Big Heegaard Distance Implies Finite Mapping Class Group

Hossein Namazi [1]

Princeton University, Fine Hall, Washington RD, Princeton, NJ 08544-1000

Abstract

We show that if M is a closed three manifold with a Heegaard splitting with sufficiently big *Heegaard distance* then the subgroup of the mapping class group of the Heegaard surface, whose elements extend to both handlebodies is finite. As a corollary, this implies that under the same hypothesis, the mapping class group of M is finite.

Key words: Heegaard splitting, Heegaard distance, curve complex

1 Introduction

It is well known that any closed orientable 3-manifold is obtained by taking two copies of a handlebody and gluing them along the boundary. Such a decomposition is called a *Heegaard splitting*. An outstanding problem in studying 3-manifolds is to obtain information about the manifold from the gluing map. One important feature is that when the gluing map is "complex", we expect to get a "rigid structure" for the manifold.

The first result in this direction is perhaps Haken's lemma [6] which proves that if the gluing map does not take any *meridian* to another meridian then the manifold is irreducible. Here by a meridian we mean a homotopically non-trivial simple closed curve on the boundary that is homotopically trivial in the handlebody. On the other hand, Casson and Gordon [2] proved that if a splitting satisfies the above assumption but image of a meridian (by the gluing map) has zero intersection with another meridian then the manifold is

Email address: `hossein@math.princeton.edu` (Hossein Namazi).

[1] Part of this work was done while the author was partially supported by a Liftoff Fellowship from the Clay Mathematics Institute.

Haken. Hempel [7] generalized this and defined a *Heegaard distance* for a Heegaard splitting. In this definition we can restate Haken's lemma by saying that nonzero Heegaard distance implies the 3-manifold is irreducible. Also Casson-Gordon's result shows a 3-manifold with a distance one Heegaard splitting is Haken. Hempel [7] conjectured that if this number is bigger than two or at least is sufficiently large, then the manifold is hyperbolic. As a matter of fact assuming the Geometrization Conjecture for 3-manifolds, this follows from known results about Heegaard distances of splittings in toroidal and Seifert fibered 3-manifolds.

However one expects to be able to use this combinatorial property of a Heegaard splitting and find a topological approach to solve this problem. One approach to understand consequences of large Heegaard distance is to show that manifolds which admit a splitting with sufficiently large Heegaard distance share many of the qualities of hyperbolic 3-manifolds.

In this article, we look at the splitting more from the point of view of the *curve complex* of the Heegaard surface and we use results of Masur-Minsky [13–15] about the coarse geometry of this complex. As a straightforward application of this machinery and provided that the Heegaard distance is sufficiently large, we prove that there are only finitely many isotopy classes of the homeomorphisms of the 3-manifold which preserve the splitting (up to isotopy).

In an application of this result, we use it to show that the mapping class group of the 3-manifold, the group of isotopy classes of self-homeomorphisms, is finite. This uses finiteness of isotopy classes of Heegaard surfaces of the same genus in an atoroidal 3-manifold. This is the content of Waldhausen's conjecture which has been proved recently by Tao Li [11].

We should remark that when M admits a hyperbolic metric, an easy application of Mostow Rigidity Theorem shows that the group of homotopy classes of self-homeomorphisms of M is finite. This group is a quotient of the mapping class group of M. Still, work of Gabai-Meyerhoff-Thurston [5] proves that for closed hyperbolic 3-manifolds the mapping class group is also finite.

Let M be a closed orientable 3-manifold with a Heegaard splitting. In other terms, $M = H^+ \cup_S H^-$ is the union of two handlebodies H^+ and H^- with the same genus, glued along their boundaries. We call $S = \partial H^+ = \partial H^-$ a *Heegaard surface* and it is defined up to isotopy.

To define the Heegaard distance for the splitting, we need to define the curve complex of S, $\mathcal{C}(S)$. This is a locally infinite simplicial complex, whose vertices are homotopy classes of homotopically nontrivial simple closed curves and its k-simplices are $(k+1)$-tuples $[\alpha_0, \ldots, \alpha_k]$ of distinct vertices that have pairwise disjoint representatives on S. The curve complex is also equipped with a path metric $d_{\mathcal{C}}$ on its one-skeleton. This metric makes every edge isometric to the unit interval and the distance between every two points is the length of the shortest path between them.

The set of meridians for each handlebody H^+ and H^- is a subset of $\mathcal{C}(S)$. We call them Δ^+ and Δ^- respectively. The *Heegaard distance* for the splitting is $d_{\mathcal{C}}(\Delta^+, \Delta^-)$, their distance in the curve complex of S.

As usual, for a surface of finite type S, its mapping class group $\mathcal{MCG}(S)$ denotes the set of isotopy classes of orientation preserving self-homemorphisms of S which fix every component of the boundary. We call each element of $\mathcal{MCG}(S)$ an automorphism of S and when there is no confusion we feel free to go back and forward between this element and a representative of this isotopy class. We also will need to consider the mapping class group of M which is the set of isotopy classes of the orientation preserving homeomorphisms of the 3-manifold M to itself and is denoted by $\mathcal{MCG}(M)$.

We say a self-homeomorphism of the boundary of a handlebody H *extends* to the handlebody if it is the restriction to ∂H of an orientation preserving self-homeomorphism of the handlebody. Suppose ϕ and ψ are two isotopic homeomorphisms of ∂H to itself which extend to H. It is not hard to see that the extensions of ϕ and ψ are isotopic in H too. To see this, take a system of compressing disks which cuts the handlebody to a ball. Since the handlebody is irreducible, we can see that extensions of ϕ and ψ to this system of disks are isotopic. After applying this isotopy, we can assume these extensions are identical on ∂H and on this system of disks. Now cut H along the system of disks; what remains is a 3-ball and two self-homeomorphisms of the 3-ball which are identical on its boundary. Any two such maps are isotopic in the interior of the ball. Thus we can extend the isotopy to the ball and therefore to the entire handlebody. Using this observation, we say an element of $\mathcal{MCG}(\partial H)$ *extends* to H if any of its representatives (and therefore all of them) extend.

