $T - A$. So $H$ is a hexagonal prism.
Suppose $D'$ is a component of $D \cap H$. So $D'$ is a disk with $D' \cap \partial_v H$ disjoint from $\partial_v A$ (the vertical edges of $H$). Let $\epsilon$ be an arc of $D' \cap \partial_v H$. (If $F$ is orientable, we choose $\epsilon \subset D' \cap Y$.) Let $E = \rho^{-1}(\epsilon)$ and let $N = N(E)$ be a closed product neighborhood of $E$. So $N$ is a rectangular solid. Three consecutive sides of $D'$ are contained in $\partial N$: two vertical sides and $\epsilon$. See Figure 8.13. For any arc $\beta$ of $D' \cap \partial_v N$ either $\beta$ can be made vertical by ambient isotopy or $\beta$ cuts a disk or bigon out of $\partial_v N$. In the latter cases we proceed as in the first half of the proof, making $D'$ vertical in $H$. Performing all of these isotopies makes $D$ vertical and the lemma is proved.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure8.13.png}
\caption{The rectangular solid $N$ contains $D' \cap N$.}
\end{figure}

9. Holes for the disk complex

Here we begin to classify holes for the disk complex, a more difficult analysis than for the arc complex. Suppose $(M, S)$ is a spotless pair. Recall that there is a natural inclusion $\nu: \mathcal{D}(M, S) \to \mathcal{C}(S)$.

Remark 9.1. Suppose $X$ is cleanly embedded in $S$. Then $X$ is a hole for $\mathcal{D}(M, S)$ if every essential disk $(D, \partial D) \subset (M, S)$ cuts $X$. Equivalently, $S - n(X)$ is incompressible in $M$. Some authors call $X$ disk-busting for $(M, S)$.

The classification of holes $X$ for $\mathcal{D}(M, S)$ breaks into three cases. Either $X$ is an annulus, $X$ compresses in $M$, or $X$ is incompressible in $M$. For each we have a theorem.

Theorem 10.1. Suppose $X$, an annulus, is a hole for $\mathcal{D}(M, S)$. Then the diameter of $X$ is at most 11.

Theorem 11.10. Suppose $X$ is a compressible hole for $\mathcal{D}(M, S)$ with diameter at least 15. Then there are essential disks $D, E \subset M$ so that

- $\partial D, \partial E \subset X$ and
- $\partial D$ and $\partial E$ fill $X$. 

Theorem 12.1. Suppose $X$ is an incompressible hole for $D(M,S)$ with diameter at least 57. Then there is an $I$–bundle $\rho_F : T \to F$ embedded in $M$ so that
- $\partial_h T \subset S$,
- $X$ is a component of $\partial_h T$,
- some component of $\partial_v T$ is boundary parallel into $S$,
- $F$ supports a pseudo-Anosov map.

These theorems have a corollary.

Corollary 9.2. If $X$ is hole for $D(M,S)$ with diameter at least 57, then $X$ has infinite diameter.

Proof. If $X$ is a hole with diameter at least 57, then either Theorem 11.10 or Theorem 12.1 applies.

If $X$ is compressible, then Dehn twists, in opposite directions about the given disks $D$ and $E$, yield a homeomorphism $f : M \to M$ so that $f|X$ is pseudo-Anosov. This follows from Thurston’s construction [47, Theorem 7]. By Lemma 2.5 the hole $X$ has infinite diameter.

If $X$ is incompressible, then $X \subset \partial_h T$, where $\rho_F : T \to F$ is the given $I$–bundle. Let $f : F \to F$ be the given pseudo-Anosov map. Now $g$, the suspension of $f$, gives a homeomorphism of $M$. Again it follows that the hole $X$ has infinite diameter. □

Applying Lemma 5.11 and Corollary 9.2 we find the following.

Proposition 9.3. If $D(M,S)$ admits a hole $X \subseteq S$ with diameter at least 57, then the inclusion $\nu : D(M,S) \to \mathcal{C}(S)$ is not a quasi-isometric embedding. □

10. Holes for the Disk Complex – Annuli

Suppose $(M,S)$ is spotless. This section is devoted to proving the following.

Theorem 10.1. Suppose $X$, an annulus, is a hole for $D(M,S)$. Then the diameter of $X$ is at most 11.

The proof shares many features with the proofs of Theorems 11.10 and 12.1. However, the exceptional definition of $\mathcal{C}(S_{0,2})$ prevents a unified approach. Fix $(M,S)$ a spotless pair. When $M$ is a solid torus, then $D(M,S)$ is at most a point and there is nothing to prove. Henceforth we assume that $M$ is not a solid torus. We begin with the following.

Claim. For all $D \in D(M,S)$, we have $|D \cap X| \geq 2$.

Proof. Since $X$ is a hole, every disk cuts $X$. Let $\alpha$ be a core curve for $X$. If $|\alpha \cap D| = 1$, then let $N = N(\alpha \cup D)$. Since $M$ is not a solid torus the disk $E = \fr(N)$ is essential. Also $E$ is disjoint from $\alpha$. Thus $E$ misses $X$, a contradiction. □

Assume, to obtain a contradiction, that $X$ has diameter at least 12. Suppose $D \in D(M,S)$ is a disk chosen to minimize $D \cap X$. Pick any disk $E \in D(M,S)$ so that $d_X(D,E) \geq 6$. Isotope $D$ and $E$ to make the boundaries tight and also tight with respect to $\partial X$. Tightening triples of curves is not canonical; nonetheless there is a tightening so that $S - (\partial D \cup \partial E \cup X)$ contains no triangles. See Figure 10.2.

After tightening $\partial D$, $\partial E$, and $\partial X$ in this way we have the following.

Claim. Suppose $\delta \subset X \cap \partial D$ and $\epsilon \subset X \cap \partial E$ are any connected components (and hence arcs). Then $|\delta \cap \epsilon| \geq 1$. 

Figure 10.2. Triangles outside of $X$ (left side) can be moved into $X$ (right side). This decreases the number of points of $D \cap E \cap (S - X)$.

**Proof.** Let $S^X$ be the annular cover of $S$ corresponding to $X$. Let $X' \subset S^X$ be the homeomorphic lift of $X$ to $S^X$. Define $\partial D|X = \kappa_X(\partial D)$ and define $\partial E|X$ similarly.

Let $\delta' \subset X' \cap (\partial D|X)$ be the homeomorphic lift of $\delta$ to $X'$. Define $\epsilon'$ similarly. Since $|\delta \cap \epsilon| = |\delta' \cap \epsilon'|$ it suffices to bound the latter from below. Note that $\delta'$ is properly embedded in $X'$ but not in $S^X$. To cure this, define $\delta^* \subset \partial D|X$ to be the properly embedded arc in $S^X$ that contains $\delta'$. Define $\epsilon^*$ similarly.

Since $d_X(D,E) = \text{diam}_X(\partial D|X \cup \partial E|X) \geq 6$ we find that $d_X(\delta^*, \epsilon^*) \geq 4$. It follows from (2.4) that $|\delta^* \cap \epsilon^*| \geq 3$.

Suppose $x, y \in \delta^* \cap \epsilon^*$ are consecutive along $\delta^*$. Note that $x$ and $y$ are contained in the preimage of $X$ but are, possibly, not contained in $X'$. See Figure 10.3.

However, if both $x$ and $y$ lie in the same component of $S^X - X'$, then either $\delta^*$ or $\epsilon^*$ shares a bigon with some lift of $\partial X$, a contradiction. Again, see Figure 10.3.

This implies that

$$|\delta' \cap \epsilon'| \geq |\delta^* \cap \epsilon^*| - 2$$

and so $|\delta' \cap \epsilon'| \geq 1$, as desired. \qed

Figure 10.3. Here is a sketch of $S^X$. The central shaded region is $X'$. The upper and lower shaded regions are other lifts of $X$ to $S^X$. These are not annuli but rather are homeomorphic to $\mathbb{R} \times I$.  

Claim. There is an outermost bigon \((B, \alpha, \beta)\) of \(E - D\) with the following properties:
- \(B \subset E\), \(B \cap D = \alpha\), and \(B \cap \partial E = \beta\);
- \(\partial \beta \subset X\);
- \(|\beta \cap X| = 2\).

It also follows that \(|D \cap X| = 2\).

Proof. Note that \(D \cap E\) is a collection of arcs and curves in \(E\). Since \(M\) is irreducible, any simple closed curve component of \(D \cap E\) can be removed by an ambient isotopy (fixing \(\partial M\) pointwise) applied to \(D\). Minimality of \(|D \cap E|\) now implies that there are no simple closed curves in \(D \cap E\).

Consider any outermost bigon \(B\) of \(E - D\), with \(\alpha\) and \(\beta\) as in the first bullet. Since \(D \cap E\) is minimal, the bigon \(B\) is a boundary compression for \(D\). Note that \(\beta\) cannot completely contain a component of \(E \cap X\) as every arc of \(E \cap X\) meets some arc of \(D \cap X\). Using this observation, Figure 10.4 shows the four possible ways \(\alpha\) may lie inside of \(E\).

![Figure 10.4](image)

Figure 10.4. The figure shows a portion of \(E\). The darker part of \(\partial E\) are the arcs of \(E \cap X\). The four arcs drawn in the interior of \(E\) are the four possibilities for the arc \(\alpha\). Note that \(\alpha\) cuts a bigon \((B, \alpha, \beta)\) off of \(E\). Thus either \(\beta\) is disjoint from \(X\), or \(\beta\) is contained in \(X\), or \(\beta\) meets \(X\) in a single subarc, or \(\beta\) meets \(X\) in two subarcs.

Note that after compression \(D_B\) is a union of two essential disks, \(D', D'' \in D(M, S)\). Suppose \(\alpha\) is one of the three unlabelled arcs depicted in Figure 10.4. It follows that either \(D'\) or \(D''\) has, after tightening, a smaller intersection with \(X\) than \(D\) does, a contradiction. We deduce that \(\alpha\) is as pictured by the labelled arc in Figure 10.4.

As \(D', D''\) cannot have a smaller intersection with \(X\) we deduce that \(|D \cap X| = 2\), proving the claim.

Using the same notation as in the proof above, let \(B\) be an outermost bigon of \(E - D\). We now study how \(\alpha \subset \partial B\) lies inside of \(D\).

Claim. The arc \(\alpha \subset D\) connects distinct components of \(D \cap X\).

Proof. Suppose not. Then there is a bigon \((C, \alpha, \gamma)\) with \(C \subset D\) and \(\gamma \subset D \cap X\). The disk \(C \cup B\) is isotopic to \(D'\) or \(D''\) and so is essential. Also, \(C \cup B\) intersects \(X\) at most once after tightening, contradicting our first claim.

We finish the proof of Theorem 10.1 by noting that \(D \cup B\) is homeomorphic to \(\Upsilon \times I\), where \(\Upsilon\) is the simplicial tree with three edges and three leaves. We may choose the homeomorphism so that \((D \cup B) \cap X = \Upsilon \times \partial I\). Since \(X\) is an annulus
there is an ambient isotopy of $D \cup B$ making $(D \cup B) \cap X$ just a pair of arcs. Recall that $D_B = D' \cup D''$. It follows that one of $D'$ or $D''$ (or both) meets $X$ in at most a single arc, contradicting our first claim. \hfill \Box

11. Holes for the disk complex – compressible

The proof of Theorem 11.10 occupies the second half of this section.

11.1. Compression sequences of essential disks. Suppose $(M, S)$ is a spotless pair. Suppose $X$ is a cleanly embedded subsurface of $S$. Suppose $D \in \mathcal{D}(M, S)$. Choose representatives so that $\partial D$ is tight with respect to $\partial X$. Suppose $D \cap \partial X \neq \emptyset$.

**Definition 11.2.** A compression sequence for the data $M, S, X, D$ is a sequence $\{\Delta_k\}_{k=1}^n$ where $\Delta_1 = \{D\}$ and where $\Delta_{k+1}$ obtained by boundary compressing $\Delta_k$ into $S - n(\partial X)$ and then tightening. Note that $\Delta_k$ is a disjoint union of exactly $k$ essential disks. We further require that every disk of every $\Delta_k$ cuts $\partial X$. A compression sequence is maximal if either
- no disk of $\Delta_n$ can be boundary compressed into $S - n(\partial X)$ or
- there is an essential disk $(E, \partial E) \subset (M, S - n(\partial X))$ disjoint from $\Delta_n$.

Such maximal sequences end essentially or end in $S - n(\partial X)$, respectively.

**Lemma 11.3.** A maximal compression sequence exists for any data $M, S, X, D$. Furthermore, some component of $S - n(\partial X)$ is compressible if and only if some (hence all) compression sequence ends in $S - n(\partial X)$.

**Proof.** All compression sequences must end, by Remark 8.5. If $Y$ is a compressible component of $S - n(\partial X)$, then, by Lemma 8.6 all compression sequences end in $S - n(\partial X)$. The backwards direction is immediate. \hfill \Box

In what follows we assume that $X$ is not an annulus or a pair of pants. Our next goal is to show that maximal sequences do not move very far in the arc and curve complex of $X$.

**Definition 11.4.** Fix $D_k \in \Delta_k$. A disjointness pair in $X$ for $D_k$ is an ordered pair $(\alpha, \beta)$ of essential arcs in $X$ where
- $\alpha \subset D_k \cap X$.
- $\beta \subset \Delta_n \cap X$, and
- $d_A(\alpha, \beta) \leq 1$.

Here $A = A(X)$.

If $\alpha \neq \alpha'$, then the two disjointness pairs $(\alpha, \beta)$ and $(\alpha', \beta)$ are distinct, even if $\alpha$ is ambiently isotopic to $\alpha'$ in $X$. We treat the second coordinate similarly. The following lemma controls how subsurface projection distance changes in maximal sequences.

**Lemma 11.5.** Suppose $D \subset \mathcal{D}(M, S)$ is an essential disk. Suppose that $X$ is a cleanly embedded subsurface of $S$. Suppose $X$ is not an annulus or pants. Suppose $D$ cuts $\partial X$. Choose a maximal sequence $\{\Delta_k\}_{k=1}^n$ for the data $M, S, X, D$. For any disk $D_k \in \Delta_k$ either
- $D_k \in \Delta_n$ or
- there are four distinct disjointness pairs $\{(\alpha_i, \beta_i)\}_{i=1}^4$ for $D_k$ in $X$, where each of the arcs $\alpha_i$ appears as the first coordinate of at most two pairs.
Proof. We induct on \( n - k \). If \( D_k \) is contained in \( \Delta_n \) there is nothing to prove. If \( D_k \) is contained in \( \Delta_{k+1} \) we are done by induction. Thus we may assume that \( D_k \) is the disk in \( \Delta_k \) which is boundary compressed at stage \( k \). Let \( D_{k+1}, D'_{k+1} \in \Delta_{k+1} \) be the two disks obtained by boundary compressing \( D_k \) along the bigon \( B \). By induction, each of \( D_{k+1} \) and \( D'_{k+1} \) either lie in \( \Delta_n \) or have disjointness pairs with the required properties. See Figure 11.6 for a picture of the pair of pants \( P \subset S \) cobounded by \( \partial D_k \) and \( \partial D_{k+1} \cup \partial D'_{k+1} \).

Choose \( \delta \) (shown as a dotted arc in Figure 11.6) to be a band sum arc for \( D_{k+1} \cup D'_{k+1} \), dual to \( B \), that minimizes \( |\delta \cap \partial X| \). Since \( B \) is a boundary compression in the complement of \( \partial X \) it follows that the band sum of \( D_{k+1} \) and \( D'_{k+1} \), along \( \delta \), is tight without any isotopy.

There are now three possibilities: neither, one, or both points of \( \partial \delta \) are contained in \( X \).

First suppose that \( X \cap \partial \delta = \emptyset \). Then every arc of \( D_{k+1} \cap X \) is parallel to an arc of \( D_k \cap X \), and similarly for \( D'_{k+1} \). If \( D_{k+1} \) and \( D'_{k+1} \) are both components of \( \Delta_n \), then choose any arcs \( \beta, \beta' \) of \( D_{k+1} \cap X \) and of \( D'_{k+1} \cap X \). Let \( \alpha, \alpha' \) be the parallel components of \( D_k \cap X \). The four disjointness pairs are then \( (\alpha, \beta), (\alpha, \beta'), (\alpha', \beta), (\alpha', \beta') \). Suppose instead \( D_{k+1} \) is not a component of \( \Delta_n \). Then \( D_k \) inherits four disjointness pairs from \( D_{k+1} \).

Second suppose that exactly one endpoint of \( \partial \delta \) meets \( X \). Breaking symmetry, suppose \( \gamma \subset D_{k+1} \) is the component of \( D_{k+1} \cap X \) meeting \( \delta \). Let \( X' \) be the component of \( X \cap P \) that contains \( \delta \). Let \( \alpha, \alpha' \) be the two components of \( D_k \cap X' \). Let \( \beta \) be any arc of \( D'_{k+1} \cap X \).

If \( D_{k+1} \notin \Delta_n \), and if \( \gamma \) is not the first coordinate of one of the four disjointness pairs for \( D_{k+1} \), then \( D_k \) inherits disjointness pairs from \( D_{k+1} \). If \( D'_{k+1} \notin \Delta_n \), then \( D_k \) inherits disjointness pairs from \( D'_{k+1} \).

Thus we may assume that both \( D_{k+1} \) and \( D'_{k+1} \) are in \( \Delta_n \) or that only \( D'_{k+1} \in \Delta_n \) while \( \gamma \) appears as the first arc of a disjointness pair for \( D_{k+1} \). In the former case the required disjointness pairs are \( (\alpha, \beta), (\alpha', \beta), (\alpha, \gamma) \) and \( (\alpha', \gamma) \). In the latter case we do not know if \( \gamma \) is allowed to appear as the second coordinate of a pair. However, we are given four disjointness pairs for \( D_{k+1} \) and are told that \( \gamma \) appears as the first coordinate of at most two of these pairs. Hence the other two pairs are inherited by \( D_k \). The pairs \( (\alpha, \beta) \) and \( (\alpha', \beta) \) give the desired conclusion.
Third suppose that the endpoints of \( \delta \) meet \( \gamma \subset D_{k+1} \) and \( \gamma' \subset D'_{k+1} \). Let \( X' \) be a component of \( X \cap P \) containing \( \gamma \). Let \( \alpha \) and \( \alpha' \) be the two arcs of \( D_k \cap X' \). Suppose both \( D_{k+1} \) and \( D'_{k+1} \) lie in \( \Delta_n \). Then the desired pairs are \((\alpha, \gamma), (\alpha', \gamma), (\alpha, \gamma') \) and \((\alpha', \gamma')\). If \( D'_{k+1} \in \Delta_n \) while \( D_{k+1} \) is not, then \( D_k \) inherits two pairs from \( D_{k+1} \). We add to these the pairs \((\alpha, \gamma') \) and \((\alpha', \gamma')\). If neither disk lies in \( \Delta_n \), then \( D_k \) inherits two pairs from each disk and the proof is complete. \( \square \)

Given a disk \( D \in \mathcal{D}(M,S) \) and a hole \( X \subset S \) our Lemma \[11.5\] adapts \( D \) to \( X \).

**Lemma 11.7.** Fix a hole \( X \) for \( \mathcal{D}(M,S) \) that is not an annulus or pants. For any disk \( D \in \mathcal{D}(M,S) \) there is a disk \( D' \in \mathcal{D}(M,S) \) with the following properties:

- \( \partial X \) and \( \partial D' \) are tight.
- If \( X \) is incompressible, then \( D' \) is not boundary compressible into \( S-n(\partial X) \) and \( d_A(D,D') \leq 3 \).
- If \( X \) is compressible, then \( \partial D' \subset X \) and \( d_{AC}(D,D') \leq 3 \).
- Thus \( d_X(D,D') \leq 6 \).

Here \( A = A(X) \) and \( AC = AC(X) \).

**Proof.** If \( \partial D \subset X \), then take \( D' = D \) and we are done. So we may assume (Remark 9.1) that \( D \) cuts \( \partial X \). By Lemma \[11.3\] there is a maximal compression sequence \( \{\Delta_k\}_{k=1}^n \) for the data \( M, S, X, D \).

