The Heegaard genus of bundles over S^1

MARK BRITTENHAM
YO'AV RIECK

1 Introduction

The purpose of this article is exploring theorems of Rubinstein and Lackenby. Rubinstein's Theorem studies the Heegaard genus of certain hyperbolic 3-manifolds that fiber over S^1 and Lackenby's Theorem studies the Heegaard genus of certain Haken manifolds. Our target audience are 3-manifold theorists with good understanding of Heegaard splittings but perhaps little experience with minimal surfaces. We will explain the background necessary for these theorem and prove them (in particular, in Section 3 we explain the main tool needed for analyzing minimal surfaces).

All manifolds considered in this paper are closed, orientable 3-manifolds and all surfaces considered are closed. By the genus of a 3-manifold M (denoted $g(M)$) we mean the genus of a minimal genus Heegaard surface for M.

A *least area* surface is a map from a surface into a Riemannian 3-manifold that minimizes the area in its homopoty class. A *minimal surface* is a critical point of the area functional. Therefore a least area surface is always minimal, as a global minimum is always a critical point. A local minimum of the area functional is called a stable minimal surface and has index zero. However, some minimal surfaces (and in particular the minimal Heegaard surfaces we will study in this paper) are unstable and have positive index. This is similar to a saddle point of the area functional. An easy example is the equatorial sphere $\{x_4 = 0\}$ in S^3 (where S^3 is the unit sphere in $\mathbb{R}^4$). One nice property that all minimal surfaces share is that their mean curvature is zero. This turns out to be equivalent to a surface being minimal. It follows that the intrinsic curvature of a minimal surface is bounded above by the curvature of the ambient manifold. Thus, the curvature of a minimal surface S in a hyperbolic manifold is bounded above by -1, and by Gauss–Bonnet the area of S is at most $2\pi\chi(S)$, where $\chi(S)$ is the Euler characteristic of S.

We assume familiarity with the basic notions of 3-manifold theory (see, for example, [7] or [9]), the basic nations about Heegaard splittings (see, for example, [21]), and Casson–Gordon's [2] concept of *strong irreducibility/weak reducibility*. A more refined notion, due to Scharlemann and Thompson, is *untelescoping* [23] (see also [20]). Untelescoping is, in essence, iterated application of weak reduction (indeed, in some cases [11] a single weak reduction does not suffice). In Section 5 we assume familiarity with this concept.

In [19] Rubinstein used minimal surfaces to study the Heegaard genus of hyperbolic manifolds that fiber over S^1, more precisely, of closed 3-manifolds that fiber over the circle with fiber a closed surface of genus g and pseudo-Anosov monodromy (say ϕ). We denote such manifold by M_ϕ or simply M when there is no place for confusion. While there exist genus two manifolds that fiber over S^1 with fiber of arbitrarily high genus (for example, consider 0-surgery on 2 bridge knots with fibered exterior [6]) Rubinstein showed that this is often not the case. A manifold that fibers over S^1 with genus g fiber has a Heegaard surface of genus $2g+1$ that is obtained by taking two disjoint fibers and tubing them together once on each side. We call this surface and surfaces obtained by stabilizing it *standard*. M has a cyclic cover of degree d (denoted M_{ϕ^d} or simply M_d), dual to the fiber, whose monodromy is ϕ^d. Rubinstein shows that for small h and large d any Heegaard surface for M_d of genus at most h is standard. In particular, the Heegaard genus of M_d (for sufficiently large d) is $2g+1$. The precise statement of Rubinstein's Theorem is:

Theorem 1.1 (Rubinstein) *Let M_ϕ be a closed orientable 3-manifold that fibers over S^1 with pseudo-Anosov monodromy ϕ. Let M_d be the d-fold cyclic cover of M_ϕ dual to the fiber.*

Then for any integer $h \geq 0$ there exists an integer $n > 0$ so that for any $d \geq n$, any Heegaard surface of genus at most h for M_d is standard.

Rubinstein's proof contains two components: the first component is a reduction to a statement about minimal surfaces. We state and prove this reduction in Section 2. It says that if M_d has the property that every minimal surface of genus at most h is disjoint from some fiber then every Heegaard surface for M_d of genus at most h is standard.

The second component of Rubinstein's proof is showing that for large enough d, this property holds for M_d; this was obtained independently by Lackenby [14, Theorem 1.9]. A statement and proof are given in Section 4; we describe it here. Let M be a hyperbolic manifold and $F \subset M$ a non-separating surface (not necessarily a fiber in a fibration over S^1). Construct the d-fold cyclic covers dual to F, denoted M_d, as follows: let M^* be M cut open along F. Then ∂M^* has two components, say F_- and F_+. The identification of F_- with F_+ in M defines a homeomorphism $h : F_- \to F_+$. We take d copies of M^* (denoted M_i^*, with boundaries denoted $F_{i,-}$ and $F_{i,+}$ $(i = 1, \ldots, d)$) and glue them together by identifying $F_{i,+}$ with $F_{i+1,-}$ (the indices are taken modulo d). The gluing maps are defined using h. The manifold obtained is M_d. In Theorem 4.1 we prove that for any M there exists n so that if $d \geq n$ then any minimal surface of genus at most h in M_d is disjoint from at least one of the preimages of F.