Now for the Heegaard splitting $M = H^+ \cup_S H^-$, consider the subgroup of $\mathcal{MCG}(S)$ whose elements extend to both H^+ and H^-. We denote this subgroup by $\Gamma(H^+, H^-)$. The above argument shows that there is a well defined map

$$\Phi : \Gamma(H^+, H^-) \to \mathcal{MCG}(M)$$

obtained by extending the automorphism of the surface to both handlebodies.

Theorem 1.1 *If M is a closed 3-manifold which admits a Heegaard splitting $M = H^+ \cup_S H^-$ with Heegaard distance bigger than n_g then $\Gamma(H^+, H^-)$, the subgroup of $\mathcal{MCG}(S)$ whose elements extend to both handlebodies, is finite.*

Here $n_g = 2K + 2\delta$ depends only on the genus of the splitting and δ and K are described in 2.

Thompson [18] and Hempel [7] showed that when M is toroidal then any Heegaard splitting of M has distance at most 2. Therefore, the manifolds satisfying the hypothesis of the above theorem are all atoroidal. On the other hand, Li [11] has proved Waldhausen's conjecture which states that an atoroidal 3-manifold has at most a finite number of splittings of the same genus up to isotopy. Using these results and as a corollary of the above theorem we have:

Corollary 1.2 *If a closed 3-manifold admits a Heegaard splitting of genus g and Heegaard distance bigger than n_g then the mapping class group of the manifold is finite.*

Proof. Let $M = H^+ \cup_S H^-$ be such a splitting. Image of S by a homeomorphism from M to itself is a Heegaard splitting. If these two homeomorphisms are isotopic, the images will be the same splittings since a splitting is defined only up to isotopy.

Once we know that the number of all such splittings (up to isotopy) is finite, we see that the subgroup of $\mathcal{MCG}(M)$ which preserves S (up to isotopy) has finite index. To prove our claim we need to show this subgroup is finite. After changing a representative of an element of this subgroup by an isotopy we can assume that it takes S to itself. By possibly taking another index two subgroup, we can assume that the induced map on S is orientation preserving and therefore has to preserve H^+ and H^- as well. We say these elements *preserve the splitting.*

Now look at the map $\Phi : \Gamma(H^+, H^-) \to \mathcal{MCG}(M)$ defined above. One can see that the image is the subgroup of $\mathcal{MCG}(M)$ consisting of all elements that preserve the splitting. To see this take a representative of an element of $\mathcal{MCG}(M)$ that preserves the splitting. After changing it by an isotopy, we can assume that it preserves S, thus induces an automorphism of the surface and an element of $\Gamma(H^+, H^-)$.

The Main Theorem above shows that if the Heegaard distance is large then $\Gamma(H^+, H^-)$ is finite and so is its image by Φ. This proves the finiteness of $\mathcal{MCG}(M)$. $\square$

In the above proof, we showed that Φ is a surjective map from $\Gamma(H^+, H^-)$ to the subgroup of $\mathcal{MCG}(M)$ whose elements preserve the splitting. In general, this map does not need to be injective. In fact we can prove the following proposition. Recall that a splitting is *reducible* when the Heegaard distance is zero, is *weakly reducible* when the Heegaard distance is one and is *strongly irreducible* otherwise.

Proposition 1 *The map*

$$\Phi : \Gamma(H^+, H^-) \to \mathcal{MCG}(M)$$

is not injective if the splitting is reducible or weakly reducible.

Proof. We treat these two cases separately but in the same spirit. When we have a reducible splitting $M = H^+ \cup_S H^-$, there exists a homotopically nontrivial simple closed curve α on S that bounds a disk $D^+ \subset H^+$ and a disk $D^- \subset H^-$. We can assume that these disks are embedded and of course their union gives an embedded sphere in M. Now take a Dehn twist along α. It is clear that it is a nontrivial element of $\Gamma(H^+, H^-)$ and generates an infinite cyclic subgroup, but its image in $\mathcal{MCG}(M)$ is a Dehn twist along an embedded sphere. In the case that M is irreducible, this sphere bounds a ball and it is not hard to see that this mapping class is isotopic to identity. Even when M is reducible and the obtained sphere is homotopically nontrivial, a standard argument using the fact that $\pi_1(\mathrm{SO}_3(\mathbb{R})) = \mathbb{Z}_2$ shows that the square of the Dehn twist about the sphere is isotopic to the identity. This shows that the kernel of Φ is infinite.

On the other hand let $M = H^+ \cup_S H^-$ be weakly reducible. In particular, we assume that M is irreducible; otherwise by Haken's lemma [6] the splitting would be reducible. There exists α and β disjoint homotopically nontrivial and nonparallel simple closed curves on S such that α bounds a disk D^+ in H^+ and β bounds a disk D^- in H^-. Also take γ to be a band sum of α and β in S. (Connect α and β with an arbitrary simple arc k. Take a regular neighborhood of $\alpha \cup \beta \cup k$; it is a 3-holded sphere and γ is the component of the boundary of this 3-holed sphere which is not homotopic to either α or β.)

Then β and γ are disjoint and homotopic in H^+, therefore there exists a properly embedded annulus $A^+ \subset H^+$ that connects them, i.e. its boundary is $\beta \cup \gamma$. In the same way, there exists a properly embedded annulus $A^- \subset H^-$ with boundary $\alpha \cup \gamma$. Using standard cut and paste arguments, we can assume that A^+ and D^+ are disjoint and also A^- and D^- are disjoint. This implies that $D^+ \cup A^- \cup A^+ \cup D^-$ is an embedded sphere in M.

