

SPHERE RECOGNITION LIES IN NP

SAUL SCHLEIMER

ABSTRACT. We prove that the three-sphere recognition problem lies in **NP**. Rubinstein [Haifa, 1992] and Thompson [Math. Res. Let., 1994] showed that the problem is decidable. Our result relies on Casson's [MSRI, 1997] version of their algorithm and recent results of Agol, Hass, and Thurston [STOC, 2002].

1. INTRODUCTION

The three-sphere recognition problem asks: given a triangulation T , is the underlying space $|T|$ homeomorphic to the three-sphere? To solve this problem, Rubinstein [23] introduced *almost normal two-spheres* (see Section 5). Thompson [26] simplified Rubinstein's proof using Gabai's technique of *thin position* [7]:

Theorem 1.1 (Rubinstein [23], Thompson [26]). *The three-sphere recognition problem lies in **EXPTIME**.*

Casson [4] then introduced the novel idea of *crushing* triangulations along normal two-spheres (see Section 14). This reduced the space complexity:

Theorem 15.1 (Casson [4]). *The three-sphere recognition problem lies in **PSPACE**.*

Following Casson's algorithm and work of Agol, Hass, and Thurston [2] (see Theorem 5.5) we will show:

Theorem 16.1. *The three-sphere recognition problem lies in **NP**.*

That is, any triangulation T of the three-sphere admits a polynomial-sized *certificate*: a proof that T is indeed a triangulation of the three-sphere (see Section 3). Theorem 16.1 has an immediate corollary:

Corollary 1.2. *The three-ball recognition problem lies in **NP**.*

Date: August 8, 2009.

This work is in the public domain.

Proof. We must produce a certificate for every triangulation T of the three-ball. (See Theorem 4.1 for a polynomial-time algorithm to verify that $|T|$ is a three-manifold.) First write down a polynomial-sized proof that $S = \partial|T|$ is a two-sphere (check connectedness and Euler characteristic.) Next, build $D(T)$: the triangulation obtained by *doubling* across S . Next, Theorem 16.1 gives a certificate that $|D(T)|$ is a three-sphere. Finally, Alexander's Theorem [11, Theorem 1.1] tells us that the two-sphere S bounds a three-ball in $|D(T)|$. \square

A *surface vector* $v(S)$ is the vector of normal coordinates of S with respect to the triangulation T (see Section 5.1). A result closely related to Corollary 1.2 is:

Corollary 1.3. *The following problem lies in **NP**: given a triangulation T of a three-manifold and $v(S)$ where S is a normal two-sphere, decide if S bounds a three-ball in $|T|$.*

Proof. Crush the triangulation T along the surface S to obtain a triangulation T' . Theorem 5.9 of Jaco and Rubinstein's paper [16] tells us that the connect sum of the components of $|T'|$ is homeomorphic to $|T|$, up to keeping track of lens space summands.

Following Casson, Barchecheat's thesis [3, page 50] gives a polynomial-time algorithm to keep track of how to reassemble the lens spaces and components of $|T'|$ in order to recover $|T|$. Left to check is that all lens spaces and components of $|T'|$, arising as submanifolds of the ball bounded by S , are three-spheres. The former are dealt with as in [3]. The latter are certified using Theorem 16.1. \square

We next state a technical result, involved in the proof of Theorem 16.1, that may be of independent interest.

Theorem 13.1. *There is a polynomial-time algorithm that, given a triangulation T of an oriented three-manifold and $v(S)$ where S is a transversely oriented almost normal surface, produces as output $v(\tilde{S})$, the vector for the normalization of S .*

Corollary 1.3, Theorem 13.1, and the bounds given by [25, Chapter 6] result in:

Corollary 1.4. *The following problem lies in **NP**: given a triangulation T of a closed orientable irreducible atoroidal three-manifold, decide if $|T|$ is a surface bundle over the circle.* \square

Corollary 1.4, unfortunately, is very far from proving that bundle recognition lies in **NP**; certifying irreducibility or atoroidality are interesting and difficult open questions. Note that certifying *zero-efficiency*

would in turn certify irreducibility. Corollary 1.4 is similar to a earlier result of Ivanov [13]. He shows that recognition of the three-sphere, amongst the class of zero-efficient triangulations (there called *irreducible Q -triangulations*), lies in **NP**.

There are other problems in three-manifold topology lying in **NP**. Hass, Lagarias, and Pippenger [10] have shown that the unknotting problem, first solved by Haken, lies in **NP**. Agol [1] has given a proof, using sutured manifold hierarchies, that the recognition of Haken manifolds lies in **NP** (see also [15]). Agol’s algorithm requires, as the base case, some version of Corollary 1.3. Agol deduces that the unknotting problem lies in **co-NP**. Also, Agol, Hass and Thurston [2] have shown that the 3-manifold knot genus problem is **NP**-complete. For a discussion of algorithmic three-manifold topology we refer the reader to either [10] or to Matveev’s book [20]

Historical note. Three-sphere recognition (Theorem 1.1) is a fundamental result in low-dimensional topology. Accordingly many expositions exist. A solution to the problem was first presented by Rubinstein [23] in a series of lectures in Haifa, Israel, in 1992. Thompson [26] gave a different proof, relying on Gabai’s technique of *thin position* [7] soon after. Rubinstein’s papers [24, 22] lay out his original ideas on the problem. Shortly after Thompson’s paper appeared, Matveev [19] gave an exposition of her algorithm using handle decompositions instead of triangulations; his paper gives a particularly elegant version of the *light-bulb* trick, replacing Thompson’s “fluorescent light bulb trick” [26, Section 3.1]. Next, Casson [4], in lectures at MSRI, California, introduced the idea of crushing triangulations along normal two-spheres. Another exposition of the Rubinstein-Thompson algorithm was given by Ivanov [13], stated in the language of *Q -triangulations*. Barchecheat’s thesis [3, Chapter 6.1] also gives an exposition of Casson’s algorithm.

Regarding the present work: the material in Section 8 to Section 11 had its genesis as Chapter 4 of my thesis [25], supervised by Andrew Casson; this material has not been otherwise published. Ian Agol, when we were both at the University of Illinois, Chicago, suggested that those techniques might bear on the computational complexity of three-sphere recognition. This paper, first posted to the arXiv in 2004, is the result. Another novelty, also introduced here, is to use the Agol-Hass-Thurston machinery [2] to produce a normalization algorithm that runs in polynomial time (Theorem 13.1).

In 2008, Sergei Ivanov [14] published another proof of Theorem 16.1. As in this paper, his argument closely follows Casson’s algorithm. Our polynomial-time normalization is, in his paper, replaced by a discussion

of *vertex fundamental surfaces* (following Hass-Lagarias-Pippinger [10]) together with the intriguing idea of crushing along almost normal two-spheres. His proof requires computation with three-cells more complicated than tetrahedra. It is an interesting and open question whether his more recent method is more efficient than ours. His paper also shows that the problem of recognizing manifolds with compressible boundary lies in **NP**.

Acknowledgements. I thank both Andrew Casson and Ian Agol many for enlightening mathematical conversations. I thank the mathematics department at UIC to its support during the writing of this paper. I thank the referees for their comments and corrections.