The proof is an area estimate. Let S be a minimal surface in a hyperbolic manifold M_d as above; denote the components of the preimage of F by $F_1, \ldots, F_n$. If S intersects every F_i we give a lower bound on its area by showing that there exists a constant $a > 0$ so that S has area at least a near every F_i that it meets. Hence if S intersects every F_i it has area at least ad. Fixing h, if $d > \frac{2\pi(2h-2)}{a}$ then S has area greater than $2\pi(2h-2)$. As mentioned above, the minimal surface S inherits a metric with curvature bounded above by -1, and by Gauss–Bonnet the area of S is at most $2\pi(2g(S)-2)$. Thus $2\pi(2h-2) <$ area of $(S) \leq 2\pi(2g(S)-2)$. Solving for $g(S)$ we see that $g(S) > h$ as required. We note that a depends only on the geometry of M.

The only tool needed for this is a simple consequence of the *Monotonicity Principle*. It says that any minimal surface in a hyperbolic ball of radius R that intersects the center of the ball has at least as much area as a hyperbolic disk of radius R. We briefly explain this in Section 3. For the purpose of illustration we give two proofs in the case that the minimal surface is a disk. One of the proofs requires the following fact: the length of a curve of a sphere or radius r that intersects every great circle is at least $2\pi r$, that is, such curve cannot be shorter than a great circle. We give two proofs of this fact in Appendices A and B.

Let N_1 and N_2 be simple manifolds with $\partial N_1 \cong \partial N_2$ a connected surface of genus $g \geq 2$ (denoted S_g). We emphasize that by $\partial N_1 \cong \partial N_2$ we only mean that the surfaces are homeomorphic.

Let M' be a manifold obtained by gluing N_1 to N_2 along the boundary. Then the image of $\partial N_1 = \partial N_2$ (denoted S) in M' is an essential surface. If $F \subset M'$ is any essential surface with $\chi(F) \geq 0$, then after isotoping F to minimize minimize $|F \cap S|$, any component of $F \cap N_1$ and $F \cap N_2$ is essential and has non-negative Euler characteristic (possibly, $F \cap S = \emptyset$). But simplicity of N_1 and N_2 implies that there are no such surfaces. We conclude that M' is a Haken manifold with no essential surfaces of non-negative Euler characteristic. By Thurston's Uniformization of Haken manifolds

M' is hyperbolic or Seifert fibered. If M' is Seifert fibered then S can be isotoped to be either vertical (that is, everywhere tangent to the fibers) or horizontal (that is, everywhere transverse to the fibers). Both cases contradict simplicity of N_1 and N_2; the details are left to the reader. We conclude that M' is hyperbolic. Note however, that the restriction of the hyperbolic metric on M' to N_1 and N_2 does not have to resemble the original metrics.

After fixing parameterizations $i_1 : S_g \to \partial N_1$ and $i_2 : S_g \to \partial N_2$ any gluing between ∂N_1 and ∂N_2 is given by a map $i_2 \circ f \circ (i_1^{-1})$ for some map $f : S_g \to S_g$.

Fix $f : S_g \to S_g$ a pseudo-Anosov map, let M_f be the bundle over S^1 with fiber S_g and monodromy f, and M_∞ the infinite cyclic cover of M_f dual to the fiber. For $n \in \mathbb{N}$, let M_n be the manifold obtained by gluing N_1 to N_2 using the map $i_2 \circ f^n \circ (i_1^{-1})$. (Note that this is *not* M_d.) Soma [28] showed that for properly chosen points $x_n \in M_n$, (M_n, x_n) converge geometrically (in the Hausdorff–Gromov sense) to M_∞. In [13] Lackenby uses an area argument to show that for fixed h and sufficiently large n every minimal surface of genus at most h in M_n is disjoint from the image of $\partial N_1 = \partial N_2$ (denoted S). This implies that any Heegaard surface of genus at most h weakly reduces to S, and in particular for sufficiently large n, by [26] $g(M_n) = g(N_1) + g(N_2) - g(S)$. In Section 5 we discuss Lackenby's Theorem, following the same philosophy we used for Theorem 1.1. Finally we mention Souto's far reaching generalization of Lackenby's Theorem [29] and a related theorem of Namazi and Souto [17]; however, a detailed discussion and the proofs of these theorems are beyond the scope of this note.

Acknowledgment. We thank Hyam Rubinstein for helpful conversations and the anonymous referee for many helpful suggestions.

2 Reduction to minimal surfaces

In this section we reduce Theorem 1.1 to a statement about minimal surfaces in M_d. We note that the result here applies to any hyperbolic bundle M, but for consistency with applications below we use the notation M_d.

Theorem 2.1 (Rubinstein) *Let M_d be a hyperbolic bundle over S^1. Assume that every minimal surface of Euler characteristic $\geq 2 - 2h$ in M_d is disjoint from some fiber.*

Then any Heegaard surface for M_d of genus at most h is standard.

Proof Let $\Sigma \subset M_d$ be a Heegaard surface of genus at most h. By destabilizing Σ if necessary we may assume Σ is not stabilized.

Assume first that Σ is strongly irreducible. Then by Pitts and Rubinstein [18] (see also [4]) one of the following holds:

(1) Σ is isotopic to a minimal surface.

(2) M_d contains a one-sided, non-orientable, incompressible surface (say H). Let H^* denote H with an open disk removed. Then Σ is isotopic to $\partial N(H^*)$. Equivalently, Σ is isotopic to the surface obtained by tubing $\partial N(H)$ once, inside $N(H)$, via a straight tube.

Both cases lead to a contradiction:

(1) Isotoped Σ to a minimal representative. Let $\gamma \subset M_d$ be a curve. Since $\Sigma \subset M_d$ is a Heegaard surface γ is freely homotopic into Σ. By assumption, Σ is disjoint from some fiber F. Thus after free homotopy $\gamma \cap F = \emptyset$, and in particular γ has algebraic intersection zero with F. But this is absurd: clearly there exist a curve γ that intersects F algebraically once.