Suppose \( n > 1 \). Lemma \[11.5\] implies that the disk \( D = D_1 \) has a disjointness pair. Thus \( d_A(D,\Delta_n) \leq 3 \). If \( X \) is incompressible, then we may take \( D' \) to be any component of \( \Delta_n \). If \( X \) is compressible, then by Lemma \[11.3\] there is a disk \( E \) compressing \( X \) and disjoint from \( \Delta_n \). It follows that \( d_{AC}(D,E) \leq 3 \). Taking \( D' = E \) proves the lemma.

If \( n = 1 \), then the proof proceeds as in the previous paragraph, without the need for disjointness pairs.

In all cases \( d_{AC}(D,D') \leq 3 \). It follows from Corollary \[11.6\] that \( d_X(D,D') \leq 6 \). \( \square \)

**Remark 11.8.** Lemma \[11.7\] is unexpected: after all, any two curves in \( \mathcal{C}(X) \) can be connected by a sequence of band sums. Thus arbitrary band sums can change the subsurface projection to \( X \). However, the sequences of band sums arising in Lemma \[11.7\] are very special. Firstly they do not cross \( \partial X \), and secondly they are “tree-like” due to the fact that every arc in \( D \) is separating.

When \( D \) is replaced by a surface with genus, then Lemma \[11.7\] does not hold in general; this is closely related to a fundamental observation of Kobayashi [27, Theorem 1] (see also [21, Theorem 1.2]). Namazi points out that even if \( D \) is only replaced by a planar surface, Lemma \[11.7\] does not hold in general.

11.9. **Classification of compressible holes.** We now prove the theorem.

**Theorem 11.10.** Suppose \( X \) is a compressible hole for \( \mathcal{D}(M,S) \) with diameter at least 15. Then there are essential disks \( D, E \in \mathcal{D}(M,S) \) so that

- \( \partial D, \partial E \subset X \) and
- \( \partial D \) and \( \partial E \) fill \( X \).

**Proof.** By Theorem \[10.7\] the subsurface \( X \) is not an annulus. Since \( \mathcal{C}(X) \) is non-empty, \( X \) is not a pants.
Choose disks $D'$ and $E'$ in $\mathcal{D}(M,S)$ so that $d_X(D',E') \geq 15$. By Lemma 11.7 there are disks $D,E$ with $\partial D, \partial E \subset X$ so that $d_X(D',D)$ and $d_X(E',E)$ are at most six. It follows from the triangle inequality that $d_X(D,E) \geq 3$. □

12. Holes for the disk complex – incompressible

This section classifies incompressible holes for the disk complex.

**Theorem 12.1.** Suppose $X$ is an incompressible hole for $\mathcal{D}(M,S)$ with diameter at least 57. Then there is an $I$–bundle $\rho_F : T \rightarrow F$ embedded in $M$ so that

- $\partial hT \subset S$,
- $X$ is a component of $\partial hT$,
- some component of $\partial vT$ is boundary parallel into $S$, and
- $F$ supports a pseudo-Anosov map.

Here is a short plan of the proof. We are given $X$, an incompressible hole for $\mathcal{D}(M,S)$. Following Lemma 11.7 we may assume that $D,E$ are essential disks, without boundary compressions into $S - n(\partial X)$, so that $d_X(D,E) \geq 45$. We examine the intersection pattern of $D$ and $E$ to find two families of rectangles $R$ and $Q$. The intersection pattern of these rectangles in $M$ will determine the desired $I$–bundle $T$. The third conclusion of the theorem follows from an outermost bigon argument. The fourth conclusion requires another application of Lemma 11.7 as well as Lemma 2.5.

12.2. Diagonals of polygons. To understand the intersection pattern of $D$ and $E$ we discuss diagonals of polygons. Let $D$ be a $2n$–sided regular polygon. Label the sides of $D$ with the letters $X$ and $Y$ in alternating fashion. Any side labeled $X$ (or $Y$) will be called an $X$ side (or a $Y$ side).

**Definition 12.3.** An arc $\gamma$ properly embedded in $D$ is a diagonal if the points of $\partial \gamma$ lie in the interiors of distinct sides of $D$. If $\gamma$ and $\gamma'$ are diagonals for $D$ that together meet three or four distinct sides, then $\gamma$ and $\gamma'$ are non-parallel.

**Lemma 12.4.** Suppose $\Gamma \subset D$ is a disjoint union of non-parallel diagonals. Then there is an $X$ side of $D$ meeting at most eight diagonals of $\Gamma$.

**Proof.** A counting argument shows that $|\Gamma| \leq 4n - 3$. If every $X$ side meets at least nine non-parallel diagonals, then $|\Gamma| \geq \frac{9}{2}n > 4n - 3$, a contradiction. □

12.5. Improving disks. Suppose now that $X$ is an incompressible hole for $\mathcal{D}(M,S)$ with diameter at least 57. By Theorem 10.1 the subsurface $X$ is not an annulus. Since $C(X)$ is non-empty, $X$ is not a pants. Let $Y = S - n(X)$.

Choose disks $D'$ and $E'$ in $\mathcal{D}(M,S)$ so that $d_X(D',E') \geq 57$. By Lemma 11.7 there are disks $D,E \in \mathcal{D}(M,S)$ that cannot be boundary compressed into $X$ or into $Y$ so that $d_X(D',D), d_X(E',E) \leq 6$. By the triangle inequality $d_X(D,E) \geq 57 - 12 = 45$.

Recall that $\partial D$ and $\partial E$ are tight with respect to $\partial X$. We may further assume that $\partial D$ and $\partial E$ are tight with respect to each other. Also, minimize the quantities $|X \cap (\partial D \cap \partial E)|$ and $|D \cap E|$ while keeping everything tight. In particular, $X - (\partial D \cup \partial E)$ has no triangle components. Now consider $D$ and $E$ as even-sided polygons, with vertices being the points $\partial D \cap \partial X$ and $\partial E \cap \partial X$ respectively. Let $\Gamma = D \cap E$.

**Claim.** $\Gamma \subset D$ is a disjoint union of diagonals.
Proof. The minimality of $|D \cap E|$ and the irreducibility of $M$ imply that $\Gamma$ contains no simple closed curves. Suppose $\gamma \subset \Gamma$ is a non-diagonal. Then there is an outermost such arc in $D$, say $\gamma' \subset \Gamma$, cutting a bigon $B$ out of $D$. Tightness implies that $B$ is a boundary compression of $E$ into $S - n(\partial X)$. But this contradicts the construction of $E$. Thus all arcs of $\Gamma$ are diagonals for $D$ and, by the same argument, for $E$. □

One possibility for $\Gamma \subset D$ is shown in Figure 12.6. By Lemma 12.4 there is a component $\alpha \subset D \cap X$ meeting at most eight distinct types of diagonal of $\Gamma$. Choose $\beta \subset E \cap X$ similarly. As $d_X(D,E) \geq 45$, applying Lemma 4.5 proves that $d_X(\alpha, \beta) \geq 41 - 4 = 37$.

Break each of $\alpha$ and $\beta$ into at most eight subarcs $\{\alpha_i\}$ and $\{\beta_j\}$ so that each subarc meets all of the diagonals of fixed type and only of that type. Let $R_i \subset D$ be the rectangle with upper boundary $\alpha_i$ and containing all of the diagonals meeting $\alpha_i$. Let $\alpha'_i$ be the lower boundary of $R_i$. Define $Q_j \subset E$ and $\beta'_j$ similarly. See Figure 12.6 for a picture of $R_i$.

![Figure 12.6](image)

**Figure 12.6.** The rectangle $R_i \subset D$ is surrounded by the dotted line. The arc $\alpha_i$ in $\partial D \cap X$ is indicated. In general the arc $\alpha'_i$ may lie in $X$ or in $Y$.

An arc $\alpha_i$ is large if there is an arc $\beta_j$ so that $|\alpha_i \cap \beta_j| \geq 3$. Note that $|\alpha_i \cap \beta_j| = |\alpha'_i \cap \beta'_j|$, so $\alpha'_i$ is large if and only if $\alpha_i$ is large. We use the same notation for $\beta_j$. Let $\Theta$ be the union of all of the large $\alpha_i$ and $\beta_j$. Thus $\Theta$ is a graph in $X$ with all vertices of valence one or four. Let $\Theta'$ be the union of the large $\alpha'_i$ and $\beta'_j$.

**Claim 12.7.** The graph $\Theta$ is non-empty.

*Proof.* If $\Theta = \emptyset$, then all $\alpha_i$ are small. Thus $|\alpha \cap \beta| \leq 128$ and so $|\sigma_X(\alpha) \cap \sigma_X(\beta)| \leq 512$. Lemma 2.3 implies that $d_X(\alpha, \beta) \leq 20$. As $d_X(\alpha, \beta) \geq 41$ this is a contradiction. □

Let $Z \subset S$ be a small regular neighborhood of $\Theta$ and define $Z'$ similarly.

**Claim 12.8.** No component of $\Theta$ or of $\Theta'$ is contained in a disk $C \subset S$. No component of $\Theta$ or of $\Theta'$ is contained in an annulus $A \subset S$ peripheral in $X$.

*Proof.* For a contradiction suppose that $W$ is a component of $Z$ contained in a disk $C \subset S$. Then there is some pair $\alpha_i, \beta_j$ cutting a bigon out of $S$. This contradicts the tightness of $\partial D$ and $\partial E$. The same holds for $Z'$.

Suppose now that some component $W$ is contained in an annulus $A$, peripheral in $X$. Thus $W$ fills $A$. Suppose $\alpha_i$ and $\beta_j$ are large and contained in $W$. By the
classification of arcs in \( A \) we deduce that either \( \alpha_i \) and \( \beta_j \) form a bigon in \( A \) or the triple \( \partial X, \alpha_i, \) and \( \beta_j \) form a triangle. Either conclusion gives a contradiction. \( \square \)

Claim 12.9. If \( W \subset Z \) is a component and \( \delta \subset \partial W \) is a component, then either \( \delta \) is inessential or peripheral in \( X \).

Proof. Suppose some \( \delta \in \partial W \) is essential and non-peripheral. Any large \( \alpha_i \) meets \( \partial W \) in at most two points, while any small \( \alpha_i \) meets \( \partial W \) in at most 32 points. Thus \(|\sigma_X(\alpha) \cap \delta| \leq 512\) and the same holds for \( \beta \). Thus \( d_X(\alpha, \beta) \leq 40 \) by the triangle inequality. As \( d_X(\alpha, \beta) \geq 41 \), this is a contradiction. \( \square \)

From Claim 12.9 we deduce the following.

Claim 12.10. The graphs \( \Theta, \Theta' \) are each connected. Also, \( \Theta \) fills \( X \). \( \square \)

There are now two possibilities: \( \Theta \) and \( \Theta' \) either intersect or are disjoint. In the first case set \( \Sigma = \Theta \cup \Theta' \) and in the second case set \( \Sigma = \Theta \). By the claims above, \( \Sigma \) is connected and fills \( X \). Let \( \mathcal{R} = \{ R_i \} \) and \( \mathcal{Q} = \{ Q_j \} \) be the collections of large rectangles.

12.11. Building the \( I \)–bundle. Section [12.5] gives us \( \Sigma, \mathcal{R}, \) and \( \mathcal{Q} \). Note that \( \mathcal{R} \cup \mathcal{Q} \) is an \( I \)–bundle and \( \Sigma \) is the component of its horizontal boundary meeting \( X \). See Figure 12.12 for an example.

![Figure 12.12. \( \mathcal{R} \cup \mathcal{Q} \) is an \( I \)–bundle: all arcs of intersection are parallel.](https://www.ams.org/journal-terms-of-use)

Let \( T_0 \) be a regular neighborhood of \( \mathcal{R} \cup \mathcal{Q} \), taken in \( M \). Thus \( T_0 \) has the structure of an \( I \)–bundle. Note that \( \partial_h T_0 \subset S \), \( \partial_h T_0 \cap X \) is a component of \( \partial_h T_0 \), and this component fills \( X \) due to Claim 12.10. We will enlarge \( T_0 \) to obtain the desired \( I \)–bundle \( T \subset M \).

Begin by enumerating all annuli \( \{ A_i \} \subset \partial_h T_0 \) with the property that some component of \( \partial A_i \) is inessential in \( S \). Suppose we have built the \( I \)–bundle \( T_i \) and are now considering the annulus \( A = A_i \). Let \( \gamma \cup \gamma' = \partial A \subset S \) with \( \gamma \) inessential in \( S \). Let \( B \subset S \) be the disk which \( \gamma \) bounds. So \( B \) is contained in \( X \) or in \( Y = S - n(X) \). By induction, no component of \( \partial_h T_i \) is contained in a disk embedded in \( S \); the base case holds by Claim 12.8. It follows that \( B \cap T_i = \partial B = \gamma \). Thus \( B \cup A \) is isotopic, fixing \( \gamma' \) pointwise, to a properly embedded disk \( B' \subset M \). As \( \gamma' \) lies in \( X \) or \( Y \), both incompressible, \( \gamma' \) must bound a disk \( C \subset S \). Again, \( C \) lies in \( X \) or in \( Y \). Note \( C \cap T_i = \partial C = \gamma' \), using the induction hypothesis. So \( B \cap C = \emptyset \).
It follows that $B \cup A \cup C$ is an embedded two-sphere in $M$. Since $M$ is irreducible $B \cup A \cup C$ bounds a three-ball $U_i$ in $M$. Choose a homeomorphism $U_i \cong B \times I$ so that $B$ is identified with $B \times \{0\}$, $C$ is identified with $B \times \{1\}$, and $A$ is identified with $\partial B \times I$. We form $T_{i+1} = T_i \cup U_i$ and note that $T_{i+1}$ still has the structure of an $I$–bundle. Recalling that $A = A_i$ we have $\partial_v T_{i+1} = \partial_v T_i - A_i$. Also $\partial_h T_{i+1} = \partial_h T_i \cup (B \cup C) \subset S$. Thus no component of $\partial_h T_{i+1}$ is contained in a disk embedded in $S$. Similarly, $\partial_h T_{i+1} \cap X$ is a component of $\partial_h T_{i+1}$; this component is connected and fills $X$.

After dealing with all of the annuli $\{A_i\}$ in this fashion we are left with an $I$–bundle $T$. Now all components of $\partial \partial T$ are essential in $S$. All of these lying in $X$ are peripheral in $X$. This is because they are disjoint from $\Sigma \subset \partial_h T$, which fills $X$. It follows that the component of $\partial_h T$ containing $\Sigma$ is isotopic to $X$.

This finishes the construction of the promised $I$–bundle $T$ and demonstrates the first two conclusions of Theorem 12.1. For future use we record the following.

**Remark 12.13.** Every curve of $\partial \partial_v T = \partial \partial_h T$ is essential in $S$.

12.14. **A vertical annulus parallel into the boundary.** Here we obtain the third conclusion of Theorem 12.1. We say that a hole $X$ for $D(M, S)$ is small if $\text{diam}_X(D(M, S)) < 61$ and large otherwise. Suppose $X$ is a large incompressible hole and $T$ is the $I$–bundle constructed in the previous section, having $X$ as a component of $\partial_v T$. We show that at least one component of $\partial_v T$ is boundary parallel into $S$.

**Claim 12.15.** All components of $\partial_v T$ are incompressible in $M$.

**Proof.** Suppose $A \subset \partial_v T$ is a compressible annulus component. By Remark 12.13 we may compress $A$ to obtain a pair of essential disks $B$ and $C$. Note that $\partial B$ is isotopic into the complement of $\partial_h T$. Thus $S - n(\partial_h T)$ compresses. So $S - n(X)$ compresses and $X$ is not a hole, a contradiction. \qed

**Claim 12.16.** Some component of $\partial_v T$ is boundary parallel.

**Proof.** Let $D \in D(M, S)$ be an essential disk that cannot be boundary compressed into $S - n(\partial X)$. Apply an ambient isotopy to $D$ that minimizes $|D \cap \partial_v T|$. Let $\Gamma = D \cap \partial_v T$. Since $X$ is an incompressible hole, $\Gamma$ is non-empty. By Claim 12.15 and because $M$ is irreducible, $\Gamma$ has no simple closed curves. Since $D$ cannot be boundary compressed into $S - n(\partial X)$, all arcs of $\Gamma$ are essential in $\partial_v T$. Let $\gamma$ be any outermost arc of $\Gamma$ in $D$. So $\gamma$ cuts a bigon $B$ off of $D$. Let $\delta$ be the closure of $\partial B - \gamma$. Let $A$ be the component of $\partial_v T$ containing $\gamma$. So $(B, \gamma, \delta) \subset (M, A, S)$. Note that the interior of $\delta$ is contained in $S - \partial_h T$. This is because $\partial_v T$ cannot be boundary compressed into $T$.

Let $C = A_B$ be the disk that results from boundary compressing the annulus $A$ along the bigon $B$. Note that $C$ is properly embedded in $(M, S)$, with $\partial C$ disjoint from $\partial_h T$. Since $(M, S)$ is spotless, $C$ is not peripheral. It follows that $C$ is inessential. Thus $C$ cuts a closed three-ball $U$ off of $M$. Since $C$ is disjoint from $T$, from Remark 12.13 we deduce that $T \cap U = \emptyset$. It follows that $A$ is boundary parallel into $S$, as desired. \qed

**Remark 12.17.** The proof of Claim 12.16 implies that the multicurve

$$\{ \partial A \mid A \subset \partial_v T \text{ is a boundary parallel into } S \}$$

is disk-busting for $(M, S)$.
12.18. **Finding a pseudo-Anosov map.** Here we obtain the fourth conclusion of Theorem 12.21: the base surface $F$ of the $I$–bundle $T$ admits a pseudo-Anosov map.

As in Section 12.25 pick essential disks $D'$ and $E'$ in $(M, S)$ so that $d_X(D', E') \geq 57$. Lemma 11.7 provides disks $D$ and $E$ which cannot be boundary compressed into $S - n(\partial X)$. Thus $D$ and $E$ cannot be boundary compressed into $\partial hT$. Also, $d_X(D, E) \geq 57 - 12 = 45$.

After isotoping $D$ to minimize intersection with $\partial hT$ it must be the case that all components of $D \cap \partial_v T$ are essential arcs in $\partial_v T$. By Lemma 8.12 there is an ambient isotopy of $D$ making $D \cap T$ vertical in $T$. The same holds for $E$. Choose $A$ and $B$, components of $D \cap T$ and of $E \cap T$. These are vertical rectangles. As usual, we use Theorem 10.1 to rule out the possibility that $X$ is an annulus. By Lemma 4.5 we have diam$_X(\pi_X(D)) \leq 2$ and so $d_X(A, B) \geq 45 - 4 = 41$.

We now begin to work in the base surface $F$; by the above, $F$ is not an annulus. Recall $\rho_F: T \to F$ is the bundle map. Take $\alpha = \rho_F(A)$ and $\beta = \rho_F(B)$. The natural map $\mathcal{C}(F) \to \mathcal{C}(X)$, defined by taking a curve to its lift, is distance non-increasing (see (6.3)). Thus $d_F(\alpha, \beta) \geq 41$. Thus, by Lemma 2.5 the subsurface $F$ supports a pseudo-Anosov map and we are done.

12.19. **Corollaries.** We now deal with the possibility of disjoint holes for the disk complex.

**Lemma 12.20.** Suppose $X$ is a large incompressible hole for $\mathcal{D}(M, S)$ supported by the $I$–bundle $\rho_F: T \to F$. Let $Y = \partial hT - X$. Let $\tau: \partial hT \to \partial hT$ be the involution switching the ends of the interval fibers. Suppose $D \in \mathcal{D}(M, S)$ is an essential disk.

- If $F$ is orientable, then $d_{\mathcal{A}(F)}(D \cap X, D \cap Y) \leq 6$.
- If $F$ is non-orientable, then $d_X(D, \mathcal{C}(X)) \leq 3$.

