STABILIZATION OF HEEGAARD SPLITTINGS

JOEL HASS, ABIGAIL THOMPSON, AND WILLIAM THURSTON

Abstract. For each \( g \geq 2 \) there is a 3-manifold with two genus \( g \) Heegaard splittings that require \( g \) stabilizations to become equivalent. Previously known examples required at most one stabilization. Control of families of Heegaard surfaces is obtained through a deformation to harmonic maps.

1. Introduction

While area minimizing surfaces have proved to be a powerful tool in the study of 3-dimensional manifolds, harmonic maps of surfaces to 3-manifolds have not been as widely applied, due to several limitations. A homotopy class of surfaces generally gives rise to a large space of harmonic maps. A harmonic map of a surface need not minimize self-intersections and may fail to be be immersed. In negatively-curved 3-manifolds there is a unique harmonic map in each conformal class of metrics on the domain, and smooth families of surfaces give rise to smooth families of harmonic maps [5], [7]. In this paper we study Heegaard splittings of 3-manifolds using families of harmonic surfaces.

A genus \( g \) Heegaard splitting of a 3-manifold \( M \) is a decomposition of \( M \) into two genus \( g \) handlebodies with a common boundary. It is described by an ordered triple \((H_1, H_2, S)\) where each of \( H_1, H_2 \) is a handlebody and the two handlebodies intersect along their common boundary \( S \), called a Heegaard surface. Two Heegaard splittings \((H_1, H_2, S)\) and \((H'_1, H'_2, S')\) of \( M \) are equivalent if an ambient isotopy of \( M \) carries \((H_1, H_2, S)\) to \((H'_1, H'_2, S')\).

Every 3-manifold has a Heegaard splitting [14], and Heegaard splittings form one of the basic structures used to analyze and understand 3-manifolds.

Corresponding to the Heegaard splitting is a family of surfaces that sweep out the manifold, starting with a core of one handlebody and ending at a core of the second. This family is geometric controlled by deforming it to a family of harmonic maps. When the manifold is negatively curved, harmonic maps of genus \( g \) surfaces have uniformly bounded area. In the manifolds we consider, the geometry forces small area surfaces to line up with small...
A stabilization of a genus $g$ Heegaard surface is a surface of genus $g+1$ obtained by adding a 1-handle whose core is parallel to the surface. Such a surface splits the manifold into two genus $g+1$ handlebodies, and thus gives a new Heegaard splitting. Two genus $g$ Heegaard splittings are $k$-stably equivalent if they become equivalent after $k$ stabilizations. Any two Heegaard splittings become equivalent after a sequence of stabilizations [23]. An upper bound on the number of stabilizations needed to make two splittings equivalent is known in some cases. If $G_p$ and $G_q$ are splittings of genus $p$ and $q$ with $p \leq q$, and $M$ is non-Haken, then Rubinstein and Scharlemann obtained an upper bound of $5p + 8q - 9$ for the genus of a common stabilization [18]. For all previously known examples of manifolds with distinct splittings, the splittings become equivalent after a single stabilization of the larger genus Heegaard surface. The question of whether a single stabilization always suffices is sometimes called the stabilization conjecture [9] (Problem 3.89), [20], [22], [21]. In Section 7 we show that this conjecture does not hold. There are pairs of genus $g$ splittings of a 3-manifold that require $g$ stabilizations to become equivalent.

**Theorem 1.1.** For each $g > 1$ there is a 3-manifold $M_g$ with two genus $g$ Heegaard splittings that require $g$ stabilizations to become equivalent.

We outline the idea in Section 2. In Section 3 we describe the construction of the 3-manifolds $M_g$. We derive isoperimetric inequalities used in the proof in Section 4 and discuss deformations of surfaces to harmonic maps in Section 5. In Section 6 we show how a Heegaard splitting gives rise to a family of Heegaard surfaces that sweep out the 3-manifold. The proof of Theorem 1.1 is given in Section 7. Finally in Section 8 we show how a somewhat weaker result can be obtained for an easily constructed class of hyperbolic manifolds.

This result was presented at the American Institute of Mathematics Conference on Triangulations, Heegaard Splittings, and Hyperbolic Geometry held in December 2007. At this conference D. Bachman announced, using different methods, examples giving a lower bound of $g - 4$ for the number of required stabilizations.

2. Outline of the Argument

Let $M_\phi$ be a hyperbolic 3-manifold that fibers over $S^1$ with monodromy $\phi$ and let $\tilde{M}_\phi$ denote its infinite cyclic cover. A pictorial representation of $M_\phi$ is given in Figure 1. Cutting open $\tilde{M}_\phi$ along a fiber gives a fundamental domain $B$ of the $\mathbb{Z}$-action on the infinite cyclic cover, which we call a block. Blocks are homeomorphic, but not isometric, to the product of a surface and an interval. They are foliated by fibers of $M_\phi$, which in $B$ we call slices. By cutting open a cyclic cover of $M_\phi$ we obtain a hyperbolic 3-manifold with as many adjacent blocks as we wish.
The manifold \( M_g \) used in our main result has pinched negative curvature. It contains two handlebodies \( H_L \) and \( H_R \) with fixed Riemannian metrics, separated by a region homeomorphic to the product of a surface of genus \( g \) with an interval. This intermediate piece is hyperbolic and isometric to \( 2n \) adjacent blocks. The first \( n \) blocks form a submanifold called \( L \) and the next \( n \) form a submanifold called \( R \), as in Figure 2. The value of \( n \) can be chosen as large as desired without changing the geometry of \( H_L \) and \( H_R \). Details are in Section 3.

\[
M_g \text{ has two obvious Heegaard splittings } E_0 = (H_L \cup L, H_R \cup R, S) \text{ and } E_1 = (H_R \cup R, H_L \cup L, -S), \text{ where } S \text{ is a surface of genus } g \text{ separating } H_L \cup L \text{ and } H_R \cup R \text{ and } -S \text{ indicates } S \text{ with reversed orientation. We will show that these splittings are not } k\text{-stably equivalent for } k < g.
\]

Let \( G_0 \) be the Heegaard splitting obtained by stabilizing \( E_0 \) \((g-1)\) times and \( G_1 \) be the Heegaard splitting obtained by stabilizing \( E_1 \) \((g-1)\) times. If \( G_0 \) and \( G_1 \) are equivalent, then there is an isotopy \( \{I_s, \ 0 \leq s \leq 1\} \) of \( M_g \) that induces a diffeomorphism of \( M_g \) carrying \( G_0 \) to \( G_1 \). A family of Heegaard splittings \( \{G_s = I_s(G_0), \ 0 \leq s \leq 1\} \) interpolates between \( G_0 \) and \( G_1 \).