(2) Similarly, any curve $\gamma \subset M_d$ is isotopic into $\partial N(H^*)$. Since $\partial N(H^*) \subset N(H)$ and $N(H)$ is an I-bundle over H, γ is isotopic into H. Since H is essential, by [25] (see also [5]) H can be isotoped to be least area and in particular minimal. Note that $2(\chi(H)-1) = 2\chi(H^*) = 2\chi(N(H^*)) = \chi(\partial N(H^*)) = \chi(\Sigma) = 2-2h$. Hence $\chi(H) = 2 - h > 2 - 2h$. By assumption H is disjoint from some fiber F. Thus γ can be homotoped to be disjoint from F, contradiction as above.

Remark It is essential to our proof that H is essential. Let $H \subset M_d$ be a non-separating surface so that $\text{cl}(M_d \setminus N(H))$ is a handlebody. Let H^* be H with n disks removed, for some $n \geq 1$. It is easy to see that $\partial N(H^*)$ is a Heegaard splitting. However, if H is compressible, or if $n > 1$, then $\partial N(H^*)$ destabilizes. (The details are left to the reader.) The converse was recently studied by Bartolini and Rubinstein [1].

Next assume that Σ is weakly reducible. By Casson and Gordon [2] a carefully chosen weak reduction of Σ yields a (perhaps disconnected) essential surface S, and every component of S has genus less than the $g(\Sigma)$ (and hence less than h). By [25] (see also [5]) S is homotopic to a least area (and hence minimal) representative. By assumption S is disjoint from some fiber, and in particular S is embedded in fiber cross $[0, 1]$. Hence S is itself a collection of (say n) fibers and Σ is obtained from S by tubing.

Note that since Σ separates so does S. We conclude that n is even. Denote the components of S by $S_1, \ldots, S_n$ and the components of M_d cut open along S by $C_i (i =$

$1, \ldots, n$) so that $\partial C_i = F_i \sqcup F_{i+1}$ (indeces taken mod n). Thus C_i is homeomorphic to fiber $\times[0,1]$. Fix i and let Σ_i be the surface obtained by pushing ∂C_i slightly into C_i and then tubing along the tubes that are contained in C_i. It is easy to see that the component of C_i cut open along Σ_i that contains ∂C_i is a compression body. The other component is homeomorphic to a component obtained by compressing one the the handlebodies of M_d cut open along Σ. Hence it is a handlebody. We conclude that Σ_i is a Heegaard splitting of C_i, and both components of ∂C_i are on the same side of Σ_i. Scharlemann and Thompson [22] call Σ_i a *type II* Heegaard splitting of C_i. By [22] either Σ_i is obtained by a single tube that is of the form $\{p\} \times [0,1]$ (for some p in the fiber) or it is stabilized. Clearly, if Σ_i is stabilized so is Σ. We conclude that Σ is obtained from S by a single, straight tube in each C_i.

We complete the proof by showing that $n = 2$. Suppose, for a contradiction, that $n > 2$. On S_1 we see two disks, say D_0 and D_1, where the tubes in C_0 and C_1 intersect it. Let F_1^* be $F_1 \setminus (\text{int} D_0 \sqcup \text{int} D_1)$. For $i = 0, 1$ let $\alpha_i \subset F_i^*$ be a properly embedded arc with $\partial\alpha_i \subset \partial D_i$ and so that $|\alpha_0 \cap \alpha_1| = 1$. Note that $\alpha_i \times [0,1]$ is a meridional disk in C_i ($i = 0, 1$) and these disks intersect once on F_1. Since $n > 2$ these disks do not have another intersection. Hence Σ destabilizes, contradicting our assumption. We conclude that $n = 2$. □

3 The Monotonicity Principle

The Monotonicity Principle studies the growth rate of minimal surfaces. All we need is the simple consequence of the Monotonicity Principle, Proposition 3.1, stated below. For illustration purposes, we give two proofs of Proposition 3.1 in the special case when the minimal surface intersects the ball in a (topological) disk. A proof for the Monotonicity Principle for annuli is given Section 6 of Lackenby's [14]. For the general case, see [27] or [3].

We will use the following facts about minimal surfaces: (1) if a minimal surface F intersects a small totally geodesic disk D and locally F is contained on one side of D then $D \subseteq F$. (2) If D is a little piece of the round sphere ∂B (for some metric ball B) and $F \cap D \neq \emptyset$, then locally $F \not\subset B$. Roughly speaking, these facts state that a minimal surface cannot have "maxima" (or, the maximum principle for minimal surfaces).

In this section we use the following notation: $B(r)$ is a hyperbolic ball of radius r, which for convenience we identify with the ball of radius r in the Poincaré ball model in $\mathbb{R}^3$, centered at $O = (0,0,0)$. The boundary of $B(r)$ is denoted $\partial B(r)$. A great circle in $\partial B(r)$ is the intersection of $B(r)$ with a totally geodesic disk that contains O, or, equivalently, the intersection of $\partial B(r)$ with a 2-dimensional subspace of $\mathbb{R}^3$. For convenience, we use the horizontal circle (which we shall call the equator) as a great circle and denote the totally geodesic disk it bounds D_0. Note that ∂D_0 separates $\partial B(r)$ into two disks which we shall call the northern and southern hemispheres, and D_0 separates $B(r)$ into two (topological) balls which we shall call the northern and southern half balls. The ball $B(r)$ is foliated by geodesic disks D_t $(-r \leq t \leq +r)$, where D_t is the intersection of $B(r)$ with the geodesic plane that is perpendicular to the z-axis and intersects it at $(0,0,t)$. Here and throughout this paper, we denote the area of a hyperbolic disk of radius r by $a(r)$. In the first proof below we use the fact that if a curve on a sphere intersects every great circle then it is at least as long as a great circle (Proposition A.1). This is an elementary fact in spherical geometry. In Appendices A and B we give two proofs of this fact, however, we encourage the reader to find her/his own proof and send it to us.