**Proof.** We repeat the proof of Lemma 11.7 with $\partial X$ everywhere replaced by $\partial \partial hT$. So there is a disk $D' \subset M$ which is tight with respect to $\partial \partial hT$ and which cannot be boundary compressed into $\partial hT$ (or into $S - \partial hT$). For any component $Z \subset \partial hT$ we have $d_{\mathcal{A}(Z)}(D, D') \leq 3$. By Lemma 8.12 an ambient isotopy (preserving $T$ setwise) makes $D' \cap T$ vertical in $T$ and we are done. \hfill \Box

Recall Lemma 7.3 all holes for the arc complex intersect. This cannot hold for the disk complex. For example let $F$ be an orientable surface with boundary and let $\rho_F: T \to F$ be the product $I$–bundle. So $M = T$ is a handlebody. Notice that both components of $\partial hT$ are holes for $\mathcal{D}(M)$. However, by the first conclusion of Lemma 12.20 $X$ and $Y$ are paired holes, in the sense of Definition 5.6. So, as with the invariant arc complex (Lemma 7.5), all holes for the disk complex interfere.

**Lemma 12.21.** Suppose $X, Z \subset S$ are large holes for $\mathcal{D}(M, S)$. If $X \cap Z = \emptyset$, then there is an $I$–bundle $T \cong F \times I$ in $M$ so that $\partial hT = X \cup Y$ and $Y \cap Z \neq \emptyset$.

**Proof.** Suppose $X \cap Z = \emptyset$. It follows from Remark 9.1 that both $X$ and $Z$ are incompressible. Let $\rho_F: T \to F$ be the $I$–bundle in $M$ with $X \subset \partial hT$, as provided by Theorem 12.1. We also have a component $A \subset \partial_v T$ so that $A$ is boundary parallel into $S$. Let $U$ be the closure of the solid torus component of $M - A$. Note that $Z$ cannot be contained in $S \cap \partial U$ because $Z$ is not an annulus (Theorem 10.1).

Let $\alpha = \rho_F(A)$. Choose any essential arc $\delta \subset F$ with both endpoints in $\alpha \subset \partial F$. It follows that $\rho_F^{-1}(\delta)$, together with two meridional disks of $U$, forms an essential disk $D$ in $(M, S)$. Let $W$ be the closure of $\partial hT \cup (\partial U - A)$. Note that $\partial D \subset W$. \hfill \Box
If $F$ is non-orientable, then $Z \cap W = \emptyset$ and we have a contradiction. Deduce that $F$ is orientable. Now, if $Z$ misses $Y$, then $Z$ misses $W$ and we again have a contradiction. It follows that $Z$ cuts $Y$ and we are done. $\Box$

13. Axioms for combinatorial complexes

The goal of this section and the next is to prove, inductively, the upper bound on distance in a combinatorial complex $G(S)$. This section presents our axioms for $G$: sufficient hypotheses for Theorem 13.1. The axioms, apart from Axiom 13.3, are quite general. Axiom 13.3 is necessary to prove hyperbolicity and greatly simplifies the recursive construction given in Section 14.

**Theorem 13.1.** Fix $S$, a compact, connected, non-simple surface. Suppose $G(S)$ is a combinatorial complex satisfying the axioms of Section 13. For any constants $c, x \geq 0$ there is a constant $A = A(c, x)$ with the following property. Suppose $X$ is a hole for $G$ with $\xi(X) = x$. Suppose $\alpha_X, \beta_X \in G$ are contained in $X$. Then

$$d_G(\alpha_X, \beta_X) \leq A \sum [d_Y(\alpha_X, \beta_X)]^c$$

where the sum is taken over all holes $Y \subseteq X$ for $G$.

The proof of this upper bound is more difficult than the proof of the lower bound, Theorem 5.14. This is because naturally occurring paths in $G$ between $\alpha_X$ and $\beta_X$ may waste time in non-holes. As a first example, consider the path in $C(S)$ obtained by taking short curves along a Teichmüller geodesic. The Teichmüller geodesic may spend time rearranging the geometry of a subsurface. Thus the path of systoles in the curve complex might be much longer than the curve complex distance between the endpoints.

In Sections 16, 17, 19 we will verify these axioms for the curve complex of a non-orientable surface, the arc complex, and the disk complex.

13.2. The axioms. Suppose $G(S)$ is a combinatorial complex.

**Axiom 13.3** (Holes interfere). All large holes for $G$ interfere.

As discussed in Lemma 5.12 this axiom is necessary to show that $G$ is Gromov hyperbolic. It also greatly simplifies the inductive proof of Theorem 13.1. The remaining axioms provide constants so that for any pair of vertices $\alpha_X, \beta_X \in G$, both contained in a hole $X$ for $G$, there is

- a marking path $\Lambda = \{\mu_n\}_{n=0}^N$,
- an accessibility interval $J_Y \subset [0, N]$ for every essential subsurface $Y \subset X$,
- a combinatorial sequence $\Gamma = \{\gamma_i\}_{i=0}^K \subset G$ that starts with $\alpha_X$, ends with $\beta_X$ and has each $\gamma_i$ contained in $X$, and
- an increasing reindexing function $r : [0, K] \rightarrow [0, N]$ with $r(0) = 0$ and $r(K) = N$ with various properties. Here are the first four axioms.

**Axiom 13.4** (Marking path).

1. The support of $\mu_{n+1}$ is contained inside the support of $\mu_n$.
2. For any subsurface $Y \subset X$, if $\pi_Y(\mu_k) \neq \emptyset$, then for all $n \leq k$ the map $n \mapsto \pi_Y(\mu_n)$ is an unparameterized quasi-geodesic with constants depending only on $G$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
The second condition is crucial and often technically difficult to obtain.

**Axiom 13.5** (Accessibility). The accessibility interval for $X$ is $J_X = [0, N]$. There is a constant $B_1$ so that the following hold.

1. If $m \in J_Y$, then $Y$ is contained in the support of $\mu_m$.
2. If $m \in J_Y$, then $i(\partial Y, \mu_m) < B_1$.
3. If $[m, n] \cap J_Y = \emptyset$, then $d_Y(\mu_m, \mu_n) < B_1$.

**Axiom 13.6** (Combinatorial). The vertex $\gamma_i$ is contained in the support of $\mu_{r(i)}$. Further, there is a constant $B_2$ so that

1. $d_Y(\gamma_i, \mu_{r(i)}) < B_2$, for every $i \in [0, K]$ and every hole $Y \subset X$, and
2. $d_G(\gamma_i, \gamma_i+1) < B_2$, for every $i \in [0, K - 1]$.

**Axiom 13.7** (Replacement). There is a constant $C_3$ with the following property. If $Y \subset X$ is a large hole for $\mathcal{G}$ and if $r(i) \in J_Y$, then there is a vertex $\gamma' \in \mathcal{G}$ so that

1. $\gamma'$ is contained in $Y$ and
2. $d_G(\gamma_i, \gamma') < C_3$.

There are two axioms left, dealing with straight and shortcut intervals. These are given in the next subsection.

### 13.8 Inductive, Electric, Shortcut, and Straight Intervals

We describe subintervals that arise when partitioning $[0, K]$, the combinatorial interval. Let $x = \xi(X)$. As discussed in Section 13.13, we choose a general upper threshold $L_2$ and, for all $y \leq x$, a lower threshold $L_1(y)$.

**Definition 13.9.** Suppose $[i, j] \subset [0, K]$ is a subinterval of the combinatorial sequence. Then $[i, j]$ is an *inductive interval* if there is a hole $Y \subset X$ so that

1. $r([i, j]) \subset J_Y$ (if $Y$ is paired, then $r([i, j]) \subset J_Y \cap J_Y'$) and
2. $d_Y(\gamma_i, \gamma_j) \geq L_1(y)$, where $y = \xi(Y)$.

When $X$ is the only relevant hole there is a simpler definition.

**Definition 13.10.** Suppose $[i, j] \subset [0, K]$ is a subinterval of the combinatorial sequence. Then $[i, j]$ is an *electric interval* if $d_Y(\gamma_i, \gamma_j) < L_2$ for all holes $Y \subset X$.

Electric intervals will be partitioned into straight and shortcut intervals.

**Definition 13.11.** Suppose $[i, j] \subset [0, K]$ is an electric interval. Then $[p, q] \subset [i, j]$ is a *straight interval* if $d_Z(\mu_{r(p)}, \mu_{r(q)}) < L_2$ for all non-holes $Z \subset X$.

**Definition 13.12.** Suppose $[i, j] \subset [0, K]$ is an electric interval. Then $[p, q] \subset [i, j]$ is a *shortcut* if there is a non-hole $Z \subset X$ so that

1. $r([p, q]) \subset J_Z$ and
2. $d_Z(\mu_{r(p)}, \mu_{r(q)}) \geq L_1(x)$, where $x = \xi(X)$.

**Axiom 13.13** (Straight). There is a constant $A = A(x)$ so that for every straight interval $[p, q]$ we have

$$d_G(\gamma_p, \gamma_q) < A d_X(\gamma_p, \gamma_q).$$

**Axiom 13.14** (Shortcut). There is a constant $C_4 = C_4(x)$ so that for every shortcut $[p, q]$ we have

$$d_G(\gamma_p, \gamma_q) < C_4.$$

The verification of the shortcut axiom often requires some of the details of the classification of holes for $\mathcal{G}$. 
13.15. **Deductions from the axioms.** Axiom \[13.4\] and Lemma \[3.12\] imply that the reverse triangle inequality holds for projections of marking paths.

**Lemma 13.16.** There is a constant \(C_5\) so that
\[
d_Y(\mu_m, \mu_n) + d_Y(\mu_n, \mu_p) < d_Y(\mu_m, \mu_p) + C_5
\]
for every essential \(Y \subset X\) and for every \(m<n<p \in [0, N]\). \(\square\)

We make \(C_5\) larger, if necessary, to arrange \(C_5 \geq \text{diam}_S \pi_S(\mu \cup \nu)\) for any adjacent markings \(\mu, \nu \in \mathcal{M}(S)\) in the marking graph. We record a consequence of Axiom \[13.5\]

**Lemma 13.17.** There is a constant \(C_1 > B_1\) with the following property. If \(\partial Y\) cuts \(Z\) and if \(m,n \in J_Y\), then \(d_Z(\mu_m, \partial Y) < C_1\) and also \(d_Z(\mu_m, \mu_n) < C_1\).

**Proof.** Part (1) of Axiom \[13.5\] says that \(Y\) is contained in the support of \(\mu_m\). Thus \(\mu_m\) cuts \(Z\). The same is true of \(\mu_n\). Part (2) of Axiom \[13.5\] says that \(v(\mu_m, \partial Y) \leq B_1\). It follows that \(\pi_Z(\mu_m)\) and \(\pi_Z(\partial Y)\) have bounded intersection. Lemma \[2.3\] gives a bound for \(d_Z(\mu_m, \partial Y)\). The triangle inequality implies that \(d_Z(\mu_m, \mu_n)\) is also bounded. \(\square\)

Part (2) of Axiom \[13.6\] and Lemma \[5.9\] imply the following.

**Lemma 13.18.** There is a constant \(C_2 > B_2\) with the following property. For any hole \(Y\) and for any \(i \in [0, K-1]\), we have \(d_Y(\gamma_i, \gamma_{i+1}) < C_2\). \(\square\)

We now have all of the constants \(C_1, C_2, C_3, C_4, C_5\) in hand. Recall that \(L_3\) is the pairing constant of Definition \[5.6\] and that \(M_0\) is the constant from the bounded geodesic image theorem \[1.7\]. We choose a lower threshold \(L_1(y)\) for all \(y \leq \xi(X)\). We choose the general upper threshold, \(L_2\) and general lower threshold \(L_0\). For all \(z < y \leq x\) we require the following:

\[
\begin{align*}
(13.19) & \quad L_0 > C_1 + 2C_2 + 2L_3, \\
(13.20) & \quad L_1(y) > M_0 + 2C_1 + 5C_2 + 2L_3 + L_0 + 2, \\
(13.21) & \quad L_1(x) > L_1(z) + 2C_1 + 4C_2 + 4L_3, \\
(13.22) & \quad L_2 > L_1(x) + 2L_3 + 6C_5 + 2C_2 + 14C_1 + 11.
\end{align*}
\]

14. **Partition and the upper bound on distance**

In this section we prove Theorem \[13.1\] by induction on \(x = \xi(X)\). The first stage of the proof is to describe the inductive partition: we partition the given interval \([0, K]\) into inductive and electric intervals. The inductive partition is closely linked with the hierarchy machine \[31\] Section 4 and with the notion of antichains introduced in \[12\] Section 5.

We next give the shortcut partition; each electric interval is divided into straight and shortcut intervals. We finally bound \(d_S(\alpha_X, \beta_X)\) from above by combining the contributions from the various intervals.

14.1. **Inductive partition.** We begin by identifying the relevant surfaces. Pick a hole \(X\) for \(G\); pick vertices \(\alpha_X, \beta_X \in G\) contained in \(X\). Define
\[
B_X = \{Y \subset X \mid Y\text{ is a hole and }d_Y(\alpha_X, \beta_X) \geq L_1(x)\}.
\]
The axioms give a combinatorial sequence $\Gamma = \{\gamma_i\}_0^K$. For any subinterval $[i, j] \subset [0, K]$ define

$$B_X(i, j) = \{Y \in B_X \mid d_Y(\gamma_i, \gamma_j) \geq L_1(x)\}.$$  

We now partition $[0, K]$ into inductive and electric intervals. Begin with the partition of one part $P_X = \{[0, K]\}$. Recursively $P_X$ is a partition of $[0, K]$ consisting of subintervals which are either inductive, electric, or undetermined. Suppose $[i, j] \in P_X$ is undetermined.

**Claim.** If $B_X(i, j)$ is empty, then $[i, j]$ is electric.

**Proof.** Since $B_X(i, j)$ is empty, every hole $Y \not\subset X$ has either $d_Y(\gamma_i, \gamma_j) < L_1(x)$ or $Y \not\in B_X$. In the former case, as $L_1(x) < L_2$, we are done.

So suppose the latter holds. By the reverse triangle inequality (Lemma 13.10)

$$d_Y(\mu_{r(i)}, \mu_{r(j)}) < d_Y(\mu_0, \mu_N) + 2C_5.$$  

Since $r(0) = 0$ and $r(K) = N$ we find

$$d_Y(\gamma_i, \gamma_j) < d_Y(\alpha_X, \beta_X) + 2C_5 + 4C_2.$$  

Thus

$$d_Y(\gamma_i, \gamma_j) < L_1(x) + 2C_5 + 4C_2 < L_2.$$  

This proves the claim. $\Box$

Thus if $B_X(i, j)$ is empty, then $[i, j] \in P_X$ is determined to be electric. Proceed on to the next undetermined subinterval. Suppose instead that $B_X(i, j)$ is non-empty. Pick a hole $Y \in B_X(i, j)$ so that $Y$ has maximal complexity $y = \xi(Y)$ amongst the elements of $B_X(i, j)$.

Let $p, q \in [i, j]$ be the first and last indices, respectively, so that $r(p), r(q) \in J_Y$. (If $Y$ is paired with $Y'$, then we take the first and last indices that, after reindexing, lie inside of $J_Y \cap J_{Y'}$.)

**Claim.** The indices $p, q$ are well-defined.

**Proof.** Since $Y \in B_X(i, j)$, we have $d_Y(\gamma_i, \gamma_j) \geq L_1(x)$. Let $m, n = r(i), r(j)$. Suppose $J_Y \cap [m, n] = \emptyset$. By part (3) of Axiom 13.5 we have $d_Y(\mu_m, \mu_n) < B_1$. Part (1) of Axiom 13.6 implies that

$$d_Y(\gamma_i, \gamma_j) < B_1 + 2B_2 < C_1 + 2C_2.$$  

This is less than $L_1(x)$ by Equation 13.20 giving a contradiction. Thus $m < \min J_Y$ and $\max J_Y < n$.

Suppose $J_Y \cap r([i, j])$ is empty. So let $h$ be the last index with $r(h) < \min J_Y$. Thus $\max J_Y < r(h+1)$. We have

$$d_Y(\mu_m, \mu_{r(h)}) < B_1 \quad \text{and} \quad d_Y(\mu_{r(h+1)}, \mu_n) < B_1.$$  

By Lemma 13.18 we have $d_Y(\gamma_h, \gamma_{h+1}) < C_2$. Applying part (1) of Axiom 13.6 repeatedly, we find

$$d_Y(\gamma_i, \gamma_j) < C_2 + 4B_2 + 2B_1 < L_1(X),$$  

with the last inequality deduced from Equation 13.20. This is a contradiction. Thus, if $Y$ is not paired, the indices $p, q$ are well-defined.

Suppose $Y$ is paired with $Y'$. Recall that measurements made in $Y$ and $Y'$ differ by at most the pairing constant $L_3$ given in Definition 5.6. Thus we may deduce, as in the previous two paragraphs, $J_{Y'} \cap r([i, j])$ is non-empty.
Suppose now, for a contradiction, that \( J_Y \cap J_{Y'} \cap r([i,j]) \) is empty. Define
\[
h = \max \{ \ell \in [i,j] \mid r(\ell) \in J_Y \}
\]
and
\[
k = \min \{ \ell \in [i,j] \mid r(\ell) \in J_{Y'} \}.
\]

Without loss of generality we may assume that \( h < k \). It follows that \( d_{Y'}(\gamma_i, \gamma_h) < C_1 + 2C_2 \). Thus \( d_Y(\gamma_i, \gamma_h) < C_1 + 2C_2 + 2L_3 \). Also, \( d_Y(\gamma_{h+1}, \gamma_j) < C_1 + 2C_2 \).

Deduce
\[
d_Y(\gamma_i, \gamma_j) < 2C_1 + 4C_2 + 2L_3 + 2 < L_1(x)
\]
with the last inequality following from (13.20). This contradicts the assumption that \( Y \in B_X(i,j) \) and we are done. \( \square \)

Claim. The interval \([p,q]\) is inductive for \( Y \).

Proof. We must check that \( d_Y(\gamma_p, \gamma_q) \geq L_1(y) \). Suppose first that \( Y \) is not paired. Then by the definition of \( p, q \), by property (3) of Axiom 13.3 and by the triangle inequality we have
\[
d_Y(\mu_r(i), \mu_r(j)) \leq d_Y(\mu_r(p), \mu_r(q)) + 2C_1.
\]
Thus by Axiom 13.6
\[
d_Y(\gamma_i, \gamma_j) \leq d_Y(\gamma_p, \gamma_q) + 2C_1 + 4C_2.
\]
Since by (13.20),
\[
L_1(y) + 2C_1 + 4C_2 < L_1(x) \leq d_Y(\gamma_i, \gamma_j),
\]
we are done.

When \( Y \) is paired the proof is similar but we must use the slightly stronger inequality \( L_1(y) + 2C_1 + 4C_2 + 4L_3 < L_1(x) \). \( \square \)

When \( B_X(i,j) \) is non-empty, these two claims give a hole \( Y \) and indices \( p, q \). We subdivide the element \([i,j] \in P_X\) into the elements \([i,p-1]\), \([p,q]\), and \([q+1,j]\). (The first or third interval, or both, may be empty.) The interval \([p,q] \in P_X\) is determined to be inductive and associated to \( Y \). Now proceed to the next undetermined element. This completes the construction of \( P_X \).

As a bit of notation, if \([i,j] \in P_X\) is associated to \( Y \subset X \) we will sometimes write \( I_Y = [i,j] \). Note that \( I_Y \) is a subinterval of the combinatorial sequence while \( J_Y \) is a subinterval of the marking path. Note that \( r(I_Y) \subset J_Y \).


**Lemma 14.3.** Suppose \( Y \) and \( Z \), both contained in \( X \), are holes for \( \mathcal{G} \). Suppose \( I_Z \) is an inductive element of \( P_X \) associated to \( Z \). Suppose \( r(I_Z) \subset J_Y \) (or \( r(I_Z) \subset J_Y \cap J_{Y'} \) if \( Y \) is paired). Then
\[ \begin{itemize} 
  \item \( Z \) is nested in \( Y \) or 
  \item \( Z \) and \( Z' \) are paired and \( Z' \) is nested in \( Y \).
\end{itemize} \]

Proof. Let \( I_Z = [i,j] \) and \( z = \xi(Z) \). Suppose first that \( \partial Y \) cuts \( Z \). By Lemma 13.17
\[
d_Z(\mu_r(i), \mu_r(j)) < C_1. \text{ Then by Axiom 13.6 } d_Z(\gamma_i, \gamma_j) < C_1 + 2C_2 < L_1(z),
\]
a contradiction.