2. SKETCH OF THE PROOF OF THE MAIN THEOREM

We closely follow Casson's algorithm [4] for recognizing the three-sphere. Fix T , a triangulation of S^3 . Produce a certificate $\{(T_i, v(S_i))\}_{i=0}^n$ as follows: The triangulation T_0 is equal to T . For every i , Lemma 5.11 provides S_i , a normal two-sphere in T_i that is not vertex linking, if such exists. If T is zero-efficient then Lemma 5.11 provides S_i , an almost normal two-sphere in T_i . Definitions are given in Section 5.

If S_i is normal apply Theorem 14.1: T_{i+1} is obtained from T_i by crushing T_i along S_i . Briefly, we cut $|T_i|$ along S_i , cone the resulting two-sphere boundary components to points, and collapse non-tetrahedral cells of the resulting cell structure to obtain the triangulation T_{i+1} . This is discussed in Section 14, below.

If S_i is almost normal then obtain T_{i+1} from T_i by deleting the component of $|T_i|$ that contains S_i . Finally, the last triangulation T_n is empty, as is S_n .

That completes the construction of the certificate. We now turn to the procedure for checking a given certificate; we cite a series of polynomial-time algorithms that verify each part of the certificate. Begin by checking that T is a triangulation of a three-manifold that is a homology three-sphere, using Theorems 4.1 and 4.2. Next, check if $T = T_0$ using Lemma 4.4. For general i , verify that S_i is a two-sphere by checking its Euler characteristic (Lemma 5.4) and checking that it is connected (Theorem 5.5). Next, if S_i is normal verify that the triangulation T_{i+1} is identical to the triangulation obtained by crushing T_i along S_i . To do this in polynomial time requires uses Theorem 14.1 and again uses Lemma 4.4. If S_i is almost normal then check that the component T' of T_i containing S_i has $|T'| \cong S^3$ using Theorems 13.1 and 11.3.

Finally, by Theorem 14.2, for every i we have that $\#|T_i| \cong \#|T_{i+1}|$ where the connect sum on the left hand side ranges over the components of $|T_i|$ while the right hand side ranges over the components of $|T_{i+1}|$. By definition the empty connect sum is S^3 , and this finishes the verification of the certificate.

3. DEFINITIONS

Complexity theory. Please consult [8, 21] for more through treatments.

A *problem* P is a function from a set of finite binary strings, the *instances*, to another set of finite binary strings, the *answers*. A problem P is a *decision problem* if the range of P is the set $\{0, 1\}$. The length of a binary string in the domain of P is the *size* of the instance. A *solution* for P is a Turing machine \mathcal{M} that, given an instance T on its tape, computes and then halts with only the answer $P(T)$ on its tape. We will engage in the usual abuse of calling such a Turing machine an *algorithm* or a *procedure*.

An algorithm runs in *polynomial time* if there is a polynomial $q: \mathbb{R} \rightarrow \mathbb{R}$ so that the Turing machine \mathcal{M} , for every instance T , halts in time at most $q(\text{size}(T))$. Computing q , bounds on q or even the degree of q is a delicate question, sensitive to the exact model of computation. Thus say that a decision problem *lies in P* if it has a polynomial-time solution.

A decision problem P lies in **NP** if there is a polynomial q with the following property: For all instances T with $P(T) = 1$ there is a proof of length at most $q(\text{size}(T))$ that $P(T) = 1$. Such a polynomial-length proof is a *certificate* for T . More concretely: Suppose that there is a polynomial q' and a Turing machine \mathcal{M}' so that, for every instance T with $P(T) = 1$, there is a string C where \mathcal{M}' run on (T, C) outputs the desired proof that $P(T) = 1$ in time less than $q'(\text{size}(T))$. Then, again, the problem P is in **NP** and we also call C a certificate for T .

A decision problem lies in **PSPACE** if there is a polynomial q so that on every instance T and for every step of the computation the distance between the first and last non-blank squares of the tape of \mathcal{M} is at most $q(\text{size}(T))$.

A decision problem lies in **EXPTIME** if there is a polynomial q so that that the Turing machine \mathcal{M} halts in time at most $\exp(q(\text{size}(T)))$ on every instance T . Note that **P** \subset **NP** \subset **PSPACE** \subset **EXPTIME**. At least one of these inclusions is strict as **P** \neq **EXPTIME**.

Topological considerations. A *model tetrahedron* τ is a copy of the regular Euclidean tetrahedron of side length one with vertices labelled

by 0, 1, 2, and 3. See Figure 1 for a picture. Label the six edges by their vertices (0, 1), (0, 2), and so on. Label the four faces by the number of the vertex they do *not* contain. The standard orientation on \mathbb{R}^3 induces an orientation on the model tetrahedron which in turn induces orientations on the faces.

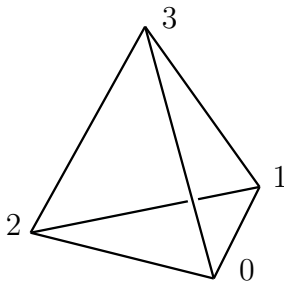


FIGURE 1. A regular Euclidean tetrahedron with all side-lengths equal to one.

A *labelled triangulation* of size n is a collection of n model tetrahedra $\{\tau_i\}_{i=1}^n$, each with a unique name, and a collection of *face pairings*.

Here a face pairing is a triple (i, j, σ) specifying a pair of tetrahedra τ_i and τ_j as well as an isometry σ from a face of τ_i to a face of τ_j . We will omit the labellings when they are clear from the context.

A triangulation is not required to be a simplicial complex. However every face must appear in exactly two face pairings or in none. We do not allow face pairings of the form (i, i, σ) .

Let $|T|$ be the *underlying topological space*; the space obtained by taking the disjoint union of the model tetrahedra and taking the quotient by the face pairings.

At this point we should fix an encoding scheme which translates triangulations into binary strings. However we will not bother to do more than remark that there are schemes which require about $n \log(n)$ bits to specify a triangulation with n tetrahedra. (This blow-up in length is due to the necessity of giving the tetrahedra unique names.) Thus it is only a slight departure from the truth to say that a triangulation T has size n when in fact its representation as a binary string is somewhat longer.

Recall that the *three-sphere* is the three-manifold

$$S^3 = \{x \in \mathbb{R}^4 \mid \|x\| = 1\}.$$

The connect sum $M \# N$ of two connected oriented three-manifolds M and N is obtained by removing an open three-ball from the interior

of each of M and N and gluing the resulting two-sphere boundary components with an orientation reverse homeomorphism. The connect sum naturally extends to a collection of connected, oriented three-manifolds; if M is the disjoint union of connected three-manifolds then $\#M$ denotes their connect sum.

Note that Alexander's Theorem [11, Theorem 1.1] implies that $M\#S^3$ is homeomorphic to M , for any three-manifold M . So we adopt the convention that the empty connect sum yields the three-sphere.