Proposition 3.1 *Let $B(R)$ be a hyperbolic ball of radius R centered at O and $F \subset M$ a minimal surface so that $O \in F$. Then the area of F is at least $a(R)$.*

Remark Lackenby's approach [14] does not require the full strength of the Monotonicity Principle. He only needs the statement for annuli, and in that case he gives a self-contained proof in Section 6 of [14].

We refer the reader to [27] or [3] for a proof. For the remainder of the section, assume $F \cap B(R)$ is topologically a disk. Then we have:

First proof Fix r, $0 < r \leq R$. Fix a great circle in $\partial B(r)$ (which for convenience we identify with the the equator). Suppose that $F \cap \partial B(r)$ is not the equator, we will show that $F \cap \partial B(r)$ intersects both the northern and southern hemispheres. Suppose for contradiction for some r this is not the case. Then one of the following holds:

(1) $F \cap \partial B(r) = \emptyset$.

(2) $F \cap \partial B(r) \neq \emptyset$ and $F \cap \partial B(r)$ does not intersect one of the two hemispheres.

Assuming Case (1) happens, and let $r' > 0$ be the largest value for which $F \cap \partial B(r') \neq \emptyset$. Then F and $\partial B(r')$ contradict fact (2) mentioned above.

Next assume Case (2) happens (say F does not intersect the southern hemisphere). Let t be the most negative value for which $F \cap D_t \neq \emptyset$. Since $O \in F$, $-r < t \leq 0$. Then by fact (1) above, F must coincide with D_t. If $t < 0$ then D_t intersects the southern hemisphere, contrary to our assumptions. Hence $t = 0$ and F is itself D_0; thus $F \cap B(r)$ is the equator, again contradicting our assumptions.

By assumption $F \cap B(R)$ is a disk and therefore $F \cap \partial B(r)$ is a circle. Clearly, a circle that intersects both the northern and the southern hemispheres must intersect the equator. We conclude that $F \cap \partial B(r)$ intersects the equator, and as the equator was chosen arbitrarily, $F \cap \partial B(r)$ intersects every great circle. By Proposition A.1 $F \cap B(r)$ is at least as long as a great circle in $\partial B(r)$. Since the intersection of a totally geodesic disk with $\partial B(r)$ is a great circle, integrating these lengths shows that the area of $F \cap B(r)$ grows at least as fast as the area of a geodesic disk, proving the proposition. □

Second proof Restricting the metric from M to F, distances can increase but cannot decrease. Therefore $F \cap \partial B(R)$ is at distance (on F) at least R from O and we conclude that F contains an entire disk of radius R. The induced metric on F has curvature at most -1 and therefore areas on F are at least as big as areas in $\mathbb{H}^2$. In particular, the disk of radius R about O has area at least $a(R)$. □

4 Main Theorem

By Theorem 2.1 the main task in proving Theorem 1.1 is showing that (for large enough d) a minimal surface of genus at most h in M_d is disjoint from some fiber F. Here we prove:

Theorem 4.1 *Let M be a compact, orientable hyperbolic manifold and $F \subset M$ a non-separating, orientable surface. Let M_d denote the cyclic cover of M dual to F of degree d (as in the introduction).*

Then for any integer $h \geq 0$ there exists a constant n so that for $d \geq n$, any minimal surface of genus at most h in M_d is disjoint from a component of the preimage of F.

Proof Fix an integer h.

Denote the distance in M by $d(\cdot,\cdot)$. Push F off itself to obtain $\widehat{F}$, a surface parallel to F and disjoint from it. For each point $p \in F$ define:

$$R(p) = \min\{\text{radius of injectivity at } p,\ d(p, \widehat{F})\}.$$

Since $\widehat{F}$ is compact $R(p) > 0$. Define:

$$R = \min\{R(p) | p \in F\}.$$

Since F is compact $R > 0$. Note that R has the following property: for any $p \in F$, the set $\{q \in M : d(p,q) < R\}$ is an embedded ball and this ball is disjoint from $\widehat{F}$. As above, let $a(R)$ denote the area of a hyperbolic disk of radius R.

Let n be the smallest integer bigger than $\frac{2\pi(2h-2)}{a(R)}$. Fix an integer $d \geq n$. Denote the preimages of F in M_d by $F_1, \dots, F_d$.

Let S be a minimal surface in M_d. Suppose S cannot be isotoped to be disjoint from the preimages of F_i for any i. We will show that $g(S) > h$, proving the theorem.