Equip S with a fixed orientation and let

$$\psi := D_\alpha D_\beta D_\gamma^{-1},$$

be a product of Dehn twists, where D_δ represents a positive Dehn twist about the simple closed curve δ. We can see that on one hand ψ extends to a Dehn twist about D^+ and a Dehn twist about A^+ in H^+ and on the other hand it extends to a Dehn twist about D^- and a Dehn twists about A^- in H^-. This implies that $\psi \in \Gamma(H^+, H^-)$; but the extension to M is a Dehn twist about the embedded sphere $D^+ \cup A^- \cup A^+ \cup D^-$. Since M is irreducible, this sphere bounds a ball and the extension of ψ to M is isotopic to the identity. □

We should point out some new developments in the study of topological properties of the Heegaard distance. Scharlemann and Tomova [17] have proved that if the Heegaard distance of a genus g Heegaard splitting is greater than $2g$ then (up to isotopy) this is the only genus g Heegaard splitting of the 3-manifold. In particular this shows that in this case the homomorphism Φ, defined above, is either surjective or its image has index two in $\mathcal{MCG}(M)$. (It might be an index two subgroup because there may exist elements of $\mathcal{MCG}(M)$ that interchange H^+ and H^-.) Also a recent work of Johnson and Rubinstein [9] is devoted to the study the map Φ. In particular, they show that if M is irreducible and atoroidal and the kernel of Φ contains a finite order element then M is Seifert fibered. As a corollary of these two results and our main theorem one can immediately see:

Corollary 1.3 *For a genus g Heegaard splitting with Heegaard distance bigger than* $\max\{n_g, 2g\}$*, the map* Φ *is an isomorphism either to* $\mathcal{MCG}(M)$ *or to an index two subgroup of* $\mathcal{MCG}(M)$*.*

In [9], the authors show that if the Heegaard distance is bigger than 4 then the kernel of Φ consists only of pseudo-Anosovs. They also classify all the cases where the distance is bigger than 2 and the kernel of Φ contains reducible elements. Furthermore they implicitly describe possible examples where the distance is 4 and the kernel of Φ is non-trivial. What remains mysterious is the presence of pseudo-Anosovs in the kernel of Φ or even in $\Gamma(H^+, H^-)$. We should remind the reader that Casson and Long [3] found an algorithm to determine if a pseudo-Anosov map on the boundary of a handlebody extends to the entire handlebody. Later Long found an example of an irreducible Heegaard splitting with a pseudo-Anosov automorphism of the Heegaard surface which extends to a self-homeomorphism of the 3-manifold.

Finally we should mention that Lustig and Moriah [12] have proved a result similar to our results by assuming that the splitting has a property which they call *double rectangle condition*. However their methods are very different and the relationship between their double rectangle condition and the Heegaard distance is not known.

To prove the main theorem we use Thurston's classification of automorphisms of surfaces for elements of $\Gamma(H^+, H^-)$. When the Heegaard distance is large we show that an element of $\Gamma(H^+, H^-)$ cannot be a pseudo-Anosov or a non-periodic reducible mapping class of S. The proof of this fact uses various properties of the curve complex and its subcomplexes Δ^+ and Δ^- and the fact that any $f \in \Gamma(H^+, H^-)$ preserves Δ^+ and Δ^-. When f is not periodic we use the iterations of the action of f on $\mathcal{C}(S)$ and find an upper bound for the curve complex distance between Δ^+ and Δ^- that depends only on the topology of S. This argument shows that if the Heegaard distance is larger than this upper bound then every element of $\Gamma(H^+, H^-)$ is periodic and has finite order. Then we invoke a theorem of Serre to show that $\mathcal{MCG}(S)$ has finite index torsion free subgroups and therefore every subgroup of $\mathcal{MCG}(S)$

has a finite index torsion free subgroup. In particular every torsion subgroup of $\mathcal{MCG}(S)$ is finite.

Note that when S is a torus, it is well known that M is S^3, $S^2 \times S^1$ or a lens space. All these cases are well understood in particular for a genus one Heegaard splitting $M = H^+ \cup_S H^-$, it is easy to see that $\Gamma(H^+, H^-) = \{1\}$ when M is not $S^2 \times S^1$ and $\Gamma(H^+, H^-) = \mathbb{Z}$ when $M \simeq S^2 \times S^1$. So we always assume that the genus of the splitting is at least two.

I want to thank Yair Minsky for all his help as a mentor, for sharing his enlightening ideas and also for laying the foundations on which the research in this paper is built. I want also to thank Andrew Casson and Saul Schleimer for interesting conversations and comments. I am also thankful to the referee for very useful comments.

2 Preliminaries

In [13], [14] and [15], Masur and Minsky studied various properties of the curve complex. The contents of this section are a few of those which we will use in the course of our proof.

In the introduction, we defined the curve complex for closed surfaces of genus at least two. We extend the definition for any finite type surface $S_{g,b}$, the compact orientable surface of genus g with b boundary components. In each case, it is a simplicial complex with a path metric. We refer to the set of vertices of $\mathcal{C}(S)$ and its one-skeleton by $\mathcal{C}_0(S)$ and $\mathcal{C}_1(S)$ respectively.

Recall that on a surface $S = S_{g,b}$ a closed curve is *essential* if it is not homotopically trivial and an essential simple closed curve is *non-peripheral* if it cannot be homotoped into ∂S. A properly embedded arc is *essential* if it cannot be homotoped (rel. ∂S) into ∂S. When $3g + b \geq 4$, we take homotopy classes of non-peripheral simple closed curves and homotopy classes of properly embedded essential arcs (rel. ∂S) as vertices of $\mathcal{C}(S)$. A $(k+1)$-tuple of different vertices makes a k-simplex if they have mutually disjoint representatives on the surface. The only other case we need is the case of a closed annulus, for which we give the definition below. In all other cases and in particular when $S = S_{0,3}$ is a three holed sphere, we define the curve complex to be empty.

Definition of the curve complex for a compact annulus. Consider a compact annulus A with its two boundaries. The vertices of $\mathcal{C}(A)$ are the homotopy classes of arcs connecting these two boundaries relative to their end points. This of course will be an uncountable set of vertices; we connect two vertices with an edge when they have representatives with disjoint interiors.