We now give a slightly non-standard definition of *compression body*. Take S a closed orientable surface. Let $C_0 = S \times [0, 1]$. Choose a disjoint collection of simple closed curves in some component of $S \times \{0\}$ and attach two-handles in the usual fashion along these curves. Cap off some (but not necessarily all) of any resulting two-sphere boundary components with three-handles. The final result, C , is a *compression body*. Set $\partial_+C = S \times \{1\}$ and set $\partial_-C = \partial C \setminus \partial_+C$. Our definition differs from others (for example [5]) in that two-sphere components in ∂_-C are allowed. The reasons for this are explained in Remark 10.3.

4. PRELIMINARIES

Here we give a few algorithms which take triangulations and check topological properties. See [10, 17, 20] for more in-depth discussions.

Theorem 4.1. *There is a polynomial-time algorithm that, given a triangulation T , decides whether or not $|T|$ is a three-manifold.*

Proof. The underlying space $|T|$ is a quotient of a disjoint union of model tetrahedra. Furthermore, the face pairings are isometries. It follows that $|T|$ is second-countable and Hausdorff.

It remains only to verify that every point $p \in |T|$ has a neighborhood homeomorphic to a three-ball. This is automatic if p lies in the interior of a tetrahedron. Since faces cannot be glued to themselves, any point in the interior of a face also has the desired neighborhood.

Now suppose that p lies in the interior of an edge. Take a sufficiently small ϵ -neighborhood about each preimage of p in the model tetrahedra. Each of these is a cone on a spherical lune. The neighborhoods fit together to form a cone on D^2 , S^2 , or \mathbb{RP}^2 . The latter may only happen at the midpoint of an edge.

Finally, suppose that p is a vertex. Again, take ϵ -neighborhoods about the preimages of p . Each is a cone on a spherical triangle. These fit together to form a cone on a surface. We will call this surface the *vertex link* of p . The point p has the desired three-ball neighborhood if and only if its vertex link is a sphere or a disk.

Thus the algorithm need only check how tetrahedra are glued around an edge and the topology of each vertex link. There are at most linearly many (in $\text{size}(T)$) edges and vertices. Checking each edge and each vertex link takes time at most polynomial (again, in $\text{size}(T)$). This is because there are at most $6 \cdot \text{size}(T)$ tetrahedra around any edge. Also, each vertex link is a union of at most $4 \cdot \text{size}(T)$ spherical triangles. \square

Recall that a three-manifold M is a *homology three-sphere* if it has the same homology groups as S^3 .

Theorem 4.2. *There is a polynomial-time algorithm that, given a triangulation T of a three-manifold, decides whether or not $|T|$ is a homology three-sphere.*

Proof. The homology groups $H_*(|T|, \mathbb{Z})$ may be read off from the Smith normal forms of the chain boundary maps: we refer the reader to [6, Section 2] for an accessible overview of algorithmic computation of homology. Finally, the Smith normal form of an integer matrix may be computed in polynomial time (see, [12]). \square

We also record for future use a few consequences of the homology three-sphere assumption:

Lemma 4.3. *If M^3 is a homology three-sphere then M is connected, closed, and orientable. Also every closed, embedded surface in M is orientable and separating. Finally, every connect summand of M is also a homology three-sphere.* \square

It follows that if N is a connect summand of a homology three-sphere and N is a lens space then $N \cong S^3$. We end this section with the simple:

Lemma 4.4. *There is a polynomial-time algorithm that, given triangulations T and T' , decides whether or not T is identical to T' .*

Proof. Recall that T and T' are labelled: all of the tetrahedra come equipped with names. To check for isomorphism simply check that every name appearing in T also appears in T' and that all of the face pairings in T and T' agree. \square

Remark 4.5. Note that the labelling is not needed to determine isomorphism of triangulations. This is because an isomorphism is determined by the image of a single tetrahedron.

5. NORMAL AND ALMOST NORMAL SURFACES

In order to study triangulations we first discuss Haken's theory of *normal surfaces*. See [10] for a detailed discussion, including references to the foundational work of Haken and Schubert. Other references on normal surfaces include [16, 3].

On a face f of the model tetrahedron τ there are three kinds of properly embedded arc with end points in distinct edges of f . These are called *normal arcs*. A simple close curve $\alpha \subset \partial\tau$ is a *normal curve* if α is transverse to the one-skeleton of τ and α is a union of normal arcs. The *length* of a normal curve α is the number of normal arcs it contains. A normal curve α is called *short* if it has length three or four.

Lemma 5.1. *Suppose that $\alpha \subset \partial\tau$ is a normal curve. The following are equivalent:*

- α misses some edge of τ^1 .
- α meets every edge of τ^1 at most once.
- α is short.

Proof. To see this, let $\{v_{(i,j)} \mid 0 \leq i < j \leq 3\}$ be the number of intersections of α with each of the six edges of τ . There are twelve inequalities $v_{(0,1)} \leq v_{(1,2)} + v_{(0,2)}$, and so on, as well as six equalities $v_{(0,1)} + v_{(1,2)} + v_{(0,2)} = 0 \pmod 2$, and so on. An easy calculation gives the desired result. \square

In a model tetrahedron there are seven types of *normal disk*, corresponding to the seven distinct short normal curves in $\partial\tau$. See Figure 2. These are the four *normal triangles* and three *normal quads*. We have triangles of type 0, 1, 2, or 3 depending on which vertex they cut off of the model tetrahedron, τ . We have quads of type 0, 1, or 2 depending on which vertex is grouped with 3 when we cut τ along the quad.

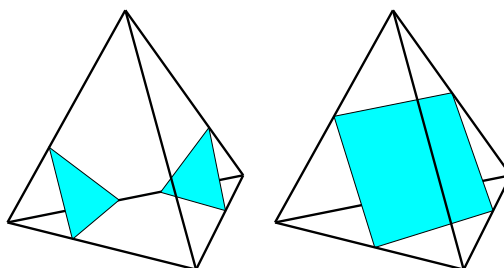


FIGURE 2. Two of the four triangles and one of the three quads.

Definition 5.2. A surface S properly embedded in $|T|$ is *normal* if $S \cap \tau$ is a collection of normal disks for every tetrahedron $\tau \in T$.

There is also the almost normal octagon and almost normal annulus, defined by Rubinstein [22]. See Figure 3 for examples. An octagon is a disk in the model tetrahedron bounded by a normal curve of length eight. An annulus is obtained by taking two disjoint normal disks and tubing them together along an arc parallel to an edge of the model tetrahedron. A surface S properly embedded in $|T|$ is *almost normal* if $S \cap \tau$ is a collection of normal disks, for every tetrahedron $\tau \in T$, except one. In the exceptional tetrahedron there is a collection of normal disks and exactly one almost normal piece.

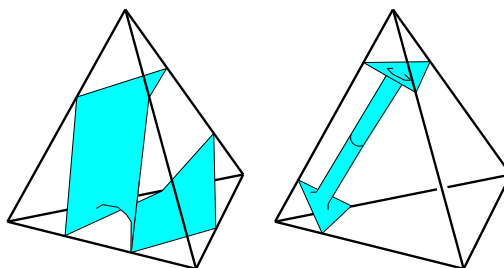


FIGURE 3. One of the three octagons and one of the 25 annuli.