Now, if \( Z \) and \( Y \) are disjoint, then by Axiom 13.3 and Definition 5.6 there are two cases. Suppose \( Y \) is paired with \( Y' \); thus \( Y' \) and \( Z \) meet. In this case we are
done, just as in the previous paragraph. Suppose instead that $Z$ is paired with $Z'$; thus $Z'$ and $Y$ meet. If $Z'$ is nested in $Y$, then we are done. If $\partial Y$ cuts $Z'$, then, as $r([i,j]) \subset J_Y$, again Axiom 13.6 and Lemma 13.17 imply that
\[ d_{Z'}(\gamma_i, \gamma_j) < C_1 + 2C_2. \]
So $d_Z(\gamma_i, \gamma_j) < C_1 + 2C_2 + 2L_3 < L_1(z)$, a contradiction.

**Remark 14.5.** It follows that for any hole $Y$ there are at most three inductive intervals in the partition $\mathcal{P}_X$, where $Y$ has projection distance greater than $L_0$.

**Proof of Proposition 14.4.** We prove the second claim. Suppose $I_Y = [p, q]$ and $I_{Y'} = [p', q']$ with $q < p'$. It follows that $[r(p), r(q')] \subset J_Y \cap J_{Y'}$. If $q + 1 = p'$, then the partition would have chosen a larger inductive interval for one of $Y$ or $Y'$. It must be the case that there is an inductive interval $I_Z \subset [q + 1, p' - 1]$ for some hole $Z$, distinct from $Y$ and $Y'$, with $\xi(Z) \geq \xi(Y)$. However, by Lemma 14.3, we find that $Z$ is nested in $Y$ or in $Y'$. It follows that $Z = Y$ or $Y'$, a contradiction.

The first statement has a similar proof.

We prove the third claim. Suppose $Z$ and $W$ are the first and last holes, if any, satisfying the hypotheses. Since $d_Y(\gamma_h, \gamma_i) \geq L_0$ we find by Axiom 13.6 that
\[ d_Y(\mu_{r(h)}, \mu_{r(i)}) \geq L_0 - 2C_2. \]
By 13.19, $L_0 - 2C_2 > C_1$ so that
\[ J_Y \cap r(I_Z) \neq \emptyset. \]
If $Y$ is paired, then, again by 13.19, we have $L_0 > C_1 + 2C_2 + 2L_3$; we also find that $J_{Y'} \cap r(I_Z) \neq \emptyset$. Symmetrically, $J_{Y'} \cap r(I_W)$ (and $J_Y \cap r(I_W)$) are also non-empty.

It follows that the interval $[i + 1, j - 1]$ between $I_Z$ and $I_W$, after applying the reindexing map, is contained in $J_Y$ (and $J_{Y'}$, if $Y$ is paired). Thus for any inductive interval $I_V = [p, q]$ between $I_Z$ and $I_W$ the associated hole $V$ is nested in $Y$ (or $V'$ is nested in $Y$), by Lemma 14.3. If $V = Y$ or $V = Y'$ there is nothing to prove. Suppose instead that $V$ (or $V'$) is strictly nested in $Y$. It follows that
\[ d_Y(\gamma_p, \gamma_q) < C_1 + 2C_2 < L_0. \]
Thus there are no inductive intervals between $I_Z$ and $I_W$ satisfying the hypotheses of the third claim.

The following lemma and proposition bound the number of inductive intervals. The discussion here is very similar to the discussion of antichains in Section 5. Our situation is complicated by the presence of non-holes and interfering holes. Suppose that $X, \alpha_X, \beta_X$ are given, as in the beginning of Section 14.1. Let $x = \xi(X)$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Lemma 14.6. Suppose $\ell \geq (3 \cdot L_2)^x$. Suppose $\{Y_i\}_{i=1}^{\ell}$ is a set of distinct holes that are strict subsurfaces of $X$, each having $d_{Y_i}(\alpha_X, \beta_X) \geq L_1(x)$. Then there is a hole $Z \subseteq X$ such that $d_Z(\alpha_X, \beta_X) \geq L_2 - 1$ and $Z$ contains at least $L_2$ of the $Y_i$.

Furthermore, for at least $L_2 - 4(C_5 + 3C_1 + 2)$ of these $Y_i$ we find that $J_{Y_i} \subseteq J_Z$. (If $Z$ is paired, then $J_{Y_i} \subseteq J_Z \cap J_{Z'}$.) Each of these $Y_i$ is disjoint from a distinct vertex $\eta_i \in [\pi_Z(\alpha_X), \pi_Z(\beta_X)]$.

Proof. Let $g_X$ be a geodesic in $C(X)$ joining $\alpha_X, \beta_X$. By the bounded geodesic image theorem (4.7), since $L_1(x) > M_0$, for every $Y_i$ there is a vertex $\omega_i \in g_X$ such that $Y_i \subset X - \omega_i$. Thus $d_X(\omega_i, \partial Y_i) \leq 1$. If there are at least $L_2$ distinct $\omega_i$, associated to distinct $Y_i$, then $d_X(\alpha_X, \beta_X) \geq L_2 - 1$. In this situation we take $Z = X$. Since $J_X = [0, N]$ we are done.

Thus assume there do not exist at least $L_2$ distinct $\omega_i$. Then there is some fixed $\omega$ among these $\omega_i$ such that at least $\frac{L_2}{L_2} \geq 3(3 \cdot L_2)^{x-1}$ of the $Y_i$ satisfy

$$Y_i \subset (X - \omega).$$

Thus one component, call it $W$, of $X - \omega$ contains at least $(3 \cdot L_2)^{x-1}$ of the $Y_i$. Let $w = \xi(W)$. Set $g_W = [\alpha_W, \beta_W]$ for $\alpha_W \in \pi_W(\beta_X)$ and $\beta_W \in \pi_W(\beta_X)$. Now, Lemma [4.5] implies that

$$d_{Y_i}(\alpha_W, \beta_W) \geq d_{Y_i}(\alpha_X, \beta_X) - 8$$

because we are projecting to nested subsurfaces. Hence $d_{Y_i}(\alpha_W, \beta_W) \geq L_1(w)$.

Again apply Theorem [4.7]. Since $L_1(w) > M_0$, for every remaining $Y_i$ there is a vertex $\eta_i \in g_W$ such that

$$Y_i \subset (W - \eta_i).$$

If there are at least $L_2$ distinct $\eta_i$, then we take $Z = W$. Otherwise we repeat the argument. Since the complexity of each successive subsurface decreases by at least 1, we must eventually find the desired $Z$ containing at least $L_2$ of the $Y_i$, each disjoint from distinct vertices of $g_Z$.

So, suppose that there are at least $L_2$ distinct $\eta_i$ associated to distinct $Y_i$ and we have taken $Z = W$. Now we must find at least $L_2 - 4(C_5 + 3C_1 + 2)$ of these $Y_i$ where $J_{Y_i} \subseteq J_Z$.

To this end we focus attention on a small subset $\{Y^j\}_{j=1}^{5} \subset \{Y_i\}$. Let $\eta_j$ be the vertex of $g_Z = g_W$ associated to $Y^j$. We choose these $Y^j$ so that

- the $\eta_j$ are arranged along $g_Z$ in order of index and
- $d_Z(\eta_j, \eta_{j+1}) > C_5 + 3C_1 + 2$, for $j = 1, 2, 3, 4$.

This is possible by [13.22] because

$$L_2 > 4(C_5 + 3C_1 + 2).$$

Set $J_j = J_{Y^j}$ and pick any indices $m_j \in J_j$. (If $Z$ is paired, then $Y^j$ is also paired; in this case pick $m_j \in J_{Y^j} \cap J_{Y^j'})$ We use $\mu(m_j)$ to denote $\mu_{m_j}$. Since $\partial Y^j$ is disjoint from $\eta_j$, Lemma [13.17] implies that

$$d_Z(\mu(m_j), \eta_j) \leq C_1 + 1.$$  

(14.7)

Since the sequence $\pi_Z(\mu_n)$ satisfies the reverse triangle inequality (Lemma [13.16]), it follows that the $m_j$ appear in $[0, N]$ in order agreeing with their index. The triangle inequality implies that

$$d_Z(\mu(m_1), \mu(m_2)) > C_1.$$
Thus Axiom 13.5 implies that $J_Z \cap [m_1, m_2]$ is non-empty. Similarly, $J_Z \cap [m_4, m_5]$ is non-empty. It follows that $[m_2, m_4] \subset J_Z$. (If $Z$ is paired, then, after applying the symmetry $\tau$ to $g_Z$, the same argument proves that $[m_2, m_4] \subset J_{Z'}$.)

Notice that $J_2 \cap J_3 = \emptyset$. For if $m \in J_2 \cap J_3$, then by (14.7) both $d_Z(\mu, \eta_2)$ and $d_Z(\mu_m, \eta_3)$ are bounded by $C_1 + 1$. It follows that
\[
d_Z(\eta_2, \eta_3) < 2C_1 + 2,
\]
a contradiction. Similarly $J_3 \cap J_4 = \emptyset$. We deduce that $J_3 \subset [m_2, m_4] \subset J_Z$. (If $Z$ is paired, then $J_3 \subset J_Z \cap J_{Z'}$.) Finally, there are at least
\[
L_2 - 4(C_5 + 3C_1 + 2)
\]
possible $Y_i$’s which satisfy the hypothesis on $Y^3$. This completes the proof. \(\square\)

Now define
\[
\mathcal{P}_{\text{ind}} = \{I \in \mathcal{P}_X \mid I \text{ is inductive}\}.
\]

**Proposition 14.8.** Let $x = \xi(X)$. We have
\[
|\mathcal{P}_{\text{ind}}| \leq A d_X(\alpha_X, \beta_X),
\]
where $A = 2(3 \cdot L_2)^x - 1$.

**Proof.** Suppose, for a contradiction, that the conclusion fails. Let $g_X = [\alpha_X, \beta_X]$ be a geodesic in $C(X)$. Then, as in the proof of Lemma 14.6 there is a vertex $\omega$ of $g_X$ and a component $W \subset X - \omega$ where at least $(3 \cdot L_2)^x - 1$ of the inductive intervals in $I_X$ have associated surfaces, $Y_i$, contained in $W$.

Since $x - 1 \geq w = \xi(W)$ we may apply Lemma 14.6 inside of $W$. So we find a surface $Z \subset W \subset X$ so that
- $Z$ contains at least $L_2$ of the $Y_i$,
- $d_Z(\alpha_X, \beta_X) \geq L_2 - 1$, and
- there are at least $L_2 - 4(C_5 + C_1 + 2C_1 + 2)$ of the $Y_i$ where $J_{Y_i} \subset J_Z$.

Since $Y_i \subset Z$ and $Y_i$ is a hole, $Z$ is also a hole. Since $L_2 > L_1(x) - 1$ it follows that $Z \in B_X$. Let $Y = \{Y_i\}$ be the set of $Y_i$ satisfying the third bullet. Let $Y^1 \in Y$ and $\eta_1 \in g_Z$ satisfy $\partial Y^1 \cap \eta_1 = \emptyset$ and $\eta_1$ is the first such. Choose $Y^2 \in Y$ and $\eta_2 \in g_Z$ similarly, so that $\eta_2$ is the last such. By Lemma 14.6
\[
(14.9) \quad d_Z(\eta_1, \eta_2) \geq L_2 - 4(C_5 + C_1 + 2C_1 + 2) - 1.
\]

Let $p = \min I_{Y^1}$ and $q = \max I_{Y^2}$. Note that $r([p, q]) \subset J_Z$. (If $Z$ is paired with $Z'$, then $r([p, q]) \subset J_Z \cap J_{Z'}$.) Again by Lemma 13.17
\[
d_Z(\mu_r(p), \partial Y^1) < C_1.
\]

It follows that
\[
d_Z(\mu_r(p), \eta_1) \leq C_1 + 1
\]
and the same bound applies to $d_Z(\mu_r(q), \eta_2)$. Combined with (14.9) we find that
\[
d_Z(\mu_r(p), \mu_r(q)) \geq L_2 - 4C_5 - 4C_1 - 10C_1 - 11.
\]

By the reverse triangle inequality (Lemma 13.16), for any $p' \leq p, q \leq q'$,
\[
d_Z(\mu_r(p'), \mu_r(q')) \geq L_2 - 6C_5 - 4C_1 - 10C_1 - 11.
\]

Finally by Axiom 13.6 and the above inequality, we have
\[
d_Z(\gamma_{p'}, \gamma_{q'}) \geq L_2 - 6C_5 - 4C_1 - 10C_1 - 11 - 2C_2.
\]
By (13.22) the right-hand side is greater than \( L_1(x) + 2L_3 \), so we deduce that \( Z \in B_X(p', q') \), for any such \( p', q' \). (When \( Z \) is paired, deduce also that \( Z' \in B_X(p', q') \).

Let \( I_V \) be the first inductive interval chosen by the procedure with the property that \( I_V \cap [p, q] \neq \emptyset \). Since \( I_{Y_1} \) and \( I_{Y_2} \) were also chosen, deduce that \( I_V \subset [p, q] \). Let \( p', q' \) be the indices so that \( V \) is chosen from \( B_X(p', q') \). Thus \( p' \leq p \) and \( q \leq q' \). However, since \( I_V \subset [p, q] \) and since \( r([p, q]) \) \( \subset J_Z \), Lemma 14.3 implies that \( V \) is strictly nested in \( Z \). (When pairing occurs we may find instead that \( V \subset Z' \) or \( V' \subset Z \).) Thus \( \xi(Z) > \xi(V) \) and we find that \( Z \) would be chosen from \( B_X(p', q') \), instead of \( V \). This is a contradiction. \( \square \)

14.10. Shortcut partition. The goal of this subsection is to prove the following.

**Proposition 14.11.** Let \( x = \xi(X) \). There is a constant \( A = A(x) \) with the following property. If \([i, j] \subset [0, K] \) is an electric interval, then

\[
d_G(\gamma_i, \gamma_j) \leq A \, d_X(\gamma_i, \gamma_j).
\]

We begin by building a partition of the given electric interval \([i, j]\) into straight and shortcut intervals. Define

\[
C_X = \{Z \subseteq X \mid Z \text{ is a non-hole and } d_Z(\mu_{r(i)}, \mu_{r(j)}) \geq L_1(x)\}.
\]

We also define, for all \([p, q] \subset [i, j]\),

\[
C_X(p, q) = \{Z \in C_X \mid J_Z \cap [r(p), r(q)] \neq \emptyset\}.
\]

Our recursion starts with the partition of one part, \( \mathcal{P}(i, j) = \{[i, j]\} \). Recursively \( \mathcal{P}(i, j) \) is a partition of \([i, j]\) into shortcut, straight, or undetermined intervals. Suppose \([p, q] \in \mathcal{P}(i, j) \) is undetermined.

**Claim.** If \( C_X(p, q) \) is empty, then \([p, q]\) is straight.

**Proof.** We show the contrapositive. Suppose \( Z \) is a non-hole with \( d_Z(\mu_{r(p)}, \mu_{r(q)}) \geq L_2 \). The reverse triangle inequality (Lemma 13.16) gives

\[
d_Z(\mu_{r(p)}, \mu_{r(q)}) < d_Z(\mu_{r(i)}, \mu_{r(j)}) + 2C_5.
\]

Since \( L_2 > L_1(x) + 2C_5 \), we find that \( Z \in C_X \). Since \( L_2 > C_1 > B_1 \), Axiom 13.5 implies that \( J_Y \cap [r(p), r(q)] \) is non-empty. Thus \( Z \in C_X(p, q) \). \( \square \)

So when \( C_X(p, q) \) is empty the interval \([p, q]\) is determined to be straight. Proceed onto the next undetermined element of \( \mathcal{P}(i, j) \). Now suppose that \( C_X(p, q) \) is non-empty. Then we choose any \( Z \in C_X(p, q) \) so that \( Z \) has maximal \( \xi(Z) \) amongst the elements of \( C_X(p, q) \).

There are two cases. Suppose \( J_Z \cap r([p, q]) \) is empty. Let \( s \in [p, q] \) be the largest integer so that \( r(s) < \min J_Z \). Remove \([p, q]\) from the partition \( \mathcal{P}(i, j) \) and add the three intervals

\[
[p, s], [s + 1/2], [s + 1, q]
\]

to \( \mathcal{P}(i, j) \). Here \([s + 1/2]\) is an interval of length zero: we call this a shortcut of length zero for \( Z \). The intervals \([p, s]\) and \([s + 1, q]\) are undetermined.

Suppose \( J_Z \cap r([p, q]) \) is non-empty. Define \( s, t \in [p, q] \) to be the largest and smallest indices in \([p, q]\) so that \( r(s), r(t) \in J_Z \). (We permit \( s = t \).) Thus \( r([s, t]) \subset J_Z \). Since \( Z \in C_X(p, q) \) it follows that \( Z \in C_X \) and so \( d_Z(\mu_{r(i)}, \mu_{r(j)}) \geq L_1(x) \). Thus \([s, t]\) is a shortcut interval for the non-hole \( Z \). So remove \([p, q]\) from the partition \( \mathcal{P}(i, j) \) and add the three intervals

\[
[p, s - 1], [s, t], [t + 1, q]
\]
to \( P(i, j) \). The intervals \([p, s - 1]\) and \([t + 1, q]\) are undetermined. This completes the recursive construction of the shortcut partition.

Define
\[
P_{\text{str}} = \{ I \in P(i, j) \mid I \text{ is straight} \}
\]
and
\[
P_{\text{short}} = \{ I \in P(i, j) \mid I \text{ is a shortcut} \}.
\]

**Proposition 14.12.** Let \( x = \xi(X) \). We have
\[
|P_{\text{short}}| \leq_A d_X(\gamma_i, \gamma_j),
\]
where \( A = 2(3 \cdot L_2)^{x-1} + 1 \).

**Proof.** The proof is identical to that of Proposition 14.8 with the caveat that in Lemma 14.6 we must use the markings \( \mu_r(i) \) and \( \mu_r(j) \) instead of the endpoints \( \gamma_i \) and \( \gamma_j \). \( \square \)

We are now equipped to give the proof of Proposition 14.11

**Proof.** Suppose \( P(i, j) \) is the given partition of the electric interval \([i, j]\) into straight and shortcut subintervals. As a bit of notation, if \([p, q] = I \in P(i, j)\), we take \( d_G(I) = d_G(\gamma_p, \gamma_q) \) and \( d_X(I) = d_X(\gamma_p, \gamma_q) \). We have
\[
(14.13) \quad d_G(\gamma_i, \gamma_j) \leq \sum_{I \in P_{\text{str}}} d_G(I) + \sum_{I \in P_{\text{short}}} d_G(I) + C_2|P(i, j)|.
\]

The last term arises from connecting left endpoints of intervals with the right endpoint of the following interval, applying Axiom 13.6 and recalling that \( B_2 < C_2 \). We now bound the three terms on the right.

We begin with the third; recall that \(|P(i, j)| = |P_{\text{short}}| + |P_{\text{str}}|\), that \(|P_{\text{str}}| \leq |P_{\text{short}}| + 1\), and that \(|P_{\text{short}}| \leq_A d_X(\gamma_i, \gamma_j)\). The second inequality follows from the construction of the partition while the last is Proposition 14.12. Thus the third term of (14.13) is quasi-bounded above by \( d_X(\gamma_i, \gamma_j) \). By Axiom 13.13 the second term of (14.13) is at most \( C_4|P_{\text{short}}|\).