The metric $d_{\mathcal{C}}$ on the one-skeleton of the curve complex is defined as before: each edge has length one and the distance between any two points in the one-skeleton is the length of the shortest path connecting them. For a compact annulus, it is not hard to see that the complex with its metric is *quasi-isometric* to $\mathbb{Z}$ with its word metric [14].

We should remark that our definition of the curve complex is a little bit different from the usual definition (cf. [13]) but one can easily see that the two definitions are naturally quasi-isometric.

From now on, we only consider compact connected orientable surfaces of finite type $S = S_{g,b}$. An *essential* subsurface of S is a subsurface Y that is injective in the level of fundamental groups and if Y is an annulus then it cannot be homotoped into ∂S. Whenever we use the word subsurface we mean an essential connected subsurface.

We say two elements $\alpha, \beta \in \mathcal{C}_0(S)$ intersect *essentially* if $d_{\mathcal{C}}(\alpha, \beta) \geq 2$ and we say $\alpha \in \mathcal{C}_0(S)$ intersects a subsurface Y *essentially* if either α represents an element of $\mathcal{C}_0(Y)$ or it essentially intersects a boundary component of Y.

By Thurston's classification of surface automorphisms, elements of $\mathcal{MCG}(S)$ are periodic, reducible or pseudo-Anosov. Periodic automorphisms have finite order and reducible automorphisms preserve a set of homotopy classes of disjoint non-peripheral simple closed curves. Finally if $\phi \in \mathcal{MCG}(S)$ is a pseudo-Anosov it has stable and unstable laminations λ^+ and λ^-. These are two transversally measured laminations which have nonzero intersection and each of them *fills* S: every measured lamination with zero intersection with λ^+ (resp. λ^-) has to have the same support as λ^+ (resp. λ^-). Also for any non-peripheral simple closed curve α, the sequence $(\phi^n(\alpha))$ converges to λ^+ as $n \to \infty$ and converges to λ^- as $n \to -\infty$, in the space of projectivized measured laminations $\mathcal{PML}(S)$. For an exposition of Thurston's theorem and more on the spaces of laminations see [1] and [4].

The following lemma shows that under iterations of action of a pseudo-Anosov an element moves further and further in the curve complex. In particular the curve complex has infinite diameter. The argument in the context of the curve complex was described by Luo who attributes the main idea in the argument to Kobayashi's paper [10].

Lemma 2.1 *Let $S = S_{g,b}$ $(3g + b \geq 4)$ be any surface and ϕ any pseudo-Anosov automorphism of S. For any $\alpha \in \mathcal{C}_0(S)$ the distance $d_{\mathcal{C}(S)}(\alpha, \phi^n(\alpha))$ goes to infinity as $n \to \infty$. In particular $\mathcal{C}_1(S)$ equipped with its path metric has infinite diameter.*

Proof. Suppose $d_{\mathcal{C}}(\alpha, \phi^{n_k}(\alpha)) \leq d$ for a subsequence of the sequence $(\phi^n(\alpha))$. Hence for every k there is a path of length d,

$$\alpha = \gamma_0^k, \gamma_1^k, \cdots, \gamma_{d-1}^k, \gamma_d^k = \phi^{n_k}(\alpha)$$

in $\mathcal{C}_0(S)$ such that $i(\gamma_{i-1}^k, \gamma_i^k) = 0$ for $1 \leq i \leq d$, where $i(\cdot, \cdot)$ represents the topological intersection. After passing to a subsequence, which we still call (n_k), we can assume that for every $0 \leq i \leq d$ the sequence (γ_i^k) converges to an element λ_i in $\mathcal{PML}(S)$. Obviously λ_0 is the element of $\mathcal{PML}$ that is supported on α and $\lambda_d = \lim_k \phi^{n_k}(\alpha) = \lambda^+$ is the stable lamination of the pseudo-Anosov ϕ.

Because $i(\gamma_{i-1}^k, \gamma_i^k) = 0$ for every k and every $1 \leq i \leq d$, by continuity of the intersection number in $\mathcal{PML}(S)$ we have $i(\lambda_{i-1}, \lambda_i) = 0$ for every $1 \leq i \leq d$. However λ^+ fills the surface S and since $i(\lambda_{d-1}, \lambda^+) = 0$, λ_{d-1} has to have the same support as λ^+ and fills S. The same argument shows that all the laminations $\lambda_{d-2}, \ldots, \lambda_1, \lambda_0$ fill, but λ_0 is supported on α and does not fill S. This is a contradiction and we have proved the lemma. $\square$

A significant achievement of Masur-Minsky in the study of the curve complex was the next theorem which states that $\mathcal{C}_1(S)$ with its metric is *δ-hyperbolic in sense of Gromov*. This means that in any geodesic triangle, each side is in a δ neighborhood of the other two.

Theorem 2.2 (Hyperbolicity, Masur-Minsky [13]) *Let S be a surface of finite type with negative Euler Characteristic and $\mathcal{C}(S)$ its curve complex. Then $\mathcal{C}_1(S)$ with its path metric is δ-hyperbolic in sense of Gromov.*

Let Y be an essential subsurface of S. Following Masur-Minsky [14], we define a projection π_Y from $\mathcal{C}_0(S)$ to subsets of $\mathcal{C}_0(Y)$ with diameter at most one. Suppose $\alpha \in \mathcal{C}_0(S)$ is given and is in minimal position with respect to Y. If α does not intersect Y or Y is a three-holed sphere, we define $\pi_Y(\alpha) = \emptyset$. If not we have two cases; first assume Y is non-annular. We define $\pi_Y(\alpha)$ to be the set of essential arcs and non-peripheral simple closed curves in $\alpha \cap Y$. This obviously has diameter ≤ 1.

