

POSITIVITY OF THE UNIVERSAL PAIRING IN 3 DIMENSIONS

DANNY CALEGARI, MICHAEL H. FREEDMAN, AND KEVIN WALKER

1. INTRODUCTION

The earliest objects of study in the theory of manifolds were the fundamental group, Poincaré duality, and by the 1930s, characteristic classes. The main theme was classification: developing invariants to distinguish one manifold from another and trying to construct manifolds with prescribed invariants. In the mid 1950s to early 1960s, Thom, Milnor, and Smale shifted the emphasis to *operations* on manifolds such as cutting, gluing and especially surgery, using powerful new structural tools such as Morse theory and cobordism. The most spectacular successes of this theory were confined to sufficiently high dimensions. Especially in 3 dimensions, arbitrary surgery or cutting and pasting is too disruptive, and incompatible with even the coarsest features of the classification theory such as the prime decomposition and JSJ theorems. Three-manifold topology developed quite independently, reaching a culmination, from the classification perspective, with Thurston and Perelman.

From a completely different direction, low-dimensional topology has been invigorated over the last two decades by ideas from physics, especially quantum field theory. Classically, one studies *fields* (e.g. functions, or sections of some bundle) and their dynamics on Euclidean space, or on some smooth manifold. From quantum physics one gets the key idea of *superposition*, i.e., that one should study complex linear combinations of fields. A combination of these two ideas (cobordism and superposition) unexpectedly resonates in low-dimensional topology, and gives rise to a host of beautiful and subtle invariants, such as the Jones polynomial, Reshetikhin-Turaev-Viro invariants, and Chern-Simons partition functions.

In fact, these invariants are perhaps *too* subtle. After twenty years of work, it is profoundly frustrating that we cannot say precisely what these invariants measure or what they distinguish, and despite some tantalizing hints (e.g. Kashaev’s conjecture), these invariants remain disconnected from the (hugely successful) Thurston theory of 3-manifolds. There are two key questions: what information does 3-dimensional quantum topology distinguish in principle, particularly in the physically motivated unitary case, and how does it relate in detail to the structure theory of 3-manifolds revealed by the geometrization program?

This paper addresses both of these questions simultaneously. Firstly, we establish positivity for the *universal manifold pairing* in three-dimensions, an abstract “universal topological quantum field theory” which is, by construction, sensitive to all details of 3-manifold topology. Three dimensions is the critical case here: in

Received by the editors February 29, 2008.

2000 *Mathematics Subject Classification*. Primary 57R56; Secondary 57M50.

The first author was partially funded by NSF grants DMS 0405491 and DMS 0707130.

©2009 American Mathematical Society
Reverts to public domain 28 years from publication

two dimensions and lower, such positivity is straightforward to establish; in four dimensions and higher, positivity fails badly (see [15] and [25]). For instance, the partition function of a unitary $(3 + 1)$ -dimensional TQFT is equal on s -cobordant 4-manifolds regardless of their Donaldson invariants ([15]). Secondly, the process of establishing positivity turns out to involve the construction of a complexity function on closed 3-manifolds which involves input from every aspect of the geometric theory of 3-manifolds, and obeys a (highly nontrivial) gluing axiom, which may be thought of as a kind of *topological Cauchy-Schwarz inequality*.

We now make this discussion more precise. We work exclusively in the category of smooth, compact, oriented manifolds. Let S denote a closed $(n - 1)$ -dimensional manifold which is not assumed to be connected. Let $\dot{\mathcal{M}}(S)$ denote the set of isomorphism classes of n -manifolds marked by an identification of their boundary with S . That is, if A, B are compact n -manifolds with $\partial A = S = \partial B$, then A and B represent the same element of $\dot{\mathcal{M}}(S)$ if and only if there is a diffeomorphism from A to B extending the identity map on their boundaries. Inspired by the superposition principle, we let $\mathcal{M}(S)$ denote the complex vector space spanned by the set $\dot{\mathcal{M}}(S)$ (i.e., finite \mathbb{C} -linear combinations of elements of $\dot{\mathcal{M}}(S)$). If S is empty, we abbreviate $\dot{\mathcal{M}}(\emptyset)$ and $\mathcal{M}(\emptyset)$ by $\dot{\mathcal{M}}$ and \mathcal{M} , respectively.

Given $A, B \in \dot{\mathcal{M}}(S)$, let \bar{B} denote B with the opposite orientation. Note that $\bar{B} \in \dot{\mathcal{M}}(\bar{S})$. By abuse of notation, we denote by AB the result of gluing A to \bar{B} by the identity map on their boundaries.

Our central object of study is the complex sesquilinear pairing

$$\mathcal{M}(S) \times \mathcal{M}(S) \longrightarrow \mathcal{M}$$

defined by the formula

$$\sum a_i A_i \times \sum b_j B_j \longrightarrow \sum a_i \bar{b}_j A_i B_j.$$

This pairing is known as the *universal pairing* associated to S , and is denoted $\langle \cdot, \cdot \rangle_S$. A pairing on a vector space is *positive* if $\langle v, v \rangle = 0$ if and only if $v = 0$. With this terminology, our main theorem is the following.

Theorem A. *For all closed, oriented surfaces S , the pairing $\langle \cdot, \cdot \rangle_S$ is positive.*

In a necessarily nonpositive pairing, a vector v satisfying $v \neq 0$ and $\langle v, v \rangle = 0$ is said to be *lightlike*.

The motivation for studying this abstract pairing comes from the examples from quantum topology briefly alluded to above, most of which are examples of *topological quantum field theories*, or TQFTs for short. In Atiyah's formulation ([2]) a TQFT is a functor Z from the category of $(n - 1)$ -manifolds and cobordisms between them, to the category of \mathbb{C} vector spaces and linear maps between them, satisfying the monoidal axiom

$$Z(S_1 \amalg S_2) = Z(S_1) \otimes Z(S_2).$$

In particular, one has $Z(\emptyset) = \mathbb{C}$. If $\emptyset \xrightarrow{A} S$ is a cobordism from \emptyset to S (i.e., an element of $\dot{\mathcal{M}}(S)$), then by abuse of notation, we define $Z(A) \in Z(S)$ to be the image of $1 \in \mathbb{C} = Z(\emptyset)$ under the map $Z(A) : Z(\emptyset) \rightarrow Z(S)$ associated to the cobordism.

Remark 1.1. For some reason, it is common practice to use the letters V and Z to denote the image of the functor on an $(n - 1)$ -manifold and an n -manifold,

respectively, so that for instance $Z(M) \in V(S)$ for $M \in \dot{\mathcal{M}}(S)$. We typically do not use this convention when discussing abstract TQFTs, but we do sometimes when discussing specific TQFTs in order to be consistent with the wider literature.

In many interesting TQFTs, the images $Z(A)$ span $Z(S)$. For example, if Z is the $SU(2)$ Chern-Simons partition function at level k where $k + 2$ is prime, then Roberts [39] has shown that the images $Z(A)$ over all $A \in \dot{\mathcal{M}}(S)$ span $Z(S)$. In this case, there is a natural pairing

$$Z(S) \times Z(\overline{S}) \rightarrow Z(\emptyset) = \mathbb{C}$$

defined on generators by composition of cobordisms:

$$\emptyset \xrightarrow{A} S \circ S \xrightarrow{B} \emptyset = \emptyset \xrightarrow{AB} \emptyset$$

and extended by sesquilinearity to

$$\langle \cdot, \cdot \rangle_Z : \mathcal{M}(S) \times \mathcal{M}(S) \rightarrow \mathbb{C}.$$

In physical quantum field theories, $Z(S)$ denotes the vector space of quantum fields on a spacelike slice S of spacetime. The vector space $Z(S)$ is naturally a (positive definite) *Hilbert space*. This motivates an additional axiom for so-called *unitary* TQFT's, that $Z(S)$ (now usually finite dimensional) should admit the natural structure of a Hilbert space, and for every nonzero $A \in \mathcal{M}(S)$, the pairing defined above should satisfy

$$\langle Z(A), Z(\overline{A}) \rangle_Z = Z(A\overline{A}) = \|A\|^2 > 0.$$

It follows that studying positivity of the universal pairing (in a given dimension) is tantamount to studying what kind of topological information in principle might be extracted from unitary TQFTs in that dimension, since lightlike vectors must map to zero in any unitary TQFT. As a specific and important example, given any compact Lie group G and level $k > 0$, the Reshetikhin-Turaev TQFT [38], denoted by $V_{G,k}$, and as reconstructed by [3] fits into the following diagram:

$$\begin{array}{ccc} \mathcal{M}(S) \times \mathcal{M}(S) & \xrightarrow{\pi} & \mathcal{M}(S) \\ \downarrow \alpha & & \downarrow \alpha \\ V_{G,k} \times V_{G,k} & \xrightarrow{A} & \mathbb{C} \end{array} \quad \begin{array}{c} \downarrow Z_{G,k} \\ \mathbb{C} \end{array}$$

where π is the universal pairing, and A is Atiyah's map for (G, k) . Since the pairing A makes $V_{G,k}$ into a Hilbert space, if v is lightlike, then necessarily $\alpha(v) = 0$.

Given a sesquilinear pairing on a complex vector space, positivity of the pairing is logically equivalent to the Cauchy-Schwarz inequality. In our context, positivity is established by a topological variant of this inequality. How can positivity fail? Only if there is a nonzero vector $v = \sum_i a_i A_i$ with the property that there are enough topological "coincidences", i.e. equalities of the form $A_i A_j = A_k A_l$ for various i, j, k, l , so that the coefficients, when collected according to diffeomorphism types, can cancel and the result of pairing v with itself is trivial. In order to rule out such a possibility we construct a complexity function, defined on all closed, oriented 3-manifolds, which is maximized *only* on terms of the form $A_i A_i$. Since the coefficients of such terms are of the form $a_i \bar{a}_i = \|a_i\|^2 > 0$, there can be no

cancellation, and the result of self-pairing is never zero. Explicitly, the complexity function we define satisfies the following gluing axiom:

Topological Cauchy-Schwarz inequality. *There is an ordered set \mathcal{O} and a complexity function c defined on the set $\dot{\mathcal{M}}$ of all closed oriented 3-manifolds, and taking values in \mathcal{O} , so that for all $A, B \in \dot{\mathcal{M}}(S)$,*

$$c(AB) \leq \max(c(AA), c(BB))$$

with equality if and only if $A = B$.

Remark 1.2. It is natural to ask to what extent this analogy with the ordinary Cauchy-Schwarz inequality can be sharpened.

Question 1.3. Is there an embedding $\mathcal{O} \rightarrow \mathbb{R}$ such that the resulting complexity function $\bar{c} : \dot{\mathcal{M}} \rightarrow \mathbb{R}$ satisfies

$$\bar{c}(AB)^2 \leq \bar{c}(AA)\bar{c}(BB)$$

for all $A, B \in \dot{\mathcal{M}}$, with equality if and only if $A = B$?

Note that for any finite subset Σ of $\dot{\mathcal{M}}$, the values of \bar{c} on Σ can be chosen to satisfy the inequality above. The subtlety of this question concerns the order structure of \mathcal{O} , particularly its accumulation points.

Although our main theorem has a very clean statement, it obscures an Augean complexity in the definition of the function c and the ordered set \mathcal{O} . These objects involve highly nontrivial input from many aspects of geometric 3-manifold topology. This complexity mirrors the complexity inherent in the geometrization theorem for 3-manifolds. The geometric classification theorem for 3-manifolds has a *hierarchical* structure. A manifold is subject to a repeated sequence of decompositions, each of which only makes sense after the manifold has been filtered through the previous decompositions (see Figure 1). The complexity function c has terms which are sensitive to all different levels of the decomposition. Only if two manifolds are identical with respect to the first few levels of the decomposition does c discriminate between them based on features which make sense at later levels. A flowchart (Figure 2) summarizing the main components of the complexity function c is included to assist the reader in understanding the logic of the argument.

Precisely because the function c involves input from many different aspects of 3-manifold topology, several steps in the proof of the topological Cauchy-Schwarz inequality are of independent interest to 3-manifold topologists. These include:

- The use of Dijkgraaf-Witten finite group untwisted TQFTs to determine handlebody and compression body splittings.
- The construction of a $\mathbb{Z}[\frac{1}{2}]$ -valued invariant of a nonclosed Seifert fibered piece X in the canonical decomposition of a prime 3-manifold; this invariant, a kind of relative Euler class, measures the degree to which the fibers of X along ∂X are twisted as they are glued to neighboring JSJ pieces.
- The proof of a volume inequality for cusped hyperbolic manifolds obtained by gluing manifolds with incompressible boundary, which generalizes a similar inequality for closed manifolds obtained recently by Agol-Storm-Thurston [1].

These and other broader implications of the work are discussed briefly as they arise, and also in an appendix.

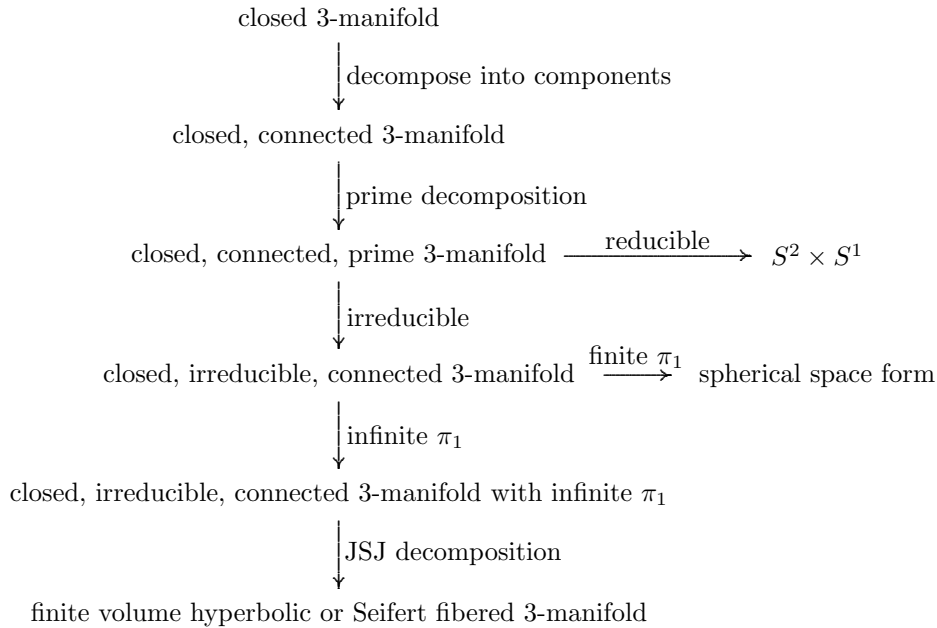


FIGURE 1. Repeated decompositions in the geometrization theorem

It should be remarked that a surprisingly large fraction of the proof involves assembling the individual complexities of hyperbolic and Seifert fibered pieces together. In some ways the study of nontrivial JSJ decompositions is an underestimated area of 3-manifold topology, with many practitioners preferring to work in either the “pure” hyperbolic or Seifert fibered (or graph manifold) world. However, there are important historical cases where the synthesis of these two worlds, Seifert fibered and hyperbolic, is a crucial and delicate issue, essential to obtaining the mathematical result. Especially noteworthy cases are the resolution of the Smith conjecture, and the cyclic surgery theorem. In both cases (as well as many others), Cameron Gordon played the critical role both in terms of mathematics and leadership, in this synthesis. Architecturally therefore, this paper should be seen as part of a tradition in 3-manifold topology of which Cameron is perhaps the most preminent champion, and it is our pleasure to acknowledge his multi-generational inspiration.

Note that the assembly of a 3-manifold from its JSJ pieces amounts to a kind of “decorated” graph theory which encompasses many subtle points. In §6.2 we obtain new positivity results for certain natural “graph TQFTs” as a warm-up to the 3-manifold case.

2. UNIVERSAL PAIRINGS

2.1. Definition.

Notation 2.1. Throughout the paper, S will usually denote a closed oriented surface, possibly disconnected. A, B, A_i, B_i and so on will denote 3-manifolds, usually compact with boundary equal to S , and M, N, M_i, N_i and so on will usually denote closed 3-manifolds. Sometimes these conventions are relativized, so that S or M

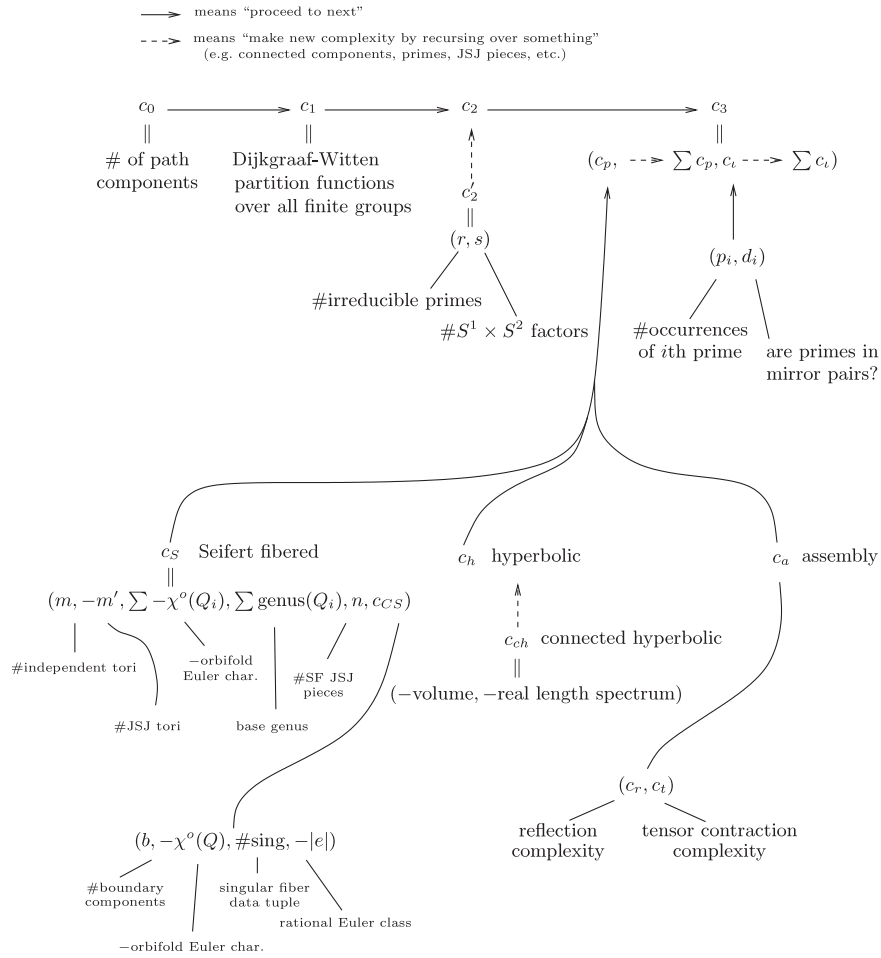
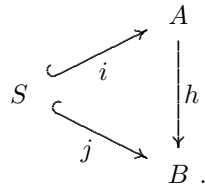


FIGURE 2. This flow chart may assist in navigating the proof.

might denote a complete manifold with a cusp, and A or B might denote a complete manifold whose boundary is a (possibly cusped) surface S .

In what follows, let S be a closed, oriented surface. We explicitly allow S to be disconnected. Let $\mathcal{M}(S)$ denote the set of smooth pairs (A, S) , where A is a compact oriented 3-manifold and $\partial A = S$, up to the equivalence relation $(A, S) \sim (B, S)$ if there is a diffeomorphism $h : A \rightarrow B$ such that $hi = j$, where $i : S \rightarrow A$ and $j : S \rightarrow B$ are the respective inclusions. That is, the following diagram commutes:



Let $\mathcal{M}(S)$ denote the \mathbb{C} -vector space spanned by the set $\dot{\mathcal{M}}(S)$. If S is empty, abbreviate $\dot{\mathcal{M}} := \dot{\mathcal{M}}(\emptyset)$ and $\mathcal{M} := \mathcal{M}(\emptyset)$.

Definition 2.2. The *universal Hermitian pairing* is the map

$$\langle \cdot, \cdot \rangle_S : \mathcal{M}(S) \times \mathcal{M}(S) \rightarrow \mathcal{M}$$

defined by the formula

$$\left\langle \sum_i a_i A_i, \sum_j b_j B_j \right\rangle_S = \sum_{i,j} a_i \bar{b}_j A_i B_j,$$

where $A_i B_j$ denotes the closed 3-manifold obtained by gluing A_i to \bar{B}_j (i.e., the manifold B_j with the opposite orientation) along S , using the *canonical* inclusion of S into either factor.

Definition 2.3. Let V, W be \mathbb{C} -vector spaces. A pairing $\langle \cdot, \cdot \rangle$ on V with values in W is *positive* if $\langle v, v \rangle = 0$ if and only if $v = 0$.

The main theorem in this paper is

Theorem A. *For all closed, oriented surfaces S , the pairing $\langle \cdot, \cdot \rangle_S$ is positive.*

The proof involves many ingredients and will occupy most of the remainder of the paper.

2.2. Complexity and diagonal dominance. The key to the proof of Theorem A is the construction of a suitable complexity function on compact, closed, oriented (not necessarily connected) 3-manifolds.

Definition 2.4. Let \mathcal{O} be an ordered set. A function $c : \dot{\mathcal{M}} \rightarrow \mathcal{O}$ is *diagonally dominant* for S if for all *distinct* pairs $A, B \in \dot{\mathcal{M}}(S)$ there is a strict inequality

$$c(AB) < \max(c(AA), c(BB)).$$

The following lemma shows how Theorem A follows from diagonal dominance.

Lemma 2.5 (Diagonal dominance implies positive). *Suppose there is a function c which is diagonally dominant for S . Then $\langle \cdot, \cdot \rangle_S$ is positive.*

Proof. Let $v = \sum_i a_i A_i$ with $a_i \neq 0$ and set $w = \langle v, v \rangle_S$. The vector w is a linear combination of manifolds of the form $M_i M_j$. By diagonal dominance, the maximum of c on the set $A_i A_j$ is realized only on manifolds of the form $A_i A_i$. Let i be such that $c(A_i A_i)$ is maximum. Let J be the set of indices j for which $A_j A_j = A_i A_i$, and observe that $A_j A_k \neq A_i A_i$ for $j \neq k$. It follows that the coefficient of w on the manifold $A_i A_i \in \mathcal{M}$ is $\sum_{j \in J} |a_j|^2 > 0$. Hence $w \neq 0$. \square

Most of the remainder of the paper will be concerned therefore with defining a suitable complexity function and proving that it satisfies diagonal dominance. In fact, the function c we define does not depend on S .

The statement that such a diagonally dominant function exists is paraphrased in the introduction as the *topological Cauchy-Schwarz inequality*, and we will sometimes use such a term interchangeably with diagonal dominance.

2.3. Overview. The complexity function c is constructed from a tuple of more specialized complexities c_0, c_1, c_2 and c_3 , each of which treats some narrower aspect of 3-manifold topology. Roughly, c_0 addresses connectivity, c_1 will tell us about the kernel of the map on fundamental groups induced by the inclusion of S into a bounding 3-manifold, c_2 is concerned with essential 2-spheres, while c_3 addresses the nature and multiplicities of the prime factors (as a mnemonic, think of c_i as addressing some aspect of π_i for each of $0 \leq i < 3$).

Write $c^0 := c_0$ and $c^1 := c_0 \times c_1$ for the lexicographic pair which first maximizes c_0 , then c_1 . Similarly, $c^2 := c_0 \times c_1 \times c_2 = c^1 \times c_2$, and finally, $c := c^3 := c_0 \times c_1 \times c_2 \times c_3 = c^2 \times c_3$.

Warning. The functions c_1, c_2 , and c_3 are themselves complicated and are built up from many simpler complexity functions.

For each $0 \leq k \leq 3$ and each complexity c^k we will prove a lemma of the form

Lemma Schema. *For all pairs $A, B \in \dot{\mathcal{M}}(S)$, either*

$$c^k(AB) < \max(c^k(AA), c^k(BB))$$

or $c^k(AB) = c^k(AA) = c^k(BB)$, and (A, S) and (B, S) are indistinguishable with respect to a certain criterion C_k .

The criteria C_k are associated to the complexities c^k in a certain way and are chosen so that if (A, S) and (B, S) are indistinguishable with respect to criteria C_k for all $0 \leq k \leq 3$, then they represent the same element of $\dot{\mathcal{M}}(S)$. Actually, this fact is absorbed into the statement of the c^3 -Lemma Schema, which is Theorem 3.27.

This fact together with Lemma Schema for $0 \leq k \leq 3$ together imply the existence of a diagonally dominant function (i.e., the topological Cauchy-Schwarz inequality), and therefore, together with Lemma 2.5, they prove Theorem A.

The complexity c_3 is itself a lexicographic tuple $c_3 = c_S \times c_h \times c_a$, where the subscripts S, h, a stand for *Seifert fibered*, *hyperbolic* and *assembly* complexities, respectively. The Lemma Schema for c_3 involves in its turn a version of the Lemma Schema for c_S, c_h, c_a . Moreover each of these individual internal complexity functions has its own internal structure involving simpler terms.

Here is the plan for the next few sections. In §3 we state and prove the Lemma Schema for c_0, c_1, c_2 . The Seifert fibered complexity c_S is treated in §4, the hyperbolic complexity c_h in §5, and the assembly complexity (which encodes the gluing up of the Seifert fibered and hyperbolic pieces) in §6.

3. PRIME DECOMPOSITION

In this section we state and prove the Lemma Schema for the complexities c_0, c_1, c_2 and the respective criteria C_0, C_1, C_2 . Then we show how these Lemma Schemas, together with a complexity function c_3 on irreducible 3-manifolds satisfying suitable properties imply Lemma 2.5 and therefore Theorem A, deferring the precise definition and proof of properties of c_3 to future sections.

3.1. The complexity c_0 . The definition of c_0 is straightforward.

Definition 3.1. Let $c_0 : \dot{\mathcal{M}} \rightarrow \mathbb{Z}$ assign to a closed 3-manifold M the number of connected components of M (an empty manifold has zero components).

Definition 3.2. Given $(A, S) \in \dot{\mathcal{M}}(S)$ the *connectivity partition* associated to A is the functor π_0 applied to the pair (A, S) .

In other words, the connectivity partition is the pair of sets $\pi_0(S), \pi_0(A)$ and the map $\pi_0(S) \rightarrow \pi_0(A)$ induced by inclusion $S \rightarrow A$. The data is equivalent to the information about which components of S are included into the same component of A , and how many closed components A has that are disjoint from S .

Criterion C_0 tries to distinguish elements of $\dot{\mathcal{M}}(S)$ using their connectivity partitions. Two elements $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ are indistinguishable with respect to C_0 if and only if they induce the same connectivity partition.

Lemma 3.3 ($c^0 = c_0$ -Lemma Schema). *Either*

$$c_0(AB) < \max(c_0(AA), c_0(BB))$$

or $c_0(AB) = c_0(AA) = c_0(BB)$ and both (A, S) and (B, S) induce the same connectivity partition.

Proof. The key observation is that c_0 is maximized only when at most one component from each of A and B contributes to a component of AB . \square

3.2. The complexity c_1 . The complexity c_1 is less trivial.

To define c_1 , begin with an ordered list F_1, F_2, F_3, \dots of all finite groups up to isomorphism. According to [10] there is a unique untwisted TQFT over \mathbb{R} associated to each F_k .

This TQFT associates to every closed surface S a finite dimensional \mathbb{R} -vector space $V_k(S)$ with an inner product $\langle \cdot, \cdot \rangle_k$, and to every element $(A, S) \in \dot{\mathcal{M}}(S)$ a vector $Z_k(A, S) \in V_k(S)$ called the (relative) *partition function* of the pair. When $S = \emptyset$ one has $V_k(\emptyset) = \mathbb{R}$, and the real number $Z_k(M) \in \mathbb{R}$ is the *partition function* of M .

The *gluing axiom* for a TQFT states that for any two elements $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ there is an equality

$$(3.2.1) \quad Z_k(AB) = \langle Z_k(A, S), Z_k(B, S) \rangle_k.$$

The crucial additional property of $\langle \cdot, \cdot \rangle_k$ is that it is positive definite (i.e., it has Euclidean signature). For a more detailed discussion of TQFTs and a more precise definition of Z_k and V_k , see Appendix A.

Definition 3.4. $c_1(M) = (Z_1(M), Z_2(M), Z_3(M), \dots)$ as a lexicographic tuple of real numbers.

Definition 3.5. Given $(A, S) \in \dot{\mathcal{M}}(S)$, the *compression set* is the set of isotopy classes of embedded loops in S which bound embedded disks in A .

Criterion C_1 tries to distinguish elements of $\dot{\mathcal{M}}(S)$ using their compression sets. Two elements $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ are indistinguishable with respect to C_1 if and only if they have the same compression sets.

Remark 3.6. Dehn's Lemma and the loop theorem (see for example [21], Chapter 4) imply that an essential simple closed curve $\gamma \subset S$ is in the compression set if and only if its image in $\pi_1(A)$ is trivial, under the homomorphism $\pi_1(S) \rightarrow \pi_1(A)$ induced by inclusion $S \rightarrow A$. From this and from Nielsen's theorem on presentations of free groups, one sees that the data of a compression set is equivalent to the functor π_1 (keeping track of basepoints, since S might be disconnected) applied to the pair (A, S) .

Remark 3.7. As far as we know, it is possible that $c_1(AB) = c_1(AA) = c_1(BB)$ already implies the existence of an isomorphism $h : \pi_1(A) \rightarrow \pi_1(B)$ making the following diagram commute:

$$\begin{array}{ccc} & & \pi_1(A) \\ & \nearrow & \downarrow h \\ \pi_1(S) & & \\ & \searrow & \downarrow \\ & & \pi_1(B) \end{array}$$

but we are not able to prove or disprove this. Essentially, the question is whether the finite group TQFTs, taken together, determine the induced map on π_1 . There seems to be no known technology that would address this question. However, if it were known to be true, the remainder of the proof could be shortened considerably.

Lemma 3.8 (c^1 -Lemma Schema). *Either*

$$c^1(AB) < \max(c^1(AA), c^1(BB))$$

or $c^1(AB) = c^1(AA) = c^1(BB)$ and both (A, S) and (B, S) induce the same compression set.

Proof. Since the inner product on $V_k(S)$ is Euclidean, the (ordinary) Cauchy-Schwarz inequality implies that either

$$(3.2.2) \quad \langle Z_k(A), Z_k(B) \rangle_k < \max(\langle Z_k(A), Z_k(A) \rangle_k, \langle Z_k(B), Z_k(B) \rangle_k)$$

or $Z_k(A) = Z_k(B)$, in which case all three inner products are equal.

Suppose there is an essential simple closed curve $\gamma \subset S$ which bounds a disk in A but not in B . By Dehn's Lemma (see Remark 3.6) the loop γ is homotopically essential in B . Since we may assume without loss of generality that $c^0(AB) = c^0(AA) = c^0(BB)$, the inclusions $S \rightarrow A, S \rightarrow B$ induce isomorphic maps on π_0 , and therefore we may restrict attention to the components A^0, B^0, S^0 containing γ . Choosing a basepoint somewhere on γ lets us define the image of γ in π_1 of these three spaces. By abuse of notation, we denote the image in $\pi_1(S^0)$ by γ , and the images in $\pi_1(A^0), \pi_1(B^0)$ by $i_*(\gamma), j_*(\gamma)$, respectively.

Claim. There is a finite group F_k and a homomorphism $\theta : \pi_1(B^0) \rightarrow F_k$ for which $\theta \circ j_*(\gamma) = x \neq 1$.

□

Proof. The fundamental group of B^0 is a free product of factors, corresponding to the prime decomposition of B^0 . The factor containing the image of γ also contains the image of $\pi_1(S^0)$. Since γ is nontrivial in S^0 , the genus of S is positive, and the rank of H_1 of this factor is positive. In particular, the factor containing the image of γ is π_1 of a Haken manifold (see [21], Chapter 13 for a discussion of Haken manifolds). A theorem of Hempel [22] states that all Haken 3-manifolds have residually finite fundamental groups, and the claim follows. □

Geometrically, the homomorphism θ defines a principal F_k bundle β' over S^0 (by the Borel construction) which can be extended as a product to a bundle β over all S . The bundle β extends over B but does not extend over A (see Figure 3). There is a natural basis for $Z_k(S)$ consisting of principal F_k bundles of S up to

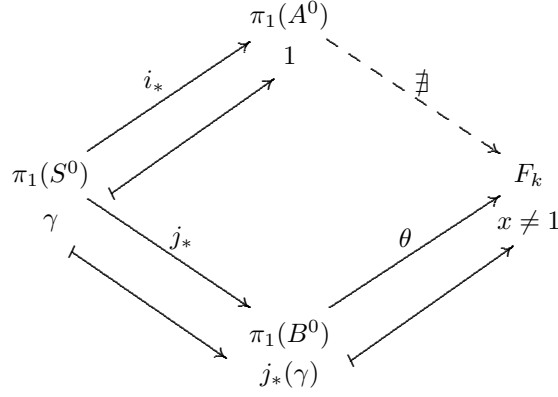


FIGURE 3. Since $\theta \circ j_*(\gamma) \neq 1$, the dashed arrow cannot be filled in to a commutative diagram.

isomorphism, such that the components of $Z_k(A) \in Z_k(S)$ count the number of extensions (divided by the order of the fiber symmetry) for each principal F_k bundle over S (see Appendix A). The component of $Z_k(B)$ in the direction β is nonzero, whereas the component of $Z_k(A)$ in the direction β is zero. In particular,

$$(3.2.3) \quad Z_k(A) \neq Z_k(B) \in V_k(S).$$

By equations (3.2.1), (3.2.2), and (3.2.3) it follows that there is an inequality

$$(3.2.4) \quad Z_k(AB) < \max(Z_k(AA), Z_k(BB))$$

and the lemma is proved.

3.3. The complexity c_2 . The complexity c_2 is a function of the decomposition of a closed 3-manifold into its irreducible factors. Similarly, two manifolds (A, S) and (B, S) in $\mathcal{M}(S)$ are indistinguishable with respect to criterion C_2 if (roughly) they have “identical” sphere and disk decompositions into irreducible pieces.

There are various ambiguities in the way in which a 3-manifold decomposes into its prime factors, one of which arises from the $S^1 \times S^2$ factors. Such factors are associated to nonseparating spheres, but the way in which they sit in the manifold is not unique. Roughly speaking, one $S^1 \times S^2$ factor can be “slid” over another; this operation is called a *handle slide*. A similar ambiguity holds for decompositions for manifolds with boundary.

If M is a closed 3-manifold, let $r(M)$ denote the number of irreducible summands in the prime decomposition of M , and let $s(M)$ denote the number of $S^1 \times S^2$ summands. The three-sphere is considered a unit, not a prime, so $r(S^3) = s(S^3) = 0$.

Define $c'_2(M)$ to be the lexicographic pair $(r(M), s(M))$, and let M_1, M_2, \dots denote the connected components of M , arranged in (any) order of nonincreasing complexity $c'_2(M_i)$.

Definition 3.9. $c_2(M) = (c'_2(M), c'_2(M_1), c'_2(M_2), \dots)$ as a lexicographic tuple of lexicographic pairs.