5.1. Weight and Euler characteristic. For an either normal or almost normal surface S take the *weight* of S , $\text{weight}(S) = |S \cap T^1|$, to be the number of intersections between S and the one-skeleton T^1 . Note that $\text{size}(S)$, the number of bits required to describe S , is about $\text{size}(T) \log(\text{weight}(S))$. To see this, record a normal surface S as a *surface vector* $v(S) \in \mathbb{Z}^{7 \cdot \text{size}(T)}$ where the first $4 \cdot \text{size}(T)$ coordinates describe the number of normal triangles of each type while the last $3 \cdot \text{size}(T)$ coordinates describe the number of normal quads of each type. At least two-thirds of these last $3 \cdot \text{size}(T)$ coordinates are zero as an embedded surface has only one kind of normal quad in each tetrahedron.

For an almost normal surface S we again record the vector $v(S)$ of numbers of normal disks, as well as the type of the almost normal piece and the name of the tetrahedron containing it. (There is a small issue when the almost normal piece is an annulus obtained by tubing a pair of normal disks of the same type. Then $v(S)$ has length $7 \cdot \text{size}(T) + 1$ as one parallel collection of normal disks is interrupted by the almost normal piece.)

Definition 5.3. An isotopy $H : |T| \times I \rightarrow |T|$ is a *normal isotopy* if, for all $s \in I$ and for every simplex σ in T , $H_s(\sigma) = \sigma$.

Two subsets of $|T|$ are *normally isotopic* if there is a normal isotopy taking one to the other. Now, if two normal (or almost normal) surfaces S and S' have the same vector then S is normally isotopic to S' .

We now have a few results concerning normal and almost normal surfaces. We assume throughout that the triangulation T has underlying space a three-manifold.

Lemma 5.4. *There is a polynomial-time algorithm that, given a triangulation T and a normal or almost normal surface vector $v(S)$, computes the weight of S and the Euler characteristic of S .*

Proof. To find the weight of S on a single edge e of T^1 count the number of normal disks meeting e (with multiplicity depending on how many times the containing tetrahedron meets e) and divide by the valency of e in T^2 , the two-skeleton.

For the Euler characteristic simply use the formula $\chi(S) = F - E + V$ and the cell structure on S coming from its being a normal surface. (If S contains an almost normal annulus then we must add a single edge running between the two boundary components of the annulus.) Counting the number of faces and edges is straight-forward. The number of vertices equals the weight.

See [2], the end of Section 5, for a more detailed discussion. □

Theorem 5.5 (Agol, Hass Thurston [2]). *There is a polynomial-time algorithm that, given a triangulation T and a normal or almost normal surface vector $v(S)$, produces the surface vectors for the connected components of S .*

A caveat is required here – if several normally isotopic copies of F appear in S then the algorithm of Theorem 5.5 produces $v(F)$ only once and also reports the number of copies. This is required if the algorithm is to have output of polynomial size on input of the form $n \cdot v(F)$.

Proof of Theorem 5.5. This is one application of the “extended counting algorithm” given in [2]. See the proof of Corollary 17 of that paper. □

5.2. Vertex linking. Fix a triangulation T of some three-manifold. Suppose $x \in |T|$ is a vertex of the triangulation. Let S be the frontier of a small regular neighborhood of x . Then S is a connected normal surface which contains no normal quads. Such a surface is called *vertex*

linking or simply a *vertex link*. A normal sphere that contains a normal quad is called *non-trivial*. If the triangulation contains no non-trivial normal two-spheres then T is *zero-efficient*.

5.3. Haken sum. Suppose S, F, G are three non-empty normal surfaces with $v(S) = v(F) + v(G)$. Then we say that F and G are *compatible*: in every tetrahedron where both F or G have quads, these quads are of the same type. After a normal isotopy of F and G we find that S is the *Haken sum* of F and G ; there is a cut-and-paste construction of F and G building S . It follows that $\chi(S) = \chi(F) + \chi(G)$.

Likewise, suppose S and F are almost normal with identical almost normal piece, G is normal, and the normal coordinates add: $v(S) = v(F) + v(G)$. Again we say that S is a Haken sum. If F contains an octagon then we may normally isotope G so that no normal disk of G meets the octagon. If F contains an annulus we may normally isotope G so that no disk of G meets the annulus in a meridian of the tube. In either case we may perform cut-and-paste and find $\chi(S) = \chi(F) + \chi(G)$.

When S is a Haken sum as above we write $S = F + G$. If S is not a Haken sum then S is *fundamental*. As an exercise:

Lemma 5.6. *If $S = F + G$, where G is a vertex link, then S is not connected.* \square

Also,

Lemma 5.7. *If $S \subset |T|$ is a fundamental normal or almost normal surface then the largest entry of $v(S)$ is at most $\exp(\text{size}(T))$.*

Proof. There is a constant c (not depending on T or S) such that the largest entry is less than $2^{c \cdot \text{size}(T)}$. This lemma is proved for normal surfaces in [10, Lemma 6.1]. In the almost normal case, when the almost normal piece is an octagon, we obtain similar bounds using the system of linear equations provided by [26, Section 5]. The case of the annulus is similar. \square

Lemma 5.8. *Suppose T is a triangulation of a homology three-sphere. Suppose T contains a non-trivial normal two-sphere. Then T contains a non-trivial normal two-sphere which is fundamental.*

Proof. This is due to Haken and Schubert; see Proposition 5.7 of [16] for a modern account. The essential points are that Euler characteristic is additive under Haken sum, that T does not contain any normal $\mathbb{R}\mathbb{P}^2$ or \mathbb{D}^2 (by Lemma 4.3), and that no summand is vertex-linking (by Lemma 5.6). \square

Lemma 5.9. *Suppose T is a zero-efficient triangulation of a homology three-sphere. Suppose T contains an almost normal two-sphere. Then T contains a fundamental almost normal two-sphere.*

Proof. This is identical to the proof of Lemma 5.8, except that S cannot have a normal two-sphere summand as T is zero-efficient. \square

Of a much different level of difficulty is Thompson’s Theorem:

Theorem 5.10 (Thompson [26]). *Suppose $|T| \cong S^3$. Suppose also that T is zero-efficient. Then T contains an almost normal two-sphere.* \square

We end this section with:

Lemma 5.11. *There is an exponential-time algorithm that, given a triangulation T of a three-manifold, produces either the surface vector of a fundamental non-trivial normal two-sphere or, if none exists, produces the surface vector of a fundamental almost normal two-sphere or, if neither exists, reports “ $|T|$ is not homeomorphic to the three-sphere”.*

Proof. We only sketch a proof – the interested reader should consult [10], [16, page 66] or [3, page 83]. If T is not zero-efficient there is a fundamental normal two-sphere. This can be found by enumerating all fundamental surfaces (a finite list, by work of Haken) and checking each surface on the list. If T is zero-efficient then no non-trivial normal two-sphere appears. However we again have that some fundamental almost normal two-sphere exists, if T contains any almost normal two-sphere. Finally, if no non-trivial normal sphere nor any almost normal sphere appears in amongst the fundamentals then, combining Lemma 5.9 and Theorem 5.10, conclude that $|T|$ is not the three-sphere.