If Y is an annulus then we need to make the definition independent of the choice among the annuli that are isotopic to Y. In order to do this equip S with a hyperbolic metric and identify the universal cover of S with $\mathbb{H}^2$. The universal cover has a compactification as a closed disk and the action of $\pi_1(S)$ extends to an action on this compactification. Take the annular cover $\widetilde{Y} = \mathbb{H}^2/\pi_1(Y)$ of S to which Y lifts homeomorphically. Note that $\pi_1(Y)$ is a cyclic subgroup of isometries of $\mathbb{H}^2$ with two fixed points at infinity. The quotient (by $\pi_1(Y)$) of the disk minus these two points is a closed annulus $\widehat{Y}$ which naturally compactifies $\widetilde{Y}$. We identify $\mathcal{C}(Y)$ with $\mathcal{C}(\widehat{Y})$ and we use $\mathcal{C}(\widehat{Y})$ to make the definition. The closures of all lifts of α to $\widehat{Y}$ provide properly embedded arcs. We define $\pi_Y(\alpha)$ to be the set of those arcs which connect the two boundary components of $\widehat{Y}$. Since α interesects Y essentially, this set cannot be empty and obviously its diameter is at most 1.

When $\pi_Y(\alpha)$ and $\pi_Y(\beta)$ are nonempty, we also define $d_Y(\alpha,\beta)$ to be the distance between $\pi_Y(\alpha)$ and $\pi_Y(\beta)$ in $\mathcal{C}(Y)$ and if either of them is empty we define $d_Y(\alpha,\beta)=\infty$.

We remark that if Y is a subsurface, any $f\in\mathcal{MCG}(S)$ acts by an isomorphism $f:\mathcal{C}(Y)\to\mathcal{C}(f(Y))$, and this fits naturally with projections via $\pi_{f(Y)}\circ f = f\circ\pi_Y$ and

$$d_Y(\alpha,\beta)=d_{f(Y)}(f(\alpha),f(\beta)). \tag{1}$$

In particular, if Y is preserved by f (up to isotopy)

$$\pi_Y\circ f = f|_Y\circ\pi_Y, \tag{2}$$

where $f|_Y$ denotes the induced action on $\mathcal{C}(Y)$. For the annular subsurfaces, we need the following lemma in parallel to lemma 2.1.

Lemma 2.3 *Let $Z\subset S$ be an essential subsurface and β a non-peripheral simple closed curve in Z. Also assume $\phi\in\mathcal{MCG}(S)$ preserves Z up to isotopy and the induced automorphism on Z is isotopic to a nonzero power of a Dehn twist along β. If $\alpha\in\mathcal{C}_0(S)$ intersects β essentially then $d_Y(\alpha,\phi^n(\alpha))$ goes to infinity as $n\to\infty$, where Y is an annular neighborhood of β.*

Proof. Suppose D_β represents the positive Dehn twist about β. For a and b distinct points in $\mathcal{C}_0(Y)$, it is not hard to see that

$$d_Y(a,b)=1+|a\cdot b|$$

where $a\cdot b$ denotes the algebraic intersection number of the interiors of a and b (cf. [14]). Using this one can see that

$$d_Y(D_\beta^n(\alpha),\alpha)\le 2+|n|,$$

for any α that intersects Y essentially. (In [14] it is claimed that the above equation is an equality when $n\neq 0$, however it is not correct in general and becomes strictly smaller when α meets β two times in different directions.)

After an isotopy, we may assume that ϕ fixes ∂Z. Also, by assumption, we have that $\phi|_Z$ is isotopic to $\psi|_Z$, a nonzero power of D_β and in particular $\phi^n(\alpha)\cap Z$ is isotopic to $\psi^n(\alpha)\cap Z$ relative to ∂Z. The above argument shows that

$$d_Y(\alpha,\psi^n(\alpha))\to\infty$$

as $n\to\infty$ and therefore we just need to show that $d_Y(\phi^n(\alpha),\psi^n(\alpha))$ is bounded independent of n.

More generally, we show that if γ and δ are representatives of arbitrary elements of $\mathcal{C}_0(S)$, put in minimal position with respect to Z, and $\gamma \cap Z$ is isotopic to $\delta \cap Z$ (rel. ∂Z) then $d_Y(\gamma, \delta) \leq 2$. If a component of $\gamma \cap Z$ which cuts β is a simple closed curve μ then μ is also isotopic to a component of $\delta \cap Z$.

If there is no such component, $\gamma \cap Z$ has a component κ that is a properly embedded arc in Z and intersects β essentially. This arc is isotopic (rel. ∂Z) to a component κ' of $\delta \cap Z$. If we take the boundary of a regular neighborhood of the union of κ and the adjacent components of ∂Z we obtain at least one non-peripheral simple closed curve μ in Z that essentially intersects β. So in either case we obtain a simple closed curve μ on Z that intersects β essentially and does not intersect (essentially) a component κ of $\gamma \cap Z$ and a component κ' of $\delta \cap Z$.

As in the definition of π_Y take $\widehat{Y}$ to be the compactification of the annular cover of S associated to $\pi_1(Y)$. Let $\widehat{\beta}$ be the lift of β which is the core of $\widehat{Y}$. There are lifts $\widehat{\gamma}$ and $\widehat{\delta}$ of γ and δ where lifts of κ and κ' intersect $\widehat{\beta}$. If $\widehat{\mu}$ is a lift of μ that intersects $\widehat{\beta}$ essentially, we can see that $\widehat{\gamma}$ and $\widehat{\mu}$ are disjoint. Similarly $\widehat{\delta}$ and $\widehat{\mu}$ are disjoint. This immediately shows that

$$d_Y(\gamma, \delta) \leq d_Y(\gamma, \widehat{\mu}) + d_Y(\widehat{\mu}, \delta) \leq 2$$

and we have proved the lemma. $\square$

The next theorem shows the significance of having projections which are far from each other. It shows that if α and β are in $\mathcal{C}_0(S)$ and have very far projections in $\mathcal{C}(Y)$, for a proper subsurface $Y \subset S$, then the geodesic connecting them in $\mathcal{C}(S)$ has distance at most one from ∂Y.