Convention 3.10. Here and throughout, we make the convention that when we compare lexicographic lists of different lengths, the shorter list is padded by $-\infty$

characters, where $-\infty$ is strictly less, in the relevant partial ordering, than any other term. So if two lists agree on their common lengths, the *longer* list has the *greater* lexicographic complexity.

In order to understand how c_2 behaves under gluing, we adopt a more graphical way of computing $r(M)$ and $s(M)$.

Definition 3.11 (Sum graph). Let M be a closed, orientable 3-manifold, and let \mathcal{S} be a pairwise disjoint collection of embedded 2-spheres in M . The *sum graph* $G(\mathcal{S})$ is a finite decorated graph, defined as follows.

- There is one edge for each 2-sphere of \mathcal{S} .
- There is one vertex for each connected component of $M \setminus \mathcal{S}$.
- Each edge is attached to the two (not necessarily distinct) components of $M \setminus \mathcal{S}$ corresponding to the two sides of the associated sphere.
- For each vertex associated to a component N of $M \setminus \mathcal{S}$, label the vertex by \overline{N} , the result of capping off the 2-sphere boundary components of N . (The bar notation has a different meaning elsewhere in this paper, but the intended meaning should always be clear from the context.)

A collection \mathcal{S} is *adequate* if all vertex labels are irreducible. The existence of an adequate collection is due to Kneser.

We define a *thin* vertex to be one whose label \overline{N} is S^3 ; other vertices are *fat*. A *thin subgraph* consists of some (any) union of edges and thin vertices.

Remark 3.12. Note that adequate collections \mathcal{S} are allowed to have “superfluous” spheres, i.e., spheres which bound 3-balls, or parallel collections of spheres.

A given manifold M admits many different adequate collections \mathcal{S} of spheres. The following theorem, due to Laudenbach, describes the equivalence relation that this ambiguity induces on sum graphs.

Theorem 3.13 (Laudenbach [27]). *If $\mathcal{S}, \mathcal{S}'$ are adequate collections of spheres for M , then $G(\mathcal{S})$ and $G(\mathcal{S}')$ are related by a sequence of edge slides and replacements of a thin subgraph with a homotopy equivalent thin subgraph. Conversely, any sequence of such moves on a sum graph $G(\mathcal{S})$ produces a sum graph $G(\mathcal{S}')$, where \mathcal{S}' is adequate if and only if \mathcal{S} is.*

Figure 4 shows two different but equivalent sum graphs associated to the same manifold.

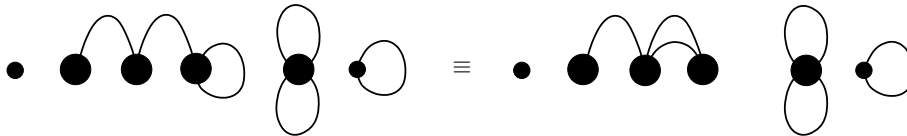


FIGURE 4. Two different but equivalent sum graphs for a manifold of four components. The first component is S^3 , and the other three components have 1, 2, and 1 factors of $S^1 \times S^2$, respectively.

The numbers $r(M)$ and $s(M)$ can be recovered from any sum graph for M .

Lemma 3.14. *Let \mathcal{S} be adequate for M . The number of fat vertices of $G(\mathcal{S})$ is equal to $r(M)$. Moreover, $s(M) = \dim(H_1(G(\mathcal{S}); \mathbb{R}))$.*

Proof. This follows immediately from the definitions. \square

To describe criterion C_2 , we must relativize the definition of sum graphs to compact 3-manifolds with boundary.

Definition 3.15 (Relative sum graph). Let A be a compact, orientable 3-manifold with boundary, and let \mathcal{D} be a pairwise disjoint collection of embedded 2-spheres and properly embedded disks in A . The *relative sum graph* $G(\mathcal{D})$ is a finite decorated graph, defined as follows.

- There is one edge for each element of \mathcal{D} ; an edge associated to a sphere is *regular*; otherwise it is a *half edge*.
- There is one vertex for each connected component of $A \setminus \mathcal{D}$.
- Edges join the (not necessarily distinct) components corresponding to their two sides in A .
- A vertex corresponding to a component disjoint from ∂A is *regular*; otherwise it is a *half vertex*.
- A regular vertex is thin or fat as before depending on whether its label is S^3 or not.
- A half vertex associated to a component N determines a label \overline{N} obtained by capping off sphere components of ∂N which are disjoint from ∂S ; a half vertex is *thin* if \overline{N} is B^3 and *fat* otherwise.

The collection \mathcal{D} is *adequate* if all regular vertex labels are irreducible, and all half vertex labels are irreducible and boundary irreducible.

The analogue of Theorem 3.13 holds for relative sum graphs:

Theorem 3.16 (Laudenbach [27]). *Two adequate collections $\mathcal{D}, \mathcal{D}'$ for M determine relative sum graphs $G(\mathcal{D}), G(\mathcal{D}')$ which are related by edge slides and homotopies of the thin part.*

See also [6] for a discussion.

Remark 3.17. Theorem 3.16 is nontrivial even (or especially) when M is a handlebody and implies (for instance) that any two meridian systems v, v' for a handlebody are related by a finite sequence of *Singer moves*; i.e., one disk in the system v is replaced by a disk contained in the complement of v (see Singer [44]).

Criterion C_2 tries to distinguish elements of $\dot{\mathcal{M}}(S)$ using their relative sum graphs. Two elements $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ are indistinguishable with respect to C_2 if there are adequate collections $\mathcal{D}_A, \mathcal{D}_B$ so that $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ are *isomorphic*.

That is,

- The families of disks in \mathcal{D}_A and \mathcal{D}_B intersect S in the same set of curves, up to isotopy; this determines a canonical isomorphism between the sets of half edges in $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$.
- There is an isomorphism between the graphs $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ which is compatible with the canonical isomorphism on half edges, and which takes fat (resp. thin) vertices to fat (resp. thin) vertices and regular (resp. half) vertices or edges to regular (resp. half) vertices or edges

In words, two elements are indistinguishable with respect to criterion C_2 if they can be cut up into irreducible pieces in a combinatorially isomorphic way (respecting

the intersection with S) and with the same S^3 and B^3 terms, but without paying attention (yet) to the homeomorphism types of the irreducible pieces which are not S^3 or B^3 .

Figure 5 shows an example of a relative sum graph.

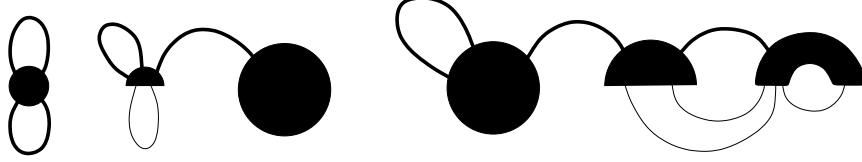


FIGURE 5. An example of a relative sum graph $G(\mathcal{D})$. From the left, the first component is an $S^1 \times S^2 \sharp S^1 \times S^2$ and the second contains an $(S^1 \times S^2 \setminus D^3) \natural S^1 \times D^2$ summand. Every edge is “thin” in the sense of Definition 3.15, but in the figure, half edges are drawn “thinner” than regular edges.

In the sequel, for the sake of legibility, we will refer to relative sum graphs simply as sum graphs.

Lemma 3.18 (c^2 -Lemma Schema). *Either*

$$c^2(AB) < \max(c^2(AA), c^2(BB))$$

or $c^2(AB) = c^2(AA) = c^2(BB)$ and there are adequate collections $\mathcal{D}_A, \mathcal{D}_B$ of disjoint spheres and disks in $(A, S), (B, S)$, respectively, so that the sum graphs $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ are isomorphic.

Proof. Let $\mathcal{D}_A, \mathcal{D}_B$ be adequate collections of disks and spheres for A and B , respectively. Without loss of generality we may assume that $c^1(AB) = c^1(AA) = c^1(BB)$ so that (A, S) and (B, S) are indistinguishable with respect to criteria C_0 and C_1 . It follows that any essential simple closed curve in S which bounds a disk in A also bounds a disk in B , so we may choose $\mathcal{D}_A, \mathcal{D}_B$ so that the disk components have the same boundaries in S , and the half edges of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ match up under the gluing.

Since the definition of adequate requires half vertex labels to be irreducible, any half vertex label whose boundary contains a sphere must be B^3 (i.e., these are exactly the set of thin half vertices). Note that if N is a component corresponding to a half vertex, then each boundary component of N is obtained from a component of $\partial N \cap S$ by filling in boundary components of $\partial N \cap S$ with disks in \mathcal{D} . Thus a boundary component of \bar{N} is a sphere if and only if the corresponding component of $\partial N \cap S$ is a planar surface. In particular, thin half vertices of $G(\mathcal{D}_A)$ precisely match up with thin half vertices of $G(\mathcal{D}_B)$.

Since disk components of \mathcal{D}_A and \mathcal{D}_B match up, the union $\mathcal{S} = \mathcal{D}_A \cup \mathcal{D}_B$ is a pairwise disjoint collection of embedded spheres in AB .

Claim. \mathcal{S} is an adequate collection of spheres for AB . □

Proof. Components of $A \setminus \mathcal{D}_A$ and $B \setminus \mathcal{D}_B$ corresponding to regular vertices are also components of $AB \setminus \mathcal{S}$, so the labels associated to these vertices of $G(\mathcal{S})$ are irreducible. So we just need to check irreducibility for labels of vertices of $G(\mathcal{S})$ corresponding to unions of half vertices of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$.

Let M_i be a set of components of $A \setminus \mathcal{D}_A$ and N_j a set of components of $B \setminus \mathcal{D}_B$ corresponding to half vertices so that the union $M = \bigcup_i M_i \bigcup_j N_j$ is a component of $AB \setminus \mathcal{S}$. We need to check that every sphere in \overline{M} is inessential. First notice that every boundary component of $N := \bigcup_i \overline{M}_i \bigcup_j \overline{N}_j$ is a sphere, obtained from exactly one disk in each of \mathcal{D}_A and \mathcal{D}_B , since the disk components glue up exactly. Then observe that \overline{M} is obtained from N by capping off each of these boundary components with a 3-ball.

Now suppose T is an embedded sphere in \overline{M} . We will show that T bounds a 3-ball in \overline{M} . Since \overline{M} is obtained from M by capping off boundary components with 3-balls, we may assume $T \subset M \subset AB$, and we put T in general position with respect to S . Since $\mathcal{D}_A, \mathcal{D}_B$ are adequate in A, B , respectively, we may boundary compress all components of $T \cap M_i$ in M_i until each component of $T \cap M_i$ is either boundary parallel into S or parallel in M_i to a disk δ_a in \mathcal{D}_A . Disks of the first type can then be pushed across S , reducing the number of intersections of $T \cap S$. On the other hand disks parallel to an element $\delta_a \in \mathcal{D}_A$ can be removed by an isotopy (working innermost to outermost) across the 3-ball in \overline{M} capping off $\delta_a \cup \delta_b$, where $\delta_b \in \mathcal{D}_B$ is the counterpart to δ_a . The isotopy results in a new $T \subset M \subset AB$ with one fewer intersection with S . So after finitely many moves of this kind, T is disjoint from S and is contained (without loss of generality) in some M_i . But \overline{M}_i is irreducible, so T bounds a 3-ball in $\overline{M}_i \subset \overline{M}$.

It follows that the sum graphs $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ for A and B can be combined to give a sum graph $G(\mathcal{S})$ for AB . More specifically,

- Each half edge of $G(\mathcal{D}_A)$ combines with its counterpart in $G(\mathcal{D}_B)$ to produce a (regular) edge of $G(\mathcal{S})$.
- Each thin half vertex of $G(\mathcal{D}_A)$ combines with its counterpart in $G(\mathcal{D}_B)$ to produce a thin vertex of $G(\mathcal{S})$.
- Two or more fat half vertices from $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ combine to produce a single fat vertex of $G(\mathcal{S})$.
- (Less interesting) The regular edges and vertices of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ become edges and vertices of $G(\mathcal{S})$, preserving fatness/thinness in the case of vertices.

Note that the third bullet point is justified by the claim above. Similarly, $G(\mathcal{D}_A)$ (respectively $G(\mathcal{D}_B)$) can be combined with its mirror image to get a sum graph $G(\mathcal{S}_{AA})$ for AA (respectively a sum graph $G(\mathcal{S}_{BB})$ for BB).

We now argue that either $G(\mathcal{D}_A)$ can be chosen to be isomorphic to $G(\mathcal{D}_B)$ or else the diagonal dominance inequality holds.

Let a_1 and b_1 be the number of fat vertices and fat half vertices of $G(\mathcal{D}_A)$, and define a_2 and b_2 similarly. Then the number of fat vertices of $G(\mathcal{S}_{AA})$ is $2a_1 + b_1$, the number of fat vertices of $G(\mathcal{S}_{BB})$ is $2a_2 + b_2$, and the number of fat vertices of $G(\mathcal{S})$ is $a_1 + a_2 + b'$, where $b' \leq \min(b_1, b_2)$. It follows that the diagonal dominance inequality holds unless $a_1 = a_2$ and $b' = b_1 = b_2$. The latter equality forces the fat half vertices of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ to have the same connectivity partition with respect to their common boundary.

Next we look at the second component of c_2 , the first Betti numbers of $G(\mathcal{S}_{AA})$, $G(\mathcal{S}_{BB})$ and $G(\mathcal{S})$.

Let $G_{1/2}(\mathcal{D}_A)$ denote the subgraph of $G(\mathcal{D}_A)$ consisting of half vertices and half edges, and define $G_{1/2}(\mathcal{D}_B)$ similarly. By the remark above, the subgraphs $G_{1/2}(\mathcal{D}_A)$ and $G_{1/2}(\mathcal{D}_B)$ are isomorphic by an isomorphism which respects the

gluing (remember that half edges correspond to disks which are paired according to Criterion C_1).

Because of the isomorphism of $G_{1/2}(\mathcal{D}_A)$ and $G_{1/2}(\mathcal{D}_B)$, the Betti numbers s can be computed from the relative first Betti numbers

$$s_A := \dim(H_1(G(\mathcal{D}_A), G_{1/2}(\mathcal{D}_A); \mathbb{R})), \quad s_B := \dim(H_1(G(\mathcal{D}_B), G_{1/2}(\mathcal{D}_B); \mathbb{R}))$$

by the formula

$$s(AA) = s_A + s_A + C, \quad s(AB) = s_A + s_B + C, \quad s(BB) = s_B + s_B + C,$$

where C is a constant depending only on the (isomorphic) graphs $G_{1/2}(\mathcal{D}_A)$ and $G_{1/2}(\mathcal{D}_B)$. If $s_A \neq s_B$, then the diagonal dominance inequality is satisfied, so assume from now on that they are equal.

The same argument can be applied to each connected component of A and B individually, using Criterion C_0 to observe that the gluing of A and B respects their connectivity partitions. In particular, connected components of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ individually have isomorphic half subgraphs, the same number of fat vertices, and the same Betti numbers. It follows that after edge slides, each component of $G(\mathcal{D}_A)$ is isomorphic to the corresponding component of $G(\mathcal{D}_B)$ by an isomorphism respecting the inclusions of S into A and B . \square

3.4. The complexity c_3 . We come now to the final piece of our complexity, c_3 . To define c_3 , we will leap ahead in the story and assume the following result, which is assembled from results proved in §§4–6 and is itself proved in §6.

Definition 3.19. Let $\dot{\mathcal{P}} \subset \dot{\mathcal{M}}$ denote the set of closed, connected, oriented, irreducible prime 3-manifolds. (We use “irreducible” here to exclude $S^1 \times S^2$.)

Theorem 3.20 (Prime complexity). *There is a complexity function $c_p : \dot{\mathcal{P}} \rightarrow \mathcal{O}_p$ for some totally ordered set \mathcal{O}_p , satisfying $c_p(M) = c_p(\overline{M})$, and such that if (A, S) and (B, S) are distinct connected, orientable, irreducible manifolds with S a non-empty incompressible surface not homeomorphic to S^2 , then*

$$c_p(AB) < \max(c_p(AA), c_p(BB)).$$

Remark 3.21. The hypotheses on (A, S) and (B, S) imply that AB, AA, BB are all in $\dot{\mathcal{P}}$.

Deferring the definition of c_p and assuming the proof of Theorem 3.20, we will now construct c_3 so that $c^3 = c^2 \times c_3$ has the final property we demand of $c := c^3$; namely,

$$c(AB) < \max(c(AA), c(BB))$$

for all *distinct* pairs $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$.

Definition 3.22. A *divisor* on $\dot{\mathcal{P}}$ is a function from $\dot{\mathcal{P}}$ to \mathbb{Z} with finite support. If Σ is a divisor, write $\Sigma \geq 0$ if all the terms are nonnegative.

We sometimes use the circumlocution “a finite subset of $\dot{\mathcal{P}}$ with multiplicity” for a divisor $\Sigma \geq 0$. We extend c_p to nonnegative divisors, as follows:

Definition 3.23. If Σ is a nonnegative divisor, let M_1, M_2, \dots be the elements of $\dot{\mathcal{P}}$ in the support of Σ listed with multiplicity, ordered so that $c_p(M_i)$ is nonincreasing. Then define \bar{c}_p to be the lexicographic list

$$\bar{c}_p(\Sigma) = (c_p(M_1), c_p(M_2), \dots).$$

Remark 3.24. Note that by convention, $\bar{c}_p(0) < \bar{c}_p(\Sigma)$ whenever $\Sigma > 0$, i.e., when the nonnegative divisor Σ is nontrivial.

The set $\dot{\mathcal{P}}$ admits an involution $\iota : \dot{\mathcal{P}} \rightarrow \dot{\mathcal{P}}$ given by reversing orientations. Let $|\dot{\mathcal{P}}|$ denote the set of orbits of $\dot{\mathcal{P}}$ under this involution. Let p_1, p_2, \dots list the elements of $|\dot{\mathcal{P}}|$. For each divisor Σ define $p_i(\Sigma)$ to be the number of $M \in \Sigma$ for which the orbit of M under ι is equal to p_i . Similarly, define $d_i(\Sigma)$ to be equal to 0 if the cardinality of p_i (as an orbit in $\dot{\mathcal{P}}$) is 1, and otherwise to be equal to minus the difference of the value of Σ on the two different elements of the orbit p_i . That is, if $P_i \in \dot{\mathcal{P}}$ is some element in the orbit p_i , define

$$d_i(\Sigma) = -|\Sigma(P_i) - \Sigma(\iota(P_i))|.$$

Observe that $d_i = 0$ if $P_i = \iota(P_i)$, so actually this formula is correct independent of the cardinality of p_i .

Definition 3.25. For each divisor $\Sigma \geq 0$ define $c_\iota(\Sigma)$ to be the lexicographic list

$$c_\iota(\Sigma) = \text{all the } (p_i(\Sigma), d_i(\Sigma))\text{'s listed in nonincreasing order.}$$

For $M \in \dot{\mathcal{M}}$, let $\Sigma(M)$ denote the nonnegative divisor counting the number of irreducible primes in the prime decomposition of M with multiplicity. Note that if M is disconnected, $\Sigma(M) = \sum_i \Sigma(M_i)$, where the M_i are the components of M .

With these preliminary definitions, we are in a position to define c_3 .

Definition 3.26. $c_3(M)$ is the lexicographical tuple

$$c_3(M) = (\bar{c}_p(\Sigma(M)), \bar{c}_p(\Sigma(M_i)), c_\iota(\Sigma(M)), c_\iota(\Sigma(M_i))).$$

Here the terms $\bar{c}_p(\Sigma(M_i))$ and $c_\iota(\Sigma(M_i))$ are shorthand for the lexicographic lists of the complexities \bar{c}_p and c_ι evaluated on the components M_i of M , ordered nonincreasingly. Note that the order of the M_i 's in the two cases might be different.

The terms c_ι become important when no nontrivial gluing as in Theorem 3.20 takes place. In this case, we need to look closely at prime multiplicity.

Theorem 3.27 (c^3 -Lemma Schema). *Let (A, S) and (B, S) be distinct elements of $\dot{\mathcal{M}}(S)$. Then*

$$(3.4.1) \quad c^3(AB) < \max(c^3(AA), c^3(BB)).$$

Proof. Suppose c_0, c_1, c_2 are all equal for AA, AB, BB so that A and B cannot be distinguished by criteria C_0, C_1, C_2 .

Claim. Either $\bar{c}_p(\Sigma(AB)) < \max(\bar{c}_p(\Sigma(AA), \Sigma(BB)))$ or else A and B pair up component by component in such a way that the fat regular vertices of the sum graphs $G(\mathcal{D}_A), G(\mathcal{D}_B)$ correspond to sets of irreducible primes with pairwise equal c_p -complexities, and the fat half vertices of $G(\mathcal{D}_A)$ and $G(\mathcal{D}_B)$ are paired by the gluing in such a way that paired half vertices have identical labels. □

Proof. For each of AA, AB, BB , let $\Sigma^{1/2}(\cdot)$ be the divisor corresponding to the set of manifolds appearing as labels of fat vertices in the sum graphs which are obtained from fat half vertices in the relative sum graphs, and let $\Sigma^r(\cdot)$ denote the other terms in $\Sigma(\cdot)$. Note that

$$\Sigma^r(\cdot) = \Sigma(\cdot) - \Sigma^{1/2}(\cdot)$$

as (nonnegative) divisors. Let $x_{AA} = \bar{c}_p(\Sigma^{1/2}(AA))$, $u_{AA} = \bar{c}_p(\Sigma^r(AA))$ and $p_{AA} = \bar{c}_p(\Sigma(AA))$, and similarly for AA replaced by AB or BB .

Ignoring the orientation of factors at the moment, there is an equality $\Sigma^r(AA) + \Sigma^r(BB) = 2\Sigma^r(AB)$, where we think of each Σ^r as a divisor on $|\dot{\mathcal{P}}|$. It follows, after interchanging the roles of A and B if necessary, that there is an inequality

$$(3.4.2) \quad u_{AA} \geq u_{AB} \geq u_{BB},$$

where both inequalities are strict unless all three terms are equal.

We next turn to the terms x_{AA} , x_{AB} and x_{BB} . These correspond to gluings of fat half vertices and satisfy the hypotheses of Theorem 3.20. By that theorem, either paired fat half vertices have labels which are diffeomorphic by a diffeomorphism preserving the intersection with S , or else

$$(3.4.3) \quad x_{AA} > x_{AB} \text{ or } x_{BB} > x_{AB}.$$

We consider these cases in turn.

Case $x_{AA} > x_{AB}$. Combining with equation (3.4.2) gives $p_{AA} > p_{AB}$ and we are done.

Case $x_{BB} > x_{AB}$, $u_{AB} = u_{BB}$. These two inequalities together imply $p_{BB} > p_{AB}$ and we are also done.

Case $x_{BB} > x_{AB}$, $u_{AB} > u_{BB}$. Combining the second inequality with equation (3.4.2) gives $u_{AA} > u_{AB}$.

Let u'_{AA} be the first term in u_{AA} which differs from a term u'_{AB} in u_{AB} , so that there is an inequality

$$(3.4.4) \quad u'_{AA} > u'_{AB}.$$

Subcase $x_{BB} > u'_{AA}$. That is, assume the first term in x_{BB} is greater than u'_{AA} . Combining with equation (3.4.4) gives $x_{BB} > u'_{AB}$ and therefore $p_{BB} > p_{AB}$ and again we are done.

Subcase $x_{BB} \leq u'_{AA}$. Since this is a subcase of the case $x_{BB} > x_{AB}$, we get $x_{AB} < u'_{AA}$. That is, the first term x'_{AB} in x_{AB} is already less than the first term in u'_{AA} in which case the u terms dominate in p . Since we are assuming throughout this case that $u_{AB} > u_{BB}$, one inequality in equation (3.4.2) is strict, and therefore the other is too; i.e. $u_{AA} > u_{AB}$. Since we have just argued that the u terms dominate in p in this subcase, we deduce $p_{AA} > p_{AB}$, so the claim is proved in this case too.

Thus the only possibility is that labels of paired fat half vertices are diffeomorphic by a diffeomorphism preserving their intersection with S , and further that $u_{AA} = u_{AB} = u_{BB}$, as claimed. \square

Repeating the preceding argument component by component, we either obtain the desired inequality in the second term of c_3 , or else there are equalities

$$\Sigma^{1/2}(A_i A_i) = \Sigma^{1/2}(A_i B_i) = \Sigma^{1/2}(B_i B_i)$$

and

$$\bar{c}_p(\Sigma^r(A_i A_i)) = \bar{c}_p(\Sigma^r(A_i B_i)) = \bar{c}_p(\Sigma^r(B_i B_i))$$

for each pair of components A_i, B_i which are glued up in AB . Since paired fat half vertices in every component are diffeomorphic by a diffeomorphism fixed on S , the

only potential difference between A and B is in the prime irreducible factors which do not intersect S ; call these the *free factors*. By uniqueness of prime factorization for 3-manifolds, A and B are equal if and only if for each pair of corresponding components A_i, B_i the set of free factors appearing in each are equal as divisors.

Ignoring questions of orientation of free factors for the moment, it is clear that either the set of free factors in A_i and in B_i are equal, or for some first $p_k \in |\dot{\mathcal{P}}|$, after possibly interchanging A and B , some component A_i has at least as many copies of p_k as any component B_j , and more copies than B_i . It follows that the maximal value of $p_k(\cdot)$ is achieved on more components of AA than of AB , so that $c_3(AB) < c_3(AA)$ with strict inequality in either the third or fourth term.

However, we are working in the category of oriented 3-manifolds, and we must take into account the fact that fat regular vertices of B contribute terms in AB and BB which appear with reversed orientation. The d_i factors in the complexity c_i favor a perfect balance in the closed manifold between primes and their orientation-reverses.

Since all factors appearing in $\Sigma^{1/2}$ are amphichiral (i.e., they are fixed points of ι), the d_i terms are maximized only when every oriented free factor in each A_i occurs in each B_i with the same multiplicity. So either $c_3(AB) < \max(c_3(AA), c_3(BB))$ with strict inequality in some d_i in the third or fourth term, or each pair A_i and B_i have exactly the same set of free factors with orientation and multiplicity. In the first case, the theorem is proved. In the second case, by what we have already proved about the $\Sigma^{1/2}$ factors, this implies that (A, S) and (B, S) are equal as elements of $\dot{\mathcal{M}}(S)$, contrary to the hypothesis.

This completes the proof of Theorem A modulo the definition of c_p and the proof of Theorem 3.20. This will occupy §§4–6.

4. SEIFERT FIBERED FACTORS

We now restrict attention to irreducible 3-manifolds with incompressible boundary.

Definition 4.1. A *Seifert fibered space* is a compact 3-manifold that admits a foliation by circles. The foliation is called a *Seifert fibration*, and the circles are called the *fibers* of the Seifert fibration. If M is a Seifert fibered space, the space obtained by quotienting the fibers to points is an orbifold Q called the *base* of the fibration. The orbifold points are also called *exceptional points*, and the fibers lying over orbifold points are called *exceptional fibers*.

Fundamental to the theory of Seifert fibered manifolds is the fact that an irreducible 3-manifold admits a natural decomposition into Seifert fibered and atoroidal pieces:

Definition 4.2. Let M be a closed, orientable, irreducible 3-manifold. The *characteristic submanifold* of M , denoted Σ , is a Seifert submanifold of M (possibly disconnected and with boundary) whose complement is atoroidal (and possibly disconnected), and which has the smallest number of boundary components.

Note that the boundary components of an orientable Seifert fibered manifold are all tori.

Theorem 4.3 (JSJ decomposition, [23], [24]). *A closed, orientable, irreducible 3-manifold has a characteristic submanifold which is unique up to isotopy.*

There is a relative version of this theorem for manifolds with incompressible boundary, which we also need.

Definition 4.4. Let M be a compact, orientable, irreducible 3-manifold with incompressible boundary S . The *characteristic submanifold* of M , denoted Σ , is a union of pieces, each of which is one of the following kinds:

- (1) a Seifert submanifold disjoint from S (call these *free factors*);
- (2) a pair (I -bundle, ∂I -bundle) over a surface, where the intersection with S is the ∂I -bundle (call these *proper I -bundles*);
- (3) a Seifert submanifold whose boundary intersects S in a union of fibered annuli and tori.

This union of pieces is uniquely determined by the property that its complement is atoroidal and acylindrical and has the smallest possible number of frontier components.

Theorem 4.5 (Relative JSJ decomposition, [23], [24]). *A compact, orientable, irreducible 3-manifold with incompressible boundary has a characteristic submanifold which is unique up to isotopy.*

Remark 4.6. There is some ambiguity in the fiber structure of a piece of the characteristic submanifold of a manifold (with or without boundary). This ambiguity is discussed in the sequel, and is resolved by certain conventions.

Remark 4.7. The third case in Definition 4.4 is somewhat nonstandard, insofar as we insist that the fibering of a Seifert submanifold X extend to a fibering of its intersection $X \cap S$. This has the following consequence: if T is a torus component of ∂X intersecting S in parallel annuli A_i , and the circle fibration of the A_i does not extend over X , we need to add a parallel copy of $T \times I$ as a piece of the characteristic submanifold which “insulates” ∂X from S . If this $T \times I$ is glued up along subannuli with nontrivial Seifert fibered pieces on the other side of S , it will survive as part of a Seifert fibered component of the JSJ decomposition of the closed manifold. Otherwise, it will be “reabsorbed” into the boundary of X .

4.1. Surfaces of finite type. The pieces which arise in the JSJ decomposition are canonical (up to isotopy), but they typically have boundary, consisting of a union of tori. It is sometimes desirable (especially when discussing the atoroidal pieces) to remove these boundary tori and consider the open manifold which is the interior of the manifold with boundary.

It is therefore convenient to extend the definition of $\dot{\mathcal{M}}(S)$ to the case that S is a surface of finite type.

Let S be an oriented surface of finite type (i.e., S is homeomorphic to a closed surface with finitely many points removed). Let $\dot{\mathcal{M}}(S)$ denote the set of smooth pairs (A, S) where

- A is an orientable 3-manifold with $\partial A = S$,
- $A \setminus S$ is homeomorphic to the interior of a compact 3-manifold \widehat{A} , and
- $\partial \widehat{A}$ decomposes as a union $\partial \widehat{A} = \partial_v \widehat{A} \cup \partial_h \widehat{A}$ (the *vertical* boundary and the *horizontal* boundary respectively) such that $\partial_h \widehat{A} = S$ and $\partial_v \widehat{A}$ is a finite union of tori and annuli whose boundary compactifies S ,

up to the equivalence relation $(A, S) \sim (B, S)$ if there is a diffeomorphism $h : A \rightarrow B$ such that $hi = j$, where $i : S \rightarrow A$ and $j : S \rightarrow B$ are the respective inclusions.

4.2. Classification and conventions. In §4.3 we will introduce the complexity term c_S which treats both closed and bounded oriented sufficiently large (to be defined below) Seifert fibered spaces, but first we must discuss the classification of Seifert fibered spaces (in order to be able to define the complexity) and describe our conventions for dealing with a few “exceptional” cases. As remarked above, the complexity must be defined for bounded as well as closed manifolds. We may assume that the gluing surface S is incompressible, since compressing disks are treated by the (earlier) complexity term c_1 .

It follows as in the proof of Theorem 3.27 that Seifert fibered pieces in the relative JSJ decomposition disjoint from the gluing surface S (i.e., the free factors) occur in either AA or BB as many times as they occur in AB , so it is only necessary to define c_S on the kinds of Seifert fibered pieces which arise by nontrivial gluing along an incompressible subsurface S .

The classification of orientable Seifert-fibered manifolds is well known. We follow the notation of Hatcher [20].

Definition 4.8 (Notation for oriented Seifert fibered manifolds). The notation $M(\pm g, b; \alpha_1/\beta_1, \dots, \alpha_k/\beta_k)$ denotes a Seifert fibered manifold specified by the following properties. The base surface B has genus g , with sign $+$ if B is orientable and $-$ otherwise. The base surface has b boundary components. The fiber structure over the base is obtained from a “model” S^1 bundle E over B with oriented total space by drilling out k fibers and gluing in solid tori with slopes $\alpha_i/\beta_i \in \mathbb{Q}$, and where $\alpha_i/\beta_i \in \mathbb{Z}$ is allowed only if $b = 0$ and $k = 1$. The “model” is unique if $b > 0$. If $b = 0$ and B is orientable, set $E = S^1 \times B$. If $b = 0$ and B is nonorientable, set $E = (S^1 \times \tilde{B})/\langle(\theta, \tilde{b}) \sim (\theta, \tilde{b}')\rangle$, where \tilde{B} is the orientation cover of B , \tilde{b}' is the covering translation of \tilde{b} , and θ is the complex conjugate of θ .

There is some redundancy in this notation, captured in the following proposition.

Proposition 4.9 (Classification theorem, [20], Prop. 2.1). *Two Seifert fiberings*

$$M(\pm g, b; \alpha_1/\beta_1, \dots, \alpha_k/\beta_k), \quad M(\pm g, b; \alpha'_1/\beta'_1, \dots, \alpha'_k/\beta'_k)$$

are isomorphic by an orientation (and fiber) preserving diffeomorphism if and only if, after possibly permuting indices, $\alpha_i/\beta_i = \alpha'_i/\beta'_i \pmod{1}$ for each i , and, in case $b = 0$, if there is an equality $\sum_i \alpha_i/\beta_i = \sum_i \alpha'_i/\beta'_i$.

If $b = 0$, the sum $\sum_i \alpha_i/\beta_i$ is an invariant, called the *Euler number* of the fibering.

Definition 4.10. A Seifert fibered space, with or without boundary, is *sufficiently large* if it contains an oriented incompressible surface (not equal to S^2 or D^2). Otherwise, the Seifert fibered space is called *small*.

Theorem 4.11 (Waldhausen, [46]). *Suppose M is a Seifert fibered space. Then all incompressible surfaces in M can be isotoped to be either horizontal (i.e. transverse to the fibers) or vertical (i.e. a union of fibers).*

The Euler number is an obstruction to the existence of a horizontal surface. In fact, the following is well known:

Lemma 4.12 ([20], Prop. 2.2). *If $b > 0$, horizontal surfaces always exist (although they may be disks if M is not sufficiently large), and if $b = 0$, then horizontal surfaces exist if and only if the Euler number is zero.*

Except for some special cases, the underlying 3-manifold of an orientable Seifert fibered space admits a unique fibering up to isomorphism:

Theorem 4.13 (Exceptional list, [20], Thm. 2.3). *Seifert fiberings of orientable Seifert manifolds are unique up to isomorphism, with the following exceptions:*

- (1) $S^1 \times D^2$,
- (2) $S^1 \tilde{\times} S^1 \tilde{\times} I$,
- (3) S^3 , $S^1 \times S^2$, lens spaces,
- (4) $M(0, 0; 1/2, -1/2, \alpha/\beta) = M(-1, 0; \beta/\alpha)$,
- (5) $S^1 \tilde{\times} S^1 \tilde{\times} S^1$.

Almost all the exceptions have base space a disk, a sphere, or a projective plane with at most 2, 3, or 1 exceptional fibers, respectively. Such manifolds are all small and will not arise by gluing irreducible, boundary irreducible components of A (B) along torus or annulus subcomponents of S (although they might arise as a factor in the relative JSJ decomposition). However, some cases require special treatment.