As presented the running time of the algorithm is unclear; it depends on the number of fundamental surfaces. However, using *vertex fundamental* surfaces and linear programming techniques Casson [4] reduces the search to take time at most a polynomial times $3^{\text{size}(T)}$. \square

6. BLOCKED SUBMANIFOLDS

Normal (and almost normal) surfaces cut a triangulated manifold into pieces. These submanifolds have natural polyhedral structures which we now investigate.

Let τ be a model tetrahedron, and suppose that $S \subset \tau$ is an embedded collection of normal disks and at most one almost normal piece. Let B be the closure of any component of $\tau \setminus S$. We call B a *block*. See Figure 4.

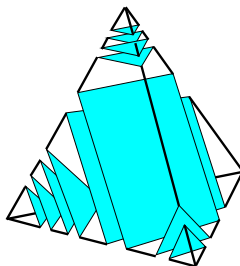


FIGURE 4. The tetrahedron τ cut along S . Note that there are two blocks of the form “normal disk cross interval”.

A block meeting exactly two normal disks of the same type is called a *product block*. All other blocks are called *core blocks*. Note that there are only seven kinds of product block possible, corresponding to the seven types of normal disks. Likewise there is a bounded number of core blocks. Five such are shown in Figure 4, but many more are possible. Most of these meet an almost normal annulus.

The components of ∂B meeting S are the *horizontal* boundary components of B , denoted $\partial_h B$. All other faces of B (the faces of B which lie in the two-skeleton) are $\partial_v B$, the *vertical* boundary.

Suppose now that T is a triangulation of a three-manifold and $S \subset |T|$ is a normal or almost normal surface. For simplicity, suppose that S is a transversely oriented and separating. Let N_S be the closure of the component of $|T| \setminus S$ into which the transverse orientation points.

Then N_S is a *blocked submanifold* of $|T|$. Note that the induced cell structure on N_S (coming from T) naturally breaks into two parts. So, let \widehat{N}_P be the union of all product blocks in N_S and let \widehat{N}_C be the union of all core blocks in N_S .

Remark 6.1. In any blocked submanifold there are at most a linear number (in $\text{size}(T)$) of core blocks. In fact there are at most six in each tetrahedron (plus possibly two more coming from the almost normal annulus). This simple observation underlies Kneser-Haken finiteness (see, for example, [9]) and is generally useful in the algorithmic setting.

Note that \widehat{N}_P and \widehat{N}_C are not necessarily submanifolds of $|T|$. To produce submanifolds take N_P to be a regular closed neighborhood of \widehat{N}_P , where the neighborhood is taken inside of N_S . Also, take N_C to be the closure of $N_S \setminus N_P$. Note the asymmetry between the definitions of N_P and N_C : we have $\widehat{N}_P \subset N_P$ while $N_C \subset \widehat{N}_C$. As above define

$\partial_h N_P = N_P \cap S$ and $\partial_v N_P = \overline{\partial N_P \setminus \partial_h N_P}$. The horizontal and vertical boundaries $\partial_h N_C$ and $\partial_v N_C$ are defined similarly.

Note also that \widehat{N}_P may be specified by describing the core blocks and by a *block vector*; a vector $v(\widehat{N}_P) \in \mathbb{Z}^{7 \cdot \text{size}(T)}$ where the first $4 \cdot \text{size}(T)$ coordinates describe the number of triangle product blocks of each type while the last $3 \cdot \text{size}(T)$ coordinates describe the number of quad product blocks of each type. (As discussed in Section 5.1, when the almost normal piece is an annulus obtained by tubing a pair of normal disks of the same type, the vector $v(\widehat{N}_P)$ has an additional coordinate.)

We now have:

Theorem 6.2. *There is a polynomial-time algorithm that, given a triangulation T and a surface vector $v(S)$ for a transversely oriented S , produces the block vector for each connected component of $\widehat{N}_P \subset N_S$.*

Proof. When S is not separating we add a parallel copy S' pushed off in the direction opposite the transverse orientation. Letting $S \times [0, 1]$ denote the product cobounded by $S \cup S'$ we may take N_S equal to the closure of $|T| \setminus S \times [0, 1]$. Also, S' is given a transverse orientation pointing into N_S . In what follows we now we assume that S is separating.

Now, every normal disk meets an edge $(0i)$ of the containing model tetrahedron, minimizing i . The transverse orientation on S is represented as a bit vector $(\epsilon_k) \in \{0, 1\}^{7 \cdot \text{size}(T)}$. If $\epsilon_k = 0$ then the disk (of the type k in S) closest to the vertex 0 has transverse orientation pointing away from 0. When $\epsilon_k = 1$ the orientation points towards the vertex 0. (The transverse orientation of the m^{th} disk of type k now depends only on the parity of m . Also, verifying that (ϵ_k) gives a consistent transverse orientation on the whole of S is simply a polynomial number of parity calculations.)

Recall that the transverse orientation points into N_S . Recall that $v(\widehat{N}_P)$ is the block vector. If $v_k(S)$ is odd then $v_k(\widehat{N}_P) = (v_k(S) - 1)/2$. If $v_k(S)$ is even then $v_k(\widehat{N}_P) = (v_k(S) - 2\epsilon_k)/2$. This completes the computation of the block vector $v(\widehat{N}_P)$. The remainder of the proof of Theorem 6.2 follows from the techniques in [2] and the fact that there are only a linear number of core blocks. \square

Remark 6.3. We also remark that, if S is connected, \widehat{N}_P has at most a linear number (in $\text{size}(T)$) of connected components. This is because $\partial_v N_P = \partial_v N_C$ and the latter has at most linearly many components. (See Remark 6.1.) This is a pleasant contrast with Theorem 5.5.

7. NORMALIZING SLOWLY

In this section we discuss a restricted version of Haken's *normalization* procedure for producing normal surfaces. This material appeared first in an unpublished preprint of mine and later in my thesis [25]. I thank Danny Calegari for reading an early version of this work. I also thank Bus Jaco for several enlightening conversations regarding Corollary 9.3.

Several authors have independently produced versions of these ideas; for example see [18, 16, 3]

Let $S \subset |T|$ be a transversely oriented, almost normal surface. Here T is a triangulation of a closed, orientable, connected three-manifold.

Definition 7.1. A compression body $C_S \subset |T|$ is *associated* to S if $\partial_+ C_S = S$, $\partial_- C_S$ is normal, the transverse orientation points into C_S , and any normal surface S' disjoint from S may be normally isotoped to one disjoint from C_S .

As a bit of notation take $\partial_- C_S = \tilde{S}$ and call this the *normalization* of S .

Theorem 11.1. *Given a transversely oriented almost normal surface S there exists an associated compression body C_S and C_S is unique (up to normal isotopy). Furthermore there is an algorithm that, given the triangulation T and the surface vector $v(S)$, computes the surface vector of $\partial_- C_S = \tilde{S}$.*

Remark 7.2. As in Theorem 6.2, when S is not separating we add a parallel copy S' and transversely orient away from the parallel region between S and S' . Henceforth, we will assume that S is in fact separating.