Theorem 2.4 (Bounded Geodesic Image, Masur-Minsky [14]) *Let Y be a proper subsurface of S which is not a three holed sphere and let g be a geodesic segment, ray or bi-infinite line in $\mathcal{C}(S)$ such that $\pi_Y(v) \neq \emptyset$ for every vertex of g.*

There is a constant L only depending on the Euler characteristic of Y, so that

$$\mathrm{diam}_Y(g) \leq L.$$

In a geodesic metric space, a *K-quasiconvex* subset is a subset such that any geodesic connecting two of its points is contained in the K-neighborhood of that subset. Recall from the introduction that when H is a handlebody, $\Delta(H)$ denotes the set of meridians of boundary of H.

Theorem 2.5 (Quasiconvexity, Masur-Minsky [15]) *If H is a handlebody with boundary S then $\Delta(H)$ is a K-quasiconvex subset of $\mathcal{C}(S)$, where K depends only on the genus of S.*

3 Proof of Theorem 1.1

As in the introduction, suppose $M = H^+ \cup_S H^-$ is a Heegaard splitting of genus g and Heegaard distance bigger than $n_g = 2K + 2\delta$, where K and δ are the constants obtained respectively in the Quasiconvexity Theorem 2.5 and in the Hyperbolicity Theorem 2.2. We always assume that $\delta \geq 2$.

Claim 2 *If $d_{\mathcal{C}}(\Delta^+, \Delta^-) > 2K + 2\delta$ then elements of $\Gamma(H^+, H^-)$ are all periodic.*

Proof. Take $\phi \in \Gamma(H^+, H^-) \leq \mathcal{MCG}(S)$ that extends to both handlebodies. Since it extends, it has to preserve the set of compressible curves for each handlebody. Thus, if we consider it as an isometry of the curve complex of S, it preserves the sub-complexes Δ^+ and Δ^-. This is all we will use to show that ϕ is periodic. Using Thurston's classification of elements of $\mathcal{MCG}(S)$, ϕ is either periodic, reducible or pseudo-Anosov and we only need to rule out pseudo-Anosov maps and nonperiodic reducibles.

• *ϕ is not a pseudo-Anosov.* Assume it is a pseudo-Anosov. Consider $\alpha \in \Delta^+$ and $\beta \in \Delta^-$ fixed. Also consider a geodesic segment l between α and β of length $m = d_{\mathcal{C}}(\alpha, \beta)$. Because of lemma 2.1,

$$d_{\mathcal{C}}(\alpha, \phi^n(\alpha)) \to \infty \text{ and } d_{\mathcal{C}}(\beta, \phi^n(\beta)) \to \infty$$

as $n \to \infty$.

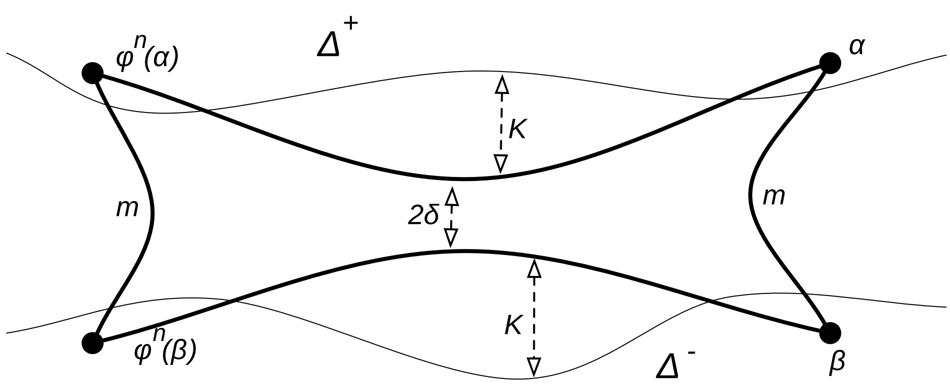

Fig. 1. Pseudo-Anosov ϕ.

Now, if we look at the rectangle with vertices α, β, $\phi^n(\alpha)$ and $\phi^n(\beta)$, it has two sides $[\alpha, \beta]$ and $[\phi^n(\alpha), \phi^n(\beta)]$ with fixed length m and the lengths of the other two sides tend to infinity as n goes to infinity. Using the Hyperbolicity

Theorem 2.2, it is easy to see that in this situation the longer sides get 2δ close.

On the other hand, α and $\phi^n(\alpha)$ are both in Δ^+ (Δ^+ is invariant under ϕ), and because Δ^+ is K-quasiconvex, Theorem 2.5, the geodesic connecting them is within at most K from Δ^+. For the same reason, the opposite side, $[\beta, \phi^n(\beta)]$, is also within K from Δ^-. So $[\alpha, \phi^n(\alpha)]$ is in the K-neighborhood of Δ^+ and $[\beta, \phi^n(\beta)]$ is in the K-neighborhood of Δ^- and we showed that they are 2δ close. Hence Δ^+ and Δ^- cannot be more than $2K + 2\delta$ apart (Figure 1). This contradicts the assumption about the Heegaard distance.

• *ϕ is not a nonperiodic reducible.* We say that an essential simple closed cure γ on S is a *reducing curve* for ϕ, if the set $\{\phi^n(\gamma)\}_{n\in\mathbb{Z}}$ consists of only a finite set of disjoint homotopy classes of curves on S. Note that these homotopy classes do not have to be disjoint on S.