Pick a point $p_i \in F_i \cap S$ ($i = 1, \dots, d$) and let B_i be the set $\{p \in M_d | d(p, p_i) < R\}$. By choice of R, for each i, B_i is an embedded ball and the preimages of $\widehat{F}$ separate these balls; hence for $i \neq j$ we see that $B_i \cap B_j = \emptyset$. $S \cap B_i$ is a minimal surface in B_i that intersects its center and by Proposition 3.1 (the Monotonicity Principle) has area at least $a(R)$. Summing these areas we see that the area of S fulfills:

$$\begin{aligned} \text{Area of } S &\geq d \cdot a(R) \\ &\geq n \cdot a(R) \\ &> \frac{2\pi(2h-2)}{a(R)} \cdot a(R) \\ &= 2\pi(2h-2). \end{aligned}$$

But a minimal surface in a hyperbolic manifold has curvature ≤ -1 and hence by the Gauss-Bonnet Theorem, the area of $S \leq -2\pi\chi(S) = 2\pi(2g(S-2)$. Hence, the genus of S is greater than h. □

Remark 4.2 (Suggested project) In Theorem 4.1 we treat the covers dual to a non-separating essential surface (denoted M_d there). In the section titled "Generalization" of [13], Lackenby shows (among other things) how to amalgamate along non-separating surfaces. Does his construction and Theorem 4.1 give useful bounds on the genus of M_d, analogous to Theorem 1.1?

5 Lackenby's Theorem

Lackenby studied the Heegaard genus of manifolds containing separating essential surfaces. Here too, the result is asymptotic. We begin by explaining the set up. Let N_1 and N_2 be simple manifolds with $\partial N_1 \cong \partial N_2$ a connected surface of genus $g \geq 2$ (that is, ∂N_1 and ∂N_2 are homeomorphic). Let S be a surface of genus g and $\psi_i : S \to \partial N_i$ parameterizations of the boundaries ($i = 1, 2$). Let $f : S \to S$ be a pseudo-Anosov map. For any n we construct the map $f_n = \psi_2 \circ f^n \circ (\psi_1)^{-1} : \partial N_1 \to \partial N_2$. By identifying ∂N_1 with ∂N_2 by the map f_n we obtain a closed hyperbolic manifold M_n. Let $S \subset M_n$ be the image of $\partial N_1 = \partial N_2$. With this we are ready to state Lackenby's Theorem:

Theorem 5.1 (Lackenby [13]) *With notation as in the previous paragraph, for any h there exists N so that for any $n \geq N$ any genus h Heegaard surface for M_n weakly reduces to S. In particular, by setting $h = g(N_1) + g(N_2) - g(S)$ we see that there exists N so that if $n \geq N$ then $g(M_n) = g(N_1) + g(N_2) - g(S)$.*

Sketch of proof As in Sections 2 and 4, the proof has two parts which we bring here as two claims:

Claim 1 *Suppose that every every minimal surface in M_n of genus at most h can be homotoped to be disjoint from S. Then any Heegaard surface of genus at most h weakly reduces to S. In particular, if $h \geq g(N_1) + g(N_2) - g(S)$ then $g(M_n) = g(N_1) + g(N_2) - g(S)$.*

Claim 2 *There exists N so that if $n \geq N$ then any minimal surface of genus at most h in M_n can be homotoped to be disjoint from S.*

Clearly, Claim 1 and 2 imply Lackenby's Theorem. We now sketch their proofs.

We paraphrase Lackenby's proof of Claim 1: let Σ be a Heegaard surface of genus at most h. Then by Scharlemann and Thompson [24] Σ untelescopes to a collection of connected surfaces F_i and Σ_j where $\cup_i F_i$ is an essential surface (with F_i its components) and Σ_j are strongly irreducible Heegaard surfaces for the components of M_n cut open along $\cup_i F_i$; in particular M_n cut open along $(\cup_i F_i) \cup (\cup_i \Sigma_j)$ consists of compression bodies and the images of the F_i's form ∂_- of these compression bodies. Since F_i and Σ_j are obtained by compressing Σ, they all have genus less than h.

By [25], [5], and [18] the surfaces F_i and Σ_j can be made minimal. We explain this process here: since F_i are essential surfaces they can be made minimal by [25] (see also [5]). Next, since the Σ_j's are strongly irreducible Heegaard surfaces for the components of M_n cut open along $\cup_i F_i$, each Σ_j can be made minimal within its component by [18] (see also [4]). Note that the surfaces F_i and Σ_j are disjointly embedded.

By assumption, S can be isotoped to be disjoint every F_i and every Σ_j. Therefore, S is an essential closed surface in a compression body and must be parallel to a component of ∂_-. Therefore, for some i, S is isotopic to F_i. In Proposition 2.13 of [12] it was shown that if Σ untelescopes to the essential surface $\cup_i F_i$, then Σ weakly reduces to any *connected separating* component of $\cup_i F_i$; therefore Σ weakly reduces to S. This proves the first part of Claim 1.

Since S is connected any minimal genus Heegaard splittings for N_1 and N_2 can be amalgamated (the converse of weak reduction [26]). By amalgamating minimal genus Heegaard surfaces we see that for any n, $g(M_n) \leq g(N_1) + g(N_2) - g(S)$. By applying the first part of Claim 1 with $h = g(N_1) + g(N_2) - g(S)$ we see that for sufficiently large n, a minimal genus Heegaard surface for M_n weakly reduces to S; by Proposition 2.8 of [12] $g(M_n) = g(N_1) + g(N_2) - g(S)$, completing the proof of Claim 1.

We now sketch the proof of Claim 2. Fix h and assume that for arbitrarily high values of n, M_n contains a minimal surface (say P_n) of genus $g(P_n) \leq h$ that cannot be homotoped to be disjoint from S. Let M_f be the bundle over S^1 with monodromy f and fix two disjoint fibers F, $\widehat{F} \subset M_f$. Let R be as in Section 4. Let M_∞ be the infinite cyclic cover dual to the fiber. Soma [28] showed that there are points $x_n \in M_n$ so that (M_n, x_n) converges in the sense of Hausdorff–Gromov to the manifold M_∞. These points are near the minimal surface S, and the picture is that M_n has a very long "neck" that looks more and more like M_∞.