The proof of this theorem is lengthy and is accordingly spread from Section 8 to Section 11. We here give the necessary definitions. In Section 8 we discuss the tightening procedure. In Section 9 we show that the tightening procedure gives an embedded isotopy. We discuss the capping off procedure in Section 10. Finally in Section 11 we prove Theorem 11.1.

7.1. Non-normal surfaces. Let S be a surface properly embedded in a triangulated three-manifold $|T|$ and suppose that S is transverse to the skeleta of T . Denote the i -skeleton of T by T^i . Generalize the notion of weight so that $\text{weight}(S) = |S \cap T^1|$.

We characterize some of the ways S can fail to be normal. A *simple curve* of S is a simple closed curve of intersection between S and the

interior of some triangular face $f \in T^2$. Also, a *bent arc* of S is a properly embedded arc of intersection between S and the interior of some triangular face $f \in T^2$ with both endpoints of the arc contained in a single edge of f . Both of these are drawn in Figure 5.

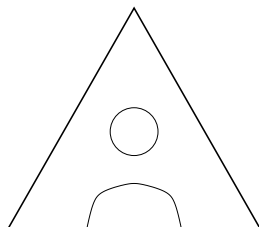


FIGURE 5. A simple curve and a bent arc.

7.2. Surgery and tightening disks.

Definition 7.3. An embedded disk $D \subset |T|$ is a *surgery disk* for S if $D \cap S = \partial D$, $D \subset T^2$ or $D \cap T^2 = \emptyset$, and $D \cap T^1 = \emptyset$.

There is a *surgery* of S along D : Remove a small neighborhood of ∂D from S and cap off the boundaries thus created with disjoint, parallel copies of D . Note that we do not require ∂D to be essential in S . A simple curve of $S \cap T^2$ is *innermost* if it is the boundary of a surgery disk embedded in a triangle of T^2 .

Definition 7.4. An embedded disk $D \subset |T|$ is a *tightening disk* for S if $\partial D = \alpha \cup \beta$ where $\alpha \cap \beta = \partial\alpha = \partial\beta$ and $D \cap S = \alpha$. Also, either $D \subset T^2$ or $D \cap T^2 = \beta$. But in either case $D \cap T^1 = \beta$, and $D \cap T^0 = \emptyset$.

There is a *tightening isotopy* of S across D : Push α along the disk D , via ambient isotopy of S supported in a small neighborhood of D , until α moves past β . This procedure reduces $\text{weight}(S)$ by exactly two. A bent arc of S is *outermost* if it lies on the boundary of a tightening disk embedded in a triangle of T^2 .

Suppose S contains an almost normal octagon, $A \subset \tau$. There are two tightening disks on opposite sides of A both giving tightening isotopies of S to a possibly non-normal surface of lesser weight. To see these disks, consult the left-hand side of Figure 3. Notice there are two edges of τ , say e and e' , so that $|A \cap e| = |A \cap e'| = 2$. The first tightening disk has boundary running along e between the points of intersection with A ; then the boundary runs along an arc in the interior of A . The second disk similarly meets e' . We arrange matters so that the tightening disks, in the interior of τ , meet each other in a single point.

The above disks are the *exceptional tightening disks* associated to A . If S contains an almost normal annulus then the tube is parallel to at least one edge of the containing tetrahedron. (See the right-hand side of Figure 3.) For every such edge there is an exceptional tightening disk. Also, the disk which surgers the almost normal annulus will be called the *exceptional surgery disk*.

8. TIGHTENING

This section discusses the *tightening* procedure which will yield an embedded isotopy. This is proved in Lemma 9.1 below.

Suppose that $S \subset |T|$ is a transversely orientable separating almost normal surface. Here T is a triangulation of a three-manifold. We wish to isotope S off of itself while steadily reducing the weight of S .

Suppose that D is an exceptional tightening disk for S . Choose the transverse orientation for S which points into the component of $|T| \setminus S$ which meets D . The *F-tightening procedure* constructs a map $\mathcal{F}: S \times [0, n] \rightarrow |T|$ as follows:

1. Let $F_0 = S$. Take $\mathcal{F}_0: S \times \{0\} \rightarrow |T|$ to be projection to the first factor. Let $D_0 = D$.
2. Now do a small normal isotopy of F_0 in the transverse direction while tightening F_0 along D_0 . This extends \mathcal{F}_0 to a map $\mathcal{F}_1: S \times [0, 1] \rightarrow |T|$, with $F_t = \mathcal{F}_1(S \times \{t\})$. Note that the surface F_1 inherits a transverse orientation from F_0 . Arrange matters so that $F_{\frac{1}{2}}$ is the only level which is not transverse to T^2 . Furthermore $F_{\frac{1}{2}}$ only has a single tangency with T^1 and this tangency occurs in the middle of $\partial D_0 \cap T^1$.
3. At stage $k \geq 1$ there are two possibilities. Suppose first that F_k has an outermost bent arc α with the transverse orientation of F_k pointing into the tightening disk D_k , which is cut out of T^2 by α . Then extend \mathcal{F}_k to $\mathcal{F}_{k+1}: S \times [0, k+1] \rightarrow |T|$ by doing a small normal isotopy of F_k in the transverse direction while tightening F_k across D_k , the k^{th} tightening disk. So $\mathcal{F}_k = \mathcal{F}_{k+1}|_{S \times [0, k]}$ and $F_t = \mathcal{F}_{k+1}(S \times \{t\})$. Note that the surface F_{k+1} inherits a transverse orientation from F_k . Arrange matters so that $F_{k+\frac{1}{2}}$ is the $k+1^{\text{th}}$ level which is not transverse to T^2 . Furthermore $F_{k+\frac{1}{2}}$ only has a single tangency with T^1 and this tangency occurs in the middle of $\partial D_k \cap T^1$.

If there is no outermost bent arc $\alpha \subset F_k$ then set $n = k$ and the procedure halts.

Remark 8.1. As $\text{weight}(F_{k+1}) = \text{weight}(F_k) - 2$ this process terminates. Note also that \mathcal{F}_n is far from unique – at any stage in the procedure there may be many tightening disks to choose from.

We will show in Lemma 9.1 that the map $\mathcal{F}_n: S \times [0, n] \rightarrow M$ is an embedding. Note that, by construction, $S = F_0 = \mathcal{F}_n(S \times \{0\})$ and in general $F_t = \mathcal{F}_n(S \times \{t\})$. To simplify notation set $\mathcal{F} = \mathcal{F}_n$.

9. TRACKING THE ISOTOPY

In this section we analyze how $\text{image}(\mathcal{F})$ intersects the skeleta of the triangulation. Let $S \subset |T|$, \mathcal{F} , \mathcal{F}_k , and F_t be as defined in Section 8.