We claim the following lemma:

Lemma 3.1 *Let ϕ be non-periodic element of $\mathcal{MCG}(S)$, γ a reducing curve for ϕ and $\alpha \in \mathcal{C}_0(S)$ arbitrary. Then for sufficiently big n, any geodesic connecting α and $\phi^n(\alpha)$ in $\mathcal{C}(S)$ has distance at most 2 from γ.*

Assume this Lemma is proved and $\phi \in \mathcal{MCG}$ is reducible, non-periodic and preserves both Δ^+ and Δ^-. Because ϕ is reducible, we know there exists an essential simple closed curve γ that satisfies the assumption in the lemma. Take any $\alpha \in \Delta^+$. Because of the above lemma, there exists $n > 0$, such that the geodesic g that connects α to $\phi^n(\alpha)$ has distance at most 2 from γ. We know α and $\phi^n(\alpha)$ both belong to Δ^+ and by Theorem 2.5, g is in the K-neighborhood of Δ^+. This implies that

$$d_{\mathcal{C}}(\gamma, \Delta^+) \le d_{\mathcal{C}}(\gamma, g) + K \le K + 2. \tag{3}$$

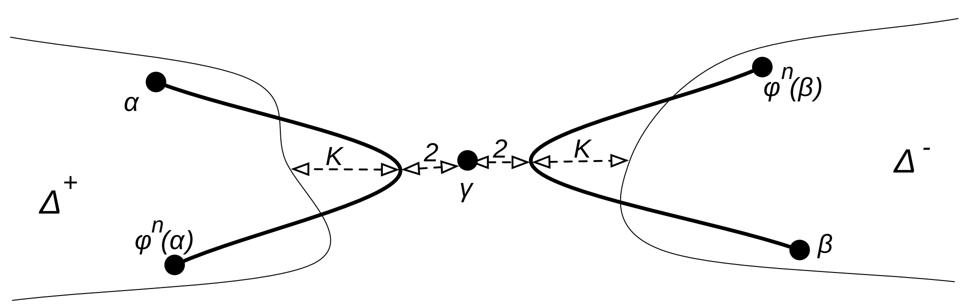

Fig. 2. Reducible ϕ

The same argument shows that

$$d_{\mathcal{C}}(\gamma, \Delta^-) \le K + 2 \tag{4}$$

and we have

$$d_{\mathcal{C}}(\Delta^+, \Delta^-) \leq 2K + 4 \leq 2K + 2\delta, \tag{5}$$

which is a contradiction (Figure 2).

So to finish the proof of the claim, we only need to prove Lemma 3.1.

Proof of Lemma 3.1. Since γ is a reducing curve, $\phi^k(\gamma)$ is isotopic to γ for some k. We say Γ is a *maximal reducing set* for ϕ if it is a maximal set of mutually disjoint and non-parallel essential simple closed curves such that every element of Γ is a reducing curve for ϕ. This maximal set exists because on a closed surface of genus g, we cannot have more than $3g - 3$ mutually disjoint and non-parallel essential simple closed curves.

The next proposition can be of an independent interest.

Proposition 3.2 *Let $\phi \in \mathcal{MCG}(S)$ be non-periodic. There exists an essential subsurface $Y \subset S$ that is not a 3-holed sphere and such that the action of ϕ on $\mathcal{C}(Y)$ is* unbounded*, i.e. for every $\alpha \in \mathcal{C}_0(S)$*

$$d_Y(\alpha, \phi^n(\alpha)) \to \infty \quad \text{as} \quad n \to \infty.$$

Moreover $Y = S$ when ϕ is a pseudo-Anosov and when ϕ is reducible with a maximal reducing set Γ, we can choose Y to be either an annular neighborhood of a component of Γ or a complementary component of Γ.

Proof. For pseudo-Anosov maps, this simply follows from lemma 2.1. So we assume ϕ is a reducible automorphism.

Suppose Γ is a maximal reducing set. Obviously we can take a power ϕ^k such that ϕ^k preserves (up to isotopy) every component of Γ and $S \setminus \Gamma$. Even more maximality of Γ shows that every component Y of $S \setminus \Gamma$ is either a 3-holed sphere or $\phi|_Y$ represents a pseudo-Anosov mapping class of Y.

Note that $\psi = \phi^k$ is non-periodic since ϕ is and Γ is a maximal reducing set for ψ. Suppose the conclusion holds for ψ, i.e. there exists a subsurface Y which is either an annular neighborhood of a component of Γ or is a component of $S \setminus \Gamma$ such that for every $\alpha \in \mathcal{C}_0(S)$

$$d_Y(\alpha, \psi^n(\alpha)) \to \infty \quad \text{as} \quad n \to \infty.$$

We show how this proves the same conclusion for ϕ and the same subsurface Y. Let $n = qk + r$, where $q \in \mathbb{Z}$ and $0 \leq r < k$. If either α or $\phi^n(\alpha)$ does not intersect Y then $d_Y(\alpha, \phi^n(\alpha)) = \infty$ and there is nothing more to do. If not since $\phi^k(Y) = Y$

$$d_Y(\phi^{qk}(\alpha), \phi^n(\alpha)) = d_{\phi^{-qk}(Y)}(\alpha, \phi^r(\alpha)) = d_Y(\alpha, \phi^r(\alpha)).$$

But r assumes only a finite set of values and therefore the above quantity is bounded independent of n. So

$$\begin{aligned} d_Y(\phi^n(\alpha), \alpha) &\geq d_Y(\alpha, \phi^{qk}(\alpha)) - d_Y(\phi^{qk}(\alpha), \phi^n(\alpha)) \\ &= d_Y(\alpha, \psi^q(\alpha) - d_Y(\alpha, \phi^r(\alpha)) \end{aligned}$$

tends to infinity as $n \to \infty$.

Hence just need to prove the conclusion for ψ. Assume there exists a component of $S \setminus \Gamma$, that is not a 3-holed sphere. Suppose Y is this component and note that by maximality of Γ, $\psi|_Y$ is a pseudo-Anosov. If α intersects Y essentially then by lemma 2.1, $d_Y((\psi|_Y)^n(\pi_Y(\alpha)), \alpha) \to \infty$ as $n \to \infty$ and because of equation (2) $(\psi|_Y)^n(\pi_Y(\alpha)) = \pi_Y(\psi^n(\alpha))$; hence

$$\lim_{n \to \infty} d_Y(\psi^n(\alpha), \alpha) = \infty.$$

If α does not intersect Y essentially then by definition $d_Y(\psi^n(\alpha), \alpha) = \infty$ which proves what we wanted.