Associated to each Heegaard splitting \( G_s \) is a family of surfaces \( F_{s,t} \) sweeping out \( M_g \) from one core to the other. In Section 5 we show that such a family of surfaces can be deformed to a family of harmonic, or energy minimizing, maps. These harmonic surfaces have area uniformly bounded by

\[
\text{Figure 1. } M_\phi \text{ and one block } B, \text{ a fundamental domain of the } \mathbb{Z}\text{-action on its infinite cyclic cover. } B \text{ is homeomorphic to } F_g \times I.
\]

\[
\text{Figure 2. } M_2, \text{ constructed with 10 blocks, 5 in each of } L \text{ and } R.
\]
a constant $A_0$. This area bound restricts the way that surfaces can divide up the volume of $M_g$. The surface cannot simultaneously split in half the volumes of $L$ and $L \cup R$. In Section 7 we use this to show that surfaces in such bounded area families cannot interpolate between $G_0$ and $G_1$. We conclude that $E_0$ and $E_1$ are not $k$-stably equivalent for $k < g$.

**Remark.** The two Heegaard splittings $E_0$ and $E_1$ of $M_g$ do become equivalent after $g$ stabilizations. After adding $g$ trivial handles, a Heegaard surface can be isotoped to a surface formed by taking two genus $g$ slices connected by a thin tube. This can be isotoped so that one slice splits the volume of $L$ in half while the second does the same to $R$, and together they bisect the volume of $L \cup R$. This type of surface arises in a family of bounded area surfaces interpolating between genus $2g$ stabilizations of $E_0$ and $E_1$.

### 3. Construction of $M_g$

In this section we describe a certain Riemannian 3-manifold $M_g$ of Heegaard genus $g$, based on the work of Namazi and Souto [12]. In Section 8 we give a simpler construction of a hyperbolic manifold that gives slightly weaker lower bounds on the number of stabilizations needed to make two Heegaard splittings equivalent.

We begin with a hyperbolic manifold $M_\phi$ that fibers over the circle with fiber a genus $g$ surface. Fix a fibration of $M_\phi$ with fibers $\{S_t, 0 \leq t \leq 1\}$ satisfying $\phi(S_0) = S_1$. Define a block $B$ to be the manifold obtained by cutting open $M_\phi$ along $S_0$ and a block manifold $B_n$ to be a union of $n$ blocks, formed by placing end to end $n$ copies of $B$. The manifold $B_n$ is topologically, though not geometrically, the product of a surface of genus $g$ and an interval. Its geometry can be obtained by cutting open the $n$-fold cover of $M_\phi$ along a lift of $S_0$. We call the fibers of $B_n$ slices and label them by $S_t$, $0 \leq t \leq n$. We label the union of all fibers between $S_{t_1}$ and $S_{t_2}$ by $B[t_1, t_2]$.

The properties we desire for the manifold $M_g$ are that it is a union of four pieces as in Figure 2. Two pieces are genus $g$ handlebodies, $H_L$ and $H_R$, and the other two $L$ and $R$ are each homeomorphic, though not isometric, to a product $F_g \times [-1, 1]$, where $F_g$ is a surface of genus $g$. $L$ and $R$ are each hyperbolic, and isometric to a block manifold $B_n$. The sectional curvatures of $M_g$ are pinched between $-1 - \epsilon_0$ and $-1 + \epsilon_0$, where $\epsilon_0 > 0$ can be chosen arbitrarily small. For our purpose we take $\epsilon_0 = 1/2$.

Namazi-Souto produced manifolds very similar to $M_g$ [12]. While the manifolds they construct are suitable for the constructions we give, the argument is simpler if we modify the metric slightly so that the middle part of $M_g$ is precisely, rather than approximately hyperbolic. We require the metric on $M_g$ to be isometric to a union of blocks on $L \cup R$. This can be arranged by perturbing the manifold $M$ in Theorem 5.1 of [12], which has sectional curvatures pinched between $-1 - \epsilon$ and $-1 + \epsilon$ where $\epsilon > 0$ can be chosen arbitrarily small. This manifold has a product region separating
two handlebodies with geometry \( \epsilon \) close (in the \( C^2 \)-metric) to the geometry of a block. A small \( C^2 \)-perturbation gives a new metric on \( M \) with sectional curvatures between \(-3/2\) and \(-1/2\) everywhere and exactly isometric to a block on an interval separating two handlebodies. Now cut open the manifold along a slice in this hyperbolic block and insert a copy of \( B_n \) to obtain \( M_g \), with \( n \) as large as desired. However large the choice of \( n \), the geometry of \( M_g \) is unchanged on the two complementary handlebodies of \( B_n \), which we call \( H_L \) and \( H_R \).

4. Isoperimetric inequalities

In this section we develop some isoperimetric inequalities for curves in surfaces and for surfaces in 3-manifolds. The following isoperimetric inequality holds for a curve in hyperbolic space \( H^2 \) [2]. A proof can be obtained by using symmetry to show that a round circle bounds at least as much area as any other curve, using no more length. Explicit formulas for length and area in hyperbolic space then imply the inequality. The curve does not need to be embedded or connected.

**Lemma 4.1.** Let \( c \) be a closed curve in \( H^2 \) with length \( L \). Then \( c \) is the boundary of a disk \( f : D^2 \to H^2 \) of area \( A \), with \( A < L \).

We now consider more general isoperimetric inequalities for curves in surfaces and surfaces in 3-manifolds. In these settings we consider the areas and volumes of 2-chains and 3-chains with boundary. We first give an isoperimetric inequality for curves in a compact family of Riemannian metrics on a surface.