Case $S^1 \times D^2$. A solid torus (case (1) in Theorem 4.13) can arise as a piece of the relative JSJ decomposition if its frontier consists of at least two annuli, or of one annulus whose core represents a proper multiple of the core of the solid torus. If A_i denote the (oriented) annuli in the frontier of the solid torus, the cores of the A_i represent multiples of the core of the solid torus in homology; define the *degree* of the solid torus to be the sum of these numbers. Choose orientations so that the degree is positive. We must have degree at least 2, or else the frontier is boundary parallel. We make the following convention about the fiber structure on a solid torus:

Convention 4.14. Let $S^1 \times D^2$ be a piece in the relative JSJ decomposition. If the piece has degree 2, and intersects S in two annuli, each homotopic to the core of the solid torus, fiber the solid torus as an I -bundle over an annulus. Otherwise, fiber the solid torus, using at most one exceptional fiber as the core, in the unique manner compatible with a fibering of its frontier (or, equivalently, its intersection with S) by circles.

It is worth remarking that the convention distinguishes between the two cases of degree 2. Both cases admit the structure of an I -bundle or an S^1 bundle compatibly with their intersection with S . The first case we give the structure of an I -bundle over an annulus (rather than a product foliation of a solid torus by circles); the second case we give the structure of an S^1 -bundle over a disk with one exceptional fiber of kind $1/2$ (rather than a twisted I -bundle over a Möbius strip).

Case $S^1 \tilde{\times} S^1 \tilde{\times} I$. A twisted I -bundle over a Klein bottle (case (2) in Theorem 4.13) can arise in the JSJ decomposition of the closed manifold either as a free factor (in which case the fiber structure is irrelevant), or by gluing two solid tori of degree 2 of the same kind. It might arise by gluing solid tori in one of two ways: either by gluing two copies of $\text{annulus} \times I$ along pairs of boundary annuli with a “twist” (that interchanges inside and outside), or by gluing two $M(0, 1; 1/2)$ along fibered annuli in their boundaries. Convention 4.14 insists in either case that the result has the fiber structure $M(0, 1; 1/2, -1/2)$. We therefore make the following:

Convention 4.15. Every $S^1 \tilde{\times} S^1 \tilde{\times} I$ in the closed manifold is given the fiber structure $M(0, 1; 1/2, -1/2)$.

An $S^1 \tilde{\times} S^1 \tilde{\times} I$ can also occur as a piece in the relative JSJ decomposition, intersecting S in a union of annuli. If these annuli are fibered compatibly with one of the fiber structures $M(0, 1; 1/2, -1/2)$ or $M(-1, 1;)$ we give the piece this fiber structure. Otherwise, the piece does not really intersect S at all, but is insulated by a $T^2 \times I$ as in Remark 4.7.

Finally,

Convention 4.16. An $S^1 \tilde{\times} S^1 \tilde{\times} I$ that occurs as a piece in the relative JSJ decomposition with its entire boundary on S is given the fiber structure $M(0, 1; 1/2, -1/2)$.

There is a further ambiguity, that a Seifert fibered manifold may admit isomorphic but nonisotopic Seifert fiberings. If M is sufficiently large, the fibers determine a central \mathbb{Z} subgroup of $\pi_1(M)$, or a normal \mathbb{Z} subgroup if the monodromy on fibers is nontrivial. If the center (after passing to a double cover if necessary) is at least 2-dimensional, M is either virtually $T^2 \times I$ or virtually T^3 . From this one can deduce the following standard fact:

Theorem 4.17 (Big center). *Let M be a sufficiently large Seifert fibered space which is not on the exceptional list from Theorem 4.13. Then either M is one of $T^2 \times I$ or T^3 , or any automorphism of M is isotopic to an automorphism which permutes the fibers, and either fixes the orientation of the base and fiber (if orientable), or simultaneously reverses both.*

The manifolds T^3 and $T^2 \times I$ will be handled as a special case; for the moment we just make the following:

Convention 4.18. A properly embedded $T^2 \times I$ factor in the relative JSJ decomposition (i.e., with $T^2 \times \partial I \subset S$) is fibered as a product I -bundle over T^2 . A nonproperly embedded factor with exactly one boundary component $\subset S$ is thought of as being fibered by circles in an unspecified manner (since any fibration on a single boundary component extends uniquely to a product fibration on the entire $T^2 \times I$).

Other than free factors, Seifert fibered spaces may arise in the pairing either by gluing two Seifert fibered spaces A and B (along part of their boundaries) so as to match the fiber structures, or by gluing two I -bundles together to match end points of intervals. In the latter case, unless the base has positive (orbifold) Euler characteristic, the result will be sufficiently large. In both cases the surface S along which the gluing occurs has a nontrivial fundamental group which injects into A (and B).

If gluing gives rise to a nonfree Seifert fibered factor, Theorem 4.11 implies that S is either a torus or a Klein bottle, or else S is horizontal. If the factor has no boundary, then if S is horizontal, the Euler number must vanish. This rules out all closed exceptional cases in Theorem 4.13 except for $S^1 \tilde{\times} S^1 \tilde{\times} S^1$, since they either contain no incompressible torus, or have nonvanishing Euler number.

This leaves the following exceptional manifolds:

$$T^2 \times I, \quad T^3, \quad S^1 \tilde{\times} S^1 \tilde{\times} S^1.$$

We will see in the proof of Lemma 4.22 that with Convention 4.18, these exceptional cases do not cause a problem. In every other case, by Theorem 4.13 and Theorem 4.17, a nonfree Seifert fibered factor arising in AA, AB or BB has a

unique Seifert fibered structure, up to (orientation-preserving) fiber-permuting automorphisms which either simultaneously preserve, or simultaneously reverse, the orientations of the fibers and the base.

4.3. Seifert fibered complexity. We are now ready to give the definition of the complexity function c_S , discussing first the complexity for *connected* Seifert fibered manifolds (possibly with boundary).

Definition 4.19. Let X be a connected, oriented, sufficiently large Seifert fibered space. Let Q denote the base orbifold of the Seifert fibering (we cannot use the notation B , for obvious reasons, so think Q for “quotient”). Define the complexity $c_{CS}(X)$ to be the ordered 4-tuple

$$c_{CS}(X) = (b, -\chi^o(Q), \#_{\text{sing}}, -|e|),$$

where b is the number of boundary components, $-\chi^o(Q)$ is (minus) the orbifold Euler characteristic of the base (see [45], Definition 13.3.3), $\#_{\text{sing}}$ depends on the singular fibers, and $-|e|$ is (minus) the absolute value of the Euler number, when $b = 0$. All the orbifolds Q which arise are “good”, meaning that they have the form $Q = \Sigma/F$, a surface Σ modulo a finite group action. The orbifold Euler characteristic of Q is defined as $\chi^o(Q) = \chi(\Sigma)/|F|$. The term $\#_{\text{sing}}$ is itself an ordered list of terms defined as follows. Choose (in advance) some total ordering of $(\mathbb{Q}/\mathbb{Z} - 0)/\pm 1$, i.e., nonzero elements of \mathbb{Q}/\mathbb{Z} up to sign. This ordering should not depend on X . For each term $\pm\alpha/\beta$ list the pair $(s(\pm\alpha/\beta), d(\pm\alpha/\beta))$, where $s(\pm\alpha/\beta)$ is the number of singular fibers of type $\pm\alpha/\beta$, and $d(\pm\alpha/\beta)$ is the *defect*, i.e.

$$d(\pm\alpha/\beta) := -|\#(\alpha/\beta \text{ fibers}) - \#(-\alpha/\beta \text{ fibers})|$$

(note that $d(\pm 1/2) = 0$). The term $\#_{\text{sing}}$ is the ordered list of ordered pairs $(s(\pm\alpha/\beta), d(\pm\alpha/\beta))$ over all $(\mathbb{Q}/\mathbb{Z} - 0)/\pm 1$.

The function c_{CS} can be extended to irreducible manifolds with boundary a union of tori by setting it equal to a formal minimum value “ $-\infty$ ” on any manifold which is not a sufficiently large Seifert fibered space. For example, $c_{CS}(X) = -\infty$ if X is a hyperbolic manifold.

The pairs (s, d) in $\#_{\text{sing}}$ should be compared with the terms $(p_i(\Sigma), d_i(\Sigma))$ in the Definition 3.25 of c_i , which serve an analogous purpose, namely to define a complexity which is maximized on the greatest number of objects appearing in orientation-reversed pairs.

Definition 4.20. Let X be a closed, connected, oriented, irreducible 3-manifold, and let X_i be the pieces of X obtained by the JSJ decomposition. Order the X_i by (decreasing) complexity c_{CS} , and let $c_S(X)$ be a tuple

$$c_S(X) = \left(m, -m', \sum_i -\chi^o(Q_i), \sum_i \text{genus}(Q_i), n, c_{CS}(X_1), \dots \right),$$

where the description of the terms is as follows. The first term m is the maximum number of independent tori in X (i.e., the maximum number of pairwise disjoint pairwise nonisotopic incompressible tori). In the second term m' is the number of tori in the JSJ decomposition of X . The third term is the sum of $-\chi^o(Q_i)$ over the Seifert fibered X_i , and the fourth term is the sum of the genera of the Q_i . The fifth term n is the number of Seifert fibered pieces in X , and the remaining terms

are the ordered list of the c_{CS} complexity of the X_i . To compare the complexity of two lists of different lengths, pad the shorter list by $-\infty$ symbols if necessary.

Given an element $(A, S) \in \dot{\mathcal{M}}(S)$, where S is incompressible and A is irreducible, by Theorem 4.5, there is a decomposition of A (along essential annuli properly embedded in A and essential tori disjoint from S) into submanifolds A_F , A_I and A_C , where each component of A_F is Seifert fibered, each component of A_I is a properly embedded essential I -bundle, and every component of A_C is atoroidal and acylindrical. Further, let $A_S := A_F \cup A_I$.

Remark 4.21. We reinterpret our conventions in this language. By Convention 4.14, a solid torus is a component of A_I if and only if it is a product annulus $\times I$ with annulus $\times \partial I$ contained in S ; otherwise it is a component of A_F . By Convention 4.16, an $S^1 \tilde{\times} S^1 \tilde{\times} I$ with entire boundary on S is a component of A_F . Finally, by Convention 4.18, a proper $T \times I$ is in A_I , whereas a torus neighborhood of a torus component of S which bounds an atoroidal, acylindrical submanifold is a component of A_F .

We are now in a position to prove diagonal dominance for c_S .

Lemma 4.22 (c_S -Lemma Schema). *Let (A, S) and (B, S) be elements of $\dot{\mathcal{M}}(S)$, where S is incompressible, and both A and B are irreducible. Then there is an inequality*

$$c_S(AB) \leq \max(c_S(AA), c_S(BB))$$

which is strict unless $S \cap A_S = S \cap B_S =: S_S$, and $(A_S, S_S) = (B_S, S_S)$ as elements of $\dot{\mathcal{M}}(S_S)$.

Remark 4.23. Two copies of surface $\times I$ can be glued together to produce a manifold with a Solv or hyperbolic structure, so it is crucial that c_S occur before c_h in our complexity.

Proof. Free factors in the JSJ decomposition of highest complexity are most common in one of the doubles, by repeating the argument in the proof of Theorem 3.27. So it suffices to restrict attention to the complexity of nonfree factors, which we now do.

The gluing of A_S to B_S does not in general match $A_S \cap S$ with $B_S \cap S$. For instance, an I -bundle component of A_S might be glued to an acylindrical component of B_C , and the union would be part of a hyperbolic piece of the JSJ decomposition of AB . The terms $m, -m'$ can almost be recovered (with the same ordering) from the later terms in the complexity, except when distinct torus boundary components of the pieces X_i are isotopic in such a way that the fibering is not compatible under the isotopy. More precisely, these terms prefer isotopic tori in different X_i to be fibered in the same way. These terms are only significant when A_F and B_F share common boundary tori in S , and are maximized when the fiberings on these tori in A_F and in B_F match up, and the pieces can be glued into a single component of the JSJ decomposition of AB .

In general, let X be a nonfree Seifert fibered component of the JSJ decomposition of AB . If $S \cap X$ is horizontal, then X is a maximal union of I -bundle pieces of A_I and B_I , glued along incompressible subsurfaces of S . Let X_h be the union of such nonfree Seifert fibered components, and let Q be the base of the fibration. Let Q_i be the bases of the fibration of the pieces in A_I and B_I which are glued

together to make X_h . Since S is incompressible, the contribution of X_h to m and m' can be recovered from the type and multiplicity of the Q_i , and the order on (m, m') is compatible with the order induced from lower complexity terms; in other words, the contribution of X_h to the m, m' terms is maximized when the lower-order complexity terms are maximized.

There is a transverse geometric structure on the Seifert fibration of (each component of) the manifold X_h , pulled back from a geometric structure on Q (see e.g. [45], §4.8 for a discussion of transverse geometric structures on Seifert fibrations). Away from the singular points, the base Q decomposes into subsurfaces Q^n where n counts the number of times the corresponding fiber intersects S (equivalently, n counts the number of I -fibers of A_I and B_I pieces which make up the given S^1 fiber). By Gauss-Bonnet there is a formula

$$\chi^\circ(Q) = \sum n\chi^\circ(Q^n).$$

Write $Q = Q_h \cup Q_e \cup Q_s$, where the subscripts denote the hyperbolic, Euclidean, and spherical components of Q . Since each component of Q is covered by an essential subsurface of S , it follows that Q_s is empty. Since each $n \geq 2$, it follows that $-\chi^\circ(Q_h)$ is maximized only when $Q_h = Q_h^2$, and Q_h is double covered by the hyperbolic part of the subsurface $A_I \cap S \cap B_I$, which must be as big as possible (and therefore the hyperbolic components of $A_I \cap S$ are equal to the hyperbolic components of $B_I \cap S$). Finally, $\sum_i \text{genus}(Q_i)$ and n are maximized only when $Q_e = Q_e^2$ and when (as above) the Euclidean part of $A_I \cap S$ and $B_I \cap S$ is as big as possible (after the hyperbolic parts are already matched up). In other words, $A_I \cap S = B_I \cap S$.

If $S \cap X$ is vertical, then X is a maximal union of Seifert fibered pieces of A_F and B_F , glued along fibered tori or annuli contained in S , possibly together with a number of $T^2 \times I$ factors in A_F or B_F with some S^1 -bundle structure (rather than with their conventional I -bundle structure). Note that an annulus component of $A_F \cap S$ cannot be contained in a torus component of $B_F \cap S$, or else a neighborhood of this torus (in A) could be glued to A_F , thereby enlarging it, contrary to the definition of a characteristic submanifold.

Let X_v be the union of such nonfree Seifert fibered components, and let Q be the base of the fibration. Then

$$(4.3.1) \quad \chi^\circ(Q) = \sum \chi^\circ(Q_i) - \#(\text{annuli}),$$

where the Q_i are the Euler characteristics of the bases of the various terms in A_F and B_F , and $\#(\text{annuli})$ counts the number of annuli in S along which they are glued. It follows that $-\chi^\circ(Q)$ is maximized when they are glued along as many annuli as possible, in particular, when the fibered annuli boundary components of the A_F and B_F pieces match up.

Gluing two fibered torus boundary components together does not change χ° , but it either increases genus or decreases m' (and therefore increases $-m'$). Gluing on a proper $T^2 \times I$ factor at both ends does not change χ° or the genus, but it *reduces* the total number of T^3 factors which occur in X_h . So if the leading terms in c_S are maximized, it follows that all proper $T^2 \times I$ factors are actually already contained in X_h , and never appear in X_v , and therefore in this case $A_F \cap S = B_F \cap S$ with the same fiber structure.

Remark 4.24. It is worth remarking at this point that any $M(-1, 1;)$ piece in the relative JSJ decomposition that is not glued up will survive as a component of the JSJ decomposition of the closed manifold, but with the “wrong” Seifert fibered structure (i.e., contrary to Convention 4.15). However, we have just showed that such unglued pieces never occur in the gluing of maximal complexity, and therefore the convention may (and will) be assumed in the sequel. The $M(-1, 1;)$ fiber structure is the right one when such pieces are glued up to other pieces in the JSJ decomposition of the closed manifold.

It follows that either the inequality is strict at one of the first four terms of c_S , or else we are in the case that $A_I \cap S = B_I \cap S$ and $A_F \cap S = B_F \cap S$, and moreover that the fiber structures on $A_F \cap S$ and on $B_F \cap S$ agree. The fifth term n in c_S is maximized only when the connectivity partitions of $A_I \cap S$ into A_I and of $A_F \cap S$ into A_F agree with that of B (see Definition 3.2 and Lemma 3.3). Therefore nonfree Seifert fibered components of AB are all made from exactly one component of each of A_I and B_I , or one component of each of A_F and B_F , whose intersections with S are equal.

When two I -bundles are glued, we have already seen that $Q = Q^2$, in other words, that every circle in the glued up component is a union of exactly two intervals, one in A and one in B . This pairing of intervals gives an identification of A_I with B_I rel their intersection with S .

It remains to check gluing of F -pieces. The situation is very much like the complexity for gluings of surfaces considered in [15]. The basic idea is that maximizing $(b, -\chi^o(Q))$ component by component (always in decreasing order) forces the most complicated F -pieces in A to be glued to pieces of the same complexity in B term by term, and by induction, every piece in A_F must be glued to a piece in B_F with the same number of boundary components and with base of the same Euler characteristic. This nearly proves the lemma, except that different topological orbifolds can have the same number of boundary components and the same $-\chi^o$. This ambiguity is resolved by further consideration of the $\#_{\text{sing}}$ term, which precisely favors the double.

There is one final piece of ambiguity for closed sufficiently large Seifert fibered spaces, namely the Euler number of the fibering. Doubled pieces in AA and BB obviously satisfy $e = 0$, and therefore maximize $-|e|$. It follows that pieces in AB must also have $e = 0$, and therefore we are done, by Theorem 4.9. \square

Remark 4.25. Note that $S^1 \tilde{\times} S^1 \tilde{\times} S^1$ factors can arise either in X_h (as a union of two copies of $S^1 \tilde{\times} S^1 \tilde{\times} I$ glued along their torus boundaries) or in X_v (by gluing two copies of $M(0, 1; 1/2, -1/2)$ along their torus boundaries). In either case, the complexity is maximized only when these factors pair up, i.e. when $(A_S, S_S) = (B_S, S_S)$.

5. HYPERBOLIC FACTORS

We assume the reader is familiar with fundamental facts from the Thurston theory of hyperbolic 3-manifolds. We use this material without comment throughout this section, justifying only facts or claims which are new or unfamiliar. A basic reference for this material is Thurston’s notes [45].

We also make some use of the theory of (stable) minimal surfaces, especially in hyperbolic 3-manifolds, and the theory of Ricci flow. A reference for the first is [8], and a reference for the second is [7].

5.1. Closed hyperbolic case. The first piece of the hyperbolic complexity is – vol, where vol denotes hyperbolic volume. For closed manifolds, the justification for using this term is the following theorem of Agol–Storm–Thurston:

Theorem 5.1 (Agol–Storm–Thurston, [1]). *Let S be a closed orientable surface so that each component has negative Euler characteristic, and let $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ be irreducible, atoroidal and acylindrical. Then AA, AB, BB admit unique complete hyperbolic structures, and either*

$$2 \operatorname{vol}(AB) > \operatorname{vol}(AA) + \operatorname{vol}(BB)$$

or else $2 \operatorname{vol}(AB) = \operatorname{vol}(AA) + \operatorname{vol}(BB)$ and S is totally geodesic in AB .

Remark 5.2. Since by hypothesis A and B are atoroidal, irreducible and acylindrical, Thurston’s hyperbolization theorem for Haken manifolds (see [32] for an exposition) implies that both A and B admit unique complete hyperbolic structures with totally geodesic boundary. Moreover, S is totally geodesic in AB if and only if the totally geodesic metrics on S inherited from the hyperbolic structures on A and B are isometric by an isometry isotopic to the identity.

A very brief outline of the proof in [1] is as follows. First, construct the hyperbolic structure on AB . Since both A and B are acylindrical and atoroidal, AB is atoroidal and such a hyperbolic structure exists, by Thurston’s hyperbolization theorem for Haken manifolds. Next, find a minimal surface representative of the surface S in the hyperbolic manifold AB . Since S is incompressible in AB , it admits a stable minimal representative in its isotopy class, by the well-known theorem of Meeks–Simon–Yau (i.e. the main theorem of [29]). Cutting AB open along S produces two Riemannian manifolds with boundary A, B . The metrics on the doubles AA, BB are not Riemannian unless S is totally geodesic, but are C^0 Riemannian, i.e., defined by a C^0 section of the bundle of symmetric 2-tensors. Since S is minimal in AB , it has vanishing mean curvature. This implies, by arguments of Miao [31] or Bray [5], arising from work on the Riemannian Penrose inequality, that the metrics on AA and BB can be uniformly approximated by C^∞ Riemannian metrics satisfying uniform pointwise scalar curvature bounds $R \geq -6$. Technology due to Miles Simon [43] lets one apply short-time Ricci flow to such singular metrics, which preserves the pointwise scalar curvature bounds. Then Perelman’s monotonicity formula for Ricci flow with surgery ([34, 35]) implies the global inequality in the theorem.

In §5.3 we will generalize this theorem to the case that S is a surface of finite type. There are a number of new analytic details which arise in this case, but we otherwise stay very close to the argument of [1].

5.2. Isometric gluing. The next piece of the hyperbolic complexity is the *length spectrum*.

Definition 5.3. Let M be a finite volume hyperbolic 3-manifold. The *real hyperbolic length spectrum* of M , denoted $\sigma(M)$, is a lexicographic tuple of real numbers $\sigma(M) = (\ell_1(M), \ell_2(M), \dots)$ which lists (with multiplicity) the length of closed geodesics in M whose complex length is real, ordered nondecreasingly by length. If this spectrum is finite (or empty), pad this tuple with a string of ∞ symbols.

Say $\sigma(M) > \sigma(N)$ if $\ell_i(M) = \ell_i(N)$ for all $i < n$ but $\ell_n(M) < \ell_n(N)$.

Warning. Note the sign convention in Definition 5.3. The manifold with the *smaller* geodesics has the *larger* value of σ .

Lemma 5.4. *Suppose $(A, S), (B, S) \in \dot{\mathcal{M}}(S)$ admit finite volume complete hyperbolic structures with totally geodesic boundary S such that the two induced hyperbolic structures on S are isometric by an isometry isotopic to the identity. Then either*

$$\sigma(AB) < \max(\sigma(AA), \sigma(BB))$$

or $(A, S) = (B, S)$.

Proof. Let ℓ_1 be the shortest length in $\sigma(AB)$. We will derive a formula for the multiplicity of ℓ_1 .

Let $\sigma(AB, \ell_1)$ denote the multiplicity of ℓ_1 in $\sigma(AB)$ and define $\sigma(AA, \ell_1)$ and $\sigma(BB, \ell_1)$ similarly.

The formula clearly will have the form

$$(5.2.1) \quad \sigma(AB, \ell_1) = k_1^A + k_1^B - k_1^S + \epsilon',$$

where k_1^A is the number of closed hyperbolic geodesics of length ℓ_1 in A (and similarly for k_1^B), k_1^S is the number of closed hyperbolic geodesics of length ℓ_1 in S , and ϵ' is the number of geodesics in AB of length ℓ_1 which are not contained on either side. The k_1^* terms come from the inclusion/exclusion formula. The ϵ' term is slightly more subtle to determine.

Define ϵ_A to be the number of proper essential arcs in A of length $\ell_1/2$, and define ϵ_B similarly. If either A or B contains an essential arc shorter than $\ell_1/2$, doubling produces a geodesic in either AA or BB with length strictly less than ℓ_1 and we would be done. So without loss of generality we can assume that every proper essential arc in either A or B of length at most $\ell_1/2$ has length exactly $\ell_1/2$, and is totally geodesic and perpendicular to the boundary. Moreover, a geodesic loop of length ℓ_1 in AB can consist of at most one arc on each side. Hence $\epsilon' \leq \min(\epsilon_A, \epsilon_B)$.

It follows that there is a formula

$$(5.2.2) \quad \sigma(AA, \ell_1) = 2k_1^A - k_1^S + \epsilon_A$$

and similarly for $\sigma(BB, \ell_1)$. Comparing equation (5.2.1) and equation (5.2.2) we deduce that either there is an inequality

$$\sigma(AB, \ell_1) < \max(\sigma(AA, \ell_1), \sigma(BB, \ell_1))$$

or else $\epsilon_A = \epsilon_B = \epsilon'$ and therefore also $k_1^A = k_1^B$.

In other words, the shortest essential geodesic arcs in A and in B have the same length and multiplicity and are paired by the gluing map $(A, B) \rightarrow AB$. Since the double of every such arc in AA or BB has complex length with zero imaginary part, the holonomies of geodesic transport along paired arcs in A and in B must be equal, or else there would be fewer arcs in AB with *complex* length ℓ_1 , and $\sigma(AB) < \max(\sigma(AA), \sigma(BB))$ already. Let Γ_1^A denote the set of geodesic arcs in A with length $\ell_1/2$ and define Γ_1^B similarly. Then by the discussion above, without loss of generality we can assume that the isometric identification of $S \subset A$ and $S \subset B$ can be extended to isometric identifications

$$N_1^A = N_1^B,$$

where N_1^A is a regular neighborhood of $S \cup \Gamma_1^A$ and N_1^B is defined similarly.

Note that the fact that the isometry can be extended over $S \cup \Gamma_1^*$ follows just by considering the real lengths of the arcs in Γ_1^* . The fact that it can be extended over a neighborhood depends on the equality of *complex* lengths. Note further that Γ_1^A and Γ_1^B might be empty.

We will define inductively systems of proper geodesic arcs

$$\Gamma_1^A \subset \Gamma_2^A \subset \Gamma_3^A \subset \cdots$$

and regular neighborhoods

$$N_p^A = \text{regular neighborhood of } S \cup \Gamma_p^A$$

and similarly for B , where for each p , the set $\Gamma_p^A \setminus \Gamma_{p-1}^A$ is the family of proper geodesic arcs in A orthogonal to S having length exactly $\ell_p/2$, where ℓ_p is the p th *distinct* term of $\sigma(AB)$ (i.e., *not* counted with multiplicity).

We fix some notation, which we use throughout the remainder of the proof. Denote by $N_p^A N_p^B$ the corresponding subset of AB , and similarly for AA and BB . If N_p^A is connected, let $[N_p^A]$ denote the covering space of A induced by the image of $\pi_1(N_p^A)$ in A . If N_p^A is disconnected, let $[N_p^A]$ denote the disjoint union of those covering spaces. The finite area boundary components of this disjoint union have an obvious identification with S , so we continue to treat $([N_p^A], S)$ as a pair. Similarly, $[N_p^A N_p^B]$ denotes the corresponding covering space of AB . Note that in our notation there is an equality

$$[N_p^A][N_p^B] = [N_p^A N_p^B]$$

and similarly for AA, BB .

By induction, assume that $\sigma(AA)$, $\sigma(BB)$ and $\sigma(AB)$ are identical length spectra with multiplicities up to ℓ_{p-1} and that the isometric identification $\partial A \rightarrow \partial B$ has been extended over $N_{p-1}^A \rightarrow N_{p-1}^B$. Since every perpendicular geodesic arc in A of length $< \ell_p/2$ can be doubled to a closed geodesic in AA of length $< \ell_p$, by the induction hypothesis the length of such a geodesic must be equal to one of the ℓ_q with $q < p$, and therefore N_{p-1}^A contains every perpendicular geodesic arc with length $< \ell_p/2$.

We write down a formula for the multiplicity $\sigma(AB, \ell_p)$ of ℓ_p in AB and similarly for AA, BB . Note that any geodesic γ in AB intersects the two sides in a collection of arcs. Each arc with length strictly less than $\ell_p/2$ can be properly homotoped into the corresponding N_{p-1} factor, as observed in the previous paragraph.

By the inclusion-exclusion formula,

$$(5.2.3) \quad \sigma(AA, \ell_p) = 2k_p^A - k_{p-1}^{AA} + m_p^A - c_{p-1}^A,$$

where

$$\begin{aligned} k_p^A &= \#\{\text{closed geodesics of length } \ell_p \text{ in } [N_{p-1}^A]A\}, \\ k_{p-1}^{AA} &= \#\{\text{closed geodesics of length } \ell_p \text{ in } [N_{p-1}^A N_{p-1}^A]\}, \\ m_p^A &= \#\{\text{perpendicular geodesic arcs of length } \ell_p/2 \text{ in } A\}, \\ c_{p-1}^A &= \#\{\text{such arcs properly homotopic into } N_{p-1}^A\}, \end{aligned}$$

and similarly for $\sigma(BB, \ell_p)$. (We remind the reader that $[N_{p-1}^A]A$ denotes the union of $[N_{p-1}^A]$ and \bar{A} along S , and similarly for other juxtapositions above.)

Note that c_{p-1}^A is also equal to the number of doubled perpendicular arcs in AA which lift to $[N_{p-1}^A N_{p-1}^A]$. Note also that, by induction, there are isometric

identifications

$$N_{p-1}^A N_{p-1}^A = N_{p-1}^B N_{p-1}^B = N_{p-1}^A N_{p-1}^B$$

and therefore there are equalities

$$k_{p-1}^{AA} = k_{p-1}^{BB} = k_{p-1}^{AB},$$

where k_{p-1}^{AB} is the number of closed geodesics of length ℓ_p in $[N_{p-1}^A N_{p-1}^B]$.

We estimate $\sigma(AB, \ell_p)$. Pairs of perpendicular geodesic arcs of length $\ell_p/2$ on different sides of AB may not match up exactly. Moreover, even if they do match up, *a priori* different arcs might have different holonomy, so the complex length of the resulting closed geodesic might have nonzero imaginary part.

It follows that there is an *inequality*

$$(5.2.4) \quad \sigma(AB, \ell_p) \leq k_p^A + k_p^B - k_{p-1}^{AB} + \min(m_p^A - c_{p-1}^A, m_p^B - c_{p-1}^B).$$

Denote $n_p^A := m_p^A - c_{p-1}^A$. It follows that either $\sigma(AB) < \max(\sigma(AA), \sigma(BB))$ and the theorem is proved, or else (after possibly interchanging A and B) there is an inequality

$$k_p^A + k_p^B + \min(n_p^A, n_p^B) \geq 2k_p^A + n_p^A \geq 2k_p^A + n_p^B$$

from which we can conclude that $k_p^A = k_p^B$ and $n_p^A = n_p^B$.

Furthermore, we deduce that the inequality in equation (5.2.4) is actually an equality, and therefore perpendicular arcs of length $\ell_p/2$ in A and B match in pairs with the same holonomy. This completes the inductive step, and lets us extend the isometric identification $\partial A \rightarrow \partial B$ to $N_p^A \rightarrow N_p^B$.

As we exhaust the length spectrum, we eventually obtain a complete generating set for the fundamental group. That is, for p sufficiently large, $\pi_1(N_p^A) \rightarrow \pi_1(A)$ is an epimorphism.

It follows that the isometry on the boundaries extends to $A \rightarrow B$, and the lemma is proved. \square

5.3. Hyperbolic case with cusps. Theorem 5.1 and Lemma 5.4 together let us define a complexity function, namely the tuple $(-\text{vol}, \sigma)$, which is diagonally dominant for *closed* hyperbolic manifolds. However, we need a generalization which is valid for *complete* hyperbolic manifolds of the kind which arise in the JSJ decomposition. Lemma 5.4 as stated applies directly to such manifolds. The following is the required generalization of Theorem 5.1 to the case of cusped manifolds.

Theorem 5.5. *Let S be an orientable surface of finite type so that each component has negative Euler characteristic, and let $(A, S), (B, S) \in \mathcal{M}(S)$ be irreducible, atoroidal and acylindrical. Then AA, AB, BB admit unique complete hyperbolic structures, and either*

$$2 \text{vol}(AB) > \text{vol}(AA) + \text{vol}(BB)$$

or else $2 \text{vol}(AB) = \text{vol}(AA) + \text{vol}(BB)$ and S is totally geodesic in AB .

Remark 5.2 applies equally well to the cusped case. Note that by a hyperbolic Dehn filling argument, the only part of this theorem which does not follow from Theorem 5.1 is the *strictness* of the inequality when S is not totally geodesic in AB .

Proof. Since A, B are acylindrical, the manifolds AA, BB and AB admit unique complete hyperbolic structures of finite volume. The manifolds have two kinds of cusps: “free cusps” which do not intersect cusps of S , and “transverse cusps” which do. The free cusps are irrelevant to the discussion and for simplicity we assume they do not exist.

In fact, as a further simplification, we assume AB has exactly one cusp T which intersects S in two cusps with opposite orientations. It will be clear from the proof in this case that multiple cusps present no additional difficulties. Here we should think of $T = \partial \overline{AB}$ where AB is homeomorphic to the interior of AB , and $T_A = \partial_v \overline{A}, T_B = \partial_v \overline{B}$ are both annuli, each with two boundary components which compactify the two cusps of S . The *meridian* on T is the (unoriented) isotopy class which is isotopic to the cores of the annuli T_A and T_B .

By [29] the surface S in AB has a least area minimal representative in its proper isotopy class. Recall that a minimal surface is said to be *stable* if the second variation of area is nonnegative for all compactly supported variations. Least area surfaces are stable. By abuse of notation, we call this surface S . Since S is stable, a fundamental estimate of Schoen applies. \square

Theorem 5.6 (Schoen [40]). *Let S be a stable minimal surface in a Riemannian 3-manifold M . Given $r \in (0, 1]$ and a point $p \in S$ such that the ball $B_r(p) \cap S$ has compact closure in S , there is a constant C depending only on the norms of R_{ijk}^l and $\nabla^m R_{ijk}^l$ on M restricted to $B_r(p)$ such that*

$$|h_{ij}|^2(p) \leq Cr^{-2},$$

where h_{ij} is the second fundamental form.

This theorem generalized to the bounded case earlier curvature estimates of Frankel [11].

For two-sided surfaces, stability is preserved under covers (see [8], p. 21 for a proof). In a complete hyperbolic 3-manifold, the norms of the curvature and its first derivatives are bounded by universal constants. So Schoen’s estimate implies a *uniform* pointwise lower bound on the sectional curvature of a complete stable minimal surface in a hyperbolic 3-manifold.

In the remainder of the proof, let A, B denote the complete Riemannian manifolds with boundary obtained from the hyperbolic 3-manifold AB by cutting along the stable minimal surface S . Let AA, BB denote the singular Riemannian manifolds obtained from A and B by metrically doubling along S . Our immediate goal is to prove short-time existence of the Ricci flow on the manifolds AA, BB , using the technology developed by Simon [43].

In fact, as has become standard in discussions of Ricci flow, following Simon [43], we use in place of Ricci flow a generalization of the DeTurck gauging (DeTurck [9]) called the *dual Ricci-Harmonic Map flow* (see Hamilton [17], §6 or Simon [43], p. 3 for a precise discussion), which is equivalent to the Ricci flow up to a diffeomorphism. In what follows, we refer simply to “flow”.