The remaining case is when all components of $S \setminus \Gamma$ are 3-holed spheres and Γ is what we call a *pants decomposition* on S. Recall that ψ fixes each component of Γ and $S \setminus \Gamma$. Since the action of ψ on each 3-holed sphere is trivial (isotopic to the identity), we deduce that ψ is a nontrivial product of powers of Dehn twists about components of Γ.

Suppose ψ contains a nonzero power of the Dehn twist about $\beta \in \Gamma$. In this case, take Y to be an annular neighborhood of β. Again if α does not intersect β, $d_Y(\alpha, \psi^n(\alpha)) = \infty$ and there is nothing left to prove. On the other hand if α intersects β essentially, then lemma 2.3 shows that $d_Y(\alpha, \psi^n(\alpha))$ tends to infinity as $n \to \infty$ and this finishes the proof. □

To prove lemma 3.1 if α intersects every component of Γ, we can use the above lemma and see that there exists a subsurface Y that is a neighborhood of a component of Γ or is the closure of a component of $S \setminus \Gamma$ such that for n sufficiently large

$$d_Y(\alpha, \phi^n(\alpha)) > L, \tag{6}$$

where L is the constant in theorem 2.4. Theorem 2.4 implies that for g, the geodesic connecting α and $\phi^n(\alpha)$ in $\mathcal{C}(S)$, there exists a vertex $v \in g$ whose projection to Y is empty. In particular v does not intersect ∂Y essentially. But $\partial Y \subset \Gamma$ and since $\gamma \in \Gamma$, $d_{\mathcal{C}}(\gamma, v) \leq 2$ which proves lemma 3.1 when α intersects every component of Γ.

If α is disjoint from a component $\beta \in \Gamma$ then

$$d_{\mathcal{C}}(\alpha, \gamma) \leq d_{\mathcal{C}}(\alpha, \beta) + d_{\mathcal{C}}(\beta, \gamma) \leq 2$$

and this concludes the proof of lemma 3.1. □

As we mentioned this also finishes the proof of the claim. $\square$

Once we know that all the elements of $\Gamma(H^+, H^-) \leq \mathcal{MCG}(S)$ are periodic, we can use a well known theorem about mapping class group of surfaces. For more on this and a proof, see [8].

Theorem 3.3 (Serre [16]) *For any $m \geq 3$, the finite index subgroup*

$$\ker(\mathcal{MCG}(S) \to \mathrm{Aut}(H_1(S, \mathbb{Z}_m)))$$

of $\mathcal{MCG}(S)$ is torsion free.

As an immediate corollary, any subgroup of $\mathcal{MCG}(S)$ has a finite index subgroup which is torsion free, say its intersection with one of the subgroups in the theorem. This shows that any torsion subgroup of $\mathcal{MCG}(S)$ has to be finite and we have proved the main theorem.

References

[1] A. J. Casson and S. A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, *London Mathematical Society Student Texts* **9**, Cambridge University Press, Cambridge, 1988.

[2] A. J. Casson and C. McA. Gordon, *Reducing Heegaard splittings*, *Topology Appl.* **27** (1987), no. 3, 275–283.

[3] A. J. Casson and D. D. Long, *Algorithmic compression of surface automorphisms*, *Invent. Math.* **81** (1985), no. 2, 295-303.

[4] Fathi, Laundenbach and Poénaru, editors, *Travaux de Thurston sur les surfaces.*, *Astérisque* **66-67** (1979).

[5] D. Gabai, R. G. Meyerhoff and N. Thurston, *Homotopy hyperbolic 3-manifolds are hyperbolic*, *Ann. of Math.* **2** (2003), no. 2, 335–431.

[6] W. Haken, *Some results on surfaces in 3-manifolds*, *Studies in Modern Topology*, Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.) (1968) 275–283.

[7] J. Hempel, *3-manifolds as viewed from the curve complex*, *Topology* **40** (2001), 631–657.

[8] N. V. Ivanov, *Subgroups of Teichmüller Modular Groups*, *Translations of Mathematical Monographs* **115**, American Mathematical Society, Providence, RI, 1992.

[9] J. Johnson and H. Rubinstein, *Mapping class groups of Heegaard splittings*, preprint, ArXiv:math.GT/0701119.

[10] T. Kobayashi, *Heights of simple loops and pseudo-Anosov homeomorphisms*, *Braids* (Santa Cruz, CA, 1986), 327–338, Contemp. Math. **78**, Amer. Math. Soc., Providence, RI, 1988.

[11] T. Li, *Heegaard surfaces and measured laminations, I: the Waldhausen conjecture*, to appear in *Invent. Math.*, ArXiv:math.GT/0408198.

[12] M. Lustig and Y. Moriah, *A finiteness result for Heegaard splittings*, *Topology* **43** (2004), no. 5, 1165–1182.

[13] H. A. Masur and Y. N. Minsky, *Geometry of the complex of curves I: Hyperbolicity*, *Invent. Math.* **138** (1999), no. 1, 103–149.

[14] H. A. Masur and Y. N. Minsky, *Geometry of the complex of curves II: Hierarchical structure*, *Geom. Funct. Anal.* **10** (2000), no. 4, 902–974.

[15] H. A. Masur and Y. N. Minsky, *Quasiconvexity in the curve complex*, *In the tradition of Ahlfors and Bers, III*, 309–320, Contemp. Math., **355**, Amer. Math. Soc., Providence, RI, 2004.

[16] J. P. Serre, *Rigidité de foncteur d'Jacobi d'échelon* $n \geq 3$, Séminaire Henri Cartan, 1960/61, Exposé 17 (Appendice), Secrétariat mathématique, Paris, 1960/61.

[17] M. Scharlemann and M. Tomova, *Alternate Heegaard genus bounds distance*, preprint, ArXive:math.GT/0501140.

[18] A. Thompson, *The disjoint curve property and genus* 2 *manifolds*, *Topology Appl.* **97** (1999), no. 3, 273–279.