By Axiom 13.13 for all \( I \in P_{\text{str}} \) we have \( d_G(I) \leq_A d_X(I) \). It follows from the reverse triangle inequality (Lemma 13.16) that
\[
\sum_{I \in P_{\text{str}}} d_X(I) \leq d_X(\gamma_i, \gamma_j) + (2C_5 + 2C_2)|P_{\text{str}}| + 2C_2.
\]

We deduce that \( \sum_{I \in P_{\text{str}}} d_G(I) \) is also quasi-bounded above by \( d_X(\gamma_i, \gamma_j) \). Thus for a somewhat larger value of \( A \) we find
\[
d_G(\gamma_i, \gamma_j) \leq_A d_X(\gamma_i, \gamma_j).
\]
This completes the proof. \( \square \)

14.14. **The upper bound.** We will need the following.

**Proposition 14.15.** Let \( x = \xi(X) \). For any \( c \geq 0 \) there is a constant \( A = A(c, x) \) with the following property. Suppose \([i, j] = I_Y \) is an inductive interval in \( P_X \). Then
\[
d_G(\gamma_i, \gamma_j) \leq_A \sum_Z [d_Z(\gamma_i, \gamma_j)]c,
\]
where \( Z \) ranges over all holes for \( G \) strictly contained in \( X \).
Proof. Let \( y = \xi(Y) \) and note \( y < x \). Axiom \([13.7]\) gives vertices \( \gamma'_i, \gamma'_j \in G \), contained in \( Y \), so that \( d_G(\gamma_i, \gamma'_i) \leq C_3 \) and the same holds for \( j \). Since projection to holes is coarsely Lipschitz (Lemma \([5.9]\)) for any hole \( Z \) we have \( d_Z(\gamma_i, \gamma'_i) \leq 2 + 2C_3 \).

Fix any \( c > 0 \). Now, since \( d_G(\gamma_i, \gamma_j) \leq d_G(\gamma'_i, \gamma'_j) + 2C_3 \), to find the required constant \( A(c, x) \) it suffices to bound \( d_G(\gamma'_i, \gamma'_j) \). Let \( c' = c + 4C_3 + 4 \). Since \( y < x \), we may apply Theorem \([13.1]\) inductively to obtain a constant \( A = A(c', y) \) with

\[
 d_G(\gamma'_i, \gamma'_j) \leq A \sum_Z \left[ d_Z(\gamma'_i, \gamma'_j) \right] c'
\]

Here \( N \) is the number of non-zero terms in the final sum. Also, the sum ranges over holes \( Z \subset Y \). We may take \( A \) somewhat larger to deal with the term \( (4C_3 + 4)N \) and include all holes \( Z \) strictly contained in \( X \) to find

\[
 d_G(\gamma_i, \gamma_j) \leq A \sum_Z \left[ d_Z(\gamma_i, \gamma_j) \right] c
\]

where the sum is over all holes \( Z \) strictly contained in \( X \). \( \square \)

14.16. Finishing the proof. Now we may finish the proof of Theorem \([13.1]\). Fix constants \( c, x \geq 0 \). Let \( X \) be any hole for \( G \) with \( \xi(X) = x \). Suppose \( \alpha_X, \beta_X \) are any vertices of \( G \) contained in \( X \). Let \( \Gamma = \{\gamma_i\}_{i=0}^K \) be the combinatorial sequence given by the axioms. Let \( P_X \) be a partition of \([0, K]\) into inductive and electric intervals. So we have

\[
(14.17) \quad d_G(\alpha_X, \beta_X) \leq \sum_{I \in P_{\text{ind}}} d_G(I) + \sum_{I \in P_{\text{ele}}} d_G(I) + C_2|P_X|.
\]

The last term arises from connecting left endpoints to right endpoints of adjacent intervals.

We bound the terms on the right-hand side; begin by noticing that \( |P_X| = |P_{\text{ind}}| + |P_{\text{ele}}|, \) \( |P_{\text{ele}}| \leq |P_{\text{ind}}| + 1, \) and \( |P_{\text{ind}}| \leq A d_X(\alpha_X, \beta_X) \). The second inequality follows from the way the partition is constructed and the last follows from Proposition \([14.8]\). Thus the third term of \((14.17)\) is quasi-bounded above by \( d_X(\alpha_X, \beta_X) \).

Next, consider the second term of \((14.17)\):

\[
\sum_{I \in P_{\text{ele}}} d_G(I) \leq A \sum_{I \in P_{\text{ele}}} d_X(I) \\
\leq d_X(\alpha_X, \beta_X) + (2C_5 + 2C_2)|P_{\text{ele}}| + 2C_2
\]

with the first inequality following from Proposition \([14.11]\) and the second from the reverse triangle inequality (Lemma \([13.16]\)).
Finally we bound the first term of (14.17). Let $c' = c + L_0$. Thus,

$$
\sum_{I \in P_{\text{ind}}} d_{G}(I) \leq \sum_{I_Y \in P_{\text{ind}}} \left( A_y' \left( \sum_{Z \subseteq X} [d_{Z}(I_Y)]c' \right) + A_y' \right)
$$

\[
\leq A'' \left( \sum_{I \in P_{\text{ind}}} \sum_{Z \subseteq X} [d_{Z}(I)]c' \right) + A'' \cdot |P_{\text{ind}}|
\]

\[
\leq A'' \left( \sum_{Z \subseteq X} \sum_{I \in P_{\text{ind}}} [d_{Z}(I)]c' \right) + A'' \cdot |P_{\text{ind}}|.
\]

Here $A_y'$ and the first inequality are given by Proposition 14.15. Also $A'' = \max \{ A_y' \mid y \leq x \}$. In the last line, each sum of the form $\sum_{I \in P_{\text{ind}}} [d_{Z}(I)]c'$ has at most three terms, by Remark 14.5 and the fact that $c' > L_0$. For the moment, fix a hole $Z$ and any three distinct elements $I, I', I'' \in P_{\text{ind}}$.

By the reverse triangle inequality (Lemma 13.16) we find that

$$
d_{Z}(I) + d_{Z}(I') + d_{Z}(I'') < d_{Z}(\alpha_X, \beta_X) + 6C_5 + 8C_2,
$$

which in turn is less than $d_{Z}(\alpha_X, \beta_X) + L_0$. It is now an exercise to show that

$$
[d_{Z}(I)]c' + [d_{Z}(I')]c' + [d_{Z}(I'')]c' < [d_{Z}(\alpha_X, \beta_X)]c' + L_0.
$$

Thus,

$$
\sum_{Z \subseteq X} \sum_{I \in P_{\text{ind}}} [d_{Z}(I)]c' \leq L_0 \cdot N + \sum_{Z \subseteq X} [d_{Z}(\alpha_X, \beta_X)]c,
$$

where $N$ is the number of non-zero terms in the final sum. Also, the sum ranges over all holes $Z \subseteq X$.

Combining the above inequalities, and increasing $A$ once again, implies that

$$
d_{G}(\alpha_X, \beta_X) \leq A \sum_{Z} [d_{Z}(\alpha_X, \beta_X)]c,
$$

where the sum ranges over all holes $Z \subseteq X$. This completes the proof of Theorem 13.1 \qed

15. BACKGROUND ON TEICHMÜLLER SPACE

Our goal in Sections 16, 17, and 19 will be to verify the axioms stated in Section 13 for the complex of curves of a non-orientable surface, for the arc complex, and for the disk complex. Here we give the necessary background on Teichmüller space. See also [37, 26].

Fix a surface $S = S_{g,n}$ of genus $g$ with $n$ punctures. Two conformal structures on $S$ are equivalent, written $\Sigma \sim \Sigma'$, if there is a conformal map $f: \Sigma \rightarrow \Sigma'$ which is isotopic to the identity. Let $\mathcal{T} = \mathcal{T}(S)$ be the Teichmüller space of $S$, the set of equivalence classes of analytically finite conformal structures $\Sigma$ on $S$.

Define the Teichmüller metric by

$$
d_{T}(\Sigma, \Sigma') = \inf_{f} \left\{ \frac{1}{2} \log K(f) \right\},
$$

where the infimum ranges over all quasiconformal maps $f: \Sigma \rightarrow \Sigma'$ isotopic to the identity and where $K(f)$ is the maximal dilatation of $f$. Recall that the infimum is
15.1. Quadratic differentials.

**Definition 15.2.** A quadratic differential \( q(z) \, dz^2 \) on \( \Sigma \) is an assignment of a holomorphic function to each coordinate chart that is a disk and of a meromorphic function to each chart that is a punctured disk. If \( z \) and \( \zeta \) are overlapping charts, then we require that

\[
q_z(z) = q_\zeta(\zeta) \left( \frac{d\zeta}{dz} \right)^2
\]

in the intersection of the charts. The meromorphic function \( q_z(z) \) has at most a simple pole at the puncture \( z = 0 \).

At any point away from the zeroes and poles of \( q \) there is a natural coordinate \( z = x + iy \) with the property that \( q_z \equiv 1 \). In this natural coordinate the foliation by lines \( y = c \) is called the horizontal foliation. The foliation by lines \( x = c \) is called the vertical foliation.

Now fix a quadratic differential \( q \) on \( \Sigma = \Sigma_0 \). Let \( x, y \) be natural coordinates for \( q \). For every \( t \in \mathbb{R} \) we obtain a new quadratic differential \( q_t \) with coordinates

\[
x_t = e^t x, \quad y_t = e^{-t} y.
\]

Also, \( q_t \) determines a conformal structure \( \Sigma_t \) on \( S \). The map \( t \mapsto \Sigma_t \) is the Teichmüller geodesic determined by \( \Sigma \) and \( q \).

15.3. Marking coming from a quadratic differential. Suppose \( \Sigma \) is an analytically finite conformal structure on \( S \). Let \( \sigma \) be the hyperbolic metric uniformizing \( \Sigma \) and note \( \sigma \) has finite area. In a slight abuse of terminology, we call the collection of shortest simple closed hyperbolic geodesics the systoles of \( \sigma \). Fix a sufficiently small constant \( \epsilon \); in particular, \( \epsilon \) is smaller than the Margulis constant. The \( \epsilon \)-thick part of Teichmüller space consists of those surfaces such that the hyperbolic systoles have length at least \( \epsilon \).

We define \( P = P(\sigma) \), a Bers pants decomposition of \( S \), as follows. Let \( \alpha_1 \) be any systole for \( \sigma \). Define \( \alpha_i \) to be any systole of \( \sigma \) restricted to \( S - (\alpha_1 \cup \ldots \cup \alpha_{i-1}) \). Continue in this fashion until \( P \) is a pants decomposition. By the collar lemma any curve with length less than the Margulis constant will necessarily be an element of \( P \).

Suppose now that \( q \) is a quadratic differential on the Riemann surface \( \Sigma \). Let \( \sigma \) be the uniformizing hyperbolic metric. Let \( P = P(\sigma) = \{ \alpha_i \} \) be a Bers pants decomposition. We must find transversals to \( P \) to obtain a complete clean marking \( \nu(q) \). Suppose \( P = \{ \alpha_i \} \). Fix \( i \) and let \( \alpha = \alpha_i \). Let \( S^\alpha \) be the annular cover of \( S \) corresponding to \( \alpha \). Note that \( q \) lifts to a singular Euclidean metric \( q^\alpha \) on \( S^\alpha \). Let \( a \) be a geodesic representative of the core curve of \( S^\alpha \) with respect to the metric \( q^\alpha \). Choose \( c \in C(S^\alpha) \) to be any geodesic arc, also with respect to \( q^\alpha \), that is perpendicular to \( a \). Let \( X_\alpha \) be the non-pants component of \( S - \nu(P - \{ \alpha \}) \). Let \( \beta \) be any curve in \( X_\alpha \) meeting \( \alpha \) minimally and so that \( d_\alpha(\beta, c) \leq 3 \). (See the discussion after the proof of Lemma 2.4 in [31].) Doing this for each \( i \) gives a complete clean marking \( \nu(q) = \{ (\alpha_i, \beta_i) \} \).

Suppose now that \( \{ \Sigma_t : t \in [-M, M] \} \) is the Teichmüller geodesic defined by the quadratic differentials \( \{ q_t \} \). Define \( \nu_t = \nu(q_t) \).
Lemma 15.4 ([11] Remark 6.2]). There is a constant $B_0 = B_0(S)$ with the following property. For any Teichmüller geodesic and for any time $t$, there is a $\delta > 0$ so that if $|t - s| \leq \delta$, then $\iota(\nu_s, \nu_t) < B_0$. \hfill \Box

Remark 15.5. Suppose $\Sigma_s$ and $\Sigma_t$ are surfaces in the $\epsilon$-thick part of $T(S)$. We take $B_0$ sufficiently large so that if $\iota(\nu_s, \nu_t) \geq B_0$, then $d_T(\Sigma_s, \Sigma_t) \geq 1$.

15.6. The marking path. We construct a path of Bers markings $\mu_n$, for $n \in [0, N] \subset \mathbb{N}$, as follows. Take $\mu_0 = \nu_{-M}$. Now suppose that $\mu_n = \nu_s$ is defined. Let $t > s$ be the first time the marking $\nu_t$ gives $\iota(\nu_s, \nu_t) \geq B_0$, if such a time exists. If so, let $\mu_{n+1} = \nu_t$. If no such time exists take $N = n$ and we are done.

Note that $\mu_n$ fills $S$ for every $n$. We now show that $\mu_n = \nu_s$ and $\mu_{n+1} = \nu_t$ have bounded intersection. By Lemma 15.4 there is a time $r$ with $s < r < t$ so that $\iota(\nu_r, \nu_t) < B_0$.

By construction

$\iota(\nu_s, \nu_r) < B_0$.

Let $\sigma$ be a hyperbolic metric on $S$ where all curves of base($\nu_r$) have length 1 and all transversals in $\nu_t$ are perpendicular to their base curves. In $\sigma$ all of the curves of $\nu_s$ and $\nu_t$ have length bounded above and below. It follows that $\iota(\nu_s, \nu_t) = \iota(\mu_n, \mu_{n+1})$ is bounded solely in terms of $B_0$. Thus there are constants $K, L$ so that $\{\mu_n\}$ is a path in the marking graph $M(S) = M_{K,L}(S)$. Note that $d_Y(\mu_n, \mu_{n+1})$ is uniformly bounded, independent of $Y \subset S$ and of $n \in [0, N]$. Theorem 6.1 of [11] says, for any subsurface $Y \subset S$, that the sequence $\{\pi_Y(\mu_n)\}$ is a uniformly quasi-geodesic.

15.7. The accessibility interval. Suppose $Y \subset S$ is an essential subsurface. Note, for any $n$, that the subsurface $Y$ is contained in the support of $\mu_n$, as the latter equals $S$.

In Section 5 of [11], Rafi defines an interval of isolation $I_Y$ inside of the parameterizing interval of the Teichmüller geodesic. Note that $I_Y$ is defined purely in terms of the geometry of the given quadratic differentials. Further, for all $t \in I_Y$ and for all components $\alpha \subset \partial Y$ the hyperbolic length of $\alpha$ in $\sigma_t$ is less than the Margulis constant. Furthermore, by Theorem 5.3 of [11], there is a constant $B_1$ so that if $[s, t] \cap I_Y = \emptyset$, then $d_Y(\nu_s, \nu_t) \leq B_1$.

We define $J_Y \subset [0, N]$ to be the subinterval of the marking path where the time corresponding to $\mu_n$ lies in $I_Y$. Finally, if $m \in J_Y$, then $\partial Y$ is contained in base($\mu_m$) and thus $\iota(\partial Y, \mu_m) \leq 2 \cdot |\partial Y|$.

15.8. The distance estimate in Teichmüller space. We end this section by quoting another result of Rafi.

Theorem 15.9 ([10] Theorem 1.1)]. Fix a surface $S$ and a constant $\epsilon > 0$. There is a constant $C_0 = C_0(S, \epsilon)$ so that for any $c > C_0$ there is a constant $A$ with the following property. Suppose $\Sigma$ and $\Sigma'$ lie in the $\epsilon$-thick part of $T(S)$. Then

$$d_T(\Sigma, \Sigma') = A \sum_{X} [d_X(\mu, \mu')]_c + \sum_{\alpha} [\log d_\alpha(\mu, \mu')]_c,$$

where $\mu$ and $\mu'$ are Bers markings on $\Sigma$ and $\Sigma'$, where $Y \subset S$ ranges over non-annular surfaces, and where $\alpha$ ranges over vertices of $C(S)$. \hfill \Box
16. Paths for the non-orientable surface

Fix $F$ a compact, connected, non-orientable surface. Let $S$ be the orientation double cover with covering map $\rho_F : S \to F$. Let $\tau : S \to S$ be the associated involution. Recall from Section 6 that $C(F)$, $C_2(F)$, and $C^\tau(S)$ are all quasi-isometric.

In this section we prove Lemma 16.8, the classification of holes for $C^\tau(S)$. This directly implies the Gromov hyperbolicity of $C(F)$; see Corollary 6.4. As a bit of practice, we also verify all of the axioms of Section 13 for $C^\tau(S)$.

16.1. The marking path. A pants decomposition $P$ for $S$ is $\tau$–invariant if $P$ is a simplex in $C^\tau(S)$.

Definition 16.2. A complete clean marking $\mu = \{(\alpha_i, \beta_i)\}$ for $S$ is $\tau$–invariant if $\text{base}(\mu)$ is $\tau$–invariant and $\tau(\{(\beta_i)\}) = \{\beta_i\}$.

Note that the condition on base($\mu$) is stronger than the condition on transversals; the latter are only required to be setwise $\tau$–invariant. We will use the extreme rigidity of Teichmüller geodesics to find a path of $\tau$–invariant markings.

Lemma 16.3. For every $\tau$–invariant hyperbolic metric $\sigma$ there is a $\tau$–invariant Bers pants decomposition $P = P(\sigma)$.

Proof. Let $P_0 = \emptyset$. Suppose $0 \leq k < \xi(S)$ curves have been chosen to form $P_k$. By induction we may assume that $P_k$ is a simplex in $C^\tau(S)$. Let $Y$ be a component of $S - P_k$ with $\xi(Y) \geq 1$. Let $\alpha$ be a systole for $Y$; so $\alpha$ is a shortest, simple, closed, essential, and non-peripheral geodesic.

Claim. Either $\tau(\alpha) = \alpha$ or $\alpha \cap \tau(\alpha) = \emptyset$.

Proof. Suppose not and take $p \in \alpha \cap \tau(\alpha)$. Then $\tau(p) \in \alpha \cap \tau(\alpha)$ as well, and, since $\tau$ has no fixed points, $p \neq \tau(p)$. The points $p$ and $\tau(p)$ divide $\alpha$ into segments $\beta$ and $\gamma$. Since $\tau$ is an isometry, we have

$$\ell_\sigma(\tau(\beta)) = \ell_\sigma(\beta) \quad \text{and} \quad \ell_\sigma(\tau(\gamma)) = \ell_\sigma(\gamma).$$

Now concatenate to obtain (possibly immersed) loops

$$\beta' = \beta \ast \tau(\beta) \quad \text{and} \quad \gamma' = \gamma \ast \tau(\gamma).$$

If $\beta'$ is null-homotopic, then $\alpha \cup \tau(\alpha)$ cuts a bigon out of $S$, contradicting our assumption that $\alpha$ was a geodesic. Suppose, by way of a contradiction, that $\beta'$ is homotopic to some boundary component $b \subset \partial Y$. Since $\tau(\beta') = \beta'$, it follows that $\tau(b)$ and $\beta'$ are also homotopic. Thus $b$ and $\tau(b)$ cobound an annulus, implying that $Y$ is an annulus, a contradiction. Thus $\beta'$ and similarly $\gamma'$ are essential and non-peripheral.