For sufficiently large n there is a ball $B(r) \subset M_n$ for arbitrarily large r that is $1 - \epsilon$ isometric to $B_\infty(r) \subset M_\infty$. Note that $B_\infty(r)$ contains arbitrarily many lifts of F separated by lifts of $\widehat{F}$. Since P_n cannot be isotoped to be disjoint from S, its image in M_∞ cannot be isotoped off the preimages of F. As in Section 4 we conclude that the images of P_n have arbitrarily high area. However, areas cannot be distorted arbitrarily by a map that is $1 - \epsilon$ close to an isometry. Hence the areas of P_n are unbounded, contradicting Gauss–Bonnet; this contradiction completes our sketch. □

In [29] Souto generalized Lackenby's result (see also a recent paper by Li [15]). Although his work is beyond the scope of this paper, we give a brief description of it here. Instead of powers of maps, Souto used a combinatorial condition on the gluings: fixing essential curves $\alpha_i \subset N_i$ $(i = 1, 2)$ and $h > 0$, Souto shows that if $\phi : N_1 \to N_2$ fulfills the condition "$d_{\mathcal{C}}(\phi(\alpha_1), \alpha_2)$ is sufficiently large" then any Heegaard splitting for $N_1 \cup_\phi N_2$ of genus at most h weakly reduces to S. The distance Souto uses—$d_{\mathcal{C}}$—is the distance in the "curve complex" and *not* the hyperbolic distance, as defined by Hempel [8], who showed (following Kobayashi [10]) that raising a fixed monodromy ϕ to a sufficiently high power does imply Souto's condition. Hence Souto's condition is indeed weaker than Lackenby's, and it is in fact too weak for us to expect Soma-type convergence to M_∞. However, using Minsky [16] Souto shows that given a sequence of manifolds M_{ϕ_n} with $d_{\mathcal{C}}(\phi_n(\alpha_1), \alpha_2) \to \infty$, the manifolds M_{ϕ_n} are "torn apart" and the cores of N_1 and N_2 become arbitrarily far apart. For a precise statement see Proposition 6 of [29]. Souto concludes that for sufficiently large n, any minimal surface for M_n that intersects both N_1 and N_2 has high area and therefore genus greater than h. Souto's Theorem now follows from Claim 1 above.

A similar result was obtained by Namazi and Souto [17] for gluing of handlebodies. They show that is N_1 and N_2 are genus g handlebodies and $\partial N_1 \to \partial N_2$ is a generic pseudo-Anosov map (for a precise definition of "generic" in this case see [17]) then for any $\epsilon > 0$ and for large enough n the manifold M_{f^n} obtained by gluing N_1 to N_2 via f^n admits a negatively curved metric with curvatures K so that $-1-\epsilon < K < -1+\epsilon$. Namazi and Souto use this metric to conclude many things about M_{f^n}, for example, the Heegaard genus and rank (that is, number of generators needed for $\pi_1(M_{f^n})$) are exactly g.

A Short curves on round spheres: take one

In this section we prove the following proposition, which is a simple exercise in spherical geometry used in Section 3. Let $S^2(r)$ be a sphere of constant curvature $+(\frac{1}{r})^2$. We isometrically identify $S^2(r)$ with $\{(x,y,z) \in \mathbb{R}^3 | x^2+y^2+z^2 = r^2\}$ and refer to it as a round sphere of radius r.

Proposition A.1 *Let $S^2(r)$ be a round sphere of radius r and $\gamma \subset S^2$ a rectifiable closed curve. Suppose $l(\gamma) \leq 2\pi r$ (the length of great circles). Then γ is disjoint from some great circle.*

Remark The proof also shows that if γ is a *smooth* curve that meets every great circle then $l(\gamma) = 2\pi r$ if and only if γ is itself a great circle.

Proof Let γ be a curve that intersects every great circle. Let $z_{\min}$ (for some $z_{\min} \in \mathbb{R}$) be the minimal value of the z-coordinate, taken over γ. Rotate $S^2(r)$ to maximize $z_{\min}$. If $z_{\min} > 0$ then γ is disjoint from the equator, contradicting our assumption. We assume from now on $z_{\min} \leq 0$.

Suppose first $z_{\min} = 0$. Suppose, for contradiction, that there exists a closed arc α on the equator so that $l(\alpha) = \pi r$ and $\alpha \cap \gamma = \emptyset$. By rotating $S^2(r)$ about the z-axis (if necessary) we may assume $\alpha = \{(x,y,0) \in S^2(r) | y \leq 0\}$. Then rotating $S^2(r)$ slightly about the x-axis pushes the points $\{(x,y,0) \in S^2(r) | y > 0\}$ above the xy-plane. By compactness of γ and α there is some ϵ so that $d(\gamma,\alpha) > \epsilon$. Hence if the rotation is small enough, no point of γ is moved to (or below) α. Thus, after rotating $S^2(r)$, $z_{\min} > 0$, contradiction. We conclude that every arc of the equator of length πr contains a point of γ. Therefore there exists a sequence of point $p_i \in \gamma \cap \{(x, y, 0)\}$ ($i = 1, \ldots, n$, for some $n \geq 2$), ordered by their order along the equator (*not* along γ), so that $d(p_i, p_{i+1})$ is at most half the equator (indices taken modulo n). The shortest path connecting p_i to p_{i+1} is an arc of the equator, and we conclude that $l(\gamma) \geq 2\pi r$ as required. If we assume, in addition, that $l(\gamma) = 2\pi r$ then either γ is itself the equator or γ consists of two arcs of great circle meeting at $c_1 \cup c_2$. Note that this can in fact happen, but then γ is not smooth. This completes the proof in the case $z_{\min} = 0$