**Lemma 4.2.** Let \((S, g_r)\) be a closed 2-dimensional surface with Riemannian metric \( g_r, r \in U \), where \( g_r \) is a family of metrics smoothly parameterized by a compact set \( U \subset \mathbb{R}^n \). There is a constant \( K \) such that for any surface \( S \) in the family \((S, g_r)\),

1. A null-homotopic curve \( c \) in \( S \) is the boundary of a disk \( f : D \to S \) with
   \[
   \text{Area}(f(D)) \leq K \cdot \text{Length}(c).
   \]

2. A collection of curves \( \{c = \cup c_i\} \), with each \( c_i \) null-homotopic in \( S \), is the boundary of a collection of disks \( \{f_i : D_i \to S\} \) with
   \[
   \sum \text{Area}(f_i(D_i)) \leq K \cdot \text{Length}(c).
   \]

3. A null-homologous curve \( c \) on \( S \) bounds a 2-chain \( X_2 \) in \( S \) with
   \[
   \text{Area}(X_2) \leq K \cdot \text{Length}(c).
   \]

**Proof.** First consider the case where \( c \) is a null-homotopic curve. For a hyperbolic surface \( S, c \) lifts to \( H^2 \) where it bounds a mapped in disk of area less than \( \text{length}(c) \) by Lemma [1]. The surface \((S, g_r)\) is conformally equivalent to a hyperbolic surface, so there is a hyperbolic metric \((S, g)\) and a diffeomorphism from \((S, g)\) to \((S, g_r)\) that stretches or compresses any
vector in the tangent space of \((S, g)\) by a factor of at most \(\lambda\). A single choice of \(\lambda\) can be made for all the surfaces in the compact family \((S, g_r)\). Lengths of curves measured in \((S, g)\) differ from lengths in \((S, g_r)\) by a factor of at most \(\lambda\) and the area of a region in \((S, g)\) differs from the area in \((S, h_r)\) by a factor of at most \(\lambda^2\). By comparing the area of a spanning disk and \(\text{Length}(c)\) in \((S, g)\), we see that in \((S, g_r)\)
\[
\text{Area}(f(D)) \leq \lambda^3 \cdot \text{Length}(c).
\]
The same argument applies for a collection of null-homotopic components.

There is a constant \(\epsilon\) that gives a lower bound for the injectivity radius for the family of metrics \(h_r\). If the length of \(c\) is less than \(\epsilon\) than each of the components of \(c\) is null-homotopic and the above bound holds for the area of a collection of spanning disks.

Now consider a general null-homologous curve \(c\) of length greater than \(\epsilon\). A small perturbation and orientation preserving cut and paste transforms \(c\) into a union of disjoint embedded components that is homologous to \(c\). A homology between this embedded curve and \(c\) has arbitrarily small area, so we can assume that \(c\) is a collection of embedded disjoint curves. Each side of \(c\) gives a 2-chain with boundary \(c\), and so \(\text{Area}(S)\) gives an upper bound to the area of a 2-chain \(X_2\) spanning \(c\). Therefore
\[
\text{Area}(X_2) \leq \text{Area}(S) \leq \frac{\text{Area}(S)}{\epsilon} \cdot \text{Length}(c).
\]
The Lemma follows with \(K = \lambda^3 + \text{Area}(S)/\epsilon\).

To simplify the next calculations, it is convenient to work with chains having \(\mathbb{Z}_2\)-coefficients. With this choice we do not need to pay attention to orientations, signs, or multiplicity other than even or odd. A 2-chain spanning a general position curve in a surface is obtained by 2-coloring the complement of the curve, dividing the surface into black and white regions separated by the curve. The area of a spanning 2-chain is the area of either the white or the black subsurface, and has a value between zero and the area of the surface. The two choices of spanning 2-chain have area that sums to the area of the entire surface. A similar statement applies for 3-chains spanning a surface in a 3-manifold.

We now analyze the geometry of surfaces in block manifolds. The metrics on the slices \(S_t\) of a block manifold, up to isometry, are parameterized by a closed interval, so Lemma 4.2 applies. Therefore there is a constant \(K_1\) independent of \(n\), such that a null-homologous curve \(c\) on a slice \(S_t\) bounds a 2-chain \(X_2\) in that slice with
\[
\text{Area}(X_2) \leq K_1 \cdot \text{Length}(c).
\]

Each slice is an incompressible surface in \(B_n\) and therefore with its induced metric has injectivity radius greater or equal to that of \(M_\phi\). Set \(\epsilon\) to be the constant
\[
\epsilon = \min_{t \in [0,1]} \{ \text{injectivity radius}(M_\phi), \text{Area}(S_t)/10K_1 \}.
\]
Now consider a proper separating surface \((F, \partial F)\) in \((B_n, \partial B_n)\). For each \(t\), the complement of \(F \cap S_t\) in the slice \(S_t\) can be 2-colored. In slices \(S_t\) where \(\text{Length}(F \cap S_t) \leq \epsilon\), each curve of \(F \cap S_t\) is null-homotopic. Lemma 4.2 implies that there is a collection of disks spanning the components of \(F \cap S_t\) whose total area is less than \(\text{Area}(S_t)/10\). Call the subsurface of \(S_t\) missing these disks the \textit{large side} of \(F \cap S_t\). Alternately, this subsurface is the largest of the connected complementary components of \(F \cap S_t\) in \(S_t\).

**Lemma 4.3.** Suppose \((F, \partial F)\) is a proper separating surface in \((B_n, \partial B_n)\) and intersects slices \(S_{t_1}, S_{t_2}\) transversely with \(\text{Length}(F \cap S_{t_1}) \leq \epsilon\) and \(\text{Length}(F \cap S_{t_2}) \leq \epsilon\). If the genus of \(F\) is less than \(g\) then an arc \(\alpha\) from the large side of \(F \cap S_{t_1}\) to the large side of \(F \cap S_{t_2}\) intersects \(F\) algebraically zero times.

**Proof.** Cap off the curves in \(F \cap S_{t_1}, F \cap S_{t_2}\), each of which is shorter than the injectivity radius, by adding disks missing the large side of each surface. These disks do not intersect the arc \(\alpha\). The resulting closed surface in the region between \(S_{t_1}\) and \(S_{t_2}\), has genus less than \(g\). Such a surface is null-homologous in the region between these two slices, which is homeomorphic to the product of a genus \(g\) surface and an interval, and therefore cannot algebraically separate them.

The next lemma gives a 3-dimensional isoperimetric inequality for surfaces in block manifolds. Note that Lemma 4.4 is false if the genus \(k \geq g\). A slice can of genus \(g\) can split the volume of \(B_n\) in half.