Let $\beta''$ and $\gamma''$ be the geodesic representatives of $\beta'$ and $\gamma'$. Since $\alpha$ and $\tau(\alpha)$ meet transversely, $\beta''$ has length strictly smaller than $2\ell_\sigma(\beta)$. Similarly the length of $\gamma''$ is strictly smaller than $2\ell_\sigma(\gamma)$. Suppose $\beta''$ is shorter than $\gamma''$. It follows that $\beta''$ is strictly shorter than $\alpha$. If $\beta''$ is embedded, then this contradicts the assumption that $\alpha$ was a systole. If $\beta''$ is not embedded, then there is an embedded curve $\beta''$ inside of a regular neighborhood of $\beta''$ which is again essential, non-peripheral, and has geodesic representative shorter than $\beta''$. This is our final contradiction and the claim is proved. \hfill \Box

If $\tau(\alpha) = \alpha$, let $P_{k+1} = P_k \cup \{\alpha\}$. If $\tau(\alpha) \neq \alpha$, then by the above claim $\alpha \cap \tau(\alpha) = \emptyset$. In this case let $P_{k+2} = P_k \cup \{\alpha, \tau(\alpha)\}$. Lemma 16.3 is proved. \hfill \Box
Recall, if $\alpha$ is a curve in $S$, then $S^\alpha$ is the corresponding annular cover. If $q$ is a quadratic differential on $S$, then $q^\alpha$ denotes the lifted metric. We pull $\alpha$ tight inside of $S^\alpha$ and define $\bot$ to be the set

$$\{ \gamma \in \mathcal{C}(S^\alpha) \mid \text{some geodesic representative of } \gamma \text{ is perpendicular to } \alpha \}. $$

**Lemma 16.4.** There is a constant $C$ with the following property. Let $q$ be a $\tau$–invariant quadratic differential and let $\sigma$ be the uniformizing hyperbolic metric. Let $P = P(\sigma) = \{ \alpha_i \}$ be a $\tau$–invariant Bers pants decomposition, as provided by Lemma 16.3. Then there are transversal curves $\beta_i$ for the $\alpha_i$ so that

- $\tau(\{ \beta_i \}) = \{ \beta_i \}$ and
- for each $i$ we have $d_{\alpha_i}(\beta_i, \bot) \leq C$.

Note that $\nu = \{ (\alpha_i, \beta_i) \}$ is a $\tau$–invariant Bers marking.

**Proof.** Fix $\alpha \in P$. Set $P_\alpha = P - \{ \alpha, \tau(\alpha) \}$. Let $X$ be the union of the non-pants components of $S - n(P_\alpha)$. There are three cases to consider depending on whether $\tau(\alpha) = \alpha$ and whether $X$ is connected.

Suppose $\tau(\alpha) \cap \alpha = \emptyset$ and $X$ is not connected. It follows that $X$ is a union of two copies of $S_{0,4}$, interchanged by $\tau$. In this case we choose a transversal $\beta$ for $\alpha$ so that $d_{\alpha}(\beta, \bot) \leq 3$.

Suppose $\tau(\alpha) \cap \alpha = \emptyset$ and $X$ is connected. Since $\tau$ fixes $X$ setwise, it cannot fix any boundary component of $X$; thus $X$ is a twice-holed torus and $X/\tau$ is a once-holed Klein bottle. In this case we choose a transversal $\beta \subset X - \tau(\alpha)$ so that $d_{\alpha}(\beta, \bot) \leq 3$.

In these two cases, add $\beta$ and $\tau(\beta)$ to the set of transversals and we are done.

Suppose $\tau(\alpha) = \alpha$. It follows that $X$ is a copy of $S_{0,4}$. Thus $X/\tau$ is a twice-holed $\mathbb{RP}^2$. There are only four essential non-peripheral curves in $X/\tau$. Two of these are cores of Möbius bands and the other two are their doubles. The cores meet in a single point. Perforce $\alpha$ is the double cover of one core, and we take $\beta$ to be the double cover of the other.

Note that $\bot$ is a $\tau$–invariant, diameter-one subset of $\mathcal{C}(S^\alpha)$. If $d_{\alpha}(\beta, \bot)$ is large, then it follows that $d_{\alpha}(\tau(\beta), \bot)$ is also large. Also, $\tau(\beta)$ twists in the opposite direction from $\beta$. Thus

$$d_{\alpha}(\beta, \tau(\beta)) - 2d_{\alpha}(\beta, \bot) = O(1)$$

and so $d_{\alpha}(\beta, \tau(\beta))$ is large, contradicting the fact that $\beta$ is $\tau$–invariant. \qed

We now turn to verifying the marking path and accessibility requirements, Axioms 13.4 and 13.5. Suppose that $\alpha, \beta \in \mathcal{C}(S)$. If $\alpha$ and $\beta$ do not fill $S$, then we may replace $S$ by the support of their union. Following Thurston [47 Theorem 7] let $q$ be the square-tiled quadratic differential, with squares associated to the points of $\alpha \cap \beta$. (See also [7] page 122.) Let $q_t$ be the image of $q$ under the Teichmüller geodesic flow.

**Lemma 16.5.** $\tau^*q_t = q_t$.

**Proof.** Note that $\tau$ preserves $\alpha$ and also $\beta$. Since $\tau$ permutes the points of $\alpha \cap \beta$ it permutes the rectangles of the singular Euclidean metric $q_t$ while preserving their vertical and horizontal foliations. Thus $\tau$ is an isometry of the metric, and the conclusion follows. \qed
Let \( \{ \Sigma_t \mid t \in [-M, M] \} \) be the Teichmüller geodesic determined by \( q_t \). Choose \( M \) so that the hyperbolic length of \( \alpha \) is less than the Margulis constant in \( \sigma_{-M} \) and the same holds for \( \beta \) in \( \sigma_M \). Also, \( \alpha \) is the shortest curve in \( \sigma_{-M} \) and similarly for \( \beta \) in \( \sigma_M \). Lemma 16.4 now gives \( \tau \)-invariant Bers markings \( \nu_t \) for every \( t \in [-M, M] \).

As in Section 15.6 we can choose a discrete subset to obtain a path in the marking graph \( \mathcal{M}(S) \). By the discussion in Section 15.6 this path satisfies Axiom 13.4. By the discussion in Section 15.7 it also satisfies Axiom 13.5.

16.6. The combinatorial sequence. The previous section gives a Teichmüller geodesic and a marking path \( \{ \mu_n \}_{n=0}^N \). Choose the combinatorial sequence by picking \( \gamma_n \in \text{base}(\mu_n) \) so that \( \gamma_n \) is a \( \tau \)-invariant curve or pair of curves and so that \( \gamma_n \) is a systole in \( \sigma_{-M} \) at the corresponding time. Note that \( \gamma_0 = \alpha \) and \( \gamma_N = \beta \). Also, the reindexing map is the identity map.

We now check Axiom 13.6. Since \( \iota(\gamma_n, \mu_{\tau(n)}) = 2 \), the first requirement is satisfied. Since \( \mu_n \) and \( \mu_{n+1} \) have bounded intersection, the same holds for \( \gamma_n \) and \( \gamma_{n+1} \). Projection to \( F \), surgery, and Lemma 2.3 imply that 
\[
d_{C^\tau}(\gamma_n, \gamma_{n+1}) \text{ is uniformly bounded.}
\]

16.7. The classification of holes. We now finish the classification of large holes for \( C^\tau(S) \). Fix \( L_0 > 3C_1 + 2C_2 + 2C_5 \). Note that these constants are available because we have verified the axioms that give them.

Lemma 16.8. Suppose \( \alpha, \beta \in C^\tau(S) \). Suppose \( Z \subset S \) has \( d_Z(\alpha, \beta) > L_0 \). Then \( Z \) is symmetric.

Proof. Let \( \{ \Sigma_t \} \) be the Teichmüller geodesic defined above and let \( \sigma_t \) be the uniformizing hyperbolic metric. Since \( L_0 > C_1 + 2C_2 \) the accessibility axiom implies that \( J_Z = [m, n] \) is non-empty. For all \( t \) in the interval of isolation \( I_Z \), we have
\[
\ell_{\sigma_t}(\delta) < \epsilon,
\]
where \( \delta \) is any component of \( \partial Z \) and \( \epsilon \) is the Margulis constant. Let \( Y = \tau(Z) \). Since \( \tau \) is an isometry (Lemma 16.5) and since the interval of isolation is metrically defined we have \( I_Y = I_Z \) and thus \( J_Y = J_Z \). Deduce that \( \partial Y \) is also short in \( \sigma_t \). The collar lemma implies that \( \partial Y \cap \partial Z = \emptyset \). If \( Y \) and \( Z \) overlap, then by Lemma 13.17 we have
\[
d_Z(\mu_m, \mu_n) < C_1.
\]
By the triangle inequality and by two applications of property (3) of Axiom 13.5 we have
\[
d_Z(\mu_0, \mu_N) < 3C_1.
\]
By the combinatorial axiom it follows that
\[
d_Z(\alpha, \beta) < 3C_1 + 2C_2,
\]
and this is a contradiction. Deduce either \( Y = Z \) or \( Y \cap Z = \emptyset \), as desired. \( \square \)

As noted in Section 6 this shows that the only hole for \( C^\tau(S) \) is \( S \) itself. Thus Axioms 13.3 and 13.7 hold vacuously.
16.9. In straight intervals. We verify Axiom\ref{ax:13.13} Suppose \([p, q]\) is a straight interval. We must show that \(d_{\mathcal{C}}(\gamma_p, \gamma_q) \leq d_S(\gamma_p, \gamma_q)\). Suppose \(\mu_p = \nu_s\) and \(\mu_q = \nu_t\); that is, \(s\) and \(t\) are times when \(\mu_p, \mu_q\) are short markings. Thus \(d_{\mathcal{C}}(\mu_p, \mu_q) \leq L_2\) for every \(Y \subseteq S\). This implies that the Teichmüller geodesic, along the straight interval, lies in the \(\epsilon\)-thick part of Teichmüller space for \(\epsilon = \epsilon(L_2)\). See \cite{Rafi11} Theorem 5.5.

Notice that \(d_{\mathcal{C}}(\gamma_p, \gamma_q) \leq C_2|p - q|\), since for all \(i \in [p, q - 1]\), \(d_{\mathcal{C}}(\gamma_i, \gamma_{i+1}) \leq C_2\). So it suffices to bound \(|p - q|\). By our choice of \(B_0\) (see Remark \ref{rem:15.3}) and because the Teichmüller geodesic lies in the thick part we find that \(|p - q| \leq d_T(\Sigma, \Sigma_t)\).

Rafi’s distance estimate (Theorem \ref{thm:15.9}) gives:

\[
d_T(\Sigma, \Sigma_t) = A \cdot d_S(\nu_s, \nu_t).
\]

Since \(\nu_s = \mu_p, \nu_t = \mu_q\), and since \(\gamma_p \in \text{base}(\mu_p), \gamma_q \in \text{base}(\mu_q)\), we deduce that

\[
d_S(\mu_p, \mu_q) \leq d_S(\gamma_p, \gamma_q) + 4.
\]

This verifies Axiom\ref{ax:13.13}.

16.10. Taking shortcuts. Finally, we verify Axiom\ref{ax:13.14} Recall that the reindexing map is the identity. Since \(S\) is the only hole, the interval \([0, N]\) is electric. Suppose \([p, q] \subset [0, N]\) is a shortcut for the non-hole \(Z \subseteq S\). Thus \(\gamma_p\) and \(\gamma_q\) are contained in \(\text{base}(\mu_p)\) and \(\text{base}(\mu_q)\), respectively. From the first half of the shortcut hypothesis (Definition\ref{def:13.12}), we deduce that \(\partial Z\) is contained in both \(\text{base}(\mu_p)\) and \(\text{base}(\mu_q)\). The second half of the shortcut hypothesis, together with Lemma\ref{lem:16.8} implies that \(\partial Z\) is symmetric, and we are done.

17. Paths for the arc complex

We verify the axioms of Section\ref{sec:13} for the arc complex \(\mathcal{A}(S, \Delta)\). It is worth pointing out that in the case of the arc complex the axioms may be verified using Teichmüller geodesics, train-track splitting sequences, quasi-Fuchsian three-manifolds, or resolutions of hierarchies. We use Teichmüller geodesics because they also deal with the non-orientable case; this is discussed at the end of the section. Train-track splittings and, presumably, quasi-Fuchsian manifolds also deal with the non-orientable case.

First note that Axiom\ref{ax:13.3} follows from Lemma\ref{lem:7.3}.

17.1. The marking path. We are given a pair of arcs \(\alpha, \beta \in \mathcal{A}(X, \Delta)\). Recall that \(\sigma_X: \mathcal{A}(X) \to \mathcal{C}(X)\) is the surgery map of Definition\ref{def:1.3} Let \(\alpha' = \sigma_X(\alpha)\) and let \(\beta' = \sigma_X(\beta)\). Note that \(\alpha'\) cuts a pants off of \(X\). By passing to a subsurface, we may assume that \(\alpha'\) and \(\beta'\) fill \(X\).

As in Section\ref{sec:16.1} let \(q\) be the square-tiled quadratic differential determined by \(\alpha'\) and \(\beta'\). As in Section\ref{sec:15.6} the differentials \(q_t\) give a marking path \(\{\mu_n\}_{n=0}^N\). This path satisfies the marking and accessibility axioms\ref{ax:13.4} \ref{ax:13.5}.

17.2. The combinatorial sequence. Let \(Y_n \subset X\) be any component of \(X - \text{base}(\mu_0)\) meeting \(\Delta\). So \(Y_n\) is a pair of pants. Let \(\gamma_n\) be any essential arc in \(Y_n\) with both endpoints in \(\Delta\). Since \(\alpha' \subset \text{base}(\mu_0)\) and \(\beta' \subset \text{base}(\mu_N)\) we may choose \(\gamma_0 = \alpha\) and \(\gamma_N = \beta\).

The reindexing map is the identity. Thus \(\varphi(\gamma_n, \mu_n) \leq 4\). This bound, the bound on \(\varphi(\mu_n, \mu_{n+1})\), and Lemma\ref{lem:13.8} imply that \(\varphi(\gamma_n, \gamma_{n+1})\) is also bounded. The usual surgery argument shows that if two arcs have bounded intersection, then they have bounded distance. This verifies Axiom\ref{ax:13.6}.
17.3. The replacement, straight, and shortcut axioms. Suppose \( Y \subset X \) is cleanly embedded and is a hole for \( \mathcal{A}(S, \Delta) \). Thus \( \Delta \subset \partial Y \). Suppose \( \gamma_n \) has \( n \in J_Y \). Thus \( \partial Y \subset \text{base}(\mu_n) \) and so \( \gamma_n \cap \partial Y = \emptyset \). Taking \( \gamma' = \gamma_n \) verifies Axiom 13.4.

Axiom 13.13 is verified as in Section 16. We now verify Axiom 13.14. Suppose \( [i, j] \subset [0, N] \) is an electric interval and \( [p, q] \subset [i, j] \) is a shortcut for a cleanly embedded non-hole \( Z \subset X \). Since \( p, q \in J_Z \), we deduce that \( \partial Z \subset \text{base}(\mu_p) \cap \text{base}(\mu_q) \). Thus \( \gamma_p \) and \( \gamma_q \) are disjoint from \( \partial Z \). There are now several cases.

If \( \ell(\gamma_p, \gamma_q) = 0 \), then we are done. If both \( \gamma_p \) and \( \gamma_q \) are contained in \( Z \), then we are done, because \( Z \) is not a hole. So suppose that \( \gamma_p \) and \( \gamma_q \) are both contained in \( Y \), a component of \( X - n(Z) \). If \( Y \) is not a hole, then we are done. Finally, suppose that \( Y \) is a hole for \( \mathcal{A}(S, \Delta) \). Since \( [i, j] \) is electric, we deduce that \( d_W(\gamma_i, \gamma_j) < L_2 \) for all holes \( W \subset X \). Lemma 13.16 gives a uniform (depending only on \( x = \xi(X) \)) upper bound on \( d_W(\gamma_p, \gamma_q) \), for all holes \( W \subset Y \). Since \( \xi(Y) < \xi(X) \) we may inductively apply Theorem 13.1 for the complex \( \mathcal{A}(Y, \Delta) \). Thus \( d_{\mathcal{A}(Y, \Delta)}(\gamma_p, \gamma_q) \) is bounded by a constant depending only on \( x \), as desired.

17.4. Non-orientable surfaces. Suppose \( F \) is a non-orientable, connected, non-simple surface with boundary. Suppose \( \Delta_F \subset F \) is a collection of boundary components. Let \( S \) be the orientation double cover and \( \tau: S \to S \) be the involution so that \( S/\tau = F \). Let \( \Delta \) be the preimage of \( \Delta_F \). Let \( \mathcal{A}(S, \Delta) \) be the invariant arc complex.

Suppose \( \alpha_F \) and \( \beta_F \) are vertices in \( \mathcal{A}(F, \Delta_F) \). Let \( \alpha, \beta \) be their preimages. Without loss of generality, we may assume that \( \sigma_F(\alpha_F) \) and \( \sigma_F(\beta_F) \) fill \( F \). Note that \( \sigma_F(\alpha_F) \) cuts a surface \( X \) off of \( F \). The surface \( X \) is either a pants or a twice-holed \( \mathbb{RP}^2 \). When \( X \) is a pants we define \( \alpha' \subset S \) to be the preimage of \( \sigma_F(\alpha_F) \). When \( X \) is a twice-holed \( \mathbb{RP}^2 \) we take \( \gamma_F \) to be a core of one of the two Möbius bands contained in \( X \) and we define \( \alpha' \) to be the preimage of \( \gamma_F \cup \sigma_F(\alpha_F) \). We define \( \beta' \) similarly. Notice that \( \alpha \) and \( \alpha' \) meet in at most four points.

We now use \( \alpha' \) and \( \beta' \) to build a \( \tau \)-invariant Teichmüller geodesic. The construction of the marking path and combinatorial sequence for \( \mathcal{A}(S, \Delta) \) is unchanged. Notice that we may choose combinatorial vertices because \( \text{base}(\mu_n) \) is \( \tau \)-invariant. There is a small annoyance: when \( X \) is a twice-holed \( \mathbb{RP}^2 \) the first vertex, \( \gamma_0 \), is disjoint from but not equal to \( \alpha \). Strictly speaking, the first and last vertices are \( \gamma_0 \) and \( \gamma_N \); our constants are stated in terms of their subsurface projection distances. However, since \( \alpha \cap \gamma_0 = \emptyset \), and the same holds for \( \beta \) and \( \gamma_N \), their subsurface projection distances are all bounded.

18. Background on train tracks

Here we give the necessary definitions and theorems regarding train tracks. The standard reference is [38]. See also [36]. We closely follow the discussion in [33].

18.1. On tracks. A generic train track \( \tau \subset S \) is a smooth, embedded trivalent graph. As usual we call the vertices switches and the edges branches. At every switch the tangents of the three branches agree. Also, there are exactly two incoming branches and one outgoing branch at each switch. See Figure 18.2 for the local model of a switch. We require every region of \( S - n(\tau) \) to have negative index.

Let \( \mathcal{B}(\tau) \) be the set of branches. A transverse measure on \( \tau \) is a function \( w: \mathcal{B} \to \mathbb{R}_{\geq 0} \) satisfying the switch conditions: at every switch the sum of the incoming
measures equals the outgoing measure. Let $P(\tau)$ be the projectivization of the cone of transverse measures. As discussed in the references, each vertex of $P(\tau)$ gives an essential, non-peripheral curve carried by $\tau$. Let $V(\tau)$ be the set of curves determined by these vertices. Thus $V(\tau)$ is a marking in the sense of Section 2.9. There are only finitely many tracks, up to the action of the mapping class group. It follows that $\iota(V(\tau))$ is uniformly bounded, depending only on the topological type of $S$.

If $\tau$ and $\sigma$ are train tracks, and $Y \subset S$ is an essential surface, then define

$$d_Y(\tau, \sigma) = d_Y(V(\tau), V(\sigma)).$$

We also adopt the notation $\pi_Y(\tau) = \pi_Y(V(\tau))$.

A train track $\sigma$ is obtained from $\tau$ by sliding if $\sigma$ and $\tau$ are related as in Figure 18.3. We say that a train track $\sigma$ is obtained from $\tau$ by splitting if $\sigma$ and $\tau$ are related as in Figure 18.4.

Recall that the number of tracks is bounded, up to the action of the mapping class group. So, if $\sigma$ is obtained from $\tau$ by either a slide or a split, then $\iota(V(\tau), V(\sigma))$ is uniformly bounded.
18.5. The marking path. We will use sequences of train tracks to define our marking path.