In the sequel we suppress BB and discuss only AA for simplicity. As in [1] and [43] we must find suitable smooth approximations AA^δ with *uniform* pointwise lower bounds for scalar curvature *independent of δ* , such that $AA^\delta \rightarrow AA$ as $\delta \rightarrow 0$. However, since the AA^δ are noncompact, we actually employ a two-parameter family of smooth approximations AA_k^δ where δ is a small positive real number and

k is one of an infinite increasing sequence of positive integers. The AA_k^δ are singular but compact approximations to AA_k , where

$$\lim_{k \rightarrow \infty} AA_k^\delta = AA^\delta \text{ for each } \delta, \lim_{\delta \rightarrow 0} AA_k^\delta = AA_k \text{ for each } k$$

as Gromov-Hausdorff limits. We assume the reader is familiar with Gromov-Hausdorff convergence and Gromov-Hausdorff limits of (pointed) metric spaces; in the sequel we usually use the term *Gromov limit* for brevity. For definitions and basic properties of Gromov convergence, see [16], §6. The AA_k turn out to be orbifolds obtained by Dehn filling k times the meridian of AA .

Let AB_k be the closed hyperbolic orbifold obtained by (orbifold) hyperbolic Dehn surgery on AB , by filling k times the meridian of T . The cusp T becomes an orbifold geodesic $\gamma_k \subset AB_k$ and the (topological) surface S fills in to become an orbifold S_k transverse to γ_k at two points.

Each surface S_k has a least area minimal orbifold representative in its isotopy class; indeed, following [14] and [19], by Selberg’s Lemma (see e.g [37]) one may pass to a finite regular manifold cover \widetilde{AB}_k of AB_k , lift S_k to a topological surface \widetilde{S}_k in \widetilde{AB}_k , and find a least area minimal representative in the cover which is disjoint from or equal to its image under every element of the deck group, or else its area could be reduced by exchange and the roundoff trick. Then this least area representative covers a least area orbifold representative in AB_k which, by abuse of notation, we denote S_k .

Cutting AB_k along S_k produces A_k, B_k and doubling these produces AA_k, BB_k . Observe that since \widetilde{S}_k is fixed by elements of the deck group which do not act freely, the geodesic γ_k is perpendicular to S_k , and AA_k and BB_k are singular orbifolds. That is, they have finite orbifold covers which are locally isometric to the double of a hyperbolic manifold along a least area minimal surface. Another way to see this is to cut \widetilde{AB}_k along \widetilde{S}_k to obtain \widetilde{A}_k and double, obtaining \widetilde{AA}_k , which isometrically orbifold covers AA_k .

Lemma 5.7. *After passing to a subsequence of integers $k \rightarrow \infty$, and after possibly replacing S by another least area minimal surface, there is convergence*

$$AB_k \rightarrow AB, S_k \rightarrow S, AA_k \rightarrow AA, BB_k \rightarrow BB$$

in the pointed Gromov-Hausdorff sense.

Proof. Thurston’s hyperbolic Dehn surgery theorem ([45], Chapter 4) implies that for k sufficiently large, AB_k is hyperbolic, and $AB_k \rightarrow AB$ as $k \rightarrow \infty$. By Schoen’s Theorem 5.6, the surfaces S_k satisfy uniform two-sided curvature bounds. Since they have bounded geometry, there is a convergent subsequence in C^∞ whose limit is a minimal surface S . The least area property is inherited by limits of minimal surfaces ([8], Chap. 1), so S is least area and may be taken to be S as above. \square

Next, we construct AA_k^δ . There are two alternate approaches.

- (1) Bray ([5], §6) uses the ODE for surface metrics \bar{g}_{ij} evolving normally in a 3-manifold

$$(5.3.1) \quad \frac{d}{dt} \bar{g}_{ij}(x, t) = 2\bar{g}_{ik}(x, t)h_j^k(x, t/\delta),$$

where $-\delta \leq t \leq \delta$ and h_j^k is antisymmetric in t/δ , to create a mirror-symmetric *plug* of thickness 2δ interpolating between A_k and its mirror

image. The h_j^k term is the second fundamental form of the $t = \text{constant}$ “slices” of the plug. This plug can be inserted between the two copies of A_k to build AA_k^δ .

- (2) Miao ([31], §3) simply mollifies the singular metric within a tubular neighborhood of uniform thickness.

Since our surfaces S_k , $k \leq \infty$, have no uniform size tubular neighborhoods to work with, we follow Bray [5] and desingularize by adding an untapered “plug” of thickness 2δ over S_k . Note that the constructions both of Bray and of Miao commute with isometries, so that in practice we perform the desingularization in the cover and then define AA_k^δ to be the quotient orbifold. Note for $\delta > 0$ that AA_k^δ is a C^∞ orbifold.

The metric and curvature uniformities of Bray’s construction are summarized in Lemma 5.8. Our Lemma 5.8 is parallel to [1], Prop. 4.1, except that we include the parameter k and note its uniformity.

Lemma 5.8. *There exists a family of C^∞ Riemannian orbifolds AA_k^δ , where $k \rightarrow \infty$, so that for each fixed k , there is $\epsilon(k, \delta)$ so that AA_k^δ is $(1 + \epsilon)$ -bi-Lipschitz to AA_k^0 , where $\epsilon \rightarrow 0$ as $\delta \rightarrow 0$. Moreover, these orbifolds satisfy the following estimates:*

- (1) *The scalar curvature satisfies $R(AA_k^\delta) \geq s(k)$, a (negative) constant independent of δ , for fixed $k < \infty$.*
- (2) *The square norms of the full Riemann curvature tensors satisfy an inequality $|R_{ijk}^l|^2 \leq c(\delta)$. That is, they are uniformly bounded as a function of position $x \in AA_k^\delta$ and $k \leq \infty$ (but not δ).*

Proof. The only new ingredient is the k -uniformity in the last assertion. The second fundamental form $h_{ij}(x, t)$ enters into the estimates of Bray (and Miao). By Theorem 5.6 the $|h_{ij}|^2$ are pointwise uniformly bounded over all complete stable minimal surfaces in hyperbolic 3-manifolds. So, the uniformity of $|h_{ij}|^2$ over position in S_k and over k accounts for the new conclusions. \square

Our next Lemma 5.9 is parallel to [1], Prop. 4.2.

Lemma 5.9. *After a conformal modification, one can further assume that the metrics on AA_k^δ satisfy pointwise estimates for scalar curvature*

$$R(AA_k^\delta) \geq -6$$

while still satisfying $AA_k^\delta \rightarrow AA_k$ in the bi-Lipschitz sense.

Essential in Simon [43] and Perelman [34, 35], and implicit in the earlier work of Hamilton, is the following principle:

Lemma 5.10. *Flow commutes with Gromov limit when flow on the limit can be defined.*

For Lemma 5.10 to be useful, it is necessary to obtain lower bounds on the time to blowup, which can be estimated uniformly. Simon [43], Thm. 5.2 says that for flow on a C^0 Riemannian manifold (M, g) , the time T to blowup for flow can be estimated in terms of $|\nabla^m R_{ijk}^l|$ of a *background metric* g' on M which is $(1 + \epsilon)$ -bi-Lipschitz close to g for some universal $\epsilon > 0$ depending only on the dimension of M (a precise statement is the first part of Lemma 5.19 in this paper). Simon’s remarkable result, derived without assuming that M is compact or even that g is better than C^0 , lets us prove a parallel to [1], §6.1.

Lemma 5.11. *There is a uniform constant $T > 0$ such that for each k and for all $\delta < \delta(k)$ flow exists for time $t \in [0, T]$ on AA_k^δ . Moreover, flow exists for time $t \in [0, T]$ on AA .*

Proof. Since, for each k and each $\epsilon' > 0$, all the manifolds AA_k^δ and AA_k are $1 + \epsilon'$ bi-Lipschitz when δ is sufficiently small, it suffices to construct for each k and for each $\epsilon > 0$, a C^∞ metric $g(k)$ on AA_k which is $(1 + \epsilon)$ -bi-Lipschitz to AA_k , and for which there are uniform pointwise bounds on $|\nabla^m R_{ijk}^l|$, depending on ϵ , but independent of k .

For each individual k , the existence of such a metric $g(k)$ is easy: any C^0 Riemannian metric (i.e., a metric defined by a C^0 symmetric bilinear form on TM) can be approximated (e.g., by mollifying in local coordinates) by some bi-Lipschitz C^∞ metric. The singular metric on AA_k (away from the orbifold locus) is of this kind, since it is obtained by doubling a genuine Riemannian metric (also see Bray [5], equation 102 and the surrounding text for an explicit estimate). After we have obtained such an approximating metric, observe since AA_k is compact, that there is some uniform pointwise bound on the curvature and all its covariant derivatives. However, the bound one gets in this case may well depend on k .

Using the fact that $AA_k \rightarrow AA$ in the Gromov sense, we see that this part of the argument works *on the thick part* of AA_k . We need to find a $(1 + \epsilon)$ -bi-Lipschitz model for the thin part of each AA_k with square curvature bounds which depend on ϵ but not on k . Then these two smooth $(1 + \epsilon)$ -bi-Lipschitz models on the overlap of the thick and thin parts can be welded together by a smooth convex combination, at the cost of possibly increasing ϵ by a bounded amount. So one just needs to choose the bi-Lipschitz constants better than necessary on each piece, so that the result of the welding is $(1 + \epsilon)$ -bi-Lipschitz. It remains to find, for any $\epsilon > 0$, a $(1 + \epsilon)$ -bi-Lipschitz model for the thin part of each AA_k with uniform square curvature bounds which depend on ϵ but not on k .

To do this, we must first prove a lemma about the geometry of the cusped least area minimal surface S deep in the thin part. We know that S corresponds to a quasi-Fuchsian group since both (A, S) and (B, S) are acylindrical. \square

Sublemma 5.12. *Let S be a quasi-Fuchsian least area embedded surface in a complete cusped hyperbolic 3-manifold M , and let \tilde{S} be a component of the preimage of S in \mathbb{H}^3 . Let $p \in S_\infty^2$ correspond to a lift of the cusp. Let B_t be a family of horoballs centered at p which are level sets of a Busemann function $b(t)$. Then there is a totally geodesic plane π through p (the osculating plane) so that the restrictions $B_t \cap \tilde{S}$ and $B_t \cap \pi$ are Hausdorff distance $o(e^{-t})$ apart. Moreover, if \tilde{S}_1, \tilde{S}_2 are different components of the preimage of S which both intersect p , their osculating planes are distinct.*

Proof. In the upper half-space model (with x, y as the horizontal coordinates and z as the vertical coordinate), put p at infinity. The horoball B_t corresponds to the set $z \geq e^t$, so we need to find a vertical plane π which is within Hausdorff distance $o(1)$ in the *Euclidean metric*, restricted to $z \geq t$; i.e., we need to show that the Euclidean Hausdorff distance between the restrictions of \tilde{S} and of π goes to 0 as $z \rightarrow \infty$.

After composing with an isometry if necessary, we can assume that \tilde{S} is stabilized by $x \rightarrow x + 1$. Let Λ be the limit set of \tilde{S} . Then Λ is the union of ∞ with a proper quasi-arc Λ_0 in the (x, y) -plane, which is invariant under $x \rightarrow x + 1$. The arc Λ_0 is

bounded in the slab $y \in [-C, C]$ for some constant C . Since minimal surfaces are contained in the convex hulls of their boundaries, \tilde{S} is also contained in the slab $y \in [-C, C]$.

A stable minimal surface whose intersection with a compact set K is trapped between two barrier stable minimal surfaces which are C^0 close on K is C^∞ close to both (in fact “stable” is superfluous here). Locally, this is just the Harnack inequality for nonnegative solutions of uniformly elliptic equations; see e.g. [8], pp. 20–21. This implies that \tilde{S} is C^∞ close to the vertical planes $y = C$ and $y = -C$ when z is sufficiently large. Note that this is C^∞ close in the *hyperbolic metric*. In the Euclidean metric at height $z = t$, the order n partial derivatives of \tilde{S} and $y = C$ are $o(t^{1-n})$ close; i.e., they differ by a term which is arbitrarily small compared to t^{1-n} as $t \rightarrow \infty$. In particular, for z sufficiently large, the tangent plane to \tilde{S} at each point is arbitrarily close to a vertical plane of the form $y = \text{constant}$, and therefore \tilde{S} is transverse to the level sets $z = \text{constant}$. Notice that this does not yet tell us that the y coordinate, thought of as a function on \tilde{S} , converges to a constant as $z \rightarrow \infty$; to establish this we must use the *periodicity* of \tilde{S} , i.e., the fact that \tilde{S} is invariant under the parabolic translation $x \rightarrow x + 1$.

Let l_t be the intersection of \tilde{S} with the horizontal plane $z = t$. Identifying $z = t$ with the (x, y) -plane, we think of l_t as the graph of a function $y = f_t(x)$ which satisfies $f_t(x + 1) = f_t(x)$. Since \tilde{S} is stable, the norm of its curvature is bounded. Since as remarked above, the tangent plane to \tilde{S} is uniformly close to vertical when z is sufficiently large, we can estimate $|\partial^2 f_t / \partial x^2| = O(1/t)$ (and $|\partial^2 f_t / \partial t^2| = O(1/t)$), and so (by the periodicity of f_t) there is an estimate $\max(f_t) - \min(f_t) = O(1/t)$. But the minimal surface $\tilde{S} \cap B_t$ is trapped in the convex hull of $\tilde{S} \cap \partial B_t = l_t \cup \infty$ for each t . So $\max(f_t)$ is monotone decreasing as $t \rightarrow \infty$ while $\min(f_t)$ is monotone increasing, and both have the same limit, which can be taken to be 0 after composing with an isometry. Hence setting π equal to the vertical plane $y = 0$ satisfies the first claim of the sublemma.

It remains to show that two different components of the preimage of S have distinct osculating planes. Let \tilde{S}_1, \tilde{S}_2 be two components of the preimage of S , and let π_1, π_2 be their osculating planes. Since the \tilde{S}_i are disjoint, the π_i can't cross, so without loss of generality, we can set π_1 to be $y = 0$ and π_2 to be $y = y_0$. We want to show $y_0 \neq 0$. Let $f_{t,i}$ be as above, and suppose without loss of generality that $f_{t,1}(x) < f_{t,2}(x)$ for each t and each x . We want to show that $y_0 > 0$.

Suppose not, so that $y_0 = 0$ and $\pi_1 = \pi_2$. Since both $f_{t,1}$ and $f_{t,2}$ are invariant under $x \rightarrow x + 1$, there is a constant $C > 0$ with $f_{t_0,2} - f_{t_0,1} > C$ for some fixed $t = t_0$. Let \tilde{S}'_2 be obtained from \tilde{S}_2 by translation $y \rightarrow y - C/2$, and let $g_{t,2} = f_{t,2} - C/2$ for each t . Then $g_{t_0,2} - f_{t_0,1} > C/2 > 0$ but by the definition of the planes π_1, π_2 and the hypothesis that they are equal, there is $t_1 > t_0$ with $g_{t_1,2} - f_{t_1,1} < 0$. Note that both \tilde{S}'_2 and \tilde{S}_1 are invariant under $x \rightarrow x + 1$, and we can assume that t_0, t_1 are big enough so that the projection of B_{t_0} is contained in the Margulis tube of the cusp of AB . In other words, the ends of $\tilde{S}_1/(x \rightarrow x + 1)$ and $\tilde{S}'_2/(x \rightarrow x + 1)$ cross in an essential loop. By cut-and-paste and the roundoff trick, we can reduce the area of S , contradicting the fact that S was least area. This contradiction shows $y_0 > 0$, as claimed. \square

In our context, the surface S indeed intersects the cusp in two components, giving rise to \tilde{S}_1, \tilde{S}_2 as above, where the thin part of A is covered by the slab contained between \tilde{S}_1 and \tilde{S}_2 in B_t . Let A_t denote this slab, and let π_t denote the slab contained between π_1 and π_2 in B_t . Since there are estimates $|f'_{t,i}| = O(1/t)$, $|\partial f_{t,i}/\partial t| = O(1/t)$ for $i = 1, 2$, the map $s : A_t \rightarrow \pi_t$ defined in coordinates by

$$s : \left(x, f_{z,1}(x) + \frac{t(y - f_{z,1}(x))}{f_{z,2}(x) - f_{z,1}(x)}, z \right) \rightarrow (x, ty_0, z)$$

is $(1 + \epsilon)$ -bi-Lipschitz, where $\epsilon \rightarrow 0$ as $t \rightarrow \infty$. In words, s fixes x and z , and for each line l which is parallel to the y -axis and intersects A_t , it takes $l \cap A_t$ to $l \cap \pi_t$ linearly. The bi-Lipschitz constant can be estimated by the ratio $(f_{z,2} - f_{z,1})/y_0$ and by $1 +$ the norm of the partial derivatives of the graphs f_1, f_2 in the x and z directions.

Doubling, there is a $(1 + \epsilon)$ -bi-Lipschitz map between the cusp of AA and the hyperbolic manifold (with horotorus boundary) obtained by doubling the quotient of the slab $\pi_t/\langle x \rightarrow x + 1 \rangle$. By Simon [43], Thm. 5.2, flow exists for a definite time $[0, T]$ on the singular manifold AA .

The components of \tilde{S}_k converge C^∞ on compact subsets to components of \tilde{S} . It follows that for large enough k , two components T_1, T_2 of \tilde{S}_k which intersect the same thin part have subdisks D_i whose boundaries are C^∞ close to a pair of curves corresponding to components of $\tilde{S} \cap \partial B_t$, where t is arbitrarily large (but fixed). In the notation of the previous paragraph, we may assume t is large enough so that $|f_{t,1}| < \epsilon$, $|f_{t,2} - y_0| < \epsilon$ where $0 < \epsilon \ll y_0$. In other words, there are round circles C_1, C_2 in \mathbb{H}^3 which are a constant distance $\kappa = y_0/t$ apart so that each ∂D_i is distance $\epsilon\kappa$ from each C_i . Let E_i be a totally geodesic disk spanning each C_i , and let E_i^\pm be almost parallel totally geodesic disks spanning nearby round circles C_i^\pm so that D_i is trapped between E_i^\pm for each i , by a barrier argument. We can choose E_i^\pm so that E_1^+ and E_1^- are much closer than E_1 and E_2 (see Figure 6). The slab between the outer disks E_1^+ and E_2^- can be foliated by totally geodesic arcs γ perpendicular to a bisecting plane such that for each γ , and each choice of signs \pm , there is an inequality

$$\frac{d(D_1 \cap \gamma, E_1^\pm \cap \gamma) + d(D_2 \cap \gamma, E_2^\pm \cap \gamma)}{d(D_1 \cap \gamma, D_2 \cap \gamma)} \leq 2\epsilon$$

(this follows just by comparing convex hulls). Since the D_i are C^∞ close to the E_i , each arc γ intersects each D_i and E_i almost orthogonally, with error term of order $O(\epsilon)$. Stretching each arc $[\gamma \cap D_1, \gamma \cap D_2]$ linearly over $[\gamma \cap E_1, \gamma \cap E_2]$ defines a $(1 + O(\epsilon))$ -bi-Lipschitz map between the slab of hyperbolic space contained between the E_i and the slab contained between the D_i . The quotient of the second slab by a rotation of order k is half of the thin part of AA_k . Doubling the quotient of the hyperbolic slab gives a suitable background metric on the thin part of AA_k , so flow exists for time $[0, T]$ on each AA_k and also on AA_k^δ for δ sufficiently small (depending on k).

Sublemma 5.12 motivates the following conjecture:

Conjecture 5.13. *Let S_1, S_2 be complete, locally least area minimal surfaces in \mathbb{H}^3 which are either equal or disjoint, and whose closures contain the same point $p \in S_\infty^2$. If the Hausdorff distance of $S_1 \cap B_t$ and $S_2 \cap B_t$ is $o(e^{-t})$, then $S_1 = S_2$.*

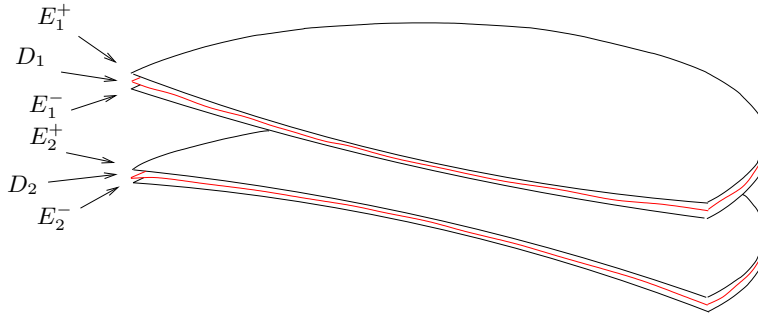


FIGURE 6. The thin slabs between E_1^+ and E_1^- , and between E_2^+ and E_2^- , trap D_1 and D_2 , and show that they cobound something close to the “thick” slab between E_1 and E_2 .

Lemma 5.10 enhances Lemma 5.11 by asserting that there is a diagonal sequence $\delta(k) \rightarrow 0$ so that the time $[0, T]$ flows on the manifolds $AA_k^{\delta(k)}$ converge pointwise and C^∞ away from time 0 to the time $[0, T]$ flow on AA .

We turn now to Hamilton’s equation [18], p. 698, for the evolution of scalar curvature R under Ricci flow “with cosmological constant”. This is simply flow with a homothetic rescaling to maintain constant volume; denote it by flow_0 . The equation is

$$(5.3.2) \quad \frac{dR}{dt} = \Delta R + 2|\text{Ric}_0|^2 + \frac{2}{3}R(R - r).$$

In the above, Ric_0 is the traceless Ricci tensor, and r is the (spatial) average scalar curvature. We will use this equation on a finite volume, but noncompact, manifold. Its original derivation, for compact manifolds, still holds. Alternatively, note that we only apply (5.3.2) to compact manifolds and their Gromov limits, so the extension also follows from Lemma 5.10.

Parallel to [1], to prove Theorem 5.5 it suffices to show, using equation (5.3.2) (or any other method), that after short-time flow_0 on AA , there is an inequality $\inf(R) > -6$. Let us review this argument. Rescaling the metric so that $\inf(R) = -6$ initially decreases volume. Then, since Perelman’s Ricci surgery is both volume decreasing and infimum scalar curvature nondecreasing (see Perelman [34, 35]), flow with surgery, normalized to $\inf(R) = -6$, monotonically reduces volume. Thus Theorem 5.5 is a consequence of the following:

Goal. For sufficiently small $t > 0$, show that $\text{flow}_0(t)(AA)$ satisfies $R > -6$.

To obtain the goal we derive a kind of parabolic maximum principle suitable to the cusped context.

As an important first step:

Lemma 5.14. For $t \in [0, T]$, flow_0 on AA satisfies $R \geq -6$.

Proof. By the convergence $AA_k^\delta \rightarrow AA$ (here $\delta = \delta(k) \rightarrow 0$ as $k \rightarrow \infty$), the inequality $R \geq -6$ for the initial metric on each AA_k^δ , and the existence and convergence of flow_0 for time $[0, T]$ for each k to flow for time $[0, T]$ on AA (i.e., Lemma 5.9, Lemma 5.10 and Lemma 5.11) it suffices to obtain this inequality termwise for each k , and take a limit. From equation (5.3.2), we see that at a negative spatial

minimum $\check{R} := \min(R)$, the right hand side is nonnegative. So, the parabolic maximum principle implies that, on the compact manifold AA_k^δ , \check{R} is strictly increasing. Taking limits we may lose strictness, but obtain the weaker inequality. \square

Assume that AA is not hyperbolic (equivalently that S is not totally geodesic, equivalently that AB is not obtained by a gluing that is isotopic to an isometry). Under this assumption we have:

Lemma 5.15. *Under flow₀, AA cannot satisfy $R(x, t) \equiv -6$ identically, at any finite time t .*

Proof. Since AA is not hyperbolic, it cannot become hyperbolic under flow in any finite time. Thus, if $R \equiv -6$, we must have $\text{Ric}_0 \neq 0$ for some (x, t) for all $t \in (0, T]$ and therefore, by analyticity, at a set of full measure of points x for each $t \in (0, T]$. At such points x , equation (5.3.2) reads

$$0 = 0 + |\text{Ric}_0|^2 + 0 > 0,$$

which is a contradiction. \square

We use the notation $AA(t)$ to denote the Riemannian manifold obtained by time t flow on AA .

Corollary 5.16. *$R(AA(t)) > -6$ for all $t \in (0, T]$.*

Proof. The inequality follows from $\inf(R) \geq -6$ (Lemma 5.14) and R not identically $\equiv -6$ (Lemma 5.15). \square

Next we need a lower bound on ΔR at points (x, t) where $R(x, t)$ is near its spatial infimum which is ≥ -6 . First, we state the principle for a single real variable.

Lemma 5.17. *Let $f : [-k, k] \rightarrow \mathbb{R}^+ \cup \{0\}$ be a nonnegative C^3 function. Let $c_3 = \max(f''')$ on $[-k, k]$. Then*

$$f(0) > -\frac{1}{4}f''(0) \left(\frac{f''(0)}{c_3} \right)^2$$

provided that $k > 10\sqrt{\frac{f(0)}{f''(0)}}$. Equivalently, if

$$k > 10\sqrt{\frac{f(0)}{f''(0)}},$$

then

$$f''(0) > -(8c_3)^{\frac{2}{3}}f(0)^{\frac{1}{3}}.$$

Proof. The worst case is when $f'(0) = 0$. In this case, $\frac{f''(0)}{2c_3}$ provides a length scale over which f decreases faster than $-\frac{1}{2}f''(0)x^2$. \square

The next lemma is a multivariable version of Lemma 5.17.

Lemma 5.18. *Let $f : M \rightarrow \mathbb{R}^+ \cup \{0\}$ be a nonnegative C^3 function on a complete Riemannian manifold with (spatially) uniform bounds on curvatures R_{ijk}^l and their first three covariant derivatives. Then there is a uniform constant c_3 , depending only on the preceding constants, so that at any point $x \in M$, any second covariant derivative satisfies $\nabla^2 f(x) > -c_3 f(x)^{\frac{1}{3}}$.*

Proof. The proof is similar to that of Lemma 5.17 but in the covariant context. One just needs to observe that the highest-order terms in ∇^2 dominate. \square

The uniform bounds required to apply Lemma 5.18 are provided by Simon [43], Thm 1.1, p. 1039, which in our context becomes:

Lemma 5.19 (Simon [43], Thm 1.1). *Let M be a manifold. Let g_0 be a complete C^0 metric and h a complete C^∞ background metric on M , which is $(1 + \epsilon)$ -bi-Lipschitz to g_0 (where ϵ is a universal constant depending on the dimension) and h satisfying uniform $\nabla^m R_{ijk}^l$ bounds, $|{}^h\nabla^m R_{ijk}^l| < k_m$. Then there exists $T > 0$, a function of the k and of dimension, and a C^∞ family of C^∞ metrics $g(t)$ for $t \in (0, T]$ solving flow, such that h remains $(1 + 2\epsilon)$ -bi-Lipschitz to the family and*

(1)

$$\limsup_{t \rightarrow 0} \sup_{\cdot \in M} |g(\cdot, t) - g_0(\cdot)| = 0,$$

(2)

$$\sup_{x \in M} |{}^h\nabla^i g|^2 \leq \frac{c_i(\dim, k_0, \dots, k_i)}{t^i},$$

where ${}^h\nabla$ denotes the covariant derivative in the h metric.

We will use $R + 6$ on AA with the background metric h as the nonnegative function in Lemma 5.18. First spatial derivatives $\nabla(\Delta R)$ of the Laplacian are estimated by third covariants $|\nabla\Delta R| = O|\nabla^3 R|$ and, by Lemma 5.11 and Lemma 5.19, are uniformly bounded in x for any fixed interval $[t_1, t_2] \subset (0, T]$. Now Lemma 5.18 yields:

Lemma 5.20. *For any fixed time interval $[t_1, t_2] \subset (0, T)$ and any $t \in [t_1, t_2]$ there is a lower bound for ΔR on $AA(t)$; precisely,*

$$\Delta R(x, t) > -c_3(R(x, t) + 6)^{\frac{1}{3}},$$

where c_3 is a positive constant independent of x or $t \in [t_1, t_2]$. \square

Now consider Hamilton's equation (5.3.2) at initial time t_1 for flow₀ on $AA(t_1)$. We set $\check{r} = \min_{t \in [t_1, t_2]} r(t)$. By Corollary 5.16 and compactness of $[t_1, t_2]$, there is a strict inequality $\check{r} > -6$.

Claim. There is an $r_0 \in (-6, \check{r}]$ with the property that at all (x, t) with $t \in [t_1, t_2]$ such that $R(x, t) < r_0$, there is an inequality

$$(5.3.3) \quad \Delta R(x, t) + \frac{2}{3}R(x, t)(R(x, t) - \check{r}) > \frac{1}{3}R(x, t)(R(x, t) - \check{r}).$$

In words, the claim says that the negative contribution of ΔR can cancel at most half of the positive contribution of the final term in Hamilton's equation (5.3.2). Assuming this claim for the moment, we complete the proof.

From the claim, throughout $[t_1, t_2]$, there is an inequality $dR(x, t)/dt \geq u$, where $u := \frac{r_0}{3}(r_0 - \check{r})$ at those points (x, t) with $R(x, t) \in [-6, r_0]$. That is, at such points, $R(x, t)$ is increasing at a definite rate. Thus, after flow₀ for time $t_2 - t_1$, starting at time t_1 we conclude:

$$(5.3.4) \quad R(x, t_2) \geq \min(r_0, (t_2 - t_1)u - 6) > -6.$$

Equation (5.3.4) implies that the (unnormalized) flow on AA instantly reduces volume:

$$\text{vol}(AA(t)) < \text{vol}(AA) \text{ for all } t \in (0, T].$$

The proof of Theorem 5.5 in the cusped case now follows exactly as in [1]. We now give the proof of the claim.

Proof. Here is how to construct r_0 so that equation (5.3.3) holds when $R(x, t) < r_0$.

We assume that we have already chosen the constant and derive (easily satisfied) conditions that it must satisfy. So assume that $R(x, t) < r_0$. We have $\Delta R > -c_3(r_0 + 6)^{\frac{1}{3}}$. Set $\epsilon' = \check{r} + 6$ and $\epsilon = r_0 + 6$. Ultimately we will choose ϵ (much smaller than ϵ') and thereby choose r_0 .

The second term of equation (5.3.3) exceeds $\frac{2r_0}{3}(r_0 - \check{r})$,

$$\frac{r_0}{3}(r_0 - \check{r}) > c_3(r_0 + 6)^{\frac{1}{3}},$$

that is,

$$(\epsilon - 6)(\epsilon - \epsilon') > 3c_3\epsilon^{\frac{1}{3}}.$$

Since we want ϵ very small, $\epsilon - 6$ is negative and bounded away from zero. So by replacing c_3 with a nearly identical constant c'_3 , we have $(\epsilon' - \epsilon) > c'_3\epsilon^{\frac{1}{3}}$, which can be rearranged as $c'_3\epsilon^{\frac{1}{3}} + \epsilon < \epsilon'$. Obviously we can find such an $\epsilon' > 0$ with this property, and then set $r_0 = \epsilon - 6$.

This completes the proof of Theorem 5.5. \square

5.4. Hyperbolic complexity. We are therefore justified in defining the hyperbolic complexity c_h as follows.

Definition 5.21. Let M be a connected, complete, finite-volume, orientable hyperbolic 3-manifold. Define $c_{ch}(M) = (-\text{vol}(M), \sigma(M))$ as a lexicographic tuple.

Definition 5.22. Let M be a complete, finite-volume, hyperbolic 3-manifold. Define $c_h(M) = \{c_{ch}(M_i)\}$ as a lexicographic tuple, where M_i are the components of M . We also use c_h to denote the ‘‘hyperbolic complexity’’ of a connected closed irreducible 3-manifold by applying c_h to the hyperbolic JSJ pieces, listed in decreasing order.

Theorem 5.23 (c_h -Lemma Schema). *Let S be an orientable surface of finite type with no sphere or torus components, and let $(A, S), (B, S)$ be distinct elements of $\mathcal{M}(S)$ which are irreducible, atoroidal and acylindrical, and such that every component has some part of their boundary on S . Then*

$$c_h(AB) < \max(c_h(AA), c_h(BB)).$$

Proof. This follows from Theorem 5.5 and Lemma 5.4. \square

6. ASSEMBLY COMPLEXITY

6.1. Introduction. Let us pause to review where we are in the proof.

We seek a complexity function c , defined on homeomorphism classes of closed 3-manifolds, such that if A and B have common boundary S , then $c(AB) \leq \max(c(AA), c(BB))$, with equality holding only if A and B are homeomorphic rel S . We define c as a lexicographic tuple, with terms introduced throughout the course of this paper. The input for this section will be the complexities c_S and c_h of Sections 4 and 5. The output of this section will be a complexity, c_p , defined on prime (actually irreducible) 3-manifolds. We assume throughout this section that AB, AA and BB are connected and irreducible. The surface S , which need not be connected, is further assumed to be incompressible in both A and B .

Assuming $c_S(AB) = c_S(AA) = c_S(BB)$ and $c_h(AB) = c_h(AA) = c_h(BB)$, we then know that these three closed 3-manifolds become homeomorphic after cutting along JSJ tori. Lemma 4.22 and Theorem 5.23 allow us to draw a similar conclusion about the nonclosed 3-manifolds A and B . We recall the framework of these arguments.

6.1.1. *Notation for pieces.* Consider the ways in which S might sit with respect to the JSJ decomposition of the glued up manifold which we refer to as AB (with the same terminology applying to AA and BB). We distinguish between two different kinds of torus components of S :

Definition 6.1. A *JSJS cusp* is a component T of S which is a JSJ torus. An *internal cusp* is a component T of S which is contained in the interior of a JSJ piece in AB , built from Seifert fibered pieces in A and B containing T in their boundary, and admitting compatible fiberings of T .

Internal cusps contribute to m ($= \#$ of independent tori) but not m' ($= \#$ of JSJ tori) in the complexity c_S defined in Definition 4.20. In the proof of Lemma 4.22 it is established that each JSJ torus is either a JSJS cusp, or (after isotopy) can be chosen to intersect S in a pair of essential circles, which cut the torus into two essential annuli which are proper in A and B , respectively.

Thus the relative JSJ decompositions of A and B are compatible with the JSJ decomposition of AB . We say that a JSJ piece of A is a *boundary JSJ piece*, or *boundary piece* for short, if it has some boundary component on S which is not a JSJS cusp. In other words, boundary pieces correspond to JSJ pieces of AA (or AB or BB) that are cut by S .

6.1.2. *Assembly complexity.* What, then, is left to do? We need to define a further complexity term c_a such that AB , AA and BB as above satisfy

$$c_a(AB) \leq \max(c_a(AA), c_a(BB))$$

with equality if and only if A and B are diffeomorphic rel S (in other words, if and only if $(A, S) = (B, S)$). The term c_a is sensitive to the way in which the JSJ pieces are *assembled*.