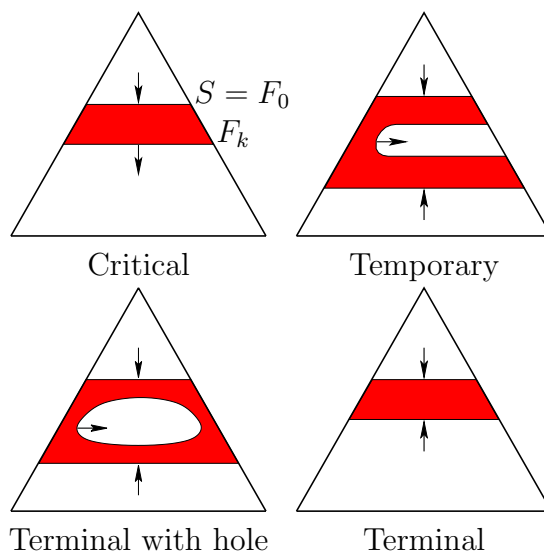


FIGURE 6. The Rectangles

Figures 6 and 7 display a few of the possible components of intersection $f \cap \text{image}(\mathcal{F}_k)$ assuming that \mathcal{F}_k is an embedding. Here f is a face of T^2 . Lemma 9.1 below shows that this collection is complete up to symmetry. Note that the normal arcs, bent arcs, and simple curves bounding the components receive a transverse orientation from S or F_k . In these figures the transverse orientation on S points into the shaded region while the transverse orientation on F_k points away. The components of intersection containing a normal arc of F_k are called *critical*. Those with a bent arc of F_k are called *temporary*. Any component containing a simple curve of F_k is called *terminal with hole*. Finally, components of $f \cap \text{image}(\mathcal{F}_k)$ which are completely disjoint from F_k are simply called *terminal*. Again, refer to Figures 6 and 7.

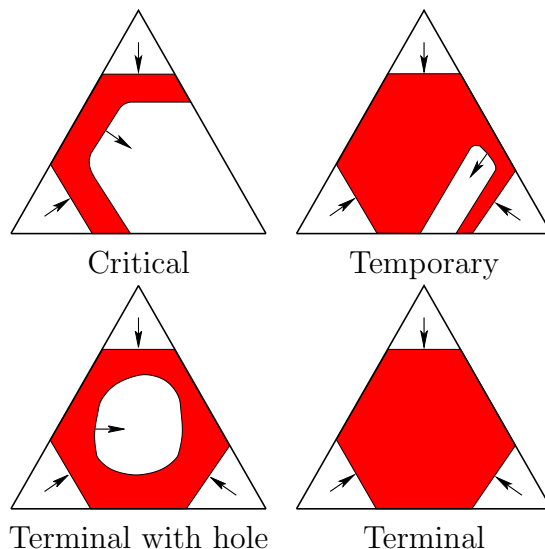


FIGURE 7. The Hexagons

The tightening procedure combines the critical components in various ways. However, a temporary component always results in a terminal (possibly with hole) and these are stable. Note also that there is a second critical rectangle with the opposite transverse orientation. The non-critical components may be foliated by the levels of \mathcal{F}_k in multiple ways, depending on the order of the tightening isotopies.

Lemma 9.1. *For every k , the map \mathcal{F}_k is an embedding. Furthermore, for $k > 0$ and for every face $f \in T^2$, the connected components of $f \cap \text{image}(\mathcal{F}_k)$ are given, up to symmetry, by Figures 6 and 7.*

Proof. Proceed by induction: Both claims are trivial for $k = 0$. Now to deal with $k = 1$. The exceptional tightening disk D_0 has interior disjoint from $S = F_0$. It follows that \mathcal{F}_1 is an embedding. To verify the second claim for $k = 1$ note that the image of $\mathcal{F}_1|S \times [0, \epsilon]$ intersects all faces $f \in T^2$ only in critical rectangles. Up to $t = \frac{1}{2}$ the image of $\mathcal{F}_1|S \times [0, t]$ intersected with f is combinatorially constant. Crossing $t = \frac{1}{2}$ adds a regular neighborhood of D_0 to the image. This only intersects f in a regular neighborhood of $\partial D_0 \cap T^1$. So the pieces of $f \cap \text{image}(\mathcal{F}_1)$ are unions of critical rectangles connected by small neighborhoods of sub-arcs of T^1 . Also these sub-arcs only meet the F_t side of the critical rectangles. As each critical rectangle meets two edges of the face f it follows that at most three critical rectangles are joined together to form a component of $f \cap \text{image}(\mathcal{F}_1)$. We list all possible cases – consulting Figures 6 and 7 will be helpful:

1. Two critical rectangles in f may be combined to produce a temporary rectangle, a terminal rectangle with a hole, or a critical hexagon.
2. Three critical rectangles in f may be combined to produce a temporary hexagon or a terminal hexagon with a hole.

Now to deal with the general case: Suppose that both hypotheses hold at stage k . Suppose that $\alpha \subset F_k$ is the bent arc on the boundary of $D_k \subset f \in T^2$, the next tightening disk in the sequence. Suppose that $\text{interior}(D_k)$ meets $\text{image}(\mathcal{F}_k)$. By the second induction hypothesis there is a component, C , of $f \cap \text{image}(\mathcal{F}_k)$ which meets $\text{interior}(D_k)$ and appears among those listed in Figures 6 and 7. Observe that each component of $f \cap \text{image}(\mathcal{F}_k)$, and hence C , meets at least two edges of f . The bent arc α meets only one edge of f . It follows that the interior of C must meet α . Thus \mathcal{F}_k was not an embedding, a contradiction.

It follows that $D_k \cap \text{image}(\mathcal{F}_k) = \alpha$. Since the $k + 1^{\text{th}}$ stage of the isotopy is supported in a small neighborhood of $F_k \cup D_k$ it follows that \mathcal{F}_{k+1} is an embedding.

Now, the transverse orientation on F_k gives rise to a transverse orientation on F_{k+1} . To verify the second hypothesis again list the possible cases:

1. Two critical rectangles in f may be combined to produce a temporary rectangle, a terminal rectangle with a hole, or a critical hexagon.
2. Three critical rectangles f may be combined to produce a temporary hexagon or a terminal hexagon with a hole.
3. A critical rectangle and critical hexagon in f may be combined to produce a temporary hexagon or a terminal hexagon with a hole.
4. A temporary component can lead only to a terminal (possibly with hole).

This completes the induction. □

Remark 9.2. By maximality of \mathcal{F} , the surface $F_n = \mathcal{F}(S \times \{n\})$ has no outermost bent arcs with outward orientation. A bent arc with inward orientation would violate the second induction hypothesis of Lemma 9.1. So F_n contains no bent arcs. F_n may contain simple curves, but the second induction hypotheses shows that all of these are innermost with transverse orientation pointing toward the bounded surgery disk.

Given that \mathcal{F} is an embedding, in the sequel $\text{image}(\mathcal{F}_k)$ is denoted by \mathcal{F}_k . Replacing S in Lemma 9.1 by a disjoint union of S with a collection of normal surfaces gives:

Corollary 9.3. *If S' is any normal surface in $|T|$ which does not intersect S then $\mathcal{F} \cap S' = \emptyset$, perhaps after a normal isotopy of S' (rel S).*

Let τ be any tetrahedron in the given triangulation T .