**Lemma 4.4.** Let \((F, \partial F) \subset (B_n, \partial B_n)\) a null-homologous proper surface in a block manifold, having genus less than \(g\) and area less than some constant \(A_0\). Then there is a constant \(B_0\), independent of \(n\), such that \(F\) bounds a 3-chain \(X\) in \(B_n\) with \(\text{Volume}(X) < B_0\). For \(n\) sufficiently large,

\[
\text{Volume}(X) < (.1)\text{Volume}(B_n).
\]

**Proof.** Since \(F\) is null-homologous, we can 2-color its complement. This gives two 3-chains with boundary \(F\), and we take \(X\) to be the one of smaller volume. The coarea formula \[6\] implies that the volume of \(X\) is comparable to the integral of its cross-sectional areas. Namely there is a constant \(C_1\) such that

\[
(1) \quad \text{Volume}(X)/C_1 < \int_0^n \text{Area}(X \cap S_t) \, dt < C_1 \cdot \text{Volume}(X).
\]

The coarea formula also gives an inequality between the area of \(F\) in \(B_n\) and the lengths of its intersections with slices. There is a constant \(C_2\) such that

\[
(2) \quad \int_0^n \text{Length}(F \cap S_t) \, dt < C_2 \cdot \text{Area}(F).
\]

The constants \(C_1\) and \(C_2\) are determined by the geometry of a single block \(B\), and do not depend on \(n\).
We estimate the volume of $X$ by integrating the area of its cross sections. To do so, we need to consistently specify for each slice $S_t$ which of the two complementary subsurfaces of $F \cap S_t$ lies in $X$. If $\text{Length}(F \cap S_t) < \epsilon$ then we take the side of $F$ containing the complementary subsurface of smaller area in $S_t$. This subsurface has area less than $K_1 \epsilon$ by Lemma 4.2 and misses the large side of $F \cap S_t$. Each such subsurface lies on the same side of $F \pmod{2}$ by Lemma 4.3. For slices in which $\text{Length}(F \cap S_t) \geq \epsilon$ the construction given in Lemma 4.2 gives a bound $K_1 \cdot \text{Length}(F \cap S_t)$ for the area of both complementary sides in $S_t$ of $F \cap S_t$, and in particular for the side lying in $X$. By Equation (1), $X$ has volume bounded above by

$$\text{Volume}(X) < C_1 \int_0^n \text{Area}(X \cap S_t) \, dt < C_1 K_1 \int_0^n \text{Length}(F \cap S_t) \, dt < C_1 K_1 C_2 A_0 = B_0.$$ 

This bound is independent of $n$, so for $n$ sufficiently large it is less than $\text{Volume}(B_n)/10$. \hfill \square

Finally we give a 3-dimensional isoperimetric inequality that implies small area surfaces in a Riemannian manifold bound regions of small volume. The argument follows the line of an argument of Meeks [13]. The result is valid for both integer and $\mathbb{Z}_2$ coefficients, though we need only the latter.

**Lemma 4.5.** Let $M$ be a compact, connected Riemannian 3-manifold and $F$ a proper surface in $M$, not necessarily embedded or connected. Given $\epsilon_0 > 0$ there is a $\delta_0 > 0$ such that if $\text{Area}(F) < \delta_0$ then $F$ bounds a 3-chain $X$ in $M$ with $\text{Volume}(X) < \epsilon_0$.

**Proof.** By cut and paste we can reduce to the case where $F$ is embedded, though not necessarily connected. The result is true for a surface in the unit ball in $\mathbb{R}^3$, where we have the inequality

$$V \leq \frac{(A)^{3/2}}{6\sqrt{\pi}}.$$ 

for the area of a surface $A$ and the volume $V$ of the region it bounds.

For a fixed Riemannian metric on a ball, there is a bound $\lambda$ to the maximum stretch or compression of the length of a vector. This implies a bound on the volume increase of a region of $\lambda^3$ and a bound on the change in the area of a surface of $\lambda^2$. It follows that

$$V \leq C \cdot A^{3/2}$$

for some constant $C$ depending only on the metric.

Now take a cell decomposition of $M$ with a single 3-cell $R$. By the coarea formula and the area bound for $F$, for sufficiently small $\delta_0$ the 2-complex forming the boundary of this 3-cell $\partial R$ can be perturbed so that the length of its intersection with $F$ is arbitrarily short. The isoperimetric inequality for the boundary 2-sphere of $R$ allows us to surger $F$ along short intersection curves with the boundary, while keeping its total area less than a constant $\delta_1$. 

satisfying $\lim_{\delta_0 \to 0} \delta_1 = 0$. The surgered surface lies in a Riemannian 3-ball in which we can apply Equation 3. The resulting small volume region in a ball gives a region $X$ in $M$ with $\partial X = F$, since $X$ has canceling boundary (mod 2) on $\partial R$. Furthermore the volume of $X$ approaches 0 as $\delta_0 \to 0$. In particular we can arrange that $\text{Volume}(X) < \epsilon_0$ by choosing $\delta_0$ sufficiently small.

□

5. Harmonic and bounded area maps

Eells and Sampson showed that a map from a Riemannian surface $F$ with metric $h$ to a negatively curved manifold $M$ is homotopic to a harmonic map [5]. This harmonic map is unique unless its image is a point or a closed geodesic [7], cases that we will not need to consider. It is obtained by deforming an initial map of a surface with a given Riemannian metric (or conformal structure) in the direction of fastest decrease for the energy, called the tension field. The resulting harmonic map depends smoothly on the metrics of both the domain and the image, as shown in Theorem 3.1 of [3]. See also [19]. A homotopy class of maps from a surface to a 3-manifold is elementary if it induces a homomorphism of fundamental groups whose image is trivial or cyclic.

Lemma 5.1. Let $\{g_u : u \in U\}$ be a family of metrics on a closed surface, parametrized by some compact set $U \subset \mathbb{R}^a$ and let $M$ be a closed manifold with negative sectional curvature. Then a smooth family of non-elementary maps $\{f_u : (F, g_u) \to M\}$ is deformable to a smooth family of harmonic maps $\{h_u : (F, g_u) \to M\}$.