Assume next $z_{\min} < 0$. Let $c_{\min}$ be the latitude of $S^2(r)$ at $z = z_{\min}$, and denote the length of $c_{\min}$ by $d_{\min}$. Suppose there is an open arc of $c_{\min}$ of length $\frac{1}{2}d_{\min}$ that does not intersect γ. Similar to above, by rotating $S^2(r)$ we may assume this arc is given by $\{(x, y, z_{\min}) \in c_{\min} | y < 0\}$. Then a tiny rotation about the x-axis increases the z-coordinate of all points $\{(x, y, z) | y \geq 0, \ z \leq 0\}$. As above ,this increases $z_{\min}$, contradicting our choice of $z_{\min}$. Therefore there is a collection of points $p_i \in \gamma \cap c_{\min}$ ($i = 1, \ldots, n$, for some $n \geq 3$), ordered by their order along the equator (*not* along $c_{\min}$), so that $d(p_i, p_{i+1}) < \frac{1}{2}d_{\min}$ (indices taken modulo n). The shortest path connecting p_i to p_{i+1} is an arc of a great circle. However, such arc has points with z-coordinate less than $z_{\min}$, and therefore cannot be a part of γ. The shortest path containing all the p_i's on the punctured sphere on $\{(x, y, z) \in S^2(r) | z \geq z_{\min}\}$ is the boundary, that is, $c_{\min}$ itself. Unfortunately, $l(c_{\min}) < 2\pi r$. Upper hemisphere to the rescue! γ must have a point with z-coordinate at least $-z_{\min}$, for otherwise rotating $S^2(r)$ by π about any horizontal axis would decrease $z_{\min}$. Then $l(\gamma)$ is at least as long as the shortest curve containing the p_i's and some point p on or above $c_{\min}$, the circle of γ at $z = z_{\min}$. Let γ be such a curve. By reordering the indices if necessary it is convenient to assume that p is between p_1 and p_2. It is clear that moving p so that its longitude is between the longitudes of p_1 and p_2 shortens γ (note that since $d(p_1, p_2) < \frac{1}{2}d_{\min}$ this is well-defined). We now see that γ intersects the equator in two point, say x_1 and x_2. Replacing the two arcs of γ above the equator by the short arc of the equator decreases length. It is not hard to see that the same hold when we replace the arc of γ below the equator with the long arc of the equator. We conclude that $l(\gamma) > l(\text{equator}) = 2\pi r$. □

B Short curves on round spheres: take two

We now give a second proof of Proposition A.1. For convenience of presentation we take S^2 to be a sphere of radius 1. Let γ be a closed curve that intersects every great circle. Every great circle is defined by two antipodal points, for example, the equator is defined by the poles. Thus, the space of great circles is $\mathbb{R}P^2$. Since S^2 has area 4π, $\mathbb{R}P^2$ has area 2π. Let $f : S^2 \to \mathbb{R}P^2$ be the "map" that assigns to a point p all the great circles that contain p; thus, for example, if p is the north pole then $f(p)$ is the projection of the equator to $\mathbb{R}P^2$.

Let C be a great circle. We claim that $\gamma \cap C$ contains at least two points of γ. (If γ is not embedded then the two may be the same point of C.) Suppose, for a contradiction, that γ meets some great circle (say the equator) in one point only (Say $(1,0,0)$). By the Jordan Curve Theorem, γ does not cross the equator. By tilting the equator slightly about the y-axis it is easy to obtain a great circle disjoint from γ. Hence we see that γ intersects every great circle at least twice. Equivalently, $f(\gamma)$ covers $\mathbb{R}P^2$ at least twice.

Let α_i be a small arc of a great circle, of length $l(\alpha_i)$; note that this length is exactly the angle α_i supports in radians. Say for convenience α_i starts at the north pole and goes towards the equator. The points that define great circles that intersect α_i are given by tilting the equator by α_i radians. This gives a set whose area is α_i/π of the total area of S^2. Since the area of S^2 is 4π, it gives a set of area $4l(\alpha_i)$. This set is invariant under the antipodal map, and so projecting to $\mathbb{R}P^2$ the area is cut by half, and we get:

$$\text{Area of } f(\alpha_i) = 2l(\alpha_i). \tag{1}$$

Fix $\epsilon > 0$. Let α be an approximation of γ by small arcs of great circles, say $\{\alpha_i\}_{i=1}^n$ are the segments of α. We require α to approximate γ well in the following two senses:

(1) $l(\alpha) \leq l(\gamma) + \epsilon$.
(2) Under f, α covers $\mathbb{R}P^2$ as well as γ does (except, perhaps, for a set of measure ϵ); *i.e.*, the area of $f(\alpha) \geq$ the area of $f(\gamma) - \epsilon$ (area measured with multiplicity).

From this we get:

$$\begin{aligned}
4\pi - \epsilon &= \text{twice the area of } \mathbb{R}P^2 - \epsilon \\
&\leq \text{the area of } f(\gamma) - \epsilon \\
&\leq \text{area of } f(\alpha) \\
&= \Sigma_{i=1}^n \text{area of } f(\alpha_i) \\
&= \Sigma_{i=1}^n 2l(\alpha_i) \\
&= 2l(\alpha) \\
&\leq 2(l(\gamma) + \epsilon).
\end{aligned}$$

(In the fifth equality we use Equation (1).) Since ϵ was arbitrary, dividing by 2 we get the desired result: $2\pi \leq l(\gamma)$.