The way in which a 3-manifold is assembled from its JSJ pieces can be encoded by a decorated graph. The vertices of the graph are labeled by (a representative of) the homeomorphism type of a JSJ piece. The edges are labeled by the gluing homeomorphisms of the cusps. Similarly, a relative JSJ decomposition can be encoded by a decorated *relative* graph, similar in many ways to the relative sum graphs defined in §3.3. Some of the vertices/edges in a relative graph are special:

- Some of the vertices (those corresponding to boundary pieces) are *half vertices*.
- Some of the edges (those corresponding to proper JSJ annuli) are *half edges*.
- There are *cut points* corresponding to tori components of S . We do not consider the cut points to be vertices.
- There are also *cut edges* which connect a cut point to an ordinary vertex (or half vertex or another cut point), which correspond to the JSJS cusps and the JSJ pieces (on one side) which they bound.

In the graphical context, two half vertices glue together to make a vertex. Two half edges are glued *lengthways* to make an edge, and two cut edges are joined *at*

their cut points to make an edge. We also allow the case of a *doubly cut edge*, which is an isolated interval component of the graph, with two cut points and no vertices. These of course correspond to $T^2 \times I$. Compare with the terminology in Definition 3.15, where half vertices corresponded to components which intersected S and half edges corresponded to disks.

There is a symmetry group associated with each vertex and half vertex, and gluing information associated to each edge or half edge. By uniqueness of the JSJ decomposition, two 3-manifolds are homeomorphic if and only if there is an isomorphism between their JSJ graphs which preserves vertex labels and is compatible with the gluing data, up to the action of the symmetry groups.

6.1.3. *Graph tensor TQFTs.* As a warm-up problem, we establish positivity for a few “toy” unitary TQFTs associated to labeled graphs with internal partial symmetries. In the simplest case, graphs are glued only along cut points; this is analogous to the case in which S consists entirely of JSJS cusp components. In the second case, graphs are glued along subgraphs, corresponding to boundary pieces of A and B .

The strategy of the proof is to build a tensor TQFT for labeled graphs with certain desirable properties. Chief among these are positivity (often called “unitarity” in the TQFT context) and genericity.

Definition 6.2. A *graph tensor TQFT* consists of the following data:

- for every edge type e , a (real or complex) vector space V^e with a positive nondegenerate inner product (symmetric in the real case, Hermitian in the complex case),
- for every vertex type v with incident edge types indexed by a (finite) set $E(v)$, a tensor T^v in $\bigotimes_{e \in E(v)} V^e$.

A labeled graph tensor TQFT is *symmetric* if there is only one edge type, and each vertex tensor is chosen to lie in the symmetric tensor power $T^v \in S^{n(v)}V$, where $n(v)$ is the number of edges incident on v .

A cut point does not get a tensor. A doubly cut edge gets the identity element of $\text{Hom}(V^e, V^e)$, where e is the edge label. Using the duality this can be denoted as $v_i \otimes v^i$, $v_i \otimes v_i$, or $v^i \otimes v^i$.

6.1.4. *Contracting tensors.* The positive nondegenerate inner product identifies V^e with $(V^e)^*$ when V^e is real, or with $(\overline{V^e})^*$ when V^e is complex. Thus the indices of T^v may be raised or lowered as desired. So if v_1, v_2 is a pair of vertices which share an edge of type e , we may contract the tensors T^{v_1} and T^{v_2} along their respective indices which correspond to the factor of type V^e . Technically, this amounts to forming $T^{v_1} \otimes T^{v_2}$, identifying an element of $V^e \otimes V^e$ inside this tensor, using the inner product to think of this as an element of $V^e \otimes (V^e)^*$, and then taking a trace to replace it with a scalar. By contracting along edges in this way, a labeled graph with “free edges” e_1, \dots, e_n gives rise to a tensor in $\bigotimes_i V^{e_i}$, and a labeled graph with no free edges gives rise to an element of \mathbb{R} or \mathbb{C} . Pairs of graphs with isomorphic sets of free edges can be glued together, and the result is a graph whose “value” is obtained by appropriately contracting the two tensors associated to the subgraphs. (Note: we use the terminology *cut edges* in §6.2 instead of *free edges*, but the meaning should be clear in each case.)

Typographically, this can be achieved by choosing an orthonormal basis for each vector space V^e , and then expressing the components of the tensors T^v in terms of this basis, as (lower) indices, taking values in the set of basis elements. Contracting is indicated by repeating an index, which in the Einstein summation convention means “sum over all possible values of this index”. So, for example, $T_{ij}T_{jk}$ means $\sum_{j=e_1}^{e_n} T_{ij}T_{jk}$.

For the sake of clarity, we give a few examples. As a simplification, assume that we are working with a symmetric labeled graph tensor TQFT. Again for simplicity, assume that all vertices are 3-valent and have the same type; denote the vertex tensor by x . Let e_1, \dots, e_n be an orthonormal basis for V . In a symmetric tensor, the positions of the indices in a given tensor are immaterial, so that for instance $x_{ijk} = x_{kij} = x_{jki}$.

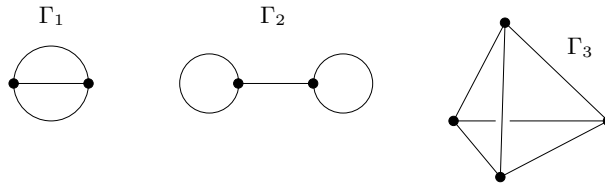


FIGURE 7

Example 6.3. Consider the three graphs in Figure 7. The theta graph Γ_1 has invariant $x_{ijk}x_{ijk}$. The barbell graph Γ_2 has invariant $x_{ij}x_{kkj}$. The tetrahedron graph Γ_3 has invariant $x_{ijk}x_{ilm}x_{jln}x_{kmn}$.

Any choice of vector spaces and tensors as above defines a TQFT, and a generic choice of tensors defines a TQFT which is powerful enough to prove a version of positivity. If vertices are decorated with symmetry groups which act on their incident edges, one must choose tensors with the same symmetries. Furthermore, edges might have “internal” symmetries, reflected in the TQFT by choosing V^e which are not just vector spaces but G -modules for various groups G , and in this case the tensors associated to vertices must be chosen in such a way as to respect this G -action. One way to do this is to replace V^e by some tensor power (on which G also acts) and choose the various tensor factors in T^v to lie in the G -invariant subspaces.

Labeled graph tensor TQFTs of progressively greater complexity are constructed in §6.2. The contents of this section are not logically necessary for the rest of the paper, but they will aid the reader considerably in following the logic in §6.3 and subsequent sections. Also, we have not attempted to be comprehensive in this motivational subsection; a paper explaining the full scope of “graph positivity” should still be written.

The construction of a suitable complexity function for JSJ graphs is modeled on the template of proof developed in §6.2. It has the following outline. First, choose a vector space for each JSJ torus, and choose compatible tensors T_X associated to each JSJ piece X . The JSJ tori have “internal” symmetries (parameterized by copies of $\mathrm{GL}(2, \mathbb{Z}) = \pm \mathrm{SL}(2, \mathbb{Z})$) and these symmetries interact with the symmetries of the JSJ pieces they bound. The major difficulty is to find suitably symmetric T_X . If T_X has too little symmetry, the partition function will not be well defined;

too much and it will not distinguish distinct glued-up manifolds. When X is hyperbolic, tensors T_X with the correct (finite) symmetries can be constructed directly. Seifert fibered pieces X present more of a problem, since their symmetries are of infinite order. So first we construct tensors \widehat{T}_X for Seifert fibered X with *too much* symmetry, which distinguishes AA , AB and BB only up to an equivalence relation slightly weaker than homeomorphism, which we call *fiber slip homeomorphism*, or *fish* for short. Then we adjust \widehat{T}_X , replacing it by a new tensor T_X which is sensitive not just to the topology of X , but the way in which it is glued to its immediate neighbors in the JSJ decomposition, to control for fish equivalence. There is one further technical point which should be mentioned here. When we come to define the assembly complexity c_a it will be a pair $c_a = (c_r, c_t)$. The first term, a “reflection symmetry” term at the JSJ tori, regularizes the picture near S and ensures the conditions under which the tensor contraction term c_t satisfies diagonal dominance.

Remark 6.4. The reader may wonder why the issue of assembly complexity cannot be treated by simply adding a term to c which “likes” lots of symmetry (e.g. by generalizing the term c_r somehow). This is a very seductive idea as it appears to offer a rapid finish. Unfortunately, one may produce examples where AB has “more” symmetry than either AA or BB for all notions of “symmetry” which we were able to quantify.

6.2. Graph positivity. As a warm-up, we now prove two versions of positivity for labeled graphs. The vertices of the graph are analogous to JSJ pieces, the edges are analogous to cusps, and a vertex label is analogous to the homeomorphism type of a JSJ piece. Accordingly, we require that the vertex label determine the valence of a vertex. We assume the set of possible vertex labels is countable and that all graphs are finite.

6.2.1. Recovering graphs from tensors. In order for our invariants to say anything about the topology and combinatorics of graphs, it is essential that the invariants are at least powerful enough to *distinguish* graphs. We prove a lemma to this effect initially in the context of edge-symmetric, labeled graph tensor TQFTs, where vertices are distinguished only by the number of incident edges, so there is exactly one vertex type for each nonnegative integer. We call this the context of “ordinary” graphs.

Lemma 6.5. *Let V be a real vector space of dimension m . Let Γ_1, Γ_2 be two (ordinary) graphs with at most m edges. Assume an identification between the cut points (boundary) of Γ_1 and Γ_2 . Then for a generic assignment of tensors $T^v \in S^{n(v)}V$ to vertex types, the invariants $T(\Gamma_1)$ and $T(\Gamma_2)$ are not equal unless Γ_1 and Γ_2 are isomorphic rel boundary.*

Proof. Fix an orthonormal basis of V . For each vertex type v , and each (unordered) set (with multiplicity) of indices for T^v , choose a variable representing the value of the given tensor entry. We distinguish these indeterminates by letters x, y, z depending on the cardinality of $n(v)$, and the indices by subscripts. So for example, x_{ijk} and y_{abcd} are examples of these variables. Then the result of tensor contraction can be expressed as a *homogeneous polynomial* in these variables, with degree equal to the number of vertices. In the Einstein summation notation, the value of $T(\Gamma_1)$ is represented by a monomial whose subscripts are indeterminate indices which appear singly or in pairs: one pair for each (uncut) edge of Γ_1 and a singleton for

each cut edge of Γ_1 . Since by hypothesis the number of edges is at most equal to the dimension of V , there is some monomial in the polynomial representing $T(\Gamma_1)$ in which each basis element appears as a subscript at most twice. By taking a vertex for each variable which appears in the monomial, an edge for each pair of variables with a common index, and a cut edge for each singleton index, we can reconstruct Γ_1 from the polynomial, and similarly for Γ_2 . This monomial contains the “gluing instructions” for the graph: if b , say, occurs as a subscript for both x and y , then some edge must join the vertices corresponding to x and y . It follows that if the entries of the various T^v are chosen algebraically independently, the values of the invariants $T(\Gamma_1)$ and $T(\Gamma_2)$ are different unless Γ_1 and Γ_2 are isomorphic relative to the identity on cut edges. \square

To put ourselves in the favorable case where the number of edges of a given graph is less than the dimension of V , we may let V_1, V_2, V_3, \dots be a sequence of vector spaces with inner products, with $\dim(V_i) = i$ for each i . Construct a symmetric labeled graph tensor TQFT T_i for each i , with algebraically generic entries. Then for any two finite graphs Γ_1, Γ_2 the strings of invariants $T_1(\Gamma_1), T_2(\Gamma_1), \dots$ and $T_1(\Gamma_2), T_2(\Gamma_2), \dots$ are equal *as strings* if and only if Γ_1 and Γ_2 are isomorphic *as graphs* (taking the cut point boundary of the graph into account, of course). The positivity theorem below is a consequence of being able to distinguish graphs via a family of tensor TQFTs.

Let C be a finite ordered set of cut points. Let $\dot{\mathcal{N}}_C$ denote the set of isomorphism classes of ordinary graphs with boundary (cut points of cut edges of the graph) identified with C . Let \mathcal{N}_C denote the vector space spanned by $\dot{\mathcal{N}}_C$. Gluing along cut points yields a pairing into closed graphs:

$$\mathcal{N}_C \otimes \mathcal{N}_C \rightarrow \mathcal{N}.$$

Theorem 6.6. *The above pairing (for ordinary graphs) is positive.*

Proof. This theorem is an easy special case of Theorem 6.8, so we defer the proof until then. \square

For a second warm-up example, let us consider a class of asymmetric graphs. These are graphs with: (1) one edge type, (2) arbitrarily many vertex types, and (3) no vertex symmetry (i.e., the edges leaving a vertex are ordered). Two graphs are considered isomorphic if and only if there is an isomorphism between the underlying unlabeled graphs which preserves vertex labels and also the ordering of the edges at each vertex. If the graphs have boundary, then the isomorphism must preserve the ordering of the cut points. Let \mathcal{N} denote the real vector space generated by finite linear combinations of isomorphism classes of such graphs.

For C a finite ordered set of cut points, let $\dot{\mathcal{N}}_C$ denote the set of isomorphism classes of labeled graphs whose “boundary” is C . Two such graphs are considered isomorphic if there is an isomorphism as above preserving labels and orderings, and which is the identity on C . Let \mathcal{N}_C denote the vector space generated by $\dot{\mathcal{N}}_C$. Gluing along cut points gives a pairing

$$\mathcal{N}_C \otimes \mathcal{N}_C \rightarrow \mathcal{N}.$$

Theorem 6.7. *The above pairing, for asymmetric vertices and gluing along cut points, is positive.*

Proof. We will construct below a family of unitary TQFTs $\{Z_i\}$, $i = 1, 2, 3, \dots$, for graphs which eventually distinguish nonisomorphic (relative) graphs in $\dot{\mathcal{N}}_C$. If A is a closed graph, define $Z(A)$ to be the infinite tuple $(Z_i(A))$, with lexicographic ordering. We will show that $Z(\cdot)$ is a diagonally dominant complexity function. If $A, B \in \dot{\mathcal{N}}_C$ are graphs with boundary C , then using only the Atiyah gluing axiom (which holds trivially for graph tensor TQFTs) and the Cauchy-Schwarz inequality, we obtain (for all i)

$$\begin{aligned} |Z_i(AB)| &= |\langle Z_i(A), Z_i(B) \rangle| \\ &\leq (|\langle Z_i(A), Z_i(A) \rangle| \cdot |\langle Z_i(B), Z_i(B) \rangle|)^{\frac{1}{2}} \\ &\leq \max(Z_i(AA), Z_i(BB)), \end{aligned}$$

with equality holding only if $Z_i(A) = Z_i(B)$. If A and B are not isomorphic, we will show below that there exists an i such that $Z_i(A) \neq Z_i(B)$. It follows that $Z(AB) < \max(Z(AA), Z(BB))$. In other words, Z satisfies the diagonal dominance inequality.

It remains to construct the TQFTs Z_i . Choose a real inner product space $V = V_i$ of dimension i . Fix an orthonormal basis for V . The vector space $V(C)$ assigned to C will be $V^{\otimes C}$. For each vertex label α of valence k , choose a tensor $T^\alpha \in V^{\otimes k}$. With respect to the basis of V , this means choosing numbers $T_{i_1 \dots i_k}^\alpha \in \mathbb{R}$. The key idea is to choose these numbers (for all vertex labels and all multi-indices) to be algebraically independent of one another. For a labeled graph G (possibly with a boundary consisting of cut points) define $Z(G) \in V(\partial G)$ to be the tensorial contraction described above. (See also for example Penrose [33].)

Suppose now that A and B are graphs with boundary C , and that $Z(A) = Z(B)$. An easy variation on the proof of Lemma 6.5 shows that if $Z_i(A) = Z_i(B)$, where $i = \dim V$ is greater than the number of edges in A and B , then A and B must be isomorphic. \square

Next we consider the case where each vertex type (label) has a specified symmetry group, i.e., a subgroup of the permutation group of the edges emanating from the vertex. These symmetry groups are analogous to symmetries of JSJ pieces. Two graphs are considered isomorphic if and only if there is an isomorphism between the underlying unlabeled graphs which preserves vertex labels and such that the induced permutation of the (ordered) edges at each vertex lies in the symmetry group for that vertex type. If the graph has boundary, the isomorphism is required to be the identity restricted to the boundary. As before we define vector spaces \mathcal{N} (for closed graphs), \mathcal{N}_C (for graphs with boundary C), and pairings $\mathcal{N}_C \otimes \mathcal{N}_C \rightarrow \mathcal{N}$.

Theorem 6.8. *The above pairing, for vertices with symmetry groups and gluing along cut points, is positive.*

Proof. The proof is similar to that of the previous theorem, except that for each vertex type we choose a tensor that has precisely the symmetry (neither more nor less) specified for that vertex type. We do this by first choosing unsymmetric tensors (with algebraically independent entries, as before), and then symmetrizing. (The symmetry group of a vertex of valence k acts on $\{1, \dots, k\}$, which induces an action on $V^{\otimes k}$.)

We must show that for graphs $A, B \in \dot{\mathcal{N}}_C$, $Z_i(A) = Z_i(B)$ for sufficiently large i only if $A \cong B$ (via allowed symmetries). $Z_i(A)$ is a contraction of symmetrized

tensors. Since symmetrizing commutes with contraction, we can also view $Z_i(A)$ as a sum of contractions of the original unsymmetric tensors. There will be one such summand for each gluing of the vertices of A which is compatible (via vertex symmetry groups) with a gluing which produces A from its vertices. It follows that if i is greater than the number of edges in A we can read off from $Z_i(A)$ (thought of as a polynomial in the unsymmetric tensor indices) the equivalence class of the adjacency matrix of the vertices of A (relative to C). By algebraic independence, $Z_i(A) = Z_i(B)$ only if corresponding polynomials are identical. It follows that the adjacency matrix of A can be transformed into that of B via vertex symmetries. In other words, $A \cong B$. \square

The third and final case of graph positivity we consider concerns gluing graphs along an intermediate subgraph H instead of gluing along edge cut points. This is analogous to the 3-manifold case where S has no JSJS cusp components; H corresponds to the unions of the JSJ pieces which intersect S , while A and B correspond to the remaining JSJ pieces. It is easy to extend the arguments given below to the case of gluing along the disjoint union of an intermediate subgraph and cut points, analogous to the case where S has both components which are non-JSJS cusps as well as components which are. In the most general case, we would allow arcs (α in Figure 8) in H with two cut points and meeting no vertices and also arcs (β and γ) in $\dot{\mathcal{N}}_C$ which meet no vertices. The basic idea is to always associate the identity $e^i \otimes e_i$ to α , $e^i \otimes e^i$ to β , and $e_i \otimes e_i$ to γ . We leave the details of this extension as an exercise for the reader.

For simplicity we assume that all vertex types are fully symmetric; the symmetry group of each vertex type is the full permutation group of the edges emanating from it. (We will see below that the symmetries of a Seifert fibered JSJ piece act as the full permutation group on its cusps.)

We assume that the boundary of H is partitioned into “upper” and “lower” cut points, both identified with C , and that H is equipped with an order two automorphism ι which permutes the upper and lower cut points and fixes all vertices and noncut edges of H (see Figure 8). This automorphism is analogous to reflecting the middle level of the 3-manifold AA (or BB) across S .

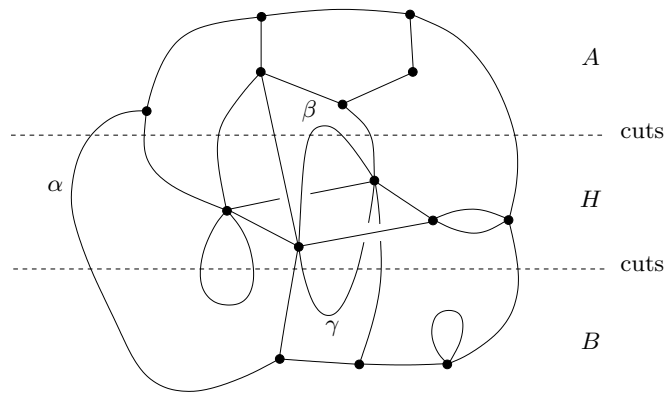


FIGURE 8

We must modify the definition of \mathcal{N}_C to take into account automorphisms of H . (So \mathcal{N}_C depends not just on C but also on H .) Two graphs in \mathcal{N}_C are considered isomorphic if there is an isomorphism of the underlying unlabeled graphs which preserves vertex labels and induces a permissible permutation of the boundary C . (Previously we required the isomorphism to be the identity on C .) A permissible permutation is one which preserves the adjacency relation to each vertex of H . Equivalently, a permissible permutation is one which extends to an automorphism of H which fixes all vertices of H and all noncut edges of H (and fixes doubly cut edges if these are permitted in H). Here we're identifying C with the upper (or lower) boundary of H . For example, the permissible permutations of C in Figure 8 can exchange cut points two and three, four and five, or six and seven (counting from the left). (If we considered the unsymmetrical context, "permissible" would have a more restricted meaning.)

Given $A, B \in \mathcal{N}_C$, define the closed graph AHB by gluing the boundary of A to the upper boundary of H and gluing the boundary of B to the lower boundary of H . This induces a pairing $\mathcal{N}_C \otimes \mathcal{N}_C \rightarrow \mathcal{N}$.

Theorem 6.9. *The above pairing, for fully symmetric vertices and gluing along a subgraph, is positive.*

Proof. As in the previous two proofs, we will construct a family of graph TQFTs whose tuple of partition functions is a diagonally dominant complexity function on closed graphs. We can think of $Z_i(H)$ as an operator

$$Z_i(H) : V_i^{\otimes C} \rightarrow V_i^{\otimes C}$$

from the Hilbert space of the lower cut points of H to the Hilbert space of the upper cut points of H . The isomorphisms used to define \mathcal{N}_C act on $V_i^{\otimes C}$ with invariant subspace $Z_i(\mathcal{N}_C) := \text{Inv}(V_i^{\otimes C}) \subset V_i^{\otimes C}$. We will show below that the vertex tensors from which the TQFT is constructed can be chosen so that the restriction of $Z_i(H)$ to $Z_i(\mathcal{N}_C)$ is strictly positive (for all i). That is, there exists an operator U such that $Z_i(H) = U^\dagger U$ and U is injective on $Z_i(\mathcal{N}_C)$. We then have (using "bra-ket" notation, and the dagger for adjoint)

$$\begin{aligned} |Z_i(AHB)| &= |\langle Z_i(A) | Z_i(H) | Z_i(B) \rangle| \\ &= |\langle Z_i(A) | U^\dagger U | Z_i(B) \rangle| \\ &= |\langle U(Z_i(A)), U(Z_i(B)) \rangle| \\ &\leq (|\langle U(Z_i(A)), U(Z_i(A)) \rangle| \cdot |\langle U(Z_i(B)), U(Z_i(B)) \rangle|)^{\frac{1}{2}} \\ &\leq \max(Z_i(AHA), Z_i(BHB)), \end{aligned}$$

with equality only if $U(Z_i(A)) = U(Z_i(B))$. Since U is injective on $Z_i(\mathcal{N}_C)$, this happens only if $Z_i(A) = Z_i(B)$. Since all entries in the symmetric tensors for each vertex type in A , B , and H may be assumed to be algebraically independent, a slight extension of the previous proof shows that this happens for all i only if $A \cong B$.

To illustrate this extension, imagine that $Z_i(\mathcal{N}_C)$ is $S^2(V)$, the symmetric square of V . This would result from two upper cut points emanating from one vertex of H . Then the symmetrized relative partition functions in $Z_i(\mathcal{N}_C)$ are

$$Z_i(A) = Z_i(A)_{p,q} + Z_i(A)_{q,p} \text{ and } Z_i(B) = Z_i(B)_{p,q} + Z_i(B)_{q,p},$$

where $\{p, q\}$ is a multi-index for $S^2(V)$. Algebraic independence ensures that these expressions are equal if and only if they have the same terms, up to permutation. In general, $Z_i(A) = Z_i(B)$ if and only if their *unsymmetrized* partition functions agree after some permissible permutation of cut edges. Then for i large enough to label all edges of A (B) differently, we conclude $A \cong B$.

It remains to show how to construct the (graph) TQFT so that $Z_i(H)$ is a positive operator on $Z_i(\mathcal{N}_C)$. In what follows, we suppress the index i .

Consider a single vertex type a with $n := n(a)$ emanating cut edges. Denote by $V^+ \subset V$ the nonnegative cone, consisting of nonnegative sums of elements from a fixed orthonormal basis e_1, \dots, e_d . We must find a tensor $T^a \in S^n V$ with the following property. Suppose v is a vertex of type a which appears as a vertex in H with k upper and lower cut edges which are interchanged by ι , and $n - 2k$ “horizontal edges” (i.e. edges joining the vertex to other vertices in H). Associated to k lower cut edges (resp. upper cut edges) there is a vector space V_l (resp. V_u) which parameterizes the possible inputs along these edges. Explicitly, each such vector space is isomorphic to the k th symmetric tensor power of V . The involution ι induces an isomorphism

$$\iota : V_u \rightarrow V_l,$$

and therefore it makes sense to say that an operator from V_l to V_u is *positive symmetric*.

The property we want is that after contracting T^a with $n - 2k$ vectors in V^+ (corresponding to input from the horizontal edges emanating from v), the result should be a positive symmetric operator from V_l to V_u . That is, for any k satisfying $2k \leq n$, and any vectors $v_j^+ \in V^+$ for $2k + 1 \leq j \leq n$,

$$(6.2.1) \quad T^a(v_{2k+1}^+, \dots, v_n^+) : V_l \rightarrow V_u \text{ is positive symmetric.}$$

Roughly, the entries of T^a (in the fixed basis e_1, \dots, e_d) are chosen by any formula which extravagantly weights index coincidences, and then perturbed to be algebraically independent. An example of what we mean by such an extravagant formula is given in the following equation:

$$(6.2.2) \quad T_{i_1, \dots, i_n}^a \approx 1 \cdot (2^{d^n})^{\# \text{ of pairs}} \cdot (2^{d^{n^2}})^{\# \text{ of triples}} \cdot (2^{d^{n^3}})^{\# \text{ of quadruples}} \dots$$

for any assignment of indices i_1, \dots, i_n chosen from the set $\{e_1, \dots, e_d\}$. Here “pairs” is the number of unordered pairs of indices with the same label, “triples” is the number of unordered triples of indices with the same label, and so on, so for example $T_{e_1, e_2, e_1, e_3, e_2, e_1, e_3, e_4}^a$ contains two pairs and one triple, and its value should have size of the order of $(2^{d^n})^2 \cdot 2^{d^{n^2}}$. There is a convenient criterion, originally due to Sylvester, to verify that a symmetric matrix M_{ij} , $1 \leq j \leq m$, acts as a positive operator.

Lemma 6.10. *M_{ij} is positive if and only if the partial determinants are all positive, i.e., $\det(M_{ij}) > 0$ for $i, j \leq l$, where $1 \leq l \leq m$.* \square

Easier than checking that T^a of the form suggested in (6.2.2) satisfies the criterion of Lemma 6.10 is to observe that the lexicographic tuple:

$$(6.2.3) \quad (\#2k \text{ tuples}, \#2k - 1 \text{ tuples}, \dots, \#\text{pairs})$$

is diagonally dominant under concatenating (symmetrized) strings of k symbols drawn from a finite set. Thus, if the finite ordered set consisting of values of the

tuple (6.2.3) is embedded in \mathbb{R}^+ with sufficiently (depending on d) rapid decay, then positivity follows from estimating determinants (as simply the “order of magnitude” of the product of diagonal elements) and then applying Lemma 6.10. In more detail, consider what we need to know to safely estimate a determinant of a matrix M simply as the product of the diagonal entries. It would suffice if all of the off-diagonal entries are much smaller than all of the diagonal entries. Unfortunately, we are not in this situation but we can easily arrange (normalizing all entries to be real and ≥ 1) that

$$M_{ij} < \frac{1}{n!} \sqrt{M_{ii}M_{jj}}$$

for all $i \neq j$ and n equal to the dimension of the matrix M . Given a permutation $\sigma \neq id$, it easily follows that

$$\prod_i M_{i,\sigma(i)} < \frac{1}{n!} \prod_i M_{ii}.$$

Thus, the diagonal term dominates the calculation of the determinant.

Similarly, when finitely many operators which weight index coincidences sufficiently extravagantly (depending on d and H) are (partially) contracted according to the combinatorics of H , the result is a *positive* operator $T(H)$ with precisely the desired symmetries. Since the property of being positive is open, we can perturb the tensors to be algebraically independent. This completes the proof. \square

6.3. Tensors for JSJ pieces. The main ingredient c_t in the assembly complexity will be similar to the graph complexity defined above. In broad outline, we will assign vector spaces to JSJ tori and appropriate tensors to JSJ pieces, then define the complexity of a closed manifold to be the result of a big tensorial contraction.

The gluing graph of JSJ pieces has three new features not present in the naked graph context:

- (1) The vertices (JSJ pieces) are now oriented manifolds, so we will use complex scalars \mathbb{C} and assign to complex conjugation the role of keeping track of orientation.
- (2) Each vertex has its own group of symmetries. Ideally, we should employ tensors with exactly the right symmetry: too little symmetry and the partition function will not be well defined, too much and subtle gluing distinctions that should distinguish A from B will be lost. In the case of hyperbolic vertices, the symmetries constitute a finite group G and we will build appropriately symmetric tensors from many smaller tensors or “micro-vertices” situated within each fundamental domain of G . Infinite symmetry in the Seifert fibered cases is a problem and requires a separate trick, namely passing to modular quotients.
- (3) The edges of the JSJ graph carry interesting information. If we assume each vertex type comes with a marking (i.e., a choice of coordinates) on (the homology of) its cusps, then “orienting an edge” is tantamount to choosing an element in $g \in \text{GL}(2, \mathbb{Z})$ which specifies the gluing (reversing the orientation sends g to g^{-1}). Thus, the vector space V on which we build tensors should have an action of $\text{GL}(2, \mathbb{Z})$ (but also a large $\text{GL}(2, \mathbb{Z})$ -invariant factor).

The fact that mapping class groups of Seifert fibered spaces are typically infinite needs to be addressed first since its solution determines the vector space V on which

we build tensors. Actually, we will construct a family of vector spaces $V_{m,m'}$ and tensor complexities $c_t^{m,m'}$, for $m, m' = 2, 3, 4$ and so on, where the vector space $V_{m,m'}$ is isomorphic to $\mathbb{C}^{m^2 m'}$ and admits a natural action of $\mathrm{GL}(2, \mathbb{Z})$ factored through the finite group $\mathrm{GL}(2, \mathbb{Z}_m)$. Explicitly, a basis for $V_{m,m'}$ is given by the elements in the finite abelian group $H_1(T^2; \mathbb{Z}_m)$ cross the set $\{1, 2, \dots, m'\}$ (here and throughout we use the “topologist’s notation” \mathbb{Z}_m for the finite cyclic group $\mathbb{Z}/m\mathbb{Z}$). The natural action of $\mathrm{GL}(2, \mathbb{Z})$ on $H_1(T; \mathbb{Z})$ induces an action of $\mathrm{GL}(2, \mathbb{Z})$ on $H_1(T^2; \mathbb{Z}_m)$ which factors through the quotient $\mathrm{GL}(2, \mathbb{Z}_m)$. While no single representation of $\mathrm{GL}(2, \mathbb{Z})$ is faithful, the intersection of the kernels is trivial, which is sufficient for our purposes. The action on $\{1, 2, \dots, m'\}$ is trivial. The numbers m and m' have no logical relation except that they both will be sent to infinity (for different reasons). For notational convenience, we set $m = m'$ and denote $V_{m,m'}$ as V_m and similarly $c_t^{m,m'}$ as c_t^m .

Unfortunately, a technical problem prevents us from getting the symmetry group exactly right. We will build in a little too much symmetry for the tensors associated to Seifert fibered JSJ pieces. The resulting partition function would not be able to distinguish manifolds which differ by a relation which we call “fiber slipping”. Finally, this deficiency is anticipated and corrected by a local modification of the tensor associated to each JSJ piece.

Let $m \geq 2$ be an integer and let V_m be the \mathbb{C} -vector space spanned by the finite set $H_1(T^2; \mathbb{Z}_m) \times \{1, 2, \dots, m\}$. Let X be a JSJ piece, which in our context is either finite-volume hyperbolic or Seifert fibered. We will associate to X a carefully chosen but suitably generic tensor T_X with n covariant indices, where n is the number of cusps of X . Reversing the orientation of X conjugates the entries of T_X . That is, $T_{\overline{X}} = \overline{T}_X$. Let $h : X \rightarrow X$ be a homeomorphism (orientation preserving or reversing). The induced action of h on T_X will factor through the action of h on $H_1(\partial X; \mathbb{Z}_m) \times \{1, 2, \dots, m\}^c$, where c is the number of cusps. The action on the set permutes factors according to the action on the end $\mathrm{end}(X)$, which, of course, embeds in the symmetric group S_c . Thus, the mapping class group of X acts on T_X but factored through the finite group $\mathrm{GL}(2, \mathbb{Z}_m) \times S_c$ for each m . If $l \in H_1(\partial X, \mathbb{Z}_m) \times \{1, 2, \dots, m\}$ is a multi-index, then symmetry demands:

$$(6.3.1) \quad h(T_l) = \begin{cases} T_{h_*(l)} & \text{if } h \text{ preserves orientation,} \\ \overline{T}_{h_*(l)} & \text{if } h \text{ reverses orientation.} \end{cases}$$

For example, consider X with three cusps reflected by h as shown in Figure 9. We use h to identify the groups at the upper and lower cusps and the “bar” on k to denote the induced involution on the middle level group to obtain:

$$(6.3.2) \quad h(T_{ijk}) = \overline{T}_{ji\bar{k}}.$$

In what follows, a *reflection symmetry* r (or just *reflection* for short) refers to an orientation-reversing involution whose fixed set $\Sigma \subset X$ is 2-dimensional. Note that when X is hyperbolic, r may be taken to be an isometry. For each SF X and with respect to any reflection symmetry r , the tensors T_X will be constructed to have the positivity property P below. Positivity is an open (convex) condition, so we may first concentrate on producing positive Hermitian tensors for each JSJ piece X and then perturb the construction to achieve genericity.

We fix m and abbreviate V_m by V . Note that V has a canonical basis, namely the set $H_1(T^2; \mathbb{Z}_m) \times \{1, \dots, m\}$. We let $V^+ \subset V$ denote the cone of nonnegative (real)

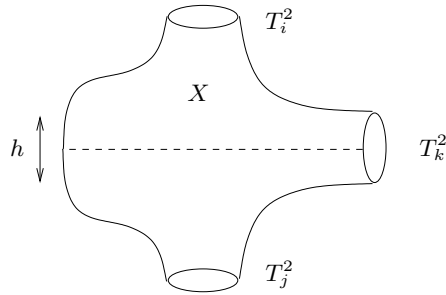


FIGURE 9

linear combinations of these basis elements. Suppose $r : X \rightarrow X$ is a reflection. Then r partitions the cusps of X into three groups: the *horizontal* cusps which the fixed set Σ meets, and two subsets, the *upper* and *lower* cusps which are disjoint from Σ , and interchanged by r . (We arbitrarily designate one side of Σ as “upper” and the other as “lower”.) Now suppose X has n cusps: k upper, k lower, and $n - 2k$ horizontal (all with respect to r). If supplied with vectors $v_{2k+1}, \dots, v_n \in V^+$ at the horizontal cusps, T_X becomes an operator

$$T_X(v_{2k+1}, \dots, v_n) : \bigotimes_k V \rightarrow \bigotimes_k V$$

from the vector space associated to the lower cusps, to the vector space associated to the upper cusps. Furthermore, the homeomorphism r induces a natural *complex-linear* isomorphism between these two vector spaces, so it makes sense (using this identification) to ask if this operator is positive, for all reflections r and all positive $\{v_i\}$ inserted at the horizontal cusps of r .