Since harmonic maps satisfy a maximal principle, they are more negatively curved than the ambient manifold wherever they are immersed [19]. The Gauss-Bonnet theorem then implies an area bound proportional to the genus. The immersion assumption can be removed by an approximation of a general map by immersions. A detailed argument is found in Theorem 3.2 of [10].

Lemma 5.2. A harmonic map $f : F \to M$ from a genus $g$ surface to a hyperbolic 3-manifold $M$ has area bounded above by $4\pi(g - 1)$. If $M$ has sectional curvatures between $-s$ and $-r$, for $0 < r < s$, then its area is bounded above by $4\pi(g - 1)/r$.

For any Riemannian metric on $F$, a basic inequality relates the energy and area of a map $f$ [3]:

$$\text{Energy}(f) \geq 2\text{Area}(f).$$

Equality occurs precisely when $f$ is almost conformal, i.e. conformal except possibly at finitely many singular points with zero derivative. In particular, equality holds for an isometric immersion, an immersion of a surface into a Riemannian manifold with the induced metric on the surface.
Lemma 5.3. A smooth family of immersions \( \{ f_{u,0} : F \to M, u \in U \} \) from a surface \( F \) to a closed negatively curved 3-manifold \( M \) is smoothly deformable to a family of harmonic maps \( \{ f_{u,1} : (F, g_u) \to M \} \) through maps \( \{ f_{u,s}, 0 \leq s \leq 1 \} \) satisfying

\[
\text{Area}(f_{u,s}) \leq \text{Area}(f_{u,0}).
\]

Proof. We begin by taking the induced metric on the surface \( F \) using each of the family of maps \( f_{u,0} \). This gives a family of isometric immersions of \( F \), for each of which the energy equals twice the area. The Eells-Sampson process of deforming along the tension field gives an energy decreasing deformation that converges to a harmonic map [5]. Since the energy is non-increasing during this flow, and the area is bounded above by twice the energy, the area of each surface in this flow is bounded above by its initial value. \( \square \)

6. Sweepouts

Consider a family of maps \( k : F \times (-1,1) \to M \) such that

\[
\lim_{t \to \pm 1} \text{Area}(k(F,t)) = 0
\]

and let \( F_t \) denote the surface \( k(F,t) \). By Lemma 4.5 we know that the volume of a 3-chain bounded by \( F_t \) also approaches zero as \( t \to \pm 1 \). It follows that there is a \( \beta > 0 \) such that one side of \( F_t \) has volume less than \( \text{Volume}(M)/10 \) if \( t \in (-1, -1+\beta] \cup [1-\beta, 1) \). In particular \( F_{-1+\beta} \) bounds a 3-chain \( C_{-1+\beta} \) of volume less than \( \text{Volume}(M)/10 \) and similarly \( F_{1-\beta} \) bounds such a chain \( C_{1-\beta} \). Consider the 3-cycle \( Z \in H_3(M; \mathbb{Z}_2) \) formed by adding the 3-chains \( k(F \times [-1+\beta, 1-\beta]), C_{-1+\beta} \) and \( C_{1-\beta} \). We say the family of maps \( k \) has degree one if \( Z \) represents the fundamental homology class of \( M \) (with \( \mathbb{Z}_2 \)-coefficients) and in that case we define it to be a sweepout of \( M \). Note that changing the value of \( \beta \) continuously changes the volume of \( Z \), while a change in the homology class of \( Z \) would cause the volume to change by the volume of \( M \). Thus the choice of \( \beta \) does not affect the question of whether \( k \) has degree one.

We digress somewhat to point out that our constructions give a new approach to constructing sweepouts of bounded area. Pitts and Rubinstein showed the existence of an unstable minimal surface in a 3-manifold using a minimax argument. The minimal surface they construct has maximal area in a 1-parameter family of surfaces obtained from a sweepout given by a strongly irreducible Heegaard splitting. Since a genus \( g \) minimal surface in a hyperbolic manifold has area less than \( 4\pi(g-1) \), this implies an area bound for each surface in the sweepout. The existence of a sweepout composed of bounded area surfaces has implications on the geometry and Heegaard genus of a 3-manifold, as noted in Rubinstein [17] and Bachman-Cooper-White [1]. Harmonic maps give an alternate way to obtain the same area bound implied by Pitts-Rubinstein, without finding a minimal surface and without assuming that a Heegaard splitting is strongly irreducible.
Theorem 6.1. If $M$ is a hyperbolic 3-manifold with a genus $g$ Heegaard splitting then $M$ has a sweepout whose surfaces each has area bounded above by $4\pi(g - 1)$.

Proof. Parametrize each surface of the Heegaard splitting to obtain a smoothly varying family of embedded maps $\{f_t : (F, g_t) \to M, -1 < t < 1\}$, where $g_t$ is the induced metric on $F$ under the map $f_t$. By Lemma 5.1, the family of maps is deformable to a smooth sweepout of harmonic maps $\{h_t : (F, g_t) \to M\}$. Each map $h_t$ has area bounded above by $4\pi(g - 1)$, by Lemma 5.2. □

Given a sweepout of a 3-manifold $M$ and two subsets $L$ and $R$, we can characterize the direction of the sweepout relative to $L$ and $R$, describing which of $L$ and $R$ is first engulfed. Let $\beta$ be a constant that satisfies the conditions of Lemma 4.5, so that for $-1 < s \leq -1 + \beta$, $F_s$ bounds a 3-chain of volume less than $\text{Volume}(M)/10$. If $-1 < t < -1 + \beta$ define $K_t$ to be the 3-chain $C_t$ with boundary $F_t$, given by the smaller volume side of $F_t$. If $t \geq -1 + \beta$ define $K_t$ to be the 3-chain obtained by adding the 3-chains $k(F \times [-1 + \beta, t])$ and $C_{-1 + \beta}$, again with boundary $F_t$. As $t$ increases on $(-1, 1)$ the volume of $K_t$ changes continuously, and satisfies

$$\lim_{t \to -1} \text{Volume}(K_t \cap (L \cup R)) = 0$$

and

$$\lim_{t \to 1} \text{Volume}(K_t \cap (L \cup R)) = \text{Volume}(L \cup R).$$

Let $Y$ denote the set of points $t \in (-1, 1)$ where $K_t$ contains half the volume of $\text{Volume}(L \cup R)$,

$$Y = \{t : \frac{\text{Volume}(K_t \cap (L \cup R))}{\text{Volume}(L \cup R)} = \frac{1}{2}\}.$$

For a generic sweepout this set is finite and odd.