**Definition 18.6.** A *sliding and splitting sequence* is a collection \( \{ \tau_n \}_{n=0}^N \) of train tracks so that \( \tau_{n+1} \) is obtained from \( \tau_n \) by a slide or a split.

The sequence \( \{ \tau_n \} \) gives a sequence of markings via the map \( \tau_n \mapsto \mu_n = V(\tau_n) \).

Note that every vertex of \( \tau_{n+1} \) has a multiple that is a sum of vertices of \( \tau_n \). Using this, it is an exercise to show that the support of \( \mu_{n+1} \) is contained within the support of \( \mu_n \). Theorem 5.5 from [33] verifies the remaining half of Axiom 13.4.

**Theorem 18.7.** Fix a surface \( S \). There is a constant \( A \) with the following property. Suppose \( \{ \tau_n \}_{n=0}^N \) is a sliding and splitting sequence of birecurrent tracks in \( S \). Suppose \( Y \subset S \) is an essential surface. Then the map \( n \mapsto \pi_Y(\tau_n) \), as parameterized by splittings, is an \( A \)-unparameterized quasi-geodesic. \( \square \)

When \( Y = S \), Theorem 18.7 is essentially due to the first author and Minsky [32, Theorem 1.3].

In Section 5.2 of [33], for every sliding and splitting sequence \( \{ \tau_n \}_{n=0}^N \) and for any essential subsurface \( X \subseteq S \) an accessible interval \( I_X \subset [0, N] \) is defined. Axiom 13.5 is now verified by Theorem 5.3 of [33].

18.8. Quasi-geodesics in the marking graph. We will also need the following.

(See [20] for closely related work.)

**Theorem 18.9** ([33, Theorem 6.1]). Fix a surface \( S \). There is a constant \( A \) with the following property. Suppose \( \{ \tau_n \}_{n=0}^N \) is a sliding and splitting sequence of birecurrent tracks, injective on slide subsequences, where \( \mu_n = V(\tau_n) \) fills \( S \), for all \( n \). Then \( \{ \mu_n \} \) is an \( A \)-quasi-geodesic in the marking graph. \( \square \)

19. Paths for the disk complex

Suppose \( (M, S) \) is a spotless pair. The goal of this section is to verify the axioms of Section 13 for the disk complex \( D(M, S) \).

19.1. Holes. The fact that all large holes interfere is recorded above as Lemma 12.21. This verifies Axioms 13.3.

19.2. The combinatorial sequence. Suppose \( D, E \in D(M, S) \) are disks contained in a compressible hole \( X \subset S \). As usual we may assume that \( D \) and \( E \) fill \( X \). Recall that if \( \tau \subset X \) is a train track, then \( V(\tau) \) is the set of vertices. We now appeal to a result of the first author and Minsky.

**Theorem 19.3** ([32, Section 4]). There exists a surgery sequence of disks \( \{ D_i \}_{i=0}^K \), a sliding and splitting sequence of birecurrent tracks \( \{ \tau_n \}_{n=0}^N \), and a reindexing function \( r : [0, K] \to [0, N] \) so that

- \( D_0 = D \),
- \( E \in \mu_N \),
- \( D_i \cap D_{i+1} = \emptyset \) for all \( i \), and
- \( \iota(\partial D_i, \mu_{r(i)}) \) is uniformly bounded for all \( i \).

Here \( \mu_n = V(\tau_n) \). \( \square \)
Remark 19.4. Note that the double-wave curve replacements of [32, Section 4] are not needed here; as $X$ is a hole, no curve of $\partial X$ compresses in $M$. It follows that consecutive disks in the surgery sequence are disjoint (as opposed to meeting at most four times). Also, in the terminology of [33], the disk $D_i$ is a wide dual for the track $\tau(\iota(i))$. Note that $\tau_n$ is recurrent because $D$ is fully dual to $\partial Y$.

Thus $\mu_n = V(\tau_n)$ will be our marking path and $D_i$ will be our combinatorial sequence. Axioms [13.4] and [13.5] were obtained in Section 18. The requirements of Axiom [13.6] are now verified by Theorem 19.3.

19.5. The replacement axiom. We turn to Axiom [13.7]. Suppose $Y \subset X$ is a large hole for $\mathcal{D}(M, S)$. Fix an index $i$ so that $n = r(i) \in J_Y$. By Axiom [13.5], we have $Y \subset \text{supp}(\mu_n)$ and also $\iota(\partial Y, \mu_n)$ is uniformly bounded. By Axiom [13.6], there is a uniform bound on $\iota(\partial D_i, \mu_n)$. It follows that there is a constant $K$ depending only on $x = \xi(X)$ so that

$$\iota(\partial D_i, \partial Y) < K.$$ 

As in Section 11.1, boundary compress $D_i$ as much as possible into $X - \partial Y$ to obtain a disk $D'$ so that either

- $D'$ cannot be boundary compressed into $X - \partial Y$ or
- $D'$ is disjoint from $\partial Y$.

We arrange matters so that every boundary compression reduces the intersection with $\partial Y$ by at least a factor of two. Thus

$$d_D(D_i, D') \leq \log_2(K).$$

Suppose $Y$ is a compressible hole. Lemma 8.6 implies $\partial D' \subset Y$ and we are done.

Remark 19.6. Note that part (1) of Axiom [13.7] cannot be obtained when $Y$ is an incompressible hole; it is impossible for any disk $D \in \mathcal{D}(M, S)$ to have $\partial D \subset Y$. We finesse this issue as follows. Suppose $Y$ is a large incompressible hole for $\mathcal{D}(M, S)$. Let $\rho_F: T \to F$ be the $F$-bundle given by Theorem 12.1. Let $\Delta \subset \partial_s T$ be the collection of annuli that are boundary parallel into $S$. Isotope the components of $\Delta$, fixing $\partial \Delta$ pointwise, to lie in $S$. Let $\delta = \rho_F(\Delta)$. We say that a disk $E \in \mathcal{D}(M, X)$ is contained in $Y$ if $E$ is ambiently isotopic to a vertical disk inside of $T$. Note that the complex of vertical rectangles in $T$, with vertical boundary in $\Delta$, is isomorphic to $A(F, \delta)$.

We now verify this form of Axiom [13.7]. Suppose $Y$ is a large incompressible hole. Let $T$ and $\Delta$ be as given in Remark 19.6. Isotope $D'$ to minimize intersection with $\partial_s T$. Let $\Gamma = \partial_s T - \Delta$. Notice that all intersections $D' \cap \Gamma$ are essential arcs in $\Gamma$. Simple closed curves are ruled out by the irreducibility of $M$. Inessential arcs are ruled out by the fact that $D'$ cannot be boundary compressed into $X - \partial Y$. Let $B$ be an outermost bigon of $D' - \Gamma$. Then Lemma 8.12 implies that $B$ is isotopic in $T$ to a vertical disk.

If $B = D'$, then we are done. If not, then let $A \in \Gamma$ be the vertical annulus meeting $B$. Let $D'' = A_F$ be the boundary compression of $A$ along $B$. Note that $D''$ is also vertical in $T$. Since $\iota(\partial D'', \partial D') \leq K - 2$ we are done.
19.7. **Straight intervals.** We now check Axiom 13.13. Suppose \([p, q] \subset [0, K]\) is a straight interval. Let \(m, n = r(p), r(q)\). Recall that \(d_Y(\mu_m, \mu_n) < L_2\) for all strict subsurfaces \(Y \subset X\). We must check that \(d_X(D_p, D_q) \leq A d_X(D_p, D_q)\). Since \(d_X(D_p, D_q) \leq C_2|q - p|\) it is enough to bound \(|q - p|\). Note that \(|q - p| \leq |n - m|\) because the reindexing map is increasing. So it is enough to bound \(|n - m|\).

Suppose \(\mu_n\) fills \(X\). Then by Theorem 18.9 the path \(\{\mu_t\}\) is a quasi-geodesic in \(\mathcal{M}(X)\). It follows that \(|n - m| \leq A d_{\mathcal{M}(X)}(\mu_m, \mu_n)\). Increasing \(A\) as needed and applying Theorem 4.13 we have

\[
d_X(\mu_m, \mu_n) \leq A \sum_Y [d_Y(\mu_m, \mu_n)] L_2
\]

and the right-hand side is quasi-bounded by \(d_X(\mu_m, \mu_n)\), which in turn is less than \(d_X(D_p, D_q) + 2C_2\), proving Axiom 13.13 when \(\mu_n\) fills \(X\).

If \(\mu_n\) does not fill \(X\), then define \(n' \in [m, n]\) to be the first index so that \(\mu_{n'}\) does not fill \(X\). Let \(q'\) be the first index so that \(r(q') \in [n', n]\). It follows from the straight hypothesis and Lemma 4.3 that \(\varepsilon(\partial D_{q'}, \partial D_q)\) is uniformly bounded. This, with the previous paragraph, verifies Axiom 13.13.

19.8. **Shortcut intervals.** Lastly we check Axiom 13.14. Suppose \([i, j] \subset [0, K]\) is electric for \(X\) and \([p, q] \subset [i, j]\) is a shortcut for \(Z \subset X\). If \(p\) is a half-integer there is nothing to prove. Note that \(X\) is a hole for \(\mathcal{D}(M, S)\) while \(Z\) is not. Let \(Y = X - Z\).

By hypothesis \(r([p, q]) \subset J_Z\). Let \(D, E = D_p, D_q\). As in the proof of the replacement axiom (Section 19.5) there is a uniform bound \(K\) so that \(\varepsilon(\partial D, \partial Z) \leq K\). Let \(D'\) be the result of maximally boundary compressing \(D\) into \(X - \partial Z\). Thus \(d_D(D, D') \leq \log_2(K)\). There is a similar disk \(E'\) for \(E\).

If \(D' \cap E' = \emptyset\), then we are done. So suppose \(D' \cap E' \neq \emptyset\). If \(Z\) or \(Y\) is compressible, then \(\partial D' \cap \partial Z = \partial E' \cap \partial Z = \emptyset\) by Lemma 11.3. If both \(\partial D'\) and \(\partial E'\) are contained in \(Z\), then we are done. If not, then both are contained in some component \(Y'\) of \(Y\). If \(Y'\) is not a hole for \(\mathcal{D}(M, S)\), then, again, we are done. Suppose \(Y'\) is a hole for \(\mathcal{D}(M, S)\). Since \(\xi(Y') < \xi(X)\) we can apply Theorem 13.1 inductively to \(\mathcal{D}(M, Y')\). Since \([i, j]\) is electric there is a sufficiently large cut-off \(c\) so that all terms on the right-hand side of the upper bound vanish. Again we are done.

We are left with the possibility that both \(Z\) and \(Y\) are incompressible. It follows that \(Z\) is a hole for \(\mathcal{D}(M, X)\). The shortcut hypothesis gives \(d_Z(\mu_{r(i)}, \mu_{r(j)}) \geq L_1(x)\). It follows that \(d_Z(D_i, D_j)\) is large and so \(Z\) is a large hole for \(\mathcal{D}(M, X)\). By Theorem 12.1 there is an \(I\)-bundle \(\rho: T \to F\) so that \(T \subset M, \partial_h T \subset X, Z\) is a component of \(\partial_h T\), and some component of \(\partial_h T\) is parallel into \(X\). As in the proof of the replacement axiom, there are disks \(D'', E''\) contained in \(T\), vertical in \(T\), and having intersection at most \(K - 1\) with \(D', E'\) respectively.

Since \(Z\) is not a hole for \(\mathcal{D}(M, S)\) there is a disk \(C\), disjoint from \(Z\), compressing \(S\) into \(M\). After performing boundary compressions we may assume that \(C \cap T = \emptyset\). Thus \(C \cap D'' = C \cap E'' = \emptyset\). This verifies the final axiom, Axiom 13.14.

It follows that the disk complex satisfies both the lower and upper bounds: Theorems 5.14 and 13.1. This can be restated as follows.
Theorem 19.9. There is a constant $C_0 = C_0(M,S)$ so that for any $c \geq C_0$ there is a constant $A$ with the following property. For any $D,E \in \mathcal{D}(M,S)$ we have

$$d_\mathcal{D}(D,E) = A \sum [d_X(D,E)]_c.$$  

The sum ranges over holes $X \subset S$ for the disk complex $\mathcal{D}(M,S)$. 

Remark 19.10. In the discussion of the shortcut axiom we had $X$ a compressible hole for $\mathcal{D}(M,S)$ and $Z \subset X$ an incompressible hole for $\mathcal{D}(M,X)$ but not a hole for $\mathcal{D}(M,S)$. In this situation Theorem 19.9 implies that the inclusion $\mathcal{D}(M,X) \to \mathcal{D}(M,S)$ is not a quasi-isometric embedding.

20. Hyperbolicity

The ideas in this section are related to the notion of “time-ordered domains” and to the hierarchy machine of [31]. See also Chapters 4 and 5 of Behrstock’s thesis [1]. As remarked above, we cannot use those tools directly as the hierarchy machine is too rigid to deal with the disk complex.

20.1. Hyperbolicity. We prove the following.

Theorem 20.2. Fix $G = G(S)$, a combinatorial complex. Suppose $G$ satisfies the axioms of Section 13. Then $G$ is Gromov hyperbolic.

As corollaries we have the following.

Theorem 20.3. The arc complex is Gromov hyperbolic. 

Theorem 20.4. The disk complex is Gromov hyperbolic.

We deduce Theorem 20.2 from the following.

Theorem 20.5. Fix $G$, a combinatorial complex. Suppose $G$ satisfies the axioms of Section 13. Then for all $A \geq 1$ there exists $\delta \geq 0$ with the following property. Suppose $T \subset G$ is a triangle of paths where the projection of any side of $T$ into any hole is an $A$–unparameterized quasi-geodesic. Then $T$ is $\delta$–slim.

Proof of Theorem 20.2. As laid out in Section 14 there is a uniform constant $A$ so that for any pair $\alpha, \beta \in G$ there is a recursively constructed path $P = \{\gamma_i\} \subset G$ so that

- for any hole $X$ for $G$, the projection $\pi_X(P)$ is an $A$–unparameterized quasi-geodesic and
- $|P| = A d_G(\alpha, \beta)$.  

So if $\alpha \cap \beta = \emptyset$, then $|P|$ is uniformly short. Also, by Theorem 20.5 triangles made of such paths are uniformly slim. Thus, by Theorem 3.15 the complex $G$ is Gromov hyperbolic.

The rest of this section is devoted to proving Theorem 20.5.

20.6. Index in a hole. For the following definitions, we assume that $\alpha$ and $\beta$ are fixed vertices of $G$. Suppose $a \in \pi_X(\alpha)$ and $b \in \pi_X(\beta)$. Let $k = [a,b]$ be any geodesic in $\mathcal{C}(X)$ connecting $a$ to $b$. Define $\rho_k : G \to k$ to be the relation $\pi_X|G : G \to \mathcal{C}(X)$ followed by taking closest points on $k$. By Lemmas 3.8 and 4.5 the diameter of $\rho_k(\gamma)$ is uniformly bounded. So we may simplify our formulas by treating $\rho_k$ as a function. Define $\text{index}_X : G \to \mathbb{N}$ to be the index in $X$: namely,

$$\text{index}_X(\sigma) = d_X(\alpha, \rho_k(\sigma)).$$
Remark 20.7. Suppose $k'$ is a different geodesic connecting $a' \in \pi_X(\alpha)$ to $b' \in \pi_X(\beta)$ and index'$_X$ is defined with respect to $k'$. Then

$$|\text{index}_X(\sigma) - \text{index}'_X(\sigma)| \leq 17\delta + 4$$

by Lemma 3.10 and Lemma 3.11. Thus, permitting a small additive error, the index depends only on $\alpha$ and $\beta$ and not on the choice of geodesic $k$. Henceforth we use the notation $k = [\pi_X(\alpha), \pi_X(\beta)]$ to denote any geodesic connecting a point of $\pi_X(\alpha)$ to a point of $\pi_X(\beta)$.

20.8. Back and sidetracking. Fix $\sigma, \tau \in \mathcal{G}$. We say that $\sigma$ precedes $\tau$ by at least $K$ in $X$ if

$$\text{index}_X(\sigma) + K \leq \text{index}_X(\tau).$$

We say that $\sigma$ precedes $\tau$ by at most $K$ if the inequality is reversed. If $\sigma$ precedes $\tau$, then we also say that $\tau$ succeeds $\sigma$.

Now take $\mathcal{P} = \{\sigma_i\}$ to be a path in $\mathcal{G}$ connecting $\alpha$ to $\beta$. Recall that we have made the simplifying assumption that $\sigma_i$ and $\sigma_{i+1}$ are disjoint.

We formalize two properties of unparameterized quasi-geodesics. The path $\mathcal{P}$ backtracks at most $K$ if for every hole $X$ and all indices $i < j$ we find that $\sigma_j$ precedes $\sigma_i$ by at most $K$. The path $\mathcal{P}$ sidetracks at most $K$ if for every hole $X$ and every index $i$ we find that

$$d_X(\sigma_i, \rho_k(\sigma_i)) \leq K,$$

where $k = [\pi_X(\alpha), \pi_X(\beta)]$.

Remark 20.9. Note that if $\mathcal{P}$ has bounded sidetracking, then one may freely use in calculations whichever of $\sigma_i$ or $\rho_k(\sigma_i)$ is more convenient.

20.10. Projection control. The following lemma, due to Behrstock [1, 4.2.1], is closely related to the notion of time-ordered domains [31, Section 4.6]. An elementary proof is given in [29, Lemma 2.5].

Lemma 20.11. There is a constant $M_1 = M_1(S)$ with the following property. Suppose $X, Y \subset S$ overlap. If $\gamma \in \mathcal{AC}(S)$ cuts both $X$ and $Y$, then either $d_X(\gamma, \partial Y) < M_1$ or $d_Y(\partial X, \gamma) < M_1$. \hfill \Box

We also require a more specialized version for the case where $X$ is nested in $Y$.

The proof is an exercise in the application of the bounded geodesic image theorem (Theorem 4.7). Recall that $M_0$ is the constant given in that theorem.

Lemma 20.12. Suppose $X \subset Y$ are nested non-simple domains. Fix $\alpha, \beta, \gamma \in \mathcal{AC}(S)$ that all cut $X$. Let $k = [\pi_Y(\alpha), \pi_Y(\beta)]$. Assume that $d_X(\alpha, \beta) \geq M_0$. We have

$$\text{index}_Y(\partial X) - 4 \leq \text{index}_Y(\gamma)$$

if $d_X(\alpha, \gamma) \geq M_0$ and

$$\text{index}_Y(\gamma) \leq \text{index}_Y(\partial X) + 4$$

if $d_X(\gamma, \beta) \geq M_0$. \hfill \Box
20.13. Finding the midpoint of a side. Fix $A \geq 1$. Let $P, Q, R$ be the sides of a triangle in $G$ with vertices at $\alpha, \beta, \gamma$. We assume that each of $P, Q$ and $R$ are $A$–unparameterized quasi-geodesics when projected to any hole.

Recall that $M_0 = M_0(S)$ and $M_1 = M_1(S)$ depend only on $\xi(S)$. We may assume that if $T \subset S$ is an essential subsurface, then $M_0(T) < M_0(S)$.

Choose $K_1 \geq \max\{M_0, 4M_1, 8\} + 8\delta$ sufficiently large so that any $A$–unparameterized quasi-geodesic in $C(X)$, for $X$ a hole, backtracks and sidetracks at most $K_1$.

Claim 20.14. If $\sigma_i$ precedes $\gamma$ in $X$ and $\sigma_j$ succeeds $\gamma$ in $Y$, both by at least $2K_1$, then $i < j$.

Proof. To begin, as $X$ and $Y$ are holes and all holes interfere, we need not consider the possibility that $X \cap Y = \emptyset$. If $X = Y$, we deduce that

$$\text{index}_X(\sigma_i) + 2K_1 \leq \text{index}_X(\gamma) \leq \text{index}_X(\sigma_j) - 2K_1.$$ 

Thus $$\text{index}_X(\sigma_i) + 4K_1 \leq \text{index}_X(\sigma_j).$$ Since $P$ backtracks at most $K_1$ we have $i < j$, as desired.