We explain the overall strategy. The gluings AB , AA , and BB along S are analogous to the graph case of Theorem 6.9, where vertices lie on the gluing set. (Both hyperbolic and SF JSJ pieces may be divided in two by S .) So, we will use some kind of $\langle A^- | H | B^- \rangle$ format to calculate c_t^m and establish a topological Cauchy-Schwarz inequality from a (weak) positivity property of H . H is an operator associated to the “middle third” while A^- (B^-) is, by abusing notation slightly by dropping the Z , the vector (covector) associated to the upper and lower thirds. We define the *geometric middle third* (GMT) as the gluing along S of the JSJ pieces incident on S together with a copy of $T^2 \times I$ (analogous to α in Figure 8) for each JSJS cusp. Thus, the GMT is made by gluing boundary pieces and adding a $T^2 \times I$ for each horizontal cusp. Recall from Definition 6.1 that the JSJS cusps are the torus components of S which are also JSJ tori of the glued-up manifold AB .

There is a very nice conjecture (appendix C) that implies the existence of “positive” tensors for hyperbolic 3-manifolds. With that input, H could be assembled from tensors corresponding to JSJ pieces of the GMT. Since we lack a proof for this conjecture we will actually define H from a smaller *algebraic middle third* (AMT). The AMT, although primarily a device for constructing the operator H , does have a geometric realization: $\text{AMT} = S \cup (\text{SF pieces of GMT glued along } S)$. Thus, the difference $\text{GMT} \setminus \text{AMT} = \bigcup_{\text{hyperbolic } X \in \text{GMT}} (X \setminus S)$. All that remains in the AMT of the hyperbolic part of the GMT is S .

We will also need a notation for the complement of the AMT: $A \cup B \setminus \text{AMT} = A^\sim \cup B^\sim$, so $A^- \subset A^\sim$ and $B^- \subset B^\sim$. Thus, the formal scheme for calculating c_t^m is more accurately written:

$$(6.3.3) \quad c_t^m(AB) = \langle c_t^m(A^\sim) | H^m | c_t^m(B^\sim) \rangle.$$

Note the close analogy with Theorem 6.9 (see Figure 10 for clarification).

What makes a calculation of this form possible is that the tensors T_X we build for hyperbolic X are actually a partial contraction of a very large tensor network N_X (enjoying exactly the same symmetry group G as X) which crosses S orthogonally. This allows T_X , for X hyperbolic and in the GMT, to be represented partly in the bra and partly in the ket of (6.3.3).

The key will be to show that H^m is strictly (“block”) positive (to be defined shortly) on the space of bras (kets) that it will ultimately be paired with in (6.3.3). This is where the symmetry issues arise. The final step will be to recover the information lost by settling for tensors T_X , X SF which are slightly too symmetric and hence only “block” positive.

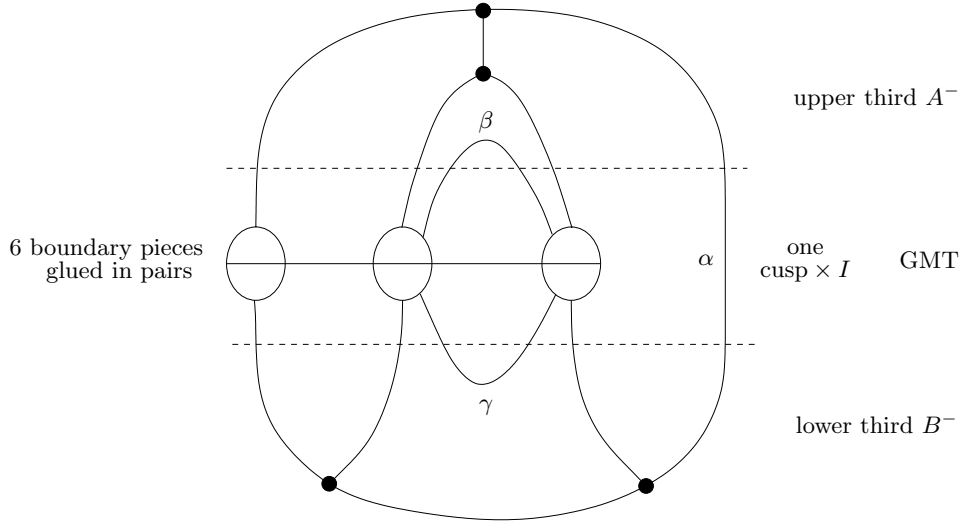


FIGURE 10. AB is decomposed into thirds $A^- \cup M \cup B^-$

We take a moment to clarify the correspondence to the graph case. JSJS cusps correspond to isolated edges in H with a cut point at either end, such as the arc α in Figure 8. Note that two cusps in the boundary of the middle third may cobound a single $T^2 \times I$ component in A^- or B^- ; these correspond to isolated edges, such as the arcs β and γ in Figure 8 in the graph case. Also as in the graph context, when we come to assign tensors to vertices, we adhere to the convention that doubly cut edges (of type α, β, γ) get identity tensors (up to index raising and lowering).

We now construct the tensors T_X for X SF. In contrast to the rigidity of hyperbolic manifolds, Seifert fibered JSJ pieces X with nontrivial boundary have infinite mapping class groups (whose elements we also refer to as *symmetries*). The restriction of the group of symmetries to the boundary of a JSJ piece is the semidirect product of two groups. The first, \tilde{P} , is an extension of \mathbb{Z}_2 by the full permutation

group P of the boundary components of the quotient orbifold Q :

$$1 \rightarrow P \rightarrow \tilde{P} \rightarrow \mathbb{Z}_2 \rightarrow 1.$$

The image in \mathbb{Z}_2 measures the action on orientation.

The second group is generated by *arcs of Dehn twist* (hereafter abbreviated to ADT) restricted to the boundary. An ADT connects a positive Dehn twist on a boundary component to a negative Dehn twist on a (possibly) distinct boundary component. The picture to have in mind is to start with a proper embedded arc in the quotient orbifold Q . Sitting above this arc is a proper annulus, which cobounds circles in two (possibly) distinct boundary components. Cut along this annulus and reglue after twisting once around each circle fiber in the annulus.

Reparameterizing a component X (e.g. by an ADT) changes the way in which X is glued to its neighbors, and how it is glued to itself (if at all). It would be desirable to find tensors T_X for Seifert fibered pieces X which are positive on exactly the subspace of vectors invariant under all such parameterizations; however, we were not able to find such tensors directly. Instead, we first define “blocky” tensors which are completely insensitive to the fiber summand of each $H_1(\partial_i X; \mathbb{Z}_m)$ index (effectively ignoring the troublesome ADTs). Second, we determine that the resulting tensors, while not leading to a strictly diagonally dominant complexity function on homeomorphism classes of manifolds, nevertheless would yield a diagonally dominant complexity function on manifolds up to the weaker relation of “fiber slip homeomorphism” (fish). Finally, we enhance our tensors $\hat{T}_X \rightsquigarrow T_X$ by letting them feel their immediate neighbors in the JSJ decomposition to resolve the fishy ambiguity. Morally, fiber slip homeomorphism is the equivalence relation obtained by failing to keep track of a certain relative Euler class (living in certain \mathbb{Z} torsors). The remedy is to find canonical coordinates on these \mathbb{Z} torsors, and favor relative Euler classes which arise by doubling.

The first step is to define “block positive” for Seifert fibered pieces.

For the moment, fix a fibered structure on the JSJ piece X . The choice of fiber structure picks out a canonical \mathbb{Z}_m summand in $H_1(\partial_i X; \mathbb{Z}_m)$. We let π be the homomorphism which quotients out this summand:

$$\begin{aligned} \pi : H_1(\partial_i X; \mathbb{Z}_m) \times \{1, 2, \dots, m\} &\rightarrow H_1(\partial_i X; \mathbb{Z}_m) / \langle [\text{fiber}] \rangle \times \{1, 2, \dots, m\} \\ &\cong \mathbb{Z}_m \times \{1, 2, \dots, m\}. \end{aligned}$$

The map π induces a map from $V_m \cong \mathbb{C}^{m^3}$ to the complex vector space, isomorphic to \mathbb{C}^{m^2} , spanned by $H_1(\partial_i X; \mathbb{Z}_m) / \langle [\text{fiber}] \rangle \times \{1, 2, \dots, m\}$. The “blocky” tensors we will construct have entries which only depend on the image of each coordinate in \mathbb{C}^{m^2} under this projection. Fix a reflection r , and partition the cusps of X into upper, lower and horizontal cusps as before, where the horizontal cusps are those meeting the fixed set Σ of r . We want \hat{T}_X to be some totally symmetric n -tensor depending only on $\pi(\text{index})$ for each index. The operator $\mathbf{O} := \hat{T}_X(v_{2k+1}, \dots, v_n)$ maps between associated symmetric powers of $S_{\text{lower}}, S_{\text{upper}}$ of V and descends to \mathbf{O}^π mapping between symmetric powers S^π of $\pi(V) := \mathbb{C}[\pi(H_1(\partial_i X; \mathbb{Z}_m))] \times \{1, 2, \dots, m\}$.

We now define block positivity as shown in Figure 11.

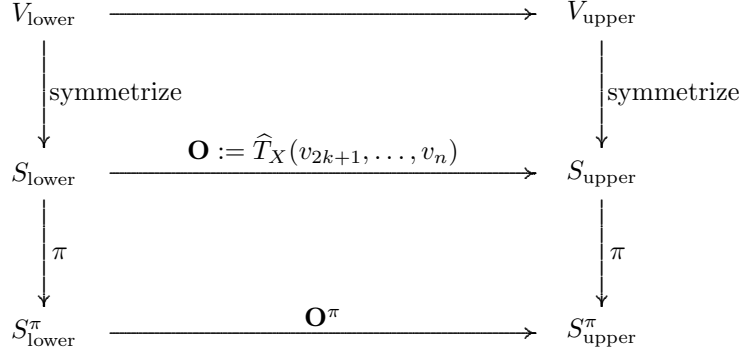


FIGURE 11

Definition 6.11. A tensor \widehat{T}_X assigned to a Seifert fibered JSJ piece X satisfies *property P* if, for all reflections r and positive inputs $\{v_{2k+1}, \dots, v_n\}$ at the horizontal cusps, the operator $\widehat{T}_X(v_{2k+1}, \dots, v_n)$ is *block positive*, i.e. if \mathbf{O}^π as above is positive.

Actually, property P will only be important for JSJ pieces which lie the middle third M of AB , AA , or BB , so it is harmless to restrict attention to those Seifert fibered X which arise by gluing two boundary pieces along (part of) S . Thus the pieces in question must be sufficiently large, and (in order to participate in forming H), must have nonempty boundary. According to the analysis of exceptional cases in §4 (Theorem 4.13, Theorem 4.17, and Lemma 4.22), the only sufficiently large JSJ piece X with boundary that might arise and that lacks a unique Seifert fibered structure is $S^1 \tilde{\times} S^1 \tilde{\times} I \cong M(+0, 1; 1/2, 1/2)$. But although this manifold admits a second nonisomorphic Seifert fibered structure, this latter structure is excluded by Convention 4.15. Thus without loss of generality we are justified in assuming a unique Seifert fibered structure on X for the purpose of defining P . This is crucial, as it justifies our use (above) of a fixed fibered structure on X , and thus makes the quotient operator \mathbf{O}^π well defined as an operator from S_{lower}^π to S_{upper}^π .

For X Seifert fibered, we must construct a generic \widehat{T}_X satisfying property P_S :

$$\widehat{T}_X \in \mathbb{C}[H_1(\partial X; \mathbb{Z}_m)] \cong \bigotimes_{i=1}^n \mathbb{C}[H_1(\partial_i X; \mathbb{Z}_m)] = \bigotimes_{i=1}^n V_m.$$

Recycling a trick from the proof of Theorem 6.9, block positivity will be ensured by choosing the entries of \widehat{T}_X according to a function which extravagantly weights coincidences. The \widehat{T}_X entry should be (approximately) given by composing the string

$$(6.3.4) \quad (\#n\text{-tuples}, \#(n-1)\text{-tuples}, \dots, \#\text{pairs})$$

into the reals \mathbb{R} by a function such as (6.2.2). Coincidence tuples are diagonally dominant functions on multi-indices under concatenation, and so are such compositions into \mathbb{R} . Here we mean k -tuples of indices which become the same in the quotient group H_1/fiber . We also insist that the entries of \widehat{T}_X are algebraically generic subject to the relation $\widehat{T}_{\overline{X}} = \widetilde{\widehat{T}}_X$ and the insensitivity to fiber coordinates.

To repeat, the entries $(\widehat{T}_X)_{i_1, \dots, i_n}$ of \widehat{T}_X , where the i_j denote elements of the various $H_1(\text{cusp}; \mathbb{Z}_m)$, depend only on the images of these i_j under the projection π . It follows that the operator \mathbf{O} , obtained by contracting \widehat{T}_X with positive vectors $v_{2k+1}, \dots, v_n \in V^+$ associated to horizontal cusps, is not positive but merely non-negative. It is the induced operator \mathbf{O}^π which is positive. The operator \mathbf{O} itself has null directions spanned by vectors of the form:

$$(6.3.5) \quad e_{b_1, f_1} \otimes \cdots \otimes e_{b_{i-1}, f_{i-1}} \otimes (e_{b_i, f_i} - e_{b_i, f'_i}) \otimes \cdots \otimes e_{b_{i+1}, f_{i+1}} \otimes e_{b_k, f_k},$$

where the pair f_i, f'_i are distinct values of the fiber coordinates in some basis for $H_1(\text{cusp}; \mathbb{Z}_m)$.

We have proven:

Lemma 6.12. *Any Seifert fibered X with unique Seifert fibering admits a tensor T_X satisfying property P .*

Next we construct the tensor T_X when X is hyperbolic. It is important that T_X be constructed from a (partial) contraction of a tensor network N defined on a fundamental domain W of X with respect to its full symmetry group G , and transported around X by the group. This way, N will enjoy the symmetries of X . We must take a little further care to ensure that N has no additional symmetry as a labeled graph. We will use our usual V_m to label all edges.

The construction is not difficult, but involves a number of steps which are described in the next several pages. Recall that the goal of this paper is to establish positivity of the universal $(2 + 1)$ -dimensional TQFT. If one knew a particular unitary TQFT that was sufficiently rich and easy to analyze, it might be possible simply to check positivity there. A genuine TQFT has the *factorization property*: whenever X can be decomposed along a surface into pieces Y and Z one has a factorization $T_X = T_Y T_Z$, a compact notation for the appropriate tensorial contraction. However, one can construct weaker substitutes for a full TQFT which have the factorization property only for certain prescribed decompositions. In a sense, we are defining a “weak TQFT,” in the style of Penrose, for hyperbolic manifolds. But it is only factorizable with respect to decompositions along the fixed sets of reflections, and not with respect to arbitrary topological decompositions.

Let G be the group of isometries of X . Since X is a finite-volume hyperbolic manifold, G is finite. Let $W \subset X$ be a piecewise smooth fundamental domain for the action of G on X . The fundamental domain W may contain cusps in its interior and also *fractional cusps*, meaning a cusp C which is “paved” by k copies of W . We think of each copy of W as containing a $1/h$ fraction of C , where h is the order of the symmetry subgroup $H \subset G$ stabilizing the cusp in question. See Figure 12.

To get the symmetries of T_X right, we initially build a tensor T_W for W (see Figure 13) with indices or “legs” exiting cusps, cusp fractions, and codimension one faces of W . T_X is assembled from $|G|$ copies of T_W (or \overline{T}_W for orientation-reversing elements of G) and additional τ -tensors described below.

We build a tensor T_W for W as follows: Begin with a finite set v of vertices geometrically positioned in the interior of W so that every point of $x \in W$ either has injectivity radius $(x) \leq \epsilon$ or $\text{dist}(x, v) < \epsilon$. Consider the graph \dot{N} consisting of the dual 1-cells to the faces of the Voronoi decomposition associated to v . \dot{N} is a fine triangulation of $\text{thick}(W)$ with (many) edges orthogonal to each face of W . Any vertex $p \in v$ whose Voronoi cell contains a cusp or cusp fraction is supplied an additional edge exiting that cusp or cusp fraction. We turn \dot{N} into a tensor

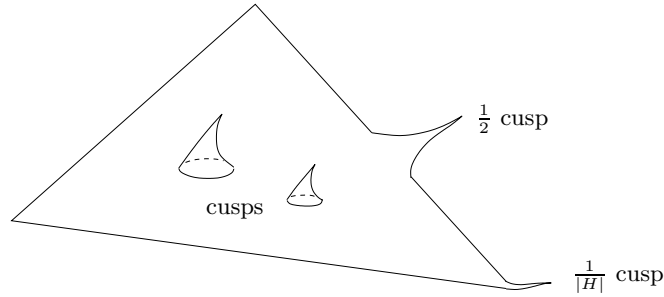


FIGURE 12. Schematic fundamental domain W , with cusps and fractional cusps.

network, now denoted N by choosing the vector space V for each edge and a distinct generic tensor T_p for each vertex $p \in v$. We take all the entries of all the T_p to be algebraically independent.

For ϵ sufficiently small, the labeled network N recovers the combinatorics of W with each face(W) marked by numbered legs from unique tensors T_p . Now let G label $|G|$ copies of N . Actually if $q \in G$ reverses orientation, we conjugate the entries of each T_p to get \overline{N} . These copies of N and \overline{N} may be assembled into a larger tensor network GN . There is now a tensor \tilde{T}_X associated to X by contracting the internal edges (those not exiting cusps or cusp fractions) of GN . It is not difficult to see that \tilde{T}_X has precisely the same symmetry group (for small ϵ) as X . In fact, such a labeled network encodes the homeomorphism type of the manifold (or orbifold) X on which it lies. Also very important to our approach is that given a reflection symmetry $r \in G$, \tilde{T}_X may be written as a contraction $\tilde{T}_X^{top} \tilde{T}_X^{bot}$ along the legs orthogonal to the fixed set of r . This “decomposability” of T_X allows us to write part of \tilde{T}_X in the bra and part in the ket when pairing to the AMT. Consult Figure 13 for a visualization of T_W as a contraction of N and Figure 14 for a visualization of $|G| = 24$ copies of N assembled into GN .

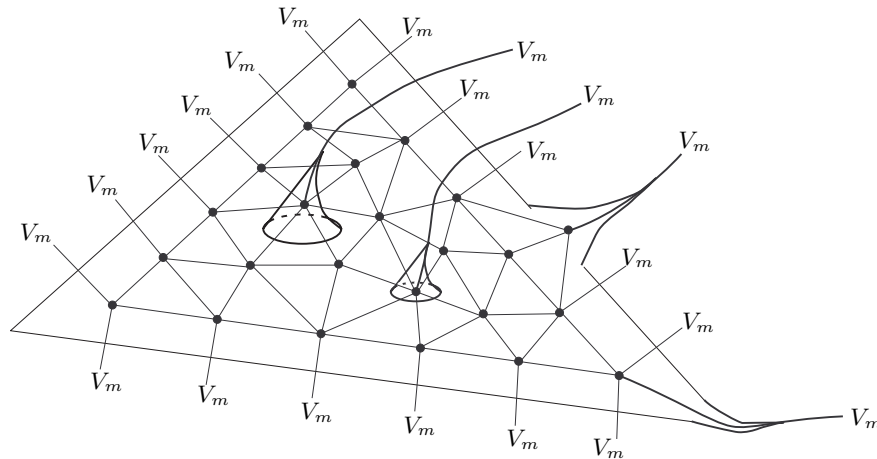


FIGURE 13. The tensor network N : one leg exits every cusp and cusp fraction; several legs exit every face.

To repeat, \tilde{T}_X is assembled from $|G|$ copies of T_W and its complex conjugate \overline{T}_W by contraction along the face-crossing indices. Figure 14 illustrates how 24 copies of W are assembled to make X , in a special case where G is equal to the symmetric group S_4 . l legs cross each codimension-1 face and additional indices, not indicated in the figure, exit each cusp and cusp fraction.

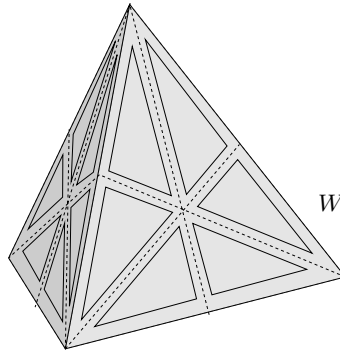


FIGURE 14. 24 copies of W are glued together by identifying 72 faces in pairs. l copies of V_m are contracted along each such face.

In more detail, label the copies of W making up X as $\{gW|g \in G\}$. G acts on the indices in the tensor associated to cusps or fractional cusps in the obvious way. That is, $H_1(C; \mathbb{Z}_m)$ serves separately as a basis for *each* fraction of the cusp C , and g acts as g_* . For instance, if we fix a basis $\{e_1, \dots, e_{m^2}\} \times \{1, 2, \dots, m\}$ for cusps and fractional cusps in the copy labeled $(\text{Id})W$, then the corresponding indices in the copy labeled gW are the basis elements $\{g_*e_1, \dots, g_*e_{m^2}\} \times \{1, 2, \dots, m\}$. (Note the inaction on the second factor.) On each copy gW we place a copy of either T_W or \overline{T}_W depending on whether g preserves or reverses orientation.

Our preliminary tensor \tilde{T}_X associated to X is the contraction of these $|G|$ copies of T_W (or \overline{T}_W) along the l legs spanning the codimension-1 faces of $\{gW\}$. For every reflection symmetry r of X with codimension-one fixed set Σ , the tensor \tilde{T}_X contracted with positive vector inputs for each horizontal (with respect to r) cusp or cusp fraction can be written in the form

$$\tilde{T}_X(v_{2k+1}, \dots, v_n) = \tilde{T}_Y^\dagger \tilde{T}_Y,$$

where \tilde{T}_Y is the operator corresponding to the “lower half” of X (again with respect to r), and therefore is nonnegative Hermitian.

The tensor \tilde{T}_X is not yet the tensor we are looking for because T_W (and similarly \overline{T}_W) has one index (ranging in V_m) for each *fractional cusp* in W , whereas we want our tensor T_X to have one index for each *cusp* of X . A cusp C stabilized by $H < G$ will contribute $|H|$ indices to \tilde{T}_X whereas *one* is the desired number. To create a tensor with the correct index set, we need to contract \tilde{T}_X with suitably chosen tensors $\tau_{H(C)}$ for each cusp C , where $H < G$ is the stabilizer.

We define $\tau_{H(C)}$ to be an element of $\text{Hom}(V_m, V_m^{\otimes |H|})$ satisfying the *reflection identity property* (RIP), explained below, and so that the tensor T_X obtained by contracting \tilde{T}_X with a copy of $\tau_{H(C)}$ for each cusp C satisfies (see Figure 15):

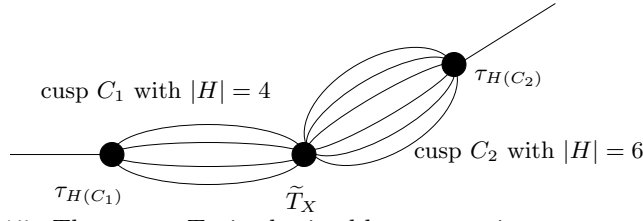


FIGURE 15. The tensor T_X is obtained by contracting $\tau_{H(C_1)}$ and $\tau_{H(C_2)}$ with \tilde{T}_X .

- (1) $\tau_{H(C)}$ is natural, so that T_X may be constructed from the homeomorphism type of X alone, without resort to any additional data such as a base point on the cusps.
- (2) $\tau_{H(C)}$ makes T_X G -invariant. (Recall that G acts by permuting cusps internally on V .) Succinctly one may say that G acts via the action on $H_1(\partial X; \mathbb{Z}_m)$. Actually, this is a corollary of item (1) above.

Conditions (1) (and (2)) hold provided $\tau_{H(C)}$ satisfies (a) invariance under the action of H on the $V_m^{\otimes |H|}$ factor and (b) naturality, in the sense that $\tau_{H(C)}$ is defined using only the action of H on C without additional choices such as a base point on C .

All the V factors in $\text{Hom}(V_m, V_m^{\otimes |H|})$ are canonically identified. If we fix a base point p in C marking one fundamental domain for the H action, then the formula

$$\tau_{H(C)}^p(e_i) = \bigotimes_{h \in H} h_*(e_i), \quad e_i = (f_j, s_k) \in H_1(C; \mathbb{Z}_m) \times \{1, 2, \dots, m\}$$

clearly satisfies condition (a). To achieve (b), set

$$\tau_{H(C)}(e_i) = \sum_{h \in H} \tau_{H(C)}^{h(p)}(e_i).$$

Note that “inputting” vectors into tensors is a special case of tensor contraction; we regard elements of V as tensors with one index. This motivates the notation in Figure 16: the small dots are 1-tensors. As illustrated in Figure 16, the vector inputs induced by $e_{i(C)}$ are r -symmetric.

Specifically, if a basis element e_i is inserted into the first index of τ_H , then the resulting $|H|$ -tensor is converted to an operator \mathcal{O} by segregating the remaining $|H|$ index slots into r -upper and r -lower groups with respect to a reflection symmetry r . RIP states that \mathcal{O} is the identity operator from the appropriate “upper space” $\mathcal{U} \cong \mathbb{C}^{h/2}$ to the appropriate “lower space” $\mathcal{L} \cong \mathbb{C}^{h/2}$, $h = |H|$, with respect to prescribed bases. Fixing a base point fundamental domain W on the r -upper side of the cusp C , let $W = W_{id}, W_{g_2}, \dots, W_{g_{h/2}}$ be the domains on the upper side. Then

$$\text{basis}(\mathcal{U}) = \{h_*e_i \otimes h_*g_{2*}(e_i) \otimes \dots \otimes h_*g_{h/2*}(e_i) | h \in H\}$$

and similarly

$$\text{basis}(\mathcal{L}) = \{rh_*e_i \otimes rh_*g_{2*}(e_i) \otimes \dots \otimes rh_*g_{h/2*}(e_i) | h \in H\}$$

with respect to domains $W_r, W_{rg_2}, \dots, W_{rg_{h/2}}$ on the lower side. Figure 16 illustrates this symmetry.

Our final tensor T_X for hyperbolic JSJ pieces is obtained from \tilde{T}_X by contracting a copy of τ_H at each cusp C (with symmetry group H).

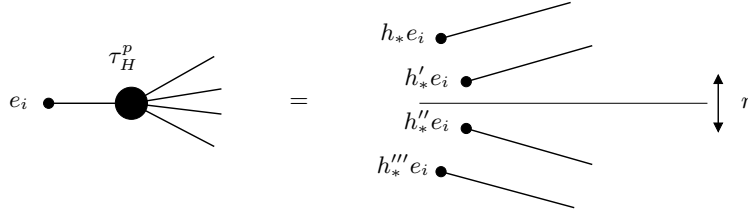


FIGURE 16. The definition of τ_H^p . Note that $rh = h'''$ and $rh' = h''$.

Having constructed T_X for X SF and τ , these pieces can now be assembled (together with identity operators on the legs of T_W (and \bar{T}_W) orthogonal to $fix(r)$ and also on vertical cusps) into the AMT operator H^m . Figure 17 organizes the definition of H into a picture.

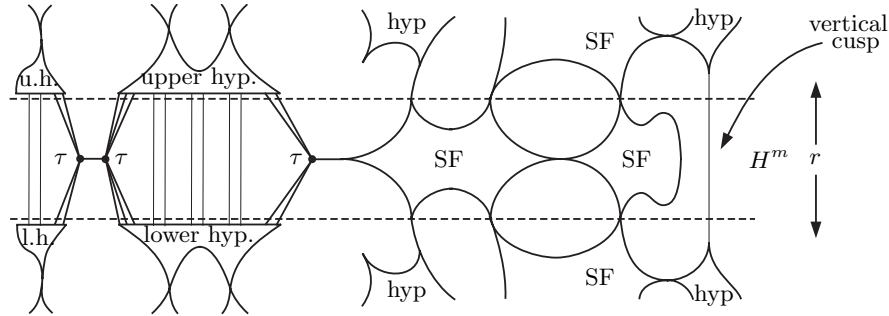


FIGURE 17. The AMT operator H^m .

In more detail, H^m is obtained by contracting geometrically adjacent tensors at the AMT. These include T_X for SF X , a τ_H for each horizontal (hyperbolic) cusp, and then (without any actual contraction) all the legs (identity operators on copies of V) which run orthogonally to $S \cup$ (hyperbolic JSJ piece) and id_V for vertical cusps. The τ_H will have its first index (slot) contracted to the first index (slot) of another $\tau_{H'}$ when two hyperbolic pieces are glued at S . Where a hyperbolic piece is glued to an SF X , τ_H will have its first index (slot) contracted with the “cusp” of the SF X .

We now identify the subspace where H^m acts as a positive operator. For each mid-level SF X , let $S_{SF X}^\pi$ denote the reduced symmetric product of the upper “cusps” of X . For each mid-level hyperbolic X , let $\mathbb{Z}_{X,r}$ denote the tensor product of the (space V associated to each of the) legs of GN orthogonal to $fix(r) \cap X = S \cap X$. For each horizontal hyperbolic cusp C , let \mathcal{U}_C be the “upper space” defined above. Let:

$$Y_{up} = \left(\bigotimes_{\text{mid level SF } X} S_{SF X}^\pi \right) \otimes \left(\bigotimes_{\text{mid level hyperbolic } X} \mathbb{Z}_{X,r} \right) \otimes \left(\bigotimes_{\text{horizontal cusps } C} \mathcal{U}_C \right) \otimes \left(\bigotimes_{\text{vertical cusps } C} V_C \right).$$

Make the analogous definition for Y_{down} , replacing \mathcal{U}_C with \mathcal{L}_C . Perfect r -symmetry of the GMT is obvious for AA and BB and will be established near the end of the

proof for AB (unless $c(AB) < \max(c(AA), c(BB))$). Assuming such symmetry of the GMT, the reflection r identifies Y_{up} and Y_{down} , so simply call them both Y .

Lemma 6.13. *If the GMT is r -symmetric, then H^m is positive on Y for all $m = 2, 3, \dots$*

Proof. The proof is extremely simple. H^m is a sum over contraction variables e_i of tensor products of operators which are either strictly positive (on $S_{SF\ X}^\pi$) or the identity (after our identification of Y_{up} and Y_{down}). Positivity is closed under both tensor product and convex combination. \square

Note. If $S_{SF\ X}^\pi$ is replaced by the ordinary symmetric power $S_{SF\ X}$ we can only conclude *block* positivity above. This point is addressed in the final paragraph of section 6.

We are almost in a position to define the complexity function c_t . Suitable tensors T_X for JSJ pieces X have been constructed when X is hyperbolic. When X is Seifert fibered, a final tweak is required because of block positivity of these tensors. This is the subject of the next subsection.

6.4. Better tensors for Seifert fibered JSJ pieces. As we noted earlier, the tensors \widehat{T}_X for Seifert fibered JSJ pieces constructed above possess too much symmetry; they are insensitive to arcs of Dehn twists. In this subsection we repair this defect at the expense of making the tensor T_X for a Seifert fibered JSJ piece slightly less local. T_X will depend on the way in which X meets its neighboring JSJ pieces, not merely on the homeomorphism type of X .

First we define the notion of *fiber slip homeomorphism*, or *fish* for short. A Seifert fibered JSJ piece X with nonempty boundary will (by Theorem 4.13) have a unique Seifert fibered structure unless $X \cong S^1 \widetilde{\times} S^1 \widetilde{\times} I$, in which case the quotient orbifold is either a disk with two orbifold points of order 2, or a Möbius band. To avoid ambiguity, by convention we always give $S^1 \widetilde{\times} S^1 \widetilde{\times} I$ the Seifert fibered structure for which the base orbifold is a disk with two order 2 points.

Definition 6.14. For a Seifert fibered JSJ piece $X \subset A$ (respectively $X \subset M$), a *fiber slip* is a regluing of A (resp. M) by a fiber-preserving homeomorphism of one or more cusps of X . Note that if two cusps of X glue to each other so as to mismatch a fiber, then there is a $\mathbb{Z} \oplus \mathbb{Z}$ -family of fiber slips at the common torus.

Definition 6.15. (A, S) and (B, S) (respectively M and M') are *fish equivalent* if after regluing the Seifert fibered JSJ pieces of (A, S) (resp. M) by fiber slips, the result is homeomorphic to (B, S) rel S (resp. homeomorphic to M').

To resolve the issue of fiber slips, we modify c_t , the partition function induced by our preliminary choice of tensors for JSJ pieces, into a more powerful tensor complexity. The tensors \widehat{T}_X previously assigned to a Seifert fibered JSJ piece X will now be perturbed, and in some cases rescaled, in a way which depends on the immediate neighbors of X in the JSJ decomposition. The perturbation T_X will depend on the homeomorphism types of the neighboring (X', C') , where X' is a JSJ piece glued along C' to (X, C) and also on the way in which C is glued to C' . Note that X' may equal X .

Consider a closed irreducible 3-manifold P and its JSJ decomposition. Corresponding to the combinatorics of the decomposition is a gluing graph \mathcal{G} whose

vertices are labeled by diffeomorphism types of JSJ pieces and whose edges correspond to gluings of cusps.

Lemma 6.16. *For each Seifert fibered piece X_i of the JSJ decomposition, let Q_i denote the underlying orbifold. A fiber slip homeomorphism of P which covers the identity on each Q_i is determined by a single integer t_i for each X_i .*

Proof. Fiber slips are of infinite order and hence cannot extend as symmetries of an adjacent hyperbolic piece. By the definition of the JSJ decomposition, when neighboring (or self-neighboring) Seifert fibered JSJ pieces meet along a cusp C , their fibers do not match in C . As a result, a fiber slip on one side will not extend to *any* diffeomorphism on the other. As a consequence, a fiber slip is determined by what it does to each Seifert fibered JSJ piece relative to its cusps. After reparameterizing a Seifert fibered piece X_i by arcs of Dehn twist, the effect of the slip may be localized to a single cusp, where it differs from the identity by some power of a Dehn twist along a fiber. Note that only an orientation of X_i , not Q_i , is required to identify the sign of t_i . \square

Remark 6.17. If we think of a Seifert fibered space as a circle bundle over an orbifold, a choice of trivialization of this bundle over each boundary component determines a relative Euler class, which can be paired with the fundamental class of the base orbifold to get a number. Two Seifert fibered spaces which differ by a fiber slip differ (after choosing trivializations on their boundaries) only in the value of this number.

It follows that the set of ways of gluing X_i to its neighbors which are in a fixed fish equivalence class is an oriented \mathbb{Z} -torsor, which we denote τ_i . It turns out that it is (almost) possible to choose a canonical basepoint for this torsor. Either one element can be unambiguously labeled 0, or else two consecutive elements can be labeled as $\{-\frac{1}{2}, +\frac{1}{2}\}$.