Definition. A point $t \in Y$ is an $L$ point if

$$\text{Volume}(K_t \cap L) > \text{Volume}(K_t \cap R)$$

and an $R$ point otherwise. A sweepout is an $LR$-sweepout if it has an odd number of $L$ points and an $RL$-sweepout if it has an odd number of $R$ points.

The Heegaard splitting $E_0$ gives rise to an $LR$-sweepout that begins with surfaces near a graph at the core of $H_L$, sweeps out $L \cup R$ with embedded slices, and ends with surfaces that collapse to a graph at the core of $H_R$. The stabilized Heegaard splitting $G_0$ gives rise to an $LR$-sweepout by embedded surfaces of genus $2g - 1$. A stabilization adds a loop to each of the graphs forming the cores of $E_0$, the two loops linking once in $M_g$, and each crossing the separating surface $S$. Between the resulting graphs is a product region which can be filled with Heegaard surfaces of the stabilized splitting. See Figure 3, which shows two cores and a surface in $G_0$ that has been stabilized once. Assuming that $G_0$ and $G_1$ are equivalent, composing the resulting
sweepout of $G_0$ with the diffeomorphisms $\{I_s, 0 \leq s \leq 1\}$ gives a family of genus $2g - 1$ sweepouts connecting $G_0$ and $G_1$.

Figure 3. Stabilized cores.

Lemma 6.1. Suppose that $E_0$ is equivalent to $E_1$ after $(g-1)$ stabilizations. Then there is a constant $A_0$, independent of the number of blocks $n$, and a family of surfaces

$$\{f_{s,t} : F \rightarrow M_g, 0 \leq s \leq 1, -1 < t < 1\}$$

satisfying the following conditions:

1. $F$ has genus $2g - 1$
2. For each $s$, $\{f_{s,t} : F \rightarrow M_g, -1 < t < 1\}$ is a sweepout.
3. $\{f_{0,t}(F), -1 < t < 1\}$ is an LR-sweepout.
4. $\{f_{1,t}(F), -1 < t < 1\}$ is an RL-sweepout.
5. Each surface in the two sweepouts $f_{0,t}(F)$ and $f_{1,t}(F)$ has area bounded above by $A_0$.
6. Each surface $\{f_{s,t}(F), 0 \leq s \leq 1, -1 < t < 1\}$ is embedded.

Proof. $M_g$ is formed from the union of a handlebody $H_L$, $n$-blocks forming $L$, $n$-blocks forming $R$, and a second handlebody $H_R$. The geometry of a single block, and of each of $H_L, H_R$, does not depend on $n$. Pick a core for each of $H_L, H_R$ and foliate the complement of this core in each handlebody by embedded Heegaard surfaces, connecting the core to the slice that forms the boundary of $H_L$. Do the same for $H_R$, and fill $L$ and $R$ with interpolating slices. This gives a foliation of the complement of the two cores in $M_g$ by genus $g$ leaves $\{L_t, -1 < t < 1\}$.

Now stabilize by adding $(g-1)$ loops to each core graph and $(g-1)$ handles to each surface between the two core graphs. By adding thin handles, this can be achieved while increasing the area of any surface by less than $a_0$ and with the additional $(g-1)$ 1-handles bounding a region of volume less than $v_0$, where the constants $a_0$ and $v_0$ can be chosen arbitrarily small, independently of $n$. For our purpose it suffices to pick $v_0 < V_B/10$, where $V_B$ is the volume of a single block, and $a_0 < 1$.

Let $F$ be a surface of genus $2g - 1$ and construct maps $\{f_{0,t} : F \rightarrow M_g, -1 < t < 1\}$ that smoothly parametrize the stabilized surfaces. We will refer to $f_{0,t}(F)$ as $F_{0,t}$. Set the value of the constant $A_0$ to be the largest area of the surfaces $\{F_{0,t}, -1 < t < 1\}$. Note that our construction gives a value for $A_0$ that is independent of the number of blocks $n$ in $M_g$.

The Heegaard splitting $G_0 = (H_1, H_2, S)$ obtained by stabilizing $E_0$ $(g-1)$ times is assumed to be equivalent to the reversed splitting $G_1 = $
uniformly bounded by a constant $A$.

Surfaces $f_{\alpha}$, $1$, $(\bar{E}$ stabilization of $\{s\}$

Fore with $L$ing $\{\cup t\}$ained by stabilizing $L$.

$RL$ not $vol$(

graphs. The surface $L$.$H$.$F$

does not give an

$(\bar{E}$, $H$, $F$, $H$, $\bar{E}$).

For any constant $\alpha > 0$ we can assume, by stretching out a collar around the invariant surface $F_{0,0}$, that $I_1$ carries $F_{0,t}$ to $F_{0,-t}$ for each $t \in [-1 + \alpha, 1 - \alpha]$. For $\alpha$ sufficiently small the embedded surfaces

\{F_{s,t}; t \in (0, -1 + \alpha] \cup [1 - \alpha, 1]\}

lie in small neighborhoods of the images of the core graphs of $G_0$ under $I_s$, have area uniformly bounded above by $a_0$, and bound submanifolds having volume less than $v_0$.

The Heegaard splitting $E_0$ gives rise to a sweepout of $M_g$ by genus $g$ surfaces $\{L_t, -1 < t < 1\}$ starting near the core of $H_L$ and ending near the core of $H_R$. This sweepout foliates the complement of the two core graphs. The surface $L_t$ bounds a 3-chain $K_t$ that fills up the side containing $\cup L_{t'} : t' < t$. $K_0$ bisects the volume of $L \cup R$ and contains $L$ but not $R$. The surfaces $L_t$, when parametrized, give an $LR$-sweepout with $Vol(K_0 \cap L) = nV_B$ while $Vol(K_0 \cap R) = 0$. The surface $F_{0,t}$, obtained by stabilizing $L_t$, bounds a 3-chain $K_{0,t}$ whose volume of intersection with $L$ and $R$ differs by less than $v_0 < v_B/10$ from that of $K_t$. Therefore $F_{0,t}$ also gives an $LR$-sweepout. By the same argument applied to the stabilization of $E_1$, we have that $F_{1,t}$ gives a $RL$-sweepout.