References

[1] Loretta Bartolini and J. Hyam Rubinstein. One-sided Heegaard splittings of non-haken 3-manifolds. Available at http://arxiv.org/abs/math.GT/0701906.

[2] A. J. Casson and C. McA. Gordon. Reducing Heegaard splittings. *Topology Appl.*, 27(3):275–283, 1987.

[3] Jaigyoung Choe. Isoperimetric inequalities of minimal submanifolds. In *Global theory of minimal surfaces*, volume 2 of *Clay Math. Proc.*, pages 325–369. Amer. Math. Soc., Providence, RI, 2005.

[4] Tobias H. Colding and Camillo De Lellis. The min-max construction of minimal surfaces. In *Surveys in differential geometry, Vol. VIII (Boston, MA, 2002)*, Surv. Differ. Geom., VIII, pages 75–107. Int. Press, Somerville, MA, 2003. Available at http://arxiv.org/abs/math.AP/0303305.

[5] Michael Freedman, Joel Hass, and Peter Scott. Least area incompressible surfaces in 3-manifolds. *Invent. Math.*, 71(3):609–642, 1983.

[6] A. Hatcher and W. Thurston. Incompressible surfaces in 2-bridge knot complements. *Invent. Math.*, 79(2):225–246, 1985.

[7] John Hempel. *3-Manifolds*. Princeton University Press, Princeton, N. J., 1976. Ann. of Math. Studies, No. 86.

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

[9] William Jaco. *Lectures on three-manifold topology*, volume 43 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, Providence, R.I., 1980.

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

[11] Tsuyoshi Kobayashi. Scharlemann-Thompson untelescoping of Heegaard splittings is finer than Casson-Gordon's. *J. Knot Theory Ramifications*, 12(7):877–891, 2003.

[12] Tsuyoshi Kobayashi and Yo'av Rieck. Heegaard genus of the connected sum of m-small knots. *Comm. Anal. Geom.*, 14(5):1033–1073, 2006. Available at http://arxiv.org/abs/math.GT/0503229.

[13] Marc Lackenby. The Heegaard genus of amalgamated 3-manifolds. *Geom. Dedicata*, 109:139–145, 2004.

[14] Marc Lackenby. Heegaard splittings, the virtually Haken conjecture and property (τ). *Invent. Math.*, 164(2):317–359, 2006.

[15] Tao Li. On the Heegaard splittings of amalgamated 3–manifolds. Available at http://arxiv.org/abs/math/0701395.

[16] Yair N. Minsky. Kleinian groups and the complex of curves. *Geom. Topol.*, 4:117–148 (electronic), 2000.

[17] Hossein Namazi and Juan Souto. Heegaard splittings and pseudo-Anosov maps. Preprint available from Souto's web page.

[18] Jon T. Pitts and J. H. Rubinstein. Existence of minimal surfaces of bounded topological type in three-manifolds. In *Miniconference on geometry and partial differential equations (Canberra, 1985)*, volume 10 of *Proc. Centre Math. Anal. Austral. Nat. Univ.*, pages 163–176. Austral. Nat. Univ., Canberra, 1986.

[19] J. Hyam Rubinstein. Minimal surfaces in geometric 3-manifolds. In *Global theory of minimal surfaces*, volume 2 of *Clay Math. Proc.*, pages 725–746. Amer. Math. Soc., Providence, RI, 2005.

[20] Toshio Saito, Martin Scharlemann, and Jennifer Schultens. Lecture notes on generalized Heegaard splittings. Available at http://arxiv.org/abs/math.GT/0504167.

[21] Martin Scharlemann. Heegaard splittings of 3-manifolds. In *Low dimensional topology*, volume 3 of *New Stud. Adv. Math.*, pages 25–39. Int. Press, Somerville, MA, 2003. Available at http://arxiv.org/abs/math.GT/0007144.

[22] Martin Scharlemann and Abigail Thompson. Heegaard splittings of (surface) $\times I$ are standard. *Math. Ann.*, 295(3):549–564, 1993.

[23] Martin Scharlemann and Abigail Thompson. Thin position for 3-manifolds. In *Geometric topology (Haifa, 1992)*, volume 164 of *Contemp. Math.*, pages 231–238. Amer. Math. Soc., Providence, RI, 1994.

[24] Martin Scharlemann and Abigail Thompson. Thin position for 3-manifolds. In *Geometric topology (Haifa, 1992)*, volume 164 of *Contemp. Math.*, pages 231–238. Amer. Math. Soc., Providence, RI, 1994.

[25] R. Schoen and Shing Tung Yau. Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature. *Ann. of Math. (2)*, 110(1):127–142, 1979.

[26] Jennifer Schultens. The classification of Heegaard splittings for (compact orientable surface) $\times S^1$. *Proc. London Math. Soc. (3)*, 67(2):425–448, 1993.

[27] Leon Simon. *Lectures on geometric measure theory*, volume 3 of *Proceedings of the Centre for Mathematical Analysis, Australian National University*. Australian National University Centre for Mathematical Analysis, Canberra, 1983.

[28] Teruhiko Soma. Volume of hyperbolic 3-manifolds with iterated pseudo-Anosov amalgamations. *Geom. Dedicata*, 90:183–200, 2002.

[29] Juan Souto. The Heegaard genus and distances in the curve complex. Preprint available from Souto's web page.

Department of Mathematics, 203 Avery Hall, University of Nebraska-Lincoln, Lincoln, NE 68588-0130

Department of mathematical Sciences, University of Arkansas, Fayetteville, AR 72701

mbrittenham2@math.unl.edu, yoav@uark.edu