Suppose instead $X \subset Y$. Since $\sigma_i$ precedes $\gamma$ in $X$, we deduce that $d_X(\alpha, \beta) \geq 2K_1 \geq M_0$ and $d_X(\alpha, \gamma) \geq 2K_1 - 2\delta \geq M_0$. Apply Lemma 20.12 to deduce that $\text{index}_Y(\partial X) - 4 \leq \text{index}_Y(\gamma)$. Since $\sigma_j$ succeeds $\gamma$ in $Y$ it follows that $\text{index}_Y(\partial X) - 4 + 2K_1 \leq \text{index}_Y(\sigma_j)$. Again using the fact that $\sigma_i$ precedes $\gamma$ in $X$ we have that $d_X(\sigma_i, \beta) \geq M_0$. We deduce from Lemma 20.12 that $\text{index}_Y(\sigma_i) \leq \text{index}_Y(\partial X) + 4$. Thus

$$\text{index}_Y(\sigma_i) - 8 + 2K_1 \leq \text{index}_Y(\sigma_j).$$

Since $P$ backtracks at most $K_1$ in $Y$ we again deduce that $i < j$. The case where $Y \subset X$ is similar.

Suppose now that $X$ and $Y$ overlap. Applying Lemma 20.11 and breaking symmetry, we may assume that $d_X(\gamma, \partial Y) < M_1$. Since $\sigma_i$ precedes $\gamma$ we have $\text{index}_X(\gamma) \geq 2K_1$. Lemma 3.10 now implies that $\text{index}_X(\partial Y) \geq 2K_1 - M_1 - 6\delta$. Thus,

$$d_X(\alpha, \partial Y) \geq 2K_1 - M_1 - 8\delta \geq M_1,$$

where the first inequality follows from Lemma 3.7.

Applying Lemma 20.11 we deduce that $d_Y(\alpha, \partial X) < M_1$. Now, since $\sigma_j$ succeeds $\gamma$ in $Y$, we find that $\text{index}_Y(\sigma_j) \geq 2K_1$. So Lemma 3.7 implies $d_Y(\alpha, \sigma_j) \geq 2K_1 - 2\delta$. The triangle inequality now gives

$$d_Y(\partial X, \sigma_j) \geq 2K_1 - M_1 - 2\delta \geq M_1.$$ 

Applying Lemma 20.11 we deduce that $d_X(\partial Y, \sigma_j) < M_1$. Thus $d_X(\gamma, \sigma_j) \leq 2M_1$. Finally, Lemma 3.10 implies that the difference in index (in $X$) between $\sigma_i$ and $\sigma_j$ is at least $2K_1 - 2M_1 - 6\delta$. Since this is greater than the backtracking constant, $K_1$, it follows that $i < j$. \qed

Let $\sigma_\alpha \in P$ be the last vertex of $P$ with the following property. There exists a hole where $\sigma_\alpha$ precedes $\gamma$ by at least $2K_1$. If no such vertex of $P$ exists, then take $\sigma_\alpha = \alpha$.

Claim 20.15. For every hole $X$, if $h = [\pi_X(\alpha), \pi_X(\beta)]$, then

$$d_X(\sigma_\alpha, \rho_h(\gamma)) \leq 3K_1 + 6\delta + 3.$$
Proof. Since $\sigma_i$ and $\sigma_{i+1}$ are disjoint we have $d_X(\sigma_i, \sigma_{i+1}) \leq 3$ and so Lemma 3.10 implies that
\[|\text{index}_X(\sigma_{i+1}) - \text{index}_X(\sigma_i)| \leq 6\delta + 3.\]
Since $\mathcal{P}$ is a path connecting $\alpha$ to $\beta$ the image $\rho_h(\mathcal{P})$ is $(6\delta + 3)$–dense in $h$. Thus, if $\text{index}_X(\sigma_\alpha) + 2K_1 + 6\delta + 3 < \text{index}_X(\gamma)$, then we have a contradiction to the definition of $\sigma_\alpha$.

On the other hand, if $\text{index}_X(\sigma_\alpha) \geq \text{index}_X(\gamma) + 2K_1$, then $\sigma_\alpha$ succeeds $\gamma$ in $X$ and $\sigma_\alpha$ precedes $\gamma$ in some hole. This directly contradicts Claim 20.14.

We deduce that the difference in index between $\sigma_\alpha$ and $\gamma$ in $X$ is at most $2K_1 + 6\delta + 3$. Finally, as $\mathcal{P}$ sidetracks by at most $K_1$ we have
\[d_X(\sigma_\alpha, \rho_h(\gamma)) \leq 3K_1 + 6\delta + 3,\]
as desired. \hfill \Box

We define $\sigma_\beta$ to be the first $\sigma_i$ to succeed $\gamma$ by at least $2K_1$; if no such vertex of $\mathcal{P}$ exists take $\sigma_\beta = \beta$. If $\alpha = \beta$, then $\sigma_\alpha = \sigma_\beta$. Otherwise, from Claim 20.14 we immediately deduce that $\sigma_\alpha$ comes before $\sigma_\beta$ in $\mathcal{P}$. A symmetric version of Claim 20.15 applies to $\sigma_\beta$: for every hole $X$,
\[d_X(\rho_h(\gamma), \sigma_\beta) \leq 3K_1 + 6\delta + 3.\]

20.16. Another side of the triangle. Recall we are also given a path $\mathcal{R} = \{\tau_i\}$ connecting $\alpha$ to $\gamma$ in $\mathcal{G}$. As before, $\mathcal{R}$ has bounded backtracking and sidetracking. Thus we again find vertices $\tau_\alpha$ and $\tau_\gamma$, the last/first to precede/succeed $\beta$ by at least $2K_1$. This is defined in terms of the closest points projection of $\beta$ to the geodesic $h = [\pi_X(\alpha), \pi_X(\gamma)]$. By Claim 20.15 for every hole $X$ the vertices $\tau_\alpha$ and $\tau_\gamma$ are close to $\rho_h(\beta)$.

By Lemma 3.9 if $k = [\pi_X(\alpha), \pi_X(\beta)]$, then $d_X(\rho_k(\gamma), \rho_h(\beta)) \leq 6\delta$. We deduce the following.

Claim 20.17. $d_X(\sigma_\alpha, \tau_\alpha) \leq 6K_1 + 18\delta + 6$. \hfill \Box

This claim and Claim 20.15 imply that the body of the triangle $\mathcal{PQR}$ has bounded size. We now show that the legs are slim.

Claim 20.18. There is a constant $N_2 = N_2(S)$ with the following property. For every $\sigma_i \leq \sigma_\alpha$ in $\mathcal{P}$ there is a $\tau_j \leq \tau_\alpha$ in $\mathcal{R}$ so that
\[d_X(\sigma_i, \tau_j) \leq N_2\]
for every hole $X$.

Proof. We only sketch the proof, as the details are similar to our previous discussion. Fix $\sigma_i \leq \sigma_\alpha$.

Suppose first that no vertex of $\mathcal{R}$ precedes $\sigma_i$ by more than $2K_1$ in any hole. So fix a hole $X$ and geodesics $k = [\pi_X(\alpha), \pi_X(\beta)]$ and $h = [\pi_X(\alpha), \pi_X(\gamma)]$. Then $\rho_k(\sigma_i)$ is within distance $2K_1$ of $\pi_X(\alpha)$. Appealing to Claim 20.17 bounded sidetracking, and hyperbolicity of $C(X)$ we find that the initial segments $[\pi_X(\alpha), \rho_k(\sigma_\alpha)] \subset k$ and $[\pi_X(\alpha), \rho_h(\tau_\alpha)] \subset h$ must fellow travel. Because of bounded backtracking along $\mathcal{P}$, $\rho_k(\sigma_i)$ lies near this initial segment of $k$. Thus by Lemma 3.11 $\rho_h(\sigma_i)$ is close to $\rho_k(\sigma_i)$ which in turn is close to $\pi_X(\sigma_i)$, because $\mathcal{P}$ has bounded sidetracking. In short, $d_X(\alpha, \sigma_i)$ is bounded for all holes $X$. Thus we may take $\tau_j = \tau_0 = \alpha$ and we are done.
Now suppose that some vertex of $\mathcal{R}$ precedes $\sigma_i$ by at least $2K_1$ in some hole $X$. Take $\tau_j$ to be the last such vertex in $\mathcal{R}$. Following the proof of Claim 20.14 shows that $\tau_j$ comes before $\tau_0$ in $\mathcal{R}$. The argument now required to bound $d_X(\sigma_i, \tau_j)$ is essentially identical to the proof of Claim 20.15.

By the distance estimate, we find that there is a uniform neighborhood of $[\sigma_0, \sigma_\alpha] \subset \mathcal{P}$, taken in $\mathcal{G}$, which contains $[\tau_0, \tau_\alpha] \subset \mathcal{R}$. Thus $\mathcal{P}\mathcal{Q}\mathcal{R}$ is slim. This completes the proof of Theorem 20.5.

21. Coarsely computing Hempel distance

We now turn to our topological application. Recall that a Heegaard splitting is a triple $(S, V, W)$ consisting of a surface and two handlebodies where $V \cap W = \partial V = \partial W = S$. Hempel [25, page 632] defines the quantity

$$d_S(V, W) = \min \{d_S(D, E) \mid D \in \mathcal{D}(V), E \in \mathcal{D}(W)\}$$

and calls it the distance of the splitting. Note that a splitting can be completely determined by giving a pair of cut systems: simplices $\alpha \subset \mathcal{D}(V)$ and $\beta \subset \mathcal{D}(W)$ where the corresponding disks cut the containing handlebody into a single three-ball. The triple $(S, \alpha, \beta)$ is a Heegaard diagram. The goal of this section is to prove the following.

**Theorem 21.1.** There is a constant $R_1 = R_1(S)$ and an algorithm that, given a Heegaard diagram $(S, \alpha, \beta)$, computes a number $N$ so that

$$|d_S(V, W) - N| \leq R_1.$$ 

Let $\rho_V : \mathcal{C}(S) \to \mathcal{D}(V)$ be the closest points relation: so

$$\rho_V(\alpha) = \{D \in \mathcal{D}(V) \mid \text{for all } E \in \mathcal{D}(V), d_S(\alpha, D) \leq d_S(\alpha, E)\}.$$ 

It suffices to show the following.

**Theorem 21.2.** There is a constant $R_0 = R_0(V)$ and an algorithm that, given an essential curve $\alpha \subset S$ and a cut system $\alpha \subset \mathcal{D}(V)$, finds a disk $C \in \mathcal{D}(V)$ so that

$$d_S(C, \rho_V(\alpha)) \leq R_0.$$ 

**Proof of Theorem 21.1** Suppose $(S, \alpha, \beta)$ is a Heegaard diagram. Using Theorem 21.2 we find a disk $D$ within distance $R_0$ of $\rho_V(\beta)$. Again using Theorem 21.2 we find a disk $E$ within distance $R_0$ of $\rho_V(\alpha)$. Notice that $E$ is defined using $D$ and not the cut system $\alpha$.

Computing distance between fixed vertices in the curve complex is algorithmic [28, 46]; thus we may compute $d_S(D, E)$. By the hyperbolicity of $\mathcal{C}(S)$ (Theorem 3.5) and by the quasi-convexity of the disk set (Theorem 4.11) this is the desired estimate, $N$.

Very briefly, the algorithm asked for in Theorem 21.2 searches an $R_2$-neighborhood in $\mathcal{M}(S)$ about a splitting sequence from $\alpha$ to $\beta$. Here are the details.

**Algorithm 21.3.** We are given $\alpha \in \mathcal{C}(S)$ and a cut system $\alpha \subset \mathcal{D}(V)$. Make $\alpha$ and $\beta$ tight. Following [32, Section 4] there is a one-switch track $\tau$ in $S = \partial V$ obtained by collapsing $\alpha$. The cut system $\alpha$ is dual to $\tau$ and also crosses the sole switch of $\tau$. Now make $\tau$ a generic track by combing away from $\beta$ [38, Proposition 1.4.1]. Note that $\alpha$ is carried by $\tau$ and so gives a transverse measure $\omega$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Lemma 21.4. There is an algorithm that, given a cut system \(\mathbb{D} \subset V\) and a curve \(\gamma \subset S\), outputs a disk \(C \subset V\) so that \(\rho(\gamma, \partial C) = \min\{\rho(\gamma, \partial E) \mid E \in \mathbb{D}(V)\}\). \(\square\)

We now discuss the constant \(R_2\). Notice that the track \(\tau_p\) is birecurrent because \(\alpha\) is fully carried and \(\mathbb{D}\) is fully dual. Thus by Theorem [18, 7] and by Morse stability, for any essential \(Y \subset S\) there is a stability constant \(M_2\) for the path \(p \mapsto \pi_Y(\mu_p)\). Let \(\delta\) be the hyperbolicity constant for \(C(S)\) (Theorem 3.5) and let \(Q\) be the quasi-convexity constant for \(\mathbb{D}(V) \subset C(S)\) (Theorem 4.11).

Since \(\rho(\mathbb{D}, \mu_0)\) is bounded we will, at the cost of an additive error, identify their images in \(C(S)\). For the purposes of the proof, for every \(p \in [0, N]\), fix \(E_p \in \rho_Y(\mu_p)\). In particular, fix \(E_0\) inside of \(\mathbb{D}\). (Note that the disks \(E_p\) are not necessarily encountered during the running of Algorithm 21.3).

Lemma 21.5. There is a constant \(R_2\) with the following property. Suppose that \(n < m\), that \(d_S(\mu_n, E_n), d_S(\mu_m, E_m) \leq M_2 + \delta + Q\), and that \(d_S(\mu_n, \mu_m) \geq 2(M_2 + \delta + Q) + 5\). Then there is a marking \(\nu \in B(\mu_n, R_2)\) and a curve \(\gamma \in \nu\) so that either

- \(\gamma\) bounds a disk in \(V\),
- \(\gamma \subset \partial Z\), where \(Z\) is a non-hole, or
- \(\gamma \subset \partial Z\), where \(Z\) is a large hole.

Proof of Lemma 21.5. Choose points \(\sigma, \sigma'\) in the \(\epsilon\)-thick part of \(\mathcal{T}(S)\) so that all curves of \(\mu_n\) have bounded length in \(\sigma\) and so that \(E_n\) has length less than the Margulis constant in \(\sigma'\). As in Section 15 there is a Teichmüller geodesic and associated markings \(\{\nu_k\}_{k=0}^K\) so that \(d_M(\nu_0, \mu_n)\) is bounded and \(E_n \in \text{base}(\nu_K)\).

Claim. There is a constant \(R_3\) so that for any small hole \(X\) we have \(d_X(\mu_n, \nu_K) < R_3\).

Proof. If \(d_X(\mu_n, \nu_K) \leq M_0\), then we are done. If the distance is greater than \(M_0\), then Theorem 4.7 gives a vertex of the \(C(S)\)-geodesic connecting \(\mu_n\) to \(E_n\) with distance at most one from \(\partial X\). It follows from the triangle inequality that every vertex of the \(C(S)\)-geodesic connecting \(\mu_m\) to \(E_m\) cuts \(X\). Thus the bounded geodesic image theorem (Theorem 4.7) implies that

\[d_X(\mu_m, E_m) < M_0.\]

Note that \(d_X(\mu_0, E_0)\) is bounded by construction. Since \(X\) is a small hole the distance \(d_X(E_p, E_q)\) is uniformly bounded for any \(p, q \in [0, m]\). Since \(p \mapsto \pi_X(\mu_p)\) is an unparameterized quasi-geodesic we deduce that \(d_X(\mu_p, E_q)\) is uniformly bounded for all \(p, q \in [0, m]\).
Using the stability of $C$ with $n < m$ let $R$ of Theorem 4.13 provides the constant terms on the right-hand side of Theorem 4.13 vanish. Again the additive error distances in small holes. As before we find a sufficiently large cut-off so that all

$$d_{Z}(\mu, \nu_{\ell}) \geq R_{3}.$$  

None of these are small holes, by the claim above. If there are no such subsurfaces, then Theorem 4.13 bounds $d_{M}(\mu, \nu_{K})$: taking the cut-off constant larger than

$$\max\{R_{3}, C_{0}, M_{2} + \delta + Q\}$$

ensures that all terms on the right-hand side vanish. In this case the additive error in Theorem 4.13 is the desired constant $R_{2}$ and the lemma is proved.

If there are such subsurfaces, then choose one, say $Z$, that minimizes $\ell = \min J_{Z}$. Thus $d_{Z}(\mu, \nu_{\ell}) < C_{1}$ for all strict non-holes and all strict large holes. Since $d_{S}(\mu, E_{n}) \leq M_{2} + \delta + Q$ and $\{\nu_{m}\}$ is an unparameterized quasi-geodesic [41 Theorem 6.1] we find that $d_{S}(\mu, \nu_{\ell})$ is uniformly bounded. The claim above bounds distances in small holes. As before we find a sufficiently large cut-off so that all terms on the right-hand side of Theorem 4.13 vanish. Again the additive error of Theorem 4.13 provides the constant $R_{2}$. Since $\partial Z \subset \text{base}(\nu_{\ell})$ the proof of Lemma 21.5 is finished.

To prove the correctness of Algorithm 21.3 it suffices to show that the disk produced is close to $\rho_{\ell}(\alpha)$. Let $m$ be the largest index so that for all $p \leq m$ we have

$$d_{S}(\mu_{p}, E_{p}) \leq M_{2} + \delta + Q.$$  

Using the stability of $p \mapsto \pi_{S}(\mu_{p})$, the hyperbolicity of $C(S)$, and the quasi-convexity of $D(V)$, we deduce that $\mu_{m+1}$ lies within distance $M_{2} + \delta$ of some vertex $v \in [\alpha, \rho_{\ell}(\alpha)]$. The remark after Lemma 13.16 implies that $d_{S}(\mu_{p}, \mu_{p+1}) \leq C_{5}$ for all $p$. By the definition of $\rho_{\ell}$ we have $d_{S}(v, \rho_{\ell}(\alpha)) \leq d_{S}(v, E_{m})$; we deduce that

$$d_{S}(\mu, \rho_{\ell}(\alpha)) \leq 2C_{5} + 3M_{2} + 3\delta + Q.$$  

Let $n < m$ be the largest index so that

$$2(M_{2} + \delta + Q) + 5 \leq d_{S}(\mu, \mu_{n}) \leq 2(M_{2} + \delta + Q) + 5 + C_{5}.$$  

If no such $n$ exists, then take $n = 0$. Lemma 21.5 implies that there is a disk $C$ with $d_{S}(C, \mu_{n}) \leq C_{5}R_{2} + C_{5} + 2$ and this disk is found during the running of Algorithm 21.3. It follows from the above inequalities that

$$d_{S}(C, \alpha) \leq C_{5}R_{2} + 5M_{2} + 5\delta + 3Q + 7 + 4C_{5} + d_{S}(\rho_{\ell}(\alpha), \alpha).$$

So the disk $D$, output by the algorithm, is at least this close to $\alpha$ in $C(S)$. Using the triangle with vertices $\alpha, \rho_{\ell}(\alpha)$, and $D$ it is an exercise to show that

$$d_{S}(D, \rho_{\ell}(\alpha)) \leq C_{5}R_{2} + 5M_{2} + 9\delta + 5Q + 7 + 4C_{5}.$$  

This completes the proof of Theorem 21.2. □
ACKNOWLEDGMENTS

The authors thank Jason Behrstock, Brian Bowditch, Yair Minsky, Lee Mosher, Hossein Namazi, and Kasra Rafi for many enlightening conversations.

The authors thank Tao Li for pointing out that their original bound, inside of Theorem 12.1, can be reduced to a constant.

REFERENCES


Department of Mathematics, University of Chicago, Chicago, Illinois 60637
E-mail address: masur@math.uic.edu

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom
E-mail address: s.schleimer@warwick.ac.uk

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use