Let $\{C\}$ denote the set of cusps of X , a Seifert fibered JSJ piece of the irreducible 3-manifold P , and assume that X has a fixed Seifert fibered structure. We need to come up with a set $\{\text{bases}\}$ of canonical base curves, one for each C which pairs nontrivially with the fiber in C . In fact, the “set” $\{\text{bases}\}$ will consist, not of a single set, but of an orbit of sets under the action of the mapping class group of X . In particular, an ADT acts on the set of choices $\{\text{bases}\}$ of canonical base curves. The elements in $\{\text{bases}\}$ are determined homologically. Assume X has q boundary components. Let K be the kernel of the map $H_1(\partial X; \mathbb{Q}) \rightarrow H_1(X; \mathbb{Q})$ induced by inclusion. Then K has dimension q . Even if the base and fiber of X are not separately orientable, we can, and do, consistently orient the fibers on $\partial X = \bigcup\{C\}$. For each $i \leq q$, let f_i denote the class in $H_1(\partial X; \mathbb{Q})$ of the oriented fiber of the i th boundary component of X . Each difference $f_i - f_{i+1}$ is in K , and together these differences span a $(q - 1)$ -dimensional subspace of K . Let $b \in H_2(X, \partial X; \mathbb{Q})$ be any class such that ∂b , together with the $f_i - f_{i+1}$, forms a basis for K , and scale b to be primitive in $H_2(X, \partial X; \mathbb{Z})$. $\partial_i b$ may be divisible in $H_1(\partial_i X; \mathbb{Z})$, so define $b_i \in H_1(\partial_i X; \mathbb{Z})$ to be a primitive fraction of $\partial_i b$. Take $\{b_i\}$ to be the desired set of basic curves. For each cusp C_i , $b_i \cap f_i = r_i \neq 0$. Replacing b_i by $-b_i$ changes r_i to $-r_i$; adding a multiple of the preimage of some $f_j - f_{j+1}$ to b obviously leaves r_i fixed. Once we fix an orientation on f and ∂X , we can resolve the ambiguity in the signs of the r_i ’s by possibly changing the sign of b_i so that each r_i is positive. This makes each b_i well defined.

The set of isotopy classes of simple essential curves in C whose intersection number with the oriented fiber f is a fixed $r > 0$ is a \mathbb{Z} -torsor τ_C . An orientation-preserving homeomorphism α of X either preserves or reverses the class of the fiber f . Since the preceding construction is homological, if $\alpha(C_i) = C_j$, then $\alpha(b_i) = \pm b_j$. Since α is orientation-preserving, either $\alpha(f) = f$ and $\alpha(b_i) = b_j$ or $\alpha(f) = -f$ and $\alpha(b_i) = -b_j$. If a cusp C is fixed by α , then $\alpha(b + kf) = \pm(b + kf)$ in the two cases, respectively. Hence the action of \mathbb{Z} on the torsor τ_C commutes with the action of the mapping class group of X .

We would like to identify a canonical basepoint (i.e., 0) in the torsor τ_C . We almost succeed; either we find a unique element, which we call 0, or we find two elements which differ by 1, which we call $-\frac{1}{2}$ and $\frac{1}{2}$ (consistent with the action of $+1$). We call this *marking the torsor* τ_C .

To mark τ_C , look at C' , the cusp to which it is glued in the JSJ decomposition. Here C' is a cusp of another JSJ piece X' , where $X = X'$ is possible. If C' lies in a Seifert fibered JSJ piece, then consider the fiber f' in C' and how it intersects with the fiber f of C under the gluing. By the definition of the JSJ decomposition, f and f' cannot be isotopic. Let b be a “base” class of C as constructed above. The function

$$g_{SF}(k) = |(b + kf) \cap f'|$$

admits a unique minimum for $k \in \mathbb{Q}$, and therefore achieves its minimal value for $k \in \mathbb{Z}$ on at most two values of k , which differ by ± 1 . We mark τ_C by $\{b + k_0 f \mid g_{SF}(k_0) \text{ is minimal}\}$.

If C' lies in a hyperbolic piece, consider

$$g_{hyp}(k) = \frac{\text{length}_{C'}^2(b + kf)}{\text{area}(C')},$$

where lengths and areas are computed with respect to some fixed Euclidean structure on C' in the similarity class induced by the hyperbolic structure on X' . Since $\text{length}^2/\text{area}$ is dimensionless, the result only depends on the similarity class. The quadratic form $g_{hyp}(k)$ is strictly convex for $k \in \mathbb{R}$ and therefore has either one or two minima $k_0 \in \mathbb{Z}$. Again, mark τ_C by $\{b + k_0 f \mid g_{hyp}(k_0) \text{ is minimal}\}$.

The previous discussion allows us to define an invariant of an oriented Seifert fibered JSJ piece X in an oriented, irreducible 3-manifold P , called the *gluing number*, and denoted $\text{gn}(P, X)$. The gluing number takes values in $\mathbb{Z}[\frac{1}{2}]$ and is computed as follows. The marking identifies each τ_C with a copy of \mathbb{Z} or $\frac{1}{2} + \mathbb{Z}$. Therefore the set of base classes in each τ_C can be thought of as a set of integers and half-integers, whose sum is $\text{gn}(P, X)$. Note that gn does not change under an orientation-preserving homeomorphism which changes f and b to $-f$ and $-b$. Moreover, adding an element $f_i - f_{i+1}$ to b adds 1 to one term in the sum and -1 to the other, and therefore does not change gn . Hence gn is well defined, independent of choices. Furthermore, a fiber slip alters gn by adding exactly the integer defined in Lemma 6.16.

We use the notation $\text{gn}(\tilde{X}, X)$ interchangeably with $\text{gn}(P, X)$, where \tilde{X} denotes the union of X with its immediate neighbors in the irreducible prime P . This notation is convenient, since it underscores the fact that gn depends on \tilde{X} (and not just X) but not necessarily on all of P .

Lemma 6.18. *If X admits an orientation-reversing homeomorphism which extends to \tilde{X} , then $\text{gn}(\tilde{X}, X) = 0$.*

Proof. We must show that gn is both chiral and an invariant. That is,

- (1) If $h : (\tilde{X}, X) \rightarrow (\tilde{Y}, Y)$ is an orientation-preserving homeomorphism, then $\text{gn}(\tilde{X}, X) = \text{gn}(\tilde{Y}, Y) \in \mathbb{Z}[\frac{1}{2}]$.
- (2) Reversing the orientation on (\tilde{X}, X) reverses the sign of $\text{gn}(\tilde{X}, X)$.

The lemma follows from (1) and (2). We remark that this is somewhat reminiscent of the classical observation that the signature of a $4k$ -dimensional manifold must vanish if it admits an orientation-reversing homeomorphism.

Property (1) is more or less proved already and is just a restatement of the fact that gn is well defined. Property (2) is more interesting. The marking of the $\mathbb{Z}[\frac{1}{2}]$ -torsor associated to a cusp C of X depends on whether the adjacent JSJ piece \bar{X}' is Seifert fibered or hyperbolic, and it depends on the unoriented class of f' in C' in the first case, and the geometry of C' in the second case. The key point is that in neither case do the functions g_{SF} and g_{hyp} depend on the *orientation* of X, C' . Reversing the orientation of X reverses the action of \mathbb{Z} on τ_C , but it does not change the basepoint(s), so it just reverses the sign of the invariant; see Figure 18. \square

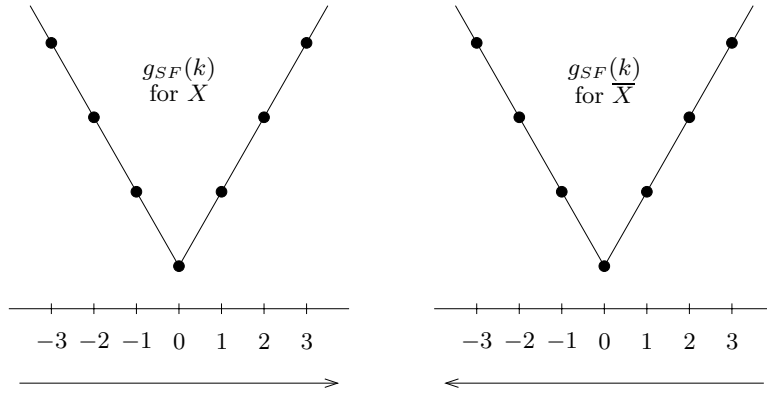


FIGURE 18

We are now almost ready to modify the tensors associated to Seifert fibered JSJ pieces. First we define the property of *domination* for Hermitian forms.

Definition 6.19. Let \mathbf{O}, \mathbf{O}' be two Hermitian forms on the same vector space. We say \mathbf{O} *dominates* \mathbf{O}' if

$$\langle x, \mathbf{O}x \rangle > \langle x, \mathbf{O}'x \rangle$$

for all nonzero x . Equivalently, \mathbf{O} dominates \mathbf{O}' if $\mathbf{O} - \mathbf{O}'$ is positive.

Now we can state the properties that we want our modified tensors to have.

Definition 6.20. For X Seifert fibered, a tensor $T_{X, \text{gn}}$ is *suitable* if it satisfies the following properties:

- (1) It has the same symmetry and positivity properties as T_X , and (compare equation (6.3.1)) if \bar{X} is the orientation reverse of X , then $T_{\bar{X}} = \bar{T}_X$.
- (2) The collection is otherwise generic, as was $\{\hat{T}_X\}$. In particular, $T_{X, n} \neq T_{X, m}$ for $n \neq m \in \mathbb{Z}[\frac{1}{2}]$.

- (3) When gn assumes the value $0 \in \mathbb{Z}$ or $\mathbb{Z}[\frac{1}{2}]$, then $T_{X,0} > T_{X,n}$, for $n \neq 0$, in the sense that for any reflection symmetry r and for any given index inputs for the horizontal cusps, the contracted operator $\mathbf{O}_{X,0}$ dominates the contracted operator $\mathbf{O}_{X,n}$ in the sense of Definition 6.19.

A simple way to satisfy condition (3) (as well as (1) and (2)), that is, to promote \widehat{T}_X to T_X , is to choose a function $\theta : \mathbb{Z}[\frac{1}{2}] \rightarrow \mathbb{R}^+$ with a single maximum at $0 \in \mathbb{Z}[\frac{1}{2}]$. First define $\tilde{T}_{X,n} = \theta(n)\widehat{T}_X$ and then obtain $T_{X,n}$ from $\tilde{T}_{X,n}$ by a fully generic (but tiny) perturbation of the entries subject only to the constraint that it should be block positive. That is, the entries (for each fixed n) depend only on the images of the indices under the quotient map π . So suitable tensors exist.

In the next subsection, we will insure that for each of the manifolds AB , AA and BB , the “middle third” consists of the same collection of JSJ pieces which have been glued together in ways which differ potentially only by fiber slip homeomorphisms. This combinatorial pattern of gluings (in each case) determines a Hermitian form, obtained by suitably contracting various tensors T_{X_i, n_i} for Seifert fibered X_i and T_{X_i} for hyperbolic X_i defined as above. The next lemma compares the operators obtained by different gluings. We suppress the parameter m as a superscript to the operator H .

Lemma 6.21. *The Hermitian form H obtained by contracting the various $T_{X_i,0}$, τ ’s, and identity operators in the AMT dominates all forms H' obtained by replacing at least one $T_{X_i,0}$ by T_{X_i, n_i} , $n_i \neq 0$, for some SF X_i .*

Proof. This follows formally from condition (3) above, provided that the factor $\theta(0)/\max_{j \neq 0} \theta(j)$ overwhelms the product of the perturbations. This may be ensured by choosing perturbations by genuine infinitesimals (in the sense of nonstandard analysis), or by extending the coefficient ring from \mathbb{C} to $\mathbb{C}[x]$, with lexicographical norm $(|a_0|, |a_1|, \dots)$ on $\sum a_i x^i \in \mathbb{C}[x]$, and taking perturbations only in the linear term (and fixing $\theta(0)/\max_{j \neq 0} \theta(j) > 1$). \square

The importance of “domination” comes from the following lemma:

Lemma 6.22. *If \mathbf{O} dominates \mathbf{O}' , then for all nonzero x, y there is an inequality*

$$\langle x, \mathbf{O}'y \rangle < \max(\langle x, \mathbf{O}x \rangle, \langle y, \mathbf{O}y \rangle).$$

Proof. By the definition of domination, there are inequalities $\langle x, \mathbf{O}x \rangle > \langle x, \mathbf{O}'x \rangle$ and $\langle y, \mathbf{O}y \rangle > \langle y, \mathbf{O}'y \rangle$, so

$$\max(\langle x, \mathbf{O}x \rangle, \langle y, \mathbf{O}y \rangle) > \max(\langle x, \mathbf{O}'x \rangle, \langle y, \mathbf{O}'y \rangle).$$

But the Cauchy-Schwarz inequality gives

$$\max(\langle x, \mathbf{O}'x \rangle, \langle y, \mathbf{O}'y \rangle) \geq \langle x, \mathbf{O}'y \rangle,$$

and we are done. \square

6.5. Assembling the assembly complexity. The assembly complexity is a lexicographic pair (c_r, c_t) , where:

- c_r is a simple symmetry based complexity, described below. Its job is to ensure that the middle thirds of AA , BB and AB are homeomorphic up to fiber slips.
- c_t is an infinite lexicographic tuple (c_t^m) , $m \geq 2$.

- c_t^m is a number obtained by contracting configurations of tensors of the kind described in §6.3 and §6.4. For each Seifert fibered JSJ piece X the tensor $T_{X,\text{gn}}$ depends on the way in which X is glued to its neighboring JSJ pieces, through the invariant gn . The purpose of this modification is to correct for the failure of \widehat{T}_X to detect fiber slips.

First we define c_r and prove an accompanying lemma. Let M be a closed irreducible 3-manifold and let T be a JSJ torus of M .

Definition 6.23. A JSJ torus T of a closed irreducible 3-manifold M is *reflective* if there is an orientation-reversing involution of T with two fixed circles, which extends to the JSJ pieces on either side of T .

Note that for the doubles AA and BB , each JSJ torus transverse to S (hereafter *vertical JSJ torus*) is reflective. The two fixed circles of the reflective involution interchanging the two sides are exactly the two components of $T \cap S$.

Definition 6.24. Let M be a closed irreducible 3-manifold. Define $c_r(M)$ to be the number of reflective JSJ tori in M .

Lemma 6.25 (c_r -Lemma Schema). *Let AB , AA and BB be closed, irreducible and connected, and suppose that their c_S and c_h complexities are equal. Then either $c_r(AB) < \max(c_r(AA), c_r(BB))$ or $c_r(AB) = c_r(AA) = c_r(BB)$ and the geometric middle thirds (GMTs) of AB , AA and BB are homeomorphic (rel boundary) up to a fiber slip homeomorphism.*

Proof. First we show that $c_r(AB) \leq \max(c_r(AA), c_r(BB))$. Define $c_r(A)$ to be the number of reflective JSJ tori in A (reflective annuli don't count). Define $c_r(B)$ similarly. For $M = AB, AA$ or BB , let $c_m(M)$ be the number of reflective tori in the middle third. Recall that these middle thirds are identical except for the gluing maps at the vertical JSJ tori, which might differ by Dehn twists parallel to S . We have

$$\begin{aligned} c_r(AB) &= c_r(A) + c_r(B) + c_m(AB), \\ c_r(AA) &= c_r(A) + c_r(A) + c_m(AA), \\ c_r(BB) &= c_r(B) + c_r(B) + c_m(BB). \end{aligned}$$

As noted above, $c_m(AA)$ and $c_m(BB)$ are both equal to the total number of vertical JSJ tori in the middle level. Thus $c_m(AB) \leq c_m(AA) = c_m(BB)$, and it follows that $c_r(AB) \leq \max(c_r(AA), c_r(BB))$.

The equality $c_r(AB) = \max(c_r(AA), c_r(BB))$ can hold only if both $c_r(A) = c_r(B)$ and $c_m(AB) = c_m(AA) = c_m(BB)$. The latter equality means that every vertical JSJ torus in the middle third of AB is reflective. We will show that this in turn implies that the middle third of AB differs from the middle third of AA (or BB) only by fiber slips.

Let T be a vertical JSJ torus in the middle third of AB , and denote the corresponding torus in AA also by T . On each side of T one finds one of the following three kinds of pieces:

- a hyperbolic JSJ piece,
- a Seifert fibered JSJ piece with fibers parallel to $S \cap T$ (horizontally Seifert fibered or HSF), or
- a Seifert fibered JSJ piece with fibers intersecting each component of $S \cap T$ once (vertically Seifert fibered or VSF).

Note that the existence of the global reflection in AA means that the fibers of a Seifert fibered JSJ piece are either parallel to $S \cap T$ or have intersection number one with each component. This property is preserved under Dehn twists parallel to S and therefore also holds in AB .

Suppose a vertical JSJ torus T in AB , AA or BB abuts two Seifert fibered pieces X, X' . Let f, f' be the class of the fibers in X, X' in the torus T . Since every vertical JSJ torus in each space is reflective, there is an involution ι of T which extends to X, X' . By uniqueness of fiber structures, the involution ι takes f to $\pm f$ and f' to $\pm f'$. Note that the pieces X, X' and whether they are HSF or VSF are the same in AB , AA and BB , though not necessarily the way they are glued. If X, X' are both VSF in AA , then the involution ι may be chosen to fix S . This implies that $f = \pm f'$, which means that X and X' can be glued along T to make a bigger JSJ piece, contrary to the defining properties of the JSJ decomposition. It follows that a vertical JSJ torus can bound a VSF piece on at most one side in AA , and therefore also in AB .

The gluing map at T gives a bijection between two copies of $\mathrm{GL}(2, \mathbb{Z})$, one from each of the two adjacent JSJ pieces. Inside each of these copies is the subset of reflections which extend over the full JSJ piece. The fact that T is reflective in AA means that, under the isomorphism of the two copies of $\mathrm{GL}(2, \mathbb{Z})$ induced by the gluing map for AA , the two sets of reflections have at least one element in common. We know that the gluing maps for AB along the middle third differ from those of AA by Dehn twists parallel to $T \cap S$. We can ask whether any of the homeomorphisms (other than the gluing map for AA) in this affine \mathbb{Z} coset give a nonempty intersection of the two sets of reflections. We will see below that if neither adjacent JSJ piece is HSF, then the answer is no, so in this case the only way for T to be reflective in AB is for the gluing map for AB to coincide with the gluing map for AA . On the other hand, if T is adjacent to an HSF JSJ piece, then all of the possible gluing maps at T are related by fiber slips, so the lemma will be proved.

Let T be a JSJ torus in the middle third, with adjacent JSJ pieces either hyperbolic or VSF. Fix coordinates so that

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \text{ denotes the canonical reflection of } AA$$

and

$$\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \text{ denotes a product of } n \text{ Dehn twists in a component of } T \cap S.$$

In these coordinates, the set of reflections of T which extend over an adjacent VSF piece must be a subset of

$$\mathcal{R}_v := \left\{ \begin{bmatrix} \pm 1 & 0 \\ 2m & \mp 1 \end{bmatrix}, m \in \mathbb{Z} \right\},$$

since such reflections must preserve the (vertical) Seifert fibers and be conjugate to the standard reflection by Dehn twisting. The set of reflections of T which extend over an adjacent hyperbolic piece must be a subset of

$$\mathcal{R}_h := \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \right\},$$

since the geometry of the cusp must be rectangular, possibly square (note that the last two matrices occur only in the case of a square cusp).

Let the gluing map at T in AB be the product of n Dehn twists parallel to $T \cap S$. Then we must have $P \in \mathcal{R}_h$ and $Q \in \mathcal{R}_h$ or $Q \in \mathcal{R}_v$ such that

$$\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} P \begin{bmatrix} 1 & -n \\ 0 & 1 \end{bmatrix} = Q.$$

It is straightforward to check that the only solution is $n = 0$. Thus the gluing maps at T for AA and AB coincide. As remarked above, if one side of T is HSF, any gluing map is a fiber slip, and the lemma is proved. \square

We now show that the complexity $c_p := (c_S, c_h, c_a)$, where $c_a = (c_r, c_t)$, satisfies the conclusion of Theorem 3.20. This is the final step in the proof of the topological Cauchy-Schwarz inequality in the guise of Theorem 3.27, and therefore also of the proof of Theorem A.

We assume without loss of generality that the terms c_S, c_h, c_r are equal for AA, AB and BB and immediately focus on $c_t = (c_t^m)$.

Contracting, as appropriate, copies of the tensors $T_{X,gn}$ for X SF, T_W for $W \subset X_{\text{hyp}}$ a fundamental domain, the various $\tau_{H(C)}$ tensors associated to horizontal hyperbolic cusps, and the identity operator on V for vertical cusps, we obtain the three pieces of $c_t^m(AB)$ (respectively $c_t^m(AA)$ and $c_t^m(BB)$). Writing the formula, just for AB , we have

$$c_t^m(AB) = \langle c_t^m(A^\sim) | H_{AB}^m | c_t^m(B^\sim) \rangle.$$

This is the tensor contraction corresponding to $A^\sim \cup \text{AMT} \cup B^\sim$. Similar formulas hold for $c_t^m(AA)$ and $c_t^m(BB)$. (Note also that our notation here differs slightly from our notation for graphs, where we wrote $G = A \cup H \cup B$. Here we write $AB = A^\sim \cup \text{AMT} \cup B^\sim$.) If $c_r(AB) = c_r(AA) = c_r(BB)$, then by Lemma 6.25, $\text{GMT}_{AB}, \text{GMT}_{AA}$ and GMT_{BB} differ only by fiber slips. Furthermore, Lemma 6.18 shows that the gn invariant of Seifert fibered pieces in M_{AA} (and similarly in M_{BB}) must vanish. Therefore by Lemma 6.21, for each m the operator $c_t^m(M_{AA})$ dominates the operator $c_t^m(M_{AB})$, unless $\text{GMT}_{AA} = \text{GMT}_{AB}$ (and similarly for GMT_{BB}).

By Lemma 6.22 it follows that there is a *strict* inequality of lexicographic strings

$$\begin{aligned} & \langle \langle c_t^m(A^\sim) | H_{AB}^m | c_t^m(B^\sim) \rangle \rangle \\ & < \max(\langle \langle c_t^m(A^\sim) | H_{AA}^m | c_t^m(A^\sim) \rangle \rangle, \langle \langle c_t^m(B^\sim) | H_{BB}^m | c_t^m(B^\sim) \rangle \rangle) \end{aligned}$$

unless $\text{GMT}_{AB} = \text{GMT}_{AA} = \text{GMT}_{BB}$ and $c_t^m(A^\sim) = c_t^m(B^\sim)$.

So to complete the proof we just need to check that $c_t^m(A^\sim) = c_t^m(B^\sim)$ implies $A^\sim = B^\sim$ (rel their gluings to the AMT). Of course, A^\sim and B^\sim are attached to different sides of $\text{AMT}_{AA} = \text{AMT}_{AB}$; we really mean that the pairs $(A^\sim, A^\sim \cup \text{AMT}_{AB})$ and $(B^\sim, B^\sim \cup \text{AMT}_{AB})$ are diffeomorphic by a diffeomorphism which restricts to the canonical involution on AMT_{AB} inherited from the identification of AMT_{AB} with AMT_{AA} .

Recall that H^m is positive on the subspace Y (Lemma 6.13). Let m be larger than the number of edges in the fine scale tensor network in A or B . Consider the most general possible basis element y belonging to Y . y is of course symmetrized on the S^π factors and also has a dictated form on the \mathcal{U}_C factors. This form will repeat a basis element whenever h and h' belonging to $H(C)$ are homotopic near the cusp. But as far as these constraints allow, y should be a tensor product of

distinct basis elements (symmetrized as required in S^π). Call these two features of y “deficiency 1” (its symmetrization) and “deficiency 2” (the repeated indices that may occur in \mathcal{U}_C). We need to show that the y components of $c_t^m(A^\sim)$ and $c_t^m(B^\sim)$ are (formally/generically) equal only if A^\sim is equivalent to B^\sim in the sense of the preceding paragraph. We call these features of y “deficiencies” because they complicate the most straightforward strategy of producing a geometric isomorphism by first matching monomials of the partition function.

Deficiency 2 is actually immaterial. To faithfully assemble copies of the fundamental domain W back into X (or the r -upper/lower half of X), we actually already have plenty of gluing information given by the face crossing legs of the small scale tensors and do not need to rely on distinctly labeled $1/2$ legs of τ tensors at the cups.

Regarding deficiency 1, the situation is quite similar to Lemma 6.9. Recall that the broad idea would be to look at a monomial where all edge colors which occur twice are distinct from each other and from the colorings occurring in y , and use it as a recipe to reconstruct A^\sim and equivalently B^\sim . But just as in Lemma 6.9 we should not expect to reconstruct A^\sim (or B^\sim) rel identity on its lower boundary but rather should expect to encounter homeomorphisms of the lower boundary which extend to admissible SF cusp permutations (which extend to the AMT). Such permutations correspond (as in the proof of Lemma 6.9) to a corresponding permutation of the monomials. So deficiency 1 is, for the most part, the familiar and expected result of the geometric symmetry of the SF cusps of the AMT. However, S^π also projects out the fiber coordinate of each SF cusp, so with the original tensors (unmodified by the gn invariant) we could only hope to identify A^\sim and B^\sim up to fish equivalence. This ambiguity is resolved in the next paragraph.

Away from the GMT we conclude that A^- and B^- are fiber slip homeomorphic (rel their gluings to the GMT). But properties (1) and (2) from Definition 6.20 (especially property (2)) show that the gluing invariants t_i of a fiber slip homeomorphism from A^- to B^- can be recovered from the properties of $T_{X,gn}$. This is the last piece of ambiguity, and therefore we conclude $A^- = B^-$ rel their gluings to the GMT, and therefore that $(A, S) = (B, S)$. This completes the proof of the main theorem.

7. EXTENSIONS OF THE MAIN THEOREM

In this section we discuss some natural variations on our main positivity theorem and sketch proofs. We write “theorem” in lower case to acknowledge that important points may have been overlooked. We hope that one or more of our readers will supply full proofs for these statements.

In a similar vein, although the proofs we give of graph positivity in §6.2 are complete, a broader formulation of the statements there would be desirable. Graduate students take note.

7.1. Unoriented case.

Theorem 7.1. *Given S , an unoriented, compact (possibly disconnected) surface without boundary, let $\dot{\mathcal{M}}_S^u$ be the set of unoriented 3-manifolds with boundary S and let \mathcal{M}_S^u be the \mathbb{C} -vector space spanned by $\dot{\mathcal{M}}_S^u$. The natural pairing $\mathcal{M}_S^u \times \mathcal{M}_S^u \rightarrow \mathcal{M}^u$ is positive.*

Proof. The key is again to define a suitable diagonally dominant complexity function c^u . If N is a closed 3-manifold, it is almost enough to take $c(\tilde{N})$ for $c^u(N)$, where \tilde{N} denotes the orientation cover of N (which consists of two copies of N when N is already oriented) and c is the complexity function which is diagonally dominant on oriented 3-manifolds.

This definition only requires two modifications. First recall the terms of the form c_i defined in §3.4. These terms count distinct elements P in the prime decomposition of \tilde{N} . The problem is that the same prime P in \tilde{N} may arise as the oriented double cover of several distinct nonorientable manifolds Q_i . An orientation-reversing involution $\theta_i : P \rightarrow P$ may be free, or have isolated fixed points giving rise to $\mathbb{R}P^3$ summands in $Q_i := P/\theta_i$. This problem is solved by keeping track of *pairs* (P, θ_i) in the definition of c_i .

Second, for a Seifert fibered JSJ piece $X \subset \tilde{N}$ we make an additional modification to the associated tensor T_X , which is involved in the construction of the c_t complexity term, defined in §6.5. Let $b_\theta(X)$ denote the number of boundary components of X/θ . Then multiply T_X by the factor $(1 + b_\theta)$.

It remains to check diagonal dominance of c^u . Suppose (A, S) and (B, S) are elements of \mathcal{M}_S^u . The orientation covers $\tilde{A}\tilde{B}$, $\tilde{A}\tilde{A}$ and $\tilde{B}\tilde{B}$ are each assembled from \tilde{A} and \tilde{B} glued along \tilde{S} . The only way for diagonal dominance to fail is if $(\tilde{A}, \tilde{S}) = (\tilde{B}, \tilde{S})$ but $(A, S) \neq (B, S)$. One way this might happen is for different free prime factors in A and B (or factors which intersect S only in S^2 or $\mathbb{R}P^2$ components) to be covered by the same prime factor in \tilde{A} and \tilde{B} ; the modification of c_t takes care of this case.

So to complete the proof, assume that (A, S_0) is nonorientable, $\mathbb{R}P^2$ -irreducible, and boundary irreducible, and that $\partial A = S_0$ is a (possibly noncompact) surface of finite type other than a disk or Möbius band. It suffices to show that there is no distinct (A', S_0) with identical orientation covers $\tilde{A}' \cong \tilde{A}$ (rel S_0). We look at the orientation involution $\theta : \tilde{A} \rightarrow \tilde{A}$. The involution θ preserves the relative JSJ decomposition of \tilde{A} . Note that the restriction of θ to \tilde{S}_0 does not depend on A or A' , but only on S . The restriction of this involution propagates in a locally rigid way across the relative JSJ pieces which make up A and A' . This is clear for a hyperbolic JSJ piece X , since an isometry is determined by its germ near some invariant subset. For a Seifert fibered JSJ piece X there may be several conjugacy classes (with respect to the mapping class group action) of extension of an involution at the cusps over X . But our modified tensors T_X give rise to c_t terms which favor involutions on both \tilde{A} and \tilde{B} which self-pair boundary components of X rather than pair distinct components. So strict diagonal dominance holds unless $b_\theta(X) = b_\theta(X')$ for all corresponding pairs X, X' of Seifert fibered JSJ pieces in \tilde{A} and \tilde{A}' . But the data of $b_\theta(X)$ is enough information to propagate θ uniquely over X , and inductively one sees that the covering involutions act on the identical pieces \tilde{A} and \tilde{A}' in the same way (rel \tilde{S}_0), and therefore A and A' are the same rel S_0 . This concludes our sketch of the proof. \square

7.2. Orbifolds. For the basic definitions and properties of orbifolds (particularly in low dimensions), see Thurston [45], Chapter 13. We remind the reader of the important distinction between good and bad orbifolds:

Definition 7.2. An orbifold M is *good* if it is (orbifold) covered by a smooth manifold.

The following important theorem of Thurston shows that bad orbifolds can be “localized” to a 2-dimensional stratum.

Theorem 7.3 (Thurston [45]). *A 3-orbifold, all of whose 2-dimensional suborbifolds are good, is itself good.*

In particular, the result of gluing two good 3-orbifolds (with boundary) along a good 2-orbifold is itself good.

In what follows, let S denote a closed oriented good 2-orbifold, and let $\dot{\mathcal{M}}_S^{or}$ be the set of isomorphism classes (rel S) of compact oriented good 3-orbifolds with boundary marked by an identification with S . As before, let \mathcal{M}_S^{or} denote the complex vector space spanned by the set $\dot{\mathcal{M}}_S^{or}$.

Theorem 7.4. *The natural pairing*

$$\mathcal{M}_S^{or} \times \mathcal{M}_S^{or} \rightarrow \mathcal{M}^{or}$$

is positive.

Proof. We sketch the modifications to c required to produce a diagonally dominant complexity function c^{or} on \mathcal{M}^{or} .

In the definition of c_1 , replace fundamental groups by orbifold fundamental groups. The Dijkgraaf-Witten TQFTs are perfectly well defined on orbifolds, and the corresponding partition functions Z_F , as a family, are maximized only when the kernels of the maps $\pi_1^{or}(S) \rightarrow \pi_1^{or}(A)$ and $\pi_1^{or}(S) \rightarrow \pi_1^{or}(B)$ induced by inclusion agree. By the equivariant loop theorem and equivariant Dehn’s Lemma (Meeks-Yau [30]), one obtains matching families of compressing elliptic orbifolds with boundary on S in A and B .

By the equivariant sphere theorem (Meeks-Simon-Yau [29]), good 3-orbifolds have prime decompositions along essential spherical 2-orbifolds. We can construct a sum graph G^{or} associated to the decomposition exactly as in §3.3 with similar conventions as in the manifold case, where only edges marked by actual 2-spheres and vertices by actual punctured 3-spheres are treated as *thin*. To define c_2^{or} , little needs to be changed. Define $r^{or}(M)$ to be the number of irreducible orbifold summands other than $S^1 \times S^2$, and $s^{or}(M)$ to be the number of $S^1 \times S^2$ summands. Close analogs of Theorem 3.13 and Lemma 3.18 continue to hold (thin edges may be slid over themselves and over “fat edges”, whereas “fat edges” are rigid and cannot be slid). As in the original proof of Lemma 3.18, *thick* glues to *thick* and *thin* to *thin*; similar algebra lets us establish a suitable diagonal dominance for c_2^{or} .

The next terms to consider are c_S and c_h (inside c_p). Irreducible (in the orbifold sense) good orbifolds have canonical JSJ decompositions along Euclidean 2-orbifolds into 3-orbifolds which are hyperbolic or covered by Seifert-fibered manifolds (Bonahon-Siebenmann [4]). These second classes of Seifert “fibered” orbifolds are somewhat annoying. A fiber-preserving involution on a Seifert fibered manifold (even an orientation-preserving one) may collapse some circle fiber into an interval with mirror endpoints. Even worse are the Euclidean 3-orbifolds, which might admit no natural fiber structure at all (even a singular one). In fact, the “miracle” of Euclidean orbifolds is that the manifolds amongst them actually *do* admit Seifert fibered structures; this is explained very carefully in Peter Scott’s well-known article on the geometries of 3-manifolds [42]. Mirror interval fibers do not cause much

trouble, but the Euclidean orbifolds are a headache (in fact for our purposes we only need to consider orbifolds which are quotients of sufficiently large manifolds; hence by Theorem 4.17, the Euclidean orbifolds are the only ones which might not admit a fiber structure).

One can deal with this issue in one of two ways. The first is to restrict the definition of \mathcal{M}_S^{gr} to allow only fibered or hyperbolic suborbifolds. The second is to deal with the unfibered Euclidean orbifolds on their own, as a separate class, distinct from the Seifert fibered and hyperbolic JSJ pieces. The good news is that unfiberable Euclidean 3-orbifolds (even those with boundary) have a rigid geometric structure, up to similarity. If we modify c_S so that the terms m, m' count the maximum number of independent Euclidean 2-orbifolds, and the number of JSJ Euclidean 2-orbifolds respectively, then maximizing $(m, -m')$ ensures that fibered Euclidean 3-orbifolds with boundary on S are glued to compatibly fibered Euclidean 3-orbifolds on the other side, and unfibered Euclidean 3-orbifolds on S are glued to unfibered Euclidean 3-orbifolds compatibly with their rigid geometric structures. A term, analogous to the length spectrum term in the definition of c_h , ensures that unfibered 3-orbifolds are identified in matching pairs by doubling along their intersection with S .