All properties now follow.

We now show that a path of sweepouts in $M_g$ whose surfaces have area uniformly bounded by a constant $A_0$ cannot start with an $LR$-sweepout and end with $RL$-sweepout, if $n$ is sufficiently large. Define

$$V_L = Vol(L) = Vol(R)$$

so that $2V_L = Vol(L \cup R)$.

Lemma 6.2. Suppose $A_0$ is a constant and that $M_g$ is constructed with $2n$-blocks. Then for $n$ sufficiently large there does not exist a smooth family of maps

$$\{h_{s,t} : F \to M_g, \ 0 \leq s \leq 1, \ -1 < t < 1\}$$

from a surface $F$ to $M_g$ satisfying the following conditions:

1. $F$ has genus less than $2g$.
2. For each fixed $s$, $h_{s,t} : F \to M_g$ is a sweepout.
3. Each surface $h_{s,t}(F)$ has area bounded above by $A_0$.
4. $\{h_{0,t}(F), -1 < t < 1\}$ is an $LR$-sweepout.
5. $\{h_{1,t}(F), -1 < t < 1\}$ is an $RL$-sweepout.
Proof. Suppose such a family exists and let \( F_{s,t} \) denote \( h_{s,t}(F) \). For each \( F_{s,t} \) we construct a 3-chain \( K_{s,t} \) with boundary \( F_{s,t} \). Pick \( \beta \) sufficiently small to satisfy the conditions of in Lemma 4.5 for each \( 0 \leq s \leq 1 \). Then for each \( s \) and each \(-1 < t_0 \leq -1 + \beta\) there is a 3-chain \( C_{s,t_0} \) with boundary \( F_{s,t_0} \) and volume less than \( V_L/10 \). If \(-1 < t < -1 + \beta\) take \( K_{s,t} \) to be equal to \( C_{s,t} \). If \(-1 + \beta < t < -1\) take \( K_{s,t} \) to be the sum of \( C_{s,-1+\beta} \) and \( \bigcup_{-1+\beta \leq t' \leq t} F_{s,t'} \).

The volume of \( K_{s,t} \) varies continuously with \( s \) and \( t \), giving a continuous function from \([0, 1] \times (-1, 1) \to \mathbb{R}\). For each fixed \( s \) the surfaces \( F_{s,t} \) sweep out \( M_g \) with degree one, so

\[
\lim_{t \to -1} \text{Volume}(K_{s,t} \cap (L \cup R)) = 0
\]

and

\[
\lim_{t \to 1} \text{Volume}(K_{s,t} \cap (L \cup R)) = 2V_L.
\]

Set

\[
Q = \{(s, t) : \frac{\text{Volume}(K_{s,t} \cap (L \cup R))}{2V_L} = \frac{1}{2}\}.
\]

If necessary, perturb the value 1/2 used to define \( Q \) to a regular value of the volume function evaluated on the rectangle. We then have that \( Q \) is a 1-manifold and that any path from the edge \( t = -1 \) to the edge \( t = 1 \) of the \((s, t)\) rectangle must cross \( Q \). It follows that there is a path contained in \( Q \) connecting the edges \( s = 0 \) and \( s = 1 \) of the \((s, t)\) rectangle, as in Figure 4.

![Figure 4](image)

**Figure 4.** Each curve in \( Q \) consists entirely of \( L \) points or \( R \) points.

We now claim that each component of \( Q \) consists entirely of \( L \) points or entirely of \( R \) points. If a component of \( Q \) has both \( L \) points and \( R \) points then there is a point \((s, t) \in Q \) on that component where

\[
\text{Volume}(K_{s,t} \cap L) = \text{Volume}(K_{s,t} \cap R)
\]

and since these two volumes sum to \( V_L \), we have

\[
(4) \quad \frac{\text{Volume}(K_{s,t} \cap L)}{2V_L} = \frac{\text{Volume}(K_{s,t} \cap R)}{2V_L} = \frac{1}{4}
\]

The area of \( F_{s,t} \) is bounded above by \( A_0 \), so Lemma 4.4 implies that for \( n \) sufficiently large, \( F_{s,t} \) bounds a 3-chain in \( L \) with volume less than \((0.1)V_L\).
Since any two 3-chains in $L$ with the same boundary have volumes in $L$ that sum to a multiple of $V_L$, we have

\[0 \leq \frac{\text{Volume}(K_{s,t} \cap L)}{V_L} \leq 0.1.\]

or

\[0.9 \leq \frac{\text{Volume}(K_{s,t} \cap L)}{V_L} \leq 1.\]

As a fraction of $2V_L$

\[0 \leq \frac{\text{Volume}(K_{s,t} \cap L)}{2V_L} \leq 0.05\]

or

\[0.45 \leq \frac{\text{Volume}(K_{s,t} \cap L)}{2V_L} \leq 0.5.\]

This contradicts Equation 4. We conclude that components of $Q$ consist either entirely of $L$ or entirely of $R$ points.

A component curve in $Q$ either meets each edge $s = 0$ and $s = 1$ once, meets one edge of the $(s,t)$ rectangle twice, or is disjoint from both. The parity of the number of $L$ points and the number of $R$ points on the two families $h_{0,t}(F)$ and $h_{1,t}(F)$ is the same for each edge $s = 0$ and $s = 1$, and thus either both are $LR$-sweepouts or both are $RL$-sweepouts. \(\square\)

We now show how to replace a family of sweepouts that starts and ends with sweepouts whose surfaces have area less than $A_0$, but with no control on the area of surfaces in the intermediate sweepouts, with a homotopic family of sweepouts whose surfaces have area uniformly bounded by $A_0$.

**Lemma 6.3.** Let \(\{f_{s,t,0} : F \to M_g, \ 0 \leq s \leq 1, \ -1 < t < 1\}\) be a family of surface maps such that

1. The genus of $F$ is less than $2g$.
2. For each $s$, \(\{f_{s,t,0} : F \to M_g, \ -1 < t < 1\}\) is a sweepout.
3. Each surface in the two sweepouts $f_{0,t,0}(F)$ and $f_{1,t,0}(F)$ has area bounded above by a constant $A_0$.
4. $f_{0,t,0}(F)$, $-1 < t < 1$ is an LR sweepout.
5. $f_{1,t,0}(F)$, $-1 < t < 1$ is an RL sweepout.
6. Each surface \(\{f_{s,t,0} : F \to M_g, \ 0 \leq s \leq 1, \ -1 < t < 1\}\) is immersed.