The c_h term requires no modification. The analogue of Theorem 5.5 holds in the orbifold context with essentially the same proof. Since all orbifolds in question are good, one can pass to manifold covers and use equivariance of Ricci flow with surgery in the cover to deduce the conclusion. Similarly, the length spectrum is defined on conjugacy classes in the orbifold fundamental group, and the analysis in Lemma 5.4 generalizes to the orbifold context.

Very little in §3.4 requires modification. The unfibered Euclidean 3-orbifolds need special treatment in §6, but the trick is to use their rigid geometric structures, and treat them in the same way that hyperbolic JSJ pieces are treated. Namely, given a Euclidean 3-orbifold X , let W be a fundamental domain for the action of the group of isometries of X , and build a tensor T_X made up from copies of tensors T_W which has exactly the same symmetries as X . Unfibered JSJ 2-orbifolds have finite symmetry groups which can be dealt with by hand. The invariant gn makes sense, and is natural and chiral, for 3-orbifolds bounded by fibered Euclidean 2-orbifolds, and a suitable generalization of Lemma 6.18 continues to hold. \square

7.3. Tangles. Finally, let S denote a closed, oriented surface with finitely many marked points $\{p\}$. Let \mathcal{M}_S^l denote the set of (relative isomorphism classes of) oriented compact manifolds A containing a tangle l with $\partial A = S$ and $\partial l = \{p\}$, and let \mathcal{M}_S^l denote the complex vector space it formally spans.

Theorem 7.5. *The natural pairing*

$$\mathcal{M}_S^l \times \mathcal{M}_S^l \rightarrow \mathcal{M}^l$$

is positive.

Proof. Define a complexity function c^l on pairs (M, l) , where M is a closed oriented 3-manifold and l is a link in M , as follows. First define a complexity c_c^l when M is connected, and then define

$$c^l(M, L) = (c_0(M), B(M, l), \{B(M_i, l_i)\}, \{c_c^l(M_i, l_i)\}),$$

where the last term in brackets is a lexicographical list of the complexities of the connected components. Here B is a “bad orbifold” weight: $B(M, l)$ denotes the

number of connected summands of (M, l) of the form $(S^1 \times S^2, S^1 \times *)$, where $*$ $\subset S^2$ is a base point and the connected sum is taken *away* from the link $S^1 \times *$, and $\{B(M_i, l_i)\}$ is the component-wise lexicographical list counting summands of this type.

To complete the definition, we must define c_c^l . The idea is that once all $(S^1 \times S^2, S^1 \times *)$ summands have been removed, the remainder (M_-, l_-) will become a *good* orbifold once any *fixed* positive integer $n \geq 2$ is attached to *each* component of l_- . Thus, it suffices to define $c_c^l(M, l) = c^{or}(M_-, l_{-,2})$, where $(M_-, l_{-,2})$ denotes the good orbifold obtained by marking all components of l_- as cone geodesics with angle π .

The number of $(S^1 \times S^2, S^1 \times *)$ in (M, l) and $\{(M_i, l_i)\}$ is maximized in the double. By Theorem 7.4, the complexity c^{or} is diagonally dominant, so

$$c^{or}((A_-, l_{A_-})(B_-, l_{B_-})) < \max(c^{or}((A_-, l_{A_-})(A_-, l_{A_-})), c^{or}((B_-, l_{B_-})(B_-, l_{B_-})))$$

unless $(A_-, l_{A_-})(B_-, l_{B_-}) \cong (A_-, l_{A_-})(A_-, l_{A_-}) \cong (B_-, l_{B_-})(B_-, l_{B_-})$ as orbifolds. Since an isomorphism of orbifolds necessarily induces an isomorphism of the orbifold loci, we obtain an isomorphism in the category of pairs (3-manifold, link). This completes the sketch. \square

APPENDIX A. FINITE GROUP TQFTS

This is a very brief summary of finite group TQFTs. For more details, see [10], [13], [36], [12] and [47]. Here we consider only the (2+1)-dimensional untwisted theories, and we ignore gluing with corners, higher codimension gluing, etc.

Let G be a finite group and BG be a classifying space for G . All constructions in this appendix depend of G , but that is left notationally implicit.

For X a manifold (of any dimension), let $\mathcal{M}(X)$ denote the space of all continuous maps $X \rightarrow BG$. There is a restriction map (fibration)

$$r : \mathcal{M}(X) \rightarrow \mathcal{M}(\partial X).$$

For $c \in \mathcal{M}(\partial X)$, let $\mathcal{M}(X; c)$ denote $r^{-1}(c)$.

Let Y be a closed 2-manifold. Define the (finite dimensional) Hilbert space $V(Y)$ to be the space of all locally constant functions $f : \mathcal{M}(Y) \rightarrow \mathbb{C}$. In other words, f assigns a number to each homotopy class of map $Y \rightarrow BG$ (equivalently, to each isomorphism class of principal G -bundles over Y). If Y is the empty 2-manifold (e.g. the boundary of a closed 3-manifold), then we have a canonical identification $V(Y) = \mathbb{C}$.

(For twisted theories, one replaces locally constant maps with maps which vary according to $f - g = \omega(h)$, where ω is a fixed 3-cocycle on BG and $h : Y \times I \rightarrow BG$ is a homotopy from g to f . To make sense of this, manifolds must be equipped with fundamental classes; see [13].)

Let $\alpha \in \pi_0(\mathcal{M}(Y))$ and let χ_α denote the function which is 1 on the path component α and 0 on the other path components of $\mathcal{M}(Y)$. The various χ_α form a basis of $V(Y)$. The inner product on $V(Y)$ is given by

$$\langle \chi_\alpha, \chi_\beta \rangle = \begin{cases} 0 & \alpha \neq \beta, \\ \frac{1}{|\pi_1(\mathcal{M}(Y); \alpha)|} & \text{otherwise.} \end{cases}$$

In particular, $\{\chi_\alpha\}$ is an orthogonal basis of $V(Y)$. Note that if Y is connected and $\rho : \pi_1(Y) \rightarrow G$ is a representative of the conjugacy class of group homomorphisms

corresponding to the path component α , then $|\pi_1(\mathcal{M}(Y); \alpha)| = |\text{stab}(\rho)|$, where $\text{stab}(\rho)$ denotes the stabilizer of ρ under the outer action of G by conjugation; i.e., $\text{stab}(\rho)$ is the set of all elements of G which commute with the image of ρ . In terms of bundles, $\pi_1(\mathcal{M}(Y); \alpha)$ and $\text{stab}(\rho)$ can be identified with the automorphisms of the G -bundle over Y corresponding to α .

Next we define the path integral $Z(M) \in V(\partial M)$ for a 3-manifold M . For $a \in \mathcal{M}(\partial M)$, define

$$Z(M)(a) = \sum_{\beta \in \pi_0(\mathcal{M}(M; a))} \frac{1}{|\pi_1(\mathcal{M}(M; a); \beta)|}.$$

In other words, we sum over the path components of the extensions of a to all of M , and each path component counts as the inverse of the size of its fundamental group. (In terms of bundles, $|\pi_1(\mathcal{M}(M; a); \beta)|$ is the number of automorphisms of the bundle corresponding to β which restrict to the identity over ∂M .)

A basic result for finite group TQFTs is the following gluing formula. Let A and B be 3-manifolds with common boundary S . Then we have $Z(A), Z(B) \in V(S)$ and

$$Z(A \cup_S B) = \langle Z(A), Z(B) \rangle.$$

The proof considers the homotopy long exact sequence of the fibration $\mathcal{M}(A \cup_S B) \rightarrow \mathcal{M}(S)$, whose fiber at $x \in \mathcal{M}(S)$ is $\mathcal{M}(A; x) \times \mathcal{M}(B; x)$. All homotopy groups in the sequence are finite, and only the six π_1 and π_0 terms are nontrivial. See [36] for details.

APPENDIX B. APPLICATION TO HEEGAARD GENUS

This appendix illustrates how the Dijkgraaf-Witten TQFTs can be used to obtain nontrivial lower bounds on the Heegaard genus.

For $i = 1, 2$, let $\gamma_i : S^1 \rightarrow S^1 \times S^2$ be disjoint nonparallel embeddings so that the complement of the images is a homology (torus $\times I$). Let S_1, S_2 be disjoint open solid torus neighborhoods of the images $\gamma_1(S^1), \gamma_2(S^1)$. Let $C = S^1 \times S^2 - S_1$ and $C' = S^1 \times S^2 - (S_1 \cup S_2)$.

For any integer n , let DC_n be obtained from two copies of C (with opposite orientations) and $2^n - 2$ copies of C' (coming in orientation-reversed pairs) glued together in the pattern

$$DC_n = C\overline{C}'C'\overline{C}' \dots C'\overline{C},$$

where C is glued to \overline{C}' along their common boundary component ∂S_1 , where a $\overline{C}'C'$ term glues the first to the second along a ∂S_2 component, and a $C'\overline{C}'$ term glues the first to the second along a ∂S_1 component. Notice that DC_n can be built from $2^n \cdot \text{const.}$ simplices, by choosing fixed triangulations of C and C' which agree on their common boundary component; see Figure 19.

Apart from the two copies of C at either end, DC_n is almost equal to the cyclic cover of order 2^{n-1} of the double DC' , where the cover unwraps the circle obtained by doubling an arc from one boundary component of C' to the other.

This appendix shows that the Heegaard genus of DC_n is of size at least $2^n \cdot \text{const.}$, where the (positive) constant depends only on γ_1, γ_2 . By the assumption on the γ_i , the manifolds DC_n are all homology $S^1 \times S^2$'s, so one cannot obtain a lower bound on their Heegaard genus from homological invariants. The point is not that this result cannot be obtained by other means (although for some cases of the above

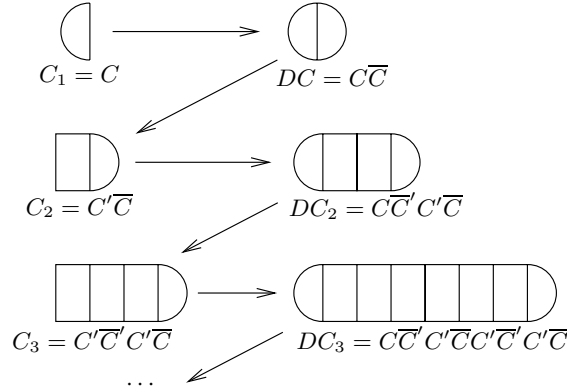


FIGURE 19. Schematic picture of manifolds constructed in proof of Theorem B.1

construction, the method of [41] also yields a lower bound which is optimal up to a constant); the point is that this estimate is obtained as a formal consequence of the properties of Dijkgraaf-Witten TQFTs. Classical 3-manifold topology enters only implicitly (but importantly) in the fact that $\pi_1(C)$ is residually finite.

For a closed 3-manifold M triangulated with T tetrahedra, there is a natural Heegaard splitting, where one side is taken to be a regular neighborhood of the 1-skeleton. The genus of the splitting surface is $T + 1$, which we take as a naive upper bound for the Heegaard genus $h(M)$ of M . We construct a family of irreducible examples of fixed homological type where this naive bound is sharp up to a constant.

Theorem B.1. *There exists a constant $c > 0$ and a sequence of irreducible 3-manifolds M_1, M_2, \dots , each of which is \mathbb{Z} -homology equivalent to $S^1 \times S^2$, such that the Heegaard genus satisfies $h(M_i) > c \cdot T(M_i)$, where T denotes the minimal number of tetrahedra in a triangulation of M_i , and $T(M_i)$ approaches infinity.*

Proof. Let $\gamma : S^1 \times D^2 \hookrightarrow S^1 \times S^2$ be an embedding with incompressible complement C , so that $S^1 \times S^2 = S \cup C$, where $S = S^1 \times D^2$. Using residual finiteness of $\pi_1(C)$ [22], let F be a finite group so that the conjugacy class of the meridian of S in $\pi_1(C)$ maps nontrivially under some homomorphism $\theta : \pi_1(C) \rightarrow F$. Let Z be the partition function (or relative partition function) of the TQFT associated to F , which is unitary by Appendix A. Since the meridian survives in at least one F -representation of $\pi_1(C)$, we have $Z(S) \neq Z(C)$. From the definition of Z , we have

$$(B.0.1) \quad Z(S^1 \times S^2) = \text{tr}(\text{Id}_{V(S^2)}) = \text{tr}(\text{Id}_{\mathbb{C}}) = 1.$$

Now, $\langle Z(S), Z(S) \rangle = Z(S^1 \times S^2) = 1$, so $|Z(S)| = 1$. But $\langle Z(C), Z(S) \rangle = Z(S^1 \times S^2) = 1$ as well. Since $Z(C) \neq Z(S)$, we conclude that $|Z(C)| = l > 1$ for some l . Furthermore, if DC denotes the double, then $|Z(DC)| = \langle Z(C), Z(C) \rangle = l^2$.

Now let $\omega : S^1 \times D^2 \hookrightarrow C$ be a fixed embedding inducing an isomorphism on integer homology. Define $C_2 = DC - \text{image}(\omega)$, where we take $\text{image}(\omega)$ to lie in the positive copy of C in DC . Then $\langle Z(S_\omega), Z(C_2) \rangle = Z(DC) = l^2$, where S_ω denotes the solid torus embedded via ω . Thus, $|Z(C_2)| \geq l^2$ and so $Z(DC_2) > l^4$.

Using another copy of $C \subset DC_2$, compose $\omega_2 : S^1 \times D^2 \hookrightarrow C \hookrightarrow DC_2$ and set $C_3 = DC_2 - \text{image}(\omega_2)$. We conclude $\langle Z(S_{\omega_2}), Z(C_3) \rangle \geq l^4$ and so $|Z(DC_3)| \geq l^8$.

Find $\omega_3 : S^1 \times D^2 \hookrightarrow DC_3$ (again, by composing ω with an inclusion $C \hookrightarrow DC_3$). Define C_4 and similarly conclude that $|Z(C_4)| > l^{16}$. By induction, we obtain a sequence $\{DC_i\}$ of \mathbb{Z} -homology $S^1 \times S^2$'s with

$$(B.0.2) \quad Z(DC_i) > l^{2^i}, \quad l > 1.$$

Associated to a finite group TQFT is a positive integer called the *total quantum dimension*, which we denote \mathcal{D} . This number is just the cardinality of the finite group G on which the TQFT is based. The only property we use is that $\mathcal{D} > 1$ for nontrivial groups, and the following lemma: \square

Lemma B.2. $|Z(\#_g S^1 \times S^2)| = \mathcal{D}^{g-1}$.

See e.g. [10].

Corollary B.3 (to lemma). $h(DC_i) > \text{const} \cdot 2^i$ for some positive constant.

Proof. Let h_g be a handlebody of genus g . By the Cauchy-Schwarz inequality,

$$Z(\#_g S^1 \times S^2) = \langle Z(h_g), Z(h_g) \rangle \geq Z(M)$$

when $h(M) \leq g$. Together with equation (B.0.2) this implies:

$$(B.0.3) \quad l^{2^i} < \mathcal{D}^{h(DC_i)-1}.$$

So, $\frac{\log(l)}{\log(\mathcal{D})} 2^i < h(DC_i) - 1$, and thus, $h(DC_i) > c' \cdot 2^i$ for some constant $c' > 0$. \square

By choosing fixed triangulations of C and $C \setminus \omega(S^1 \times D^2)$ with common boundaries, we see that there is a constant c'' so that

$$(B.0.4) \quad T(DC_i) < c'' \cdot 2^i.$$

Combining Corollary B.3 and (B.0.4), we have

$$(B.0.5) \quad h(DC_i) > \frac{c'}{c''} T(DC_i).$$

Combinatorial methods for obtaining lower bounds on the Heegaard genus include extensions due to Schultens [41] of the connect sum formula $h(M\#N) = h(M) + h(N)$, and Lackenby's sweepout method [26], partially summarized below.

Suppose a closed hyperbolic manifold M has a family of finite covers M_i of degree d_i with a uniform *spectral gap*; i.e.,

$$\lambda_1(M_i) \geq \epsilon > 0$$

for some ϵ , where λ_1 denotes the first positive eigenvalue of the Laplacian on functions (see [28], especially Chapter 4, for definitions and an introduction). Obtain an index one minimal Heegaard surface $\Sigma_i \subset M_i$ by using a minimax "sweepout" of M_i . A minimal surface in a hyperbolic manifold has curvature bounded above by -1 , so by Gauss-Bonnet, $-\chi(\Sigma_i) \geq \frac{\text{area}(\Sigma_i)}{2\pi}$. Moreover, since Σ_i is the maximal area surface in a minimax sweepout, there is an inequality

$$\text{area}(\Sigma_i) \geq \text{area}(\Sigma'_i),$$

where Σ'_i is one of the sweepout surfaces which evenly divides the volume of M_i . Cheeger's isoperimetric constant c (see [28], Chapter 4), defined by

$$c = \inf_{\text{separating } \Sigma'} \frac{\text{area}(\Sigma')}{\min(\text{vol}(\text{component } M - \Sigma'))},$$

satisfies $\lambda_1 \cdot \text{const.} < c < \sqrt{\lambda_1} \cdot \text{const.}$. Therefore we may conclude that $-\chi(\Sigma_i) > d_i \cdot \text{const.}$. In other words, the ratio of the Heegaard genus to volume (or to $T(M_i)$) is bounded below uniformly for the family $\{M_i\}$; one says the $\{M_i\}$ have a positive *Heegaard gradient*.

These three methods for bounding $h(M)$ from below appear quite different. It would be interesting to see if they can be usefully combined. There is no shortage of interesting objectives for sufficiently powerful lower bounds on h . Lackenby has pointed out that any hyperbolic 3-manifold exhibiting a positive Heegaard gradient for all finite covers would be a counterexample to Thurston’s virtually fibered conjecture.

APPENDIX C. CONJECTURE

First consider the following conjecture in linear algebra. We call it the “quantum maxflow/mincut”. Let N be a v -valent graph with k input cut edges and l output cut edges. At each vertex of N we presume that the incoming edges are locally ordered $1, 2, \dots, k$. Given this data on N , picking a v -index tensor W (on some vector space V) transforms N into a linear map $\theta_{N,W} : V^{\otimes k} \rightarrow V^{\otimes l}$. Simply insert a copy of W at each vertex according to the local edge labeling and contract where appropriate.

Conjecture C.1 (QMF/MC). *$\theta_{N,W}$ is an injection for generic W if and only if N admits k edge-disjoint paths from input to output. Furthermore, $\text{image}(\theta_{N,W})$ is generic with respect to any fixed basis for V . That is, if $h \leq l - k$ and basis vectors \vec{b} are inserted into h output slots not on the k edge disjoint paths, then the resulting map $\theta_{N,W,\vec{b}} : V^{\otimes k} \rightarrow V^{\otimes(l-h)}$ is also an injection for generic W .*

Remark C.2. (1) The “only if” part follows from the ordinary maxflow/mincut theorem. If the k edge-disjoint paths do not exist, then there is a j -cut, $j < k$, separating input from output and factoring $\theta_{N,W}$ through $V^{\otimes j}$.

(2) The nonsingularity of $\theta_{N,W}$ is an algebraic condition, so, fixing N , it will hold generically for W provided it holds for a single W .

(3) The condition that W becomes an injective operator for every division of the index (slots) into $v = a+b$, $a \leq b$, is a finite intersection of generic conditions and therefore generic. As a consequence, the conjecture is true for *monotone* networks N . These are networks N admitting a function to the interval $\phi : N \rightarrow [0, 1]$, $\phi(\text{input}) = 0$, $\phi(\text{output}) = 1$, with $|\phi^{-1}(y)| \geq |\phi^{-1}(x)|$ for all $0 \leq x < y \leq 1$, provided $\phi(x) \neq \phi(w)$, w a vertex of N .

In section 6 we constructed a tensor T_X with positivity property P for X SF. In our proof, the distinction between algebraic and geometric middle thirds, $\text{AMT} \subset \text{GMT}$, was necessitated by a lack of a similar tensor for the hyperbolic JSJ pieces. (In an early draft of this paper, we thought we had constructed such tensors but relied on a mistaken “proof” of QMF/MC.) The existence of such tensors is, for hyperbolic JSJ pieces, an independently interesting question.

Given a finite volume hyperbolic three-manifold X , consider a tensor T_X living on $Q^m = \bigotimes_{\text{cusps } C} V_m(C)$, where $V_m(C)$ is as in section 6. The important properties of $V_m(C)$ are that (1) it is canonically identified to the cusp so that actions on the cusps induce actions on Q , (2) the action is “eventually faithful” as $m \rightarrow \infty$ and

also has many orbits ($\geq m$), and (3) $V_m(C)$ has a natural basis which we regard as generating a positive cone. We assume $V_m(C)$ has these properties:

Definition C.3. T_X has property P_h if

- (1) every symmetry of T_X is induced by a unique symmetry of X ,
- (2) for each reflection symmetry r (of X), each operator \mathcal{O} from “lower” to “upper” cusp spaces induced by inserting (positive) basis vectors into the cusps meeting $fix(r)$ is positive: $\langle x|\mathcal{O}|x\rangle > 0$ for $|x\rangle \neq 0$ in the lower cusp space and $\langle x| = r_*|x\rangle$ the corresponding element of the upper cusp space.

Lemma C.4. *QMF/MC implies that for each hyperbolic finite volume 3-manifold X , there exists a tensor T_X satisfying property P_h .*

Proof sketch. Let G be the symmetry group of X and W a fundamental domain. Build a tensor T_W as indicated in Figure 20, which has a V_m index (drawn as an edge) exiting each interior cusp of W and also a V_m index exiting each fractional cusp located on a singular stratum of W . At every codimension-one face of W the tensor T_W should have l indices, each taking values in V_m . Here, the number l of “legs” is chosen sufficiently large; we will say just how large in what follows. Pictorially, we think of T_W as a high valence vertex in W with l legs exiting every face and one leg exiting each cusp and cusp fraction.

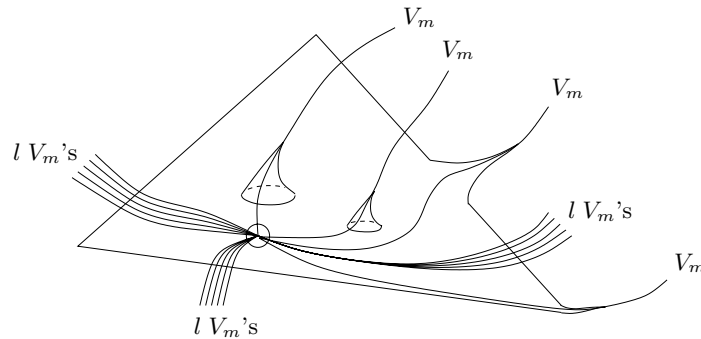
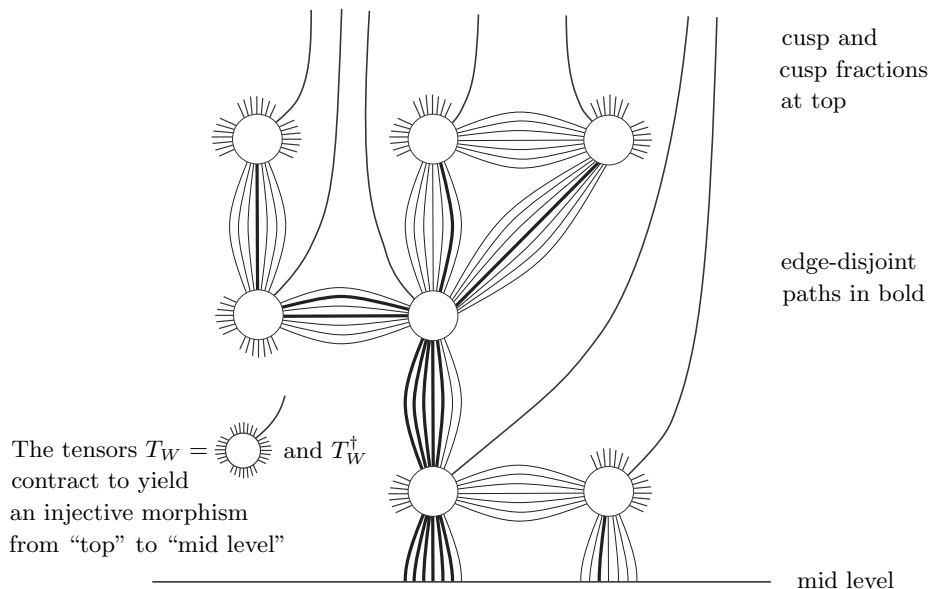


FIGURE 20. This diagram represents an alternative construction of T_W given Conjecture C.1. One leg exits every cusp and fractional cusp; l legs exit every face.

Use r to define upper and lower halves of X . Consider the network of tensors coming from the copies of W in the upper half of X ; see Figure 21. We think of this network of tensors as an operator U from the upper (possibly fractional) cusp edges to the “middle” edges along the codimension-1 fixed submanifold of r . Choose l large enough so that we can draw mutually edge-disjoint paths from each upper cusp edge to the middle level. The input from r -invariant cusps corresponds to inserting some positive basic vectors at the nodes closest to the mid level in Figure 21. The QMF/MC implies that a generic choice of T_W yields an injective operator U when inserted into the network. Thus, $\mathcal{O} = U^\dagger U$ is strictly positive. \square

FIGURE 21. In this figure, $l = 7$.

ACKNOWLEDGMENTS

We would like to thank Ian Agol, Daryl Cooper, Nathan Dunfield, Cameron Gordon, Alexei Kitaev, Sadayoshi Kojima, Marc Lackenby, Darren Long, Shigenori Matsumoto, John Morgan, Marty Scharlemann, and the anonymous referee for numerous helpful comments, suggestions and corrections. We would especially like to single out Alexei Kitaev for thanks for posing questions which were the initial stimulus for this work.

REFERENCES

- [1] Ian Agol, Peter A. Storm, and William P. Thurston. Lower bounds on volumes of hyperbolic Haken 3-manifolds. *J. Amer. Math. Soc.*, 20(4):1053–1077 (electronic), 2007. With an appendix by Nathan Dunfield. MR2328715 (2008i:53086)
- [2] Michael Atiyah. *The geometry and physics of knots*. Lezioni Lincee. [Lincei Lectures]. Cambridge University Press, Cambridge, 1990. MR1078014 (92b:57008)
- [3] C. Blanchet, N. Habegger, G. Masbaum, and P. Vogel. Topological quantum field theories derived from the Kauffman bracket. *Topology*, 34(4):883–927, 1995. MR1362791 (96i:57015)
- [4] F. Bonahon and L. C. Siebenmann. The characteristic toric splitting of irreducible compact 3-orbifolds. *Math. Ann.*, 278(1-4):441–479, 1987. MR909236 (90a:57017)
- [5] Hubert L. Bray. Proof of the Riemannian Penrose inequality using the positive mass theorem. *J. Differential Geom.*, 59(2):177–267, 2001. MR1908823 (2004j:53046)
- [6] A. J. Casson and C. McA. Gordon. Reducing Heegaard splittings. *Topology Appl.*, 27(3):275–283, 1987. MR918537 (89c:57020)
- [7] Bennett Chow and Dan Knopf. *The Ricci flow: An introduction*, volume 110 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2004. MR2061425 (2005e:53101)

- [8] Tobias H. Colding and William P. Minicozzi, II. *Minimal surfaces*, volume 4 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 1999. MR1683966 (2002b:49072)
- [9] Dennis M. DeTurck. Deforming metrics in the direction of their Ricci tensors. *J. Differential Geom.*, 18(1):157–162, 1983. MR697987 (85j:53050)
- [10] Robbert Dijkgraaf and Edward Witten. Topological gauge theories and group cohomology. *Comm. Math. Phys.*, 129(2):393–429, 1990. MR1048699 (91g:81133)
- [11] Theodore Frankel. Applications of Duschek’s formula to cosmology and minimal surfaces. *Bull. Amer. Math. Soc.*, 81:579–582, 1975. MR0362166 (50:14608)
- [12] Daniel S. Freed. Higher algebraic structures and quantization. *Comm. Math. Phys.*, 159(2):343–398, 1994. MR1256993 (95c:58034)
- [13] Daniel S. Freed and Frank Quinn. Chern-Simons theory with finite gauge group. *Comm. Math. Phys.*, 156(3):435–472, 1993. MR1240583 (94k:58023)
- [14] Michael Freedman, Joel Hass, and Peter Scott. Least area incompressible surfaces in 3-manifolds. *Invent. Math.*, 71(3):609–642, 1983. MR695910 (85e:57012)
- [15] Michael H. Freedman, Alexei Kitaev, Chetan Nayak, Johannes K. Slingerland, Kevin Walker, and Zhenghan Wang. Universal manifold pairings and positivity. *Geom. Topol.*, 9:2303–2317 (electronic), 2005. MR2209373 (2006k:57080)
- [16] Mikhael Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981. MR623534 (83b:53041)
- [17] Richard S. Hamilton. The formation of singularities in the Ricci flow. In *Surveys in differential geometry, Vol. II (Cambridge, MA, 1993)*, pages 7–136. Int. Press, Cambridge, MA, 1995. MR1375255 (97e:53075)
- [18] Richard S. Hamilton. Non-singular solutions of the Ricci flow on three-manifolds. *Comm. Anal. Geom.*, 7(4):695–729, 1999. MR1714939 (2000g:53034)
- [19] Joel Hass and Peter Scott. The existence of least area surfaces in 3-manifolds. *Trans. Amer. Math. Soc.*, 310(1):87–114, 1988. MR965747 (90c:53022)
- [20] Allen Hatcher. Notes on Basic 3-Manifold Topology. Available from the author’s website, 2000.
- [21] John Hempel. *3-Manifolds*. Ann. of Math. Studies, No. 86. Princeton University Press, Princeton, N. J., 1976. MR0415619 (54:3702)
- [22] John Hempel. Residual finiteness for 3-manifolds. In *Combinatorial group theory and topology (Alta, Utah, 1984)*, volume 111 of *Ann. of Math. Stud.*, pages 379–396. Princeton Univ. Press, Princeton, NJ, 1987. MR895623 (89b:57002)
- [23] William H. Jaco and Peter B. Shalen. Seifert fibered spaces in 3-manifolds. *Mem. Amer. Math. Soc.*, 21(220):viii+192, 1979. MR539411 (81c:57010)
- [24] Klaus Johannson. *Homotopy equivalences of 3-manifolds with boundaries*, volume 761 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979. MR551744 (82c:57005)
- [25] Matthias Kreck and Peter Teichner. Positivity of topological field theories in dimension at least 5. *J. Topol.*, 1(3):663–670, 2008. MR2417448 (2009d:81332)
- [26] Marc Lackenby. Heegaard splittings, the virtually Haken conjecture and property (τ) . *Invent. Math.*, 164(2):317–359, 2006. MR2218779 (2007c:57030)
- [27] F. Laudenbach. Sur les 2-sphères d’une variété de dimension 3. *Ann. of Math. (2)*, 97:57–81, 1973. MR0314054 (47:2606)
- [28] Alexander Lubotzky. *Discrete groups, expanding graphs and invariant measures*, volume 125 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1994. With an appendix by Jonathan D. Rogawski. MR1308046 (96g:22018)
- [29] William Meeks, III, Leon Simon, and Shing Tung Yau. Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature. *Ann. of Math. (2)*, 116(3):621–659, 1982. MR678484 (84f:53053)
- [30] William H. Meeks, III and Shing Tung Yau. The equivariant Dehn’s lemma and loop theorem. *Comment. Math. Helv.*, 56(2):225–239, 1981. MR630952 (83b:57006)
- [31] Pengzi Miao. Positive mass theorem on manifolds admitting corners along a hypersurface. *Adv. Theor. Math. Phys.*, 6(6):1163–1182 (2003), 2002. MR1982695 (2005a:53065)
- [32] Jean-Pierre Otal. Thurston’s hyperbolization of Haken manifolds. In *Surveys in differential geometry, Vol. III (Cambridge, MA, 1996)*, pages 77–194. Int. Press, Boston, MA, 1998. MR1677888 (2000b:57025)

- [33] Roger Penrose. *Structure of space-time; in Battelle rencontres. 1967 lectures in mathematics and physics*. Edited by Cecile M. DeWitt and John A. Wheeler. W. A. Benjamin, Inc., New York-Amsterdam, 1968. MR0353936 (50:6418)
- [34] Grisha Perelman. The entropy formula for the Ricci flow and its geometric applications, 2002.
- [35] Grisha Perelman. Ricci flow with surgery on three-manifolds, 2003.
- [36] Frank Quinn. Lectures on axiomatic topological quantum field theory. In *Geometry and quantum field theory (Park City, UT, 1991)*, volume 1 of *IAS/Park City Math. Ser.*, pages 323–453. Amer. Math. Soc., Providence, RI, 1995. MR1338394 (96e:57021)
- [37] M. S. Raghunathan. *Discrete subgroups of Lie groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68. Springer-Verlag, New York, 1972. MR0507234 (58:22394a)
- [38] N. Reshetikhin and V. G. Turaev. Invariants of 3-manifolds via link polynomials and quantum groups. *Invent. Math.*, 103(3):547–597, 1991. MR1091619 (92b:57024)
- [39] Justin Roberts. Irreducibility of some quantum representations of mapping class groups. *J. Knot Theory Ramifications*, 10(5):763–767, 2001. Knots in Hellas '98, Vol. 3 (Delphi). MR1839700 (2002f:57065)
- [40] Richard Schoen. Estimates for stable minimal surfaces in three-dimensional manifolds. In *Seminar on minimal submanifolds*, volume 103 of *Ann. of Math. Stud.*, pages 111–126. Princeton Univ. Press, Princeton, NJ, 1983. MR795231 (86j:53094)
- [41] Jennifer Schultens. Heegaard genus formula for Haken manifolds. *Geom. Dedicata*, 119:49–68, 2006. MR2247647 (2007e:57020)
- [42] Peter Scott. The geometries of 3-manifolds. *Bull. London Math. Soc.*, 15(5):401–487, 1983. MR705527 (84m:57009)
- [43] Miles Simon. Deformation of C^0 Riemannian metrics in the direction of their Ricci curvature. *Comm. Anal. Geom.*, 10(5):1033–1074, 2002. MR1957662 (2003j:53107)
- [44] J. Singer. Three-dimensional manifolds and their Heegaard diagrams. *Trans. Amer. Math. Soc.*, 35:88–111, 1933. MR1501673
- [45] William P. Thurston. Geometry and topology of 3-manifolds (a.k.a. Thurston's notes), 1979.
- [46] Friedhelm Waldhausen. Eine Klasse von 3-dimensionalen Mannigfaltigkeiten. I, II. *Invent. Math.* 3 (1967), 308–333; *ibid.*, 4:87–117, 1967. MR0235576 (38:3880)
- [47] Kevin Walker. TQFTs. Preprint, available at <http://canyon23.net/math/>.

DEPARTMENT OF MATHEMATICS, CALTECH, PASADENA, CALIFORNIA 91125
E-mail address: `dannyc@its.caltech.edu`

MICROSOFT STATION Q, UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CALIFORNIA 93106
E-mail address: `michaelf@microsoft.com`

MICROSOFT STATION Q, UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CALIFORNIA 93106
E-mail address: `kevin@canyon23.net`