Then if $n$ is sufficiently large, there is a family of maps \(\{f_{s,t,u} : F \to M_g, \ 0 \leq s \leq 1, \ -1 \leq t \leq 1, \ 0 \leq u \leq 1\}\) such that

1. For each fixed $s, u$; \(\{f_{s,t,u} : F \to M_g, \ -1 \leq t \leq 1\}\) is a sweepout.
2. Each surface $f_{s,t,1}(F)$ is harmonic, with area at most $8\pi(2g - 1)$.
3. $f_{0,t,1}(F)$, $-1 < t < 1$ is an LR sweepout.
4. $f_{1,t,1}(F)$, $-1 < t < 1$ is an RL sweepout.
Proof. We define a family of Riemannian metrics $h_{s,t}$ on $F$ for each $0 \leq s \leq 1$, $-1 < t < 1$ by taking the induced metric pulled back from $M_g$ by the immersion $f_{s,t,0}$. Then each map $f_{s,t,0}$ is conformal and has energy equal to twice its area. We flow the family $f_{s,t,0}$ to a family of harmonic maps $f_{s,t,1}$ as in Lemma 5.3 with the area of each harmonic surface $f_{s,t,1}(F)$ less than $A_0$. Each surface in the sweepout $f_{0,t,0}(F)$ has area bounded above by $A_0$, and this area bound is maintained for each surface $\{f_{0,t,u}(F) \mid 0 \leq u \leq 1\}$ in the homotopy to harmonic maps, as in Lemma 5.2. Similarly the homotopy to harmonic maps of the sweepout $f_{1,t,0}(F)$ given by $\{f_{1,t,u}(F) \mid 0 \leq u \leq 1\}$ has surfaces whose area is uniformly bounded above by $A_0$. Thus if $n$ is sufficiently large, the sweepout $\{f_{0,t,1}(F) \mid -1 < t < 1\}$ is an $LR$-sweepout, by Lemma 6.2 and similarly $\{f_{1,t,1}(F) \mid -1 < t < 1\}$ is an $RL$-sweepout. □

7. Inverting Heegaard surfaces

We now prove the main result:

Proof of Theorem 1.1. We begin with the two genus $g$ Heegaard foliations, $E_0$ and $E_1$ of $M_g$. We will show that these splittings are not $k$-stably equivalent for $k < g$ if $n$ is sufficiently large. It suffices to take $k = g - 1$.

Assume to the contrary that for all $n$, $G_0$ and $G_1$ are equivalent after $g - 1$ stabilizations. For $n$ large, Lemma 6.1 implies there is a family of sweepouts $\{f_{s,t,0} : F \to M_g \mid 0 \leq s \leq 1, -1 < t < 1\}$ of genus $2g - 1$ with $\{f_{0,t,0}(F), -1 < t < 1\}$ an $LR$-sweepout, $\{f_{1,t,0}(F), -1 < t < 1\}$ an $RL$-sweepout and that each surface in the two sweepouts $f_{0,t,0}(F)$ and $f_{1,t,0}(F)$ has area bounded above by $A_0$. Lemma 6.3 implies that this family can be deformed to a new family of sweepouts $\{f_{s,t,1} : F \to M_g \mid 0 \leq s \leq 1, -1 < t < 1\}$ in which $\{f_{0,t,1}(F), -1 < t < 1\}$ is an $LR$-sweepout, $\{f_{1,t,1}(F), -1 < t < 1\}$ is an $RL$-sweepout and each surface $f_{s,t,1}(F)$ has area bounded above by $A_0$. This is a path of sweepouts all of whose surfaces have area less than $A_0$ that connects an $LR$-sweepout to a $RL$-sweepout. Lemma 6.2 states that no such path of sweepouts can exist, contradicting the assumption that fewer than $g$ stabilizations can make $E_0$ and $E_1$ equivalent. □

8. A Hyperbolic Example

The Riemannian manifolds $M_g$ used in our construction were negatively curved, but not hyperbolic. We now show how to construct a family of hyperbolic manifolds that give somewhat weaker lower bounds on the number of stabilizations required to make two genus-$g$ splittings equivalent. Note that block manifolds cannot be isometrically embedded into a hyperbolic 3-manifold in which their slices are Heegaard surfaces, because the fibers of a surface bundle lift to planes in the universal cover. However there is an isometric embedding of block manifolds into a hyperbolic manifold in
which slices are separating incompressible surfaces. These surfaces become Heegaard surfaces after two stabilizations.

Let $N_0$ be a hyperbolic 3-manifold that is a union of two $I$-bundles over a non-orientable surface, glued along their common genus $k$ boundary surface, where $k$ is an integer greater than one. Such hyperbolic manifolds are double covered by a hyperbolic surface bundle over $S^1$. Some explicit examples can be found in [16]. The boundary surface of each $I$-bundle has a neighborhood isometric to a neighborhood of a fiber in its double cover. We can cut open along this fiber and insert a block manifold with arbitrarily many blocks to obtain a hyperbolic manifold $N$, still constructed as a union of two $I$-bundles. We let $S$ denote a surface separating the two $I$-bundles in the center of the block manifold.

Removing a neighborhood of an interval fiber from an $I$-bundle with a genus $k$ boundary surface results in a handlebody of genus $k + 1$. Thus $N$ has a genus $g = k + 2$ Heegaard surface $S'$ obtained by adding a 1-handle to each side of $S$, with the core of each 1-handle an interval in each $I$-bundle. The two orderings of these handlebodies give rise to two Heegaard splittings of $N$, and to corresponding sweepouts that fill the two $I$-bundles in opposite order. The arguments used on $M_g$ in Section 7 now apply to show that $N$ has two genus $g$ splittings that require no less than $g - 4$ stabilizations to become equivalent.

References


Department of Mathematics, University of California, Davis California 95616
E-mail address: hass@math.ucdavis.edu

Department of Mathematics, University of California, Davis California 95616
E-mail address: thompson@math.ucdavis.edu

Department of Mathematics, Cornell University, Ithaca, NY 14853
E-mail address: wpt@math.cornell.edu