Lemma 9.4. *For all $k \geq 1$, $\tau \setminus \mathcal{F}_k$ is a disjoint collection of balls.*

Proof. Again we use induction. Our induction hypothesis is as follows: $\tau \setminus \mathcal{F}_k$ is a disjoint collection of balls, unless $k = 0$, and τ contains the almost normal annulus of S . (In this situation $\tau \setminus \mathcal{F}_0$ is a disjoint collection of balls and one solid torus $\mathbb{D}^2 \times S^1$.)

The base case is trivial. Suppose B is a component of $\tau \setminus \mathcal{F}_k$. There are now two cases to consider. Either B is cut by an exceptional tightening disk or it is not. Assume the latter. Then B is a three-ball by induction and after the $k+1^{\text{th}}$ stage of the isotopy $B \cap \mathcal{F}_{k+1}$ is a regular neighborhood (in B) of a collection of disjoint arcs and disks in ∂B . Hence $B \setminus \mathcal{F}_{k+1}$ is still a ball.

If B is adjacent to the almost normal piece of F_0 then let D_0 be the exceptional tightening disk. Set $B_\epsilon = B \setminus \text{neigh}(D_0)$. Each component of B_ϵ is a ball, and the argument of the above paragraph shows that they persist in the complement of \mathcal{F}_1 . \square

A similar induction argument proves:

Lemma 9.5. *For all $k \geq 1$, $\tau \cap \mathcal{F}_k$ is a disjoint collection of handlebodies.* \square

This lemma is not used in what follows and its proof is accordingly left to the interested reader. Recall that $\partial \mathcal{F}_k = S \cup F_k$. A trivial corollary of Lemma 9.4 is:

Corollary 9.6. *For all k , the connected components of $\tau \cap F_k$ are planar.*

The connected components of $\tau \cap F_n$ warrant closer attention:

Lemma 9.7. *Each component of $\tau \cap F_n$ has at most one normal curve boundary component. This normal curve must be short.*

Proof. Let $\tau \in T$ be a tetrahedron. Let P be a connected component of $\tau \cap F_n$. By Lemma 9.1 the boundary ∂P is a collection of simple curves and normal curves in $\partial \tau$. Let α be any normal curve in ∂P . Let $\{\alpha_j\}$ be the normal arcs of α .

Claim. α has length three or four.

Proof of Claim. Call the collection of critical rectangles and hexagons (in $\partial\tau \cap \mathcal{F}$) meeting α the *support* of α . To prove the claim we have two cases. First suppose that only critical rectangles support α . So α is normally isotopic to a normal curve $\beta \subset \partial\tau \cap S$. The first step of the tightening procedure prevents β from being a boundary of the almost normal piece of S . It follows that α must be short.

Otherwise α_1 , say, is on the boundary of a critical hexagon $h \subset f$. Let β be a normal curve of S meeting h and let $\beta_1 \subset \beta$ be one of the normal arcs in ∂h . Let e be the edge of f which α_1 does not meet. This edge is partitioned into three pieces; $e_h \subset h$, e' , and e'' . We may assume that β_1 separates e_h from e' . See Figure 8.

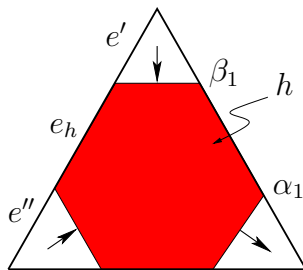


FIGURE 8. The normal arcs α_1 and β_1 are on the boundary of the critical hexagon h . Note that β does not meet e' or $\text{interior}(h)$.

Note that a normal curve of length ≤ 8 has no parallel normal arcs in a single face. Thus β meets e' exactly once at an endpoint of e' . Since α and β do not cross it follows that β separates α from e' in $\partial\tau$.

Similarly, α is separated from e'' . Thus α does not meet e at all. By Lemma 5.1 the normal curve α is short. This finishes the proof of the claim. \square

Claim. The component $P \subset \tau \cap F_n$ has at most one boundary component which is a normal curve.

Proving this will complete the lemma. So suppose that ∂P contains two normal curves: α and β . Let A be the annulus cobounded by α and β in $\partial\tau$, the boundary of the model tetrahedron.

Suppose now that the transverse orientation F_n induces on α points away from A . Thus A and the support of α intersect. There are several cases to examine, depending on the length of α and the components of the support of α .

1. Suppose α has length three:

- (a) If only critical rectangles support α then a normal triangle of S separates α and β .
 - (b) If one critical hexagon and two critical rectangles support α then the almost normal octagon and the exceptional tightening disk together separate α and β . (See left hand side of Figure 9.)
 - (c) If two critical hexagons and one critical rectangle support α then either a normal triangle or normal quad of S separates α and β . (See right hand side of Figure 9.)
 - (d) If only critical hexagons support α then a normal triangle of S separates α and β .
2. Suppose α has length four:
- (a) If only critical rectangles support α then a normal quad of S separates α and β .
 - (b) If one critical hexagon and three critical rectangles support α then S could not have been an almost normal surface. (See left hand side of Figure 10.)
 - (c) If two critical hexagons and two critical rectangles support α then a normal triangle of S separates α and β . (See right hand side of Figure 10.)

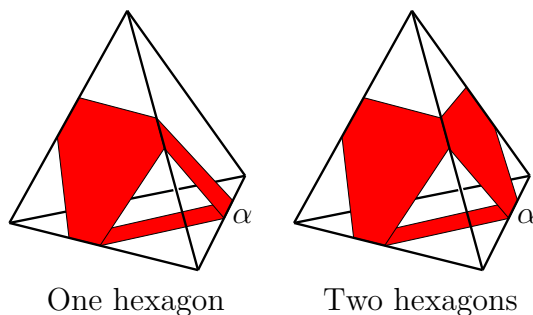


FIGURE 9. Diagrams for cases (1b) and (1c).

When α has length four it cannot be supported by more than two critical hexagons.

To recap: in all cases except 1(b) and 2(b), the support of α (possibly together with a terminal rectangle or hexagon) closes up, implying the existence of a normal disk of S with boundary a core curve of the annulus A . As this disk lies in S observe that $S \cap P \neq \emptyset$ and thus $S \cap F_n \neq \emptyset$. This contradicts the fact that \mathcal{F} is an embedding (Lemma 9.1). Case 1(b) is similar, except that the support of α meets other critical or terminal components to form the octagon piece of S .

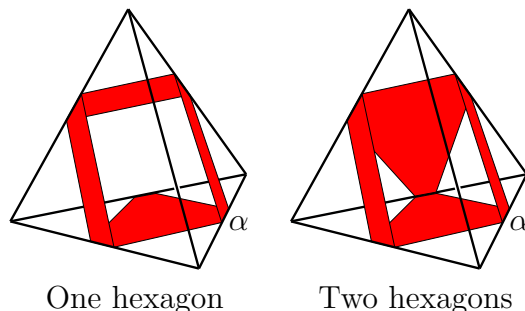


FIGURE 10. Diagrams for cases (2b) and (2c).

So P must intersect either S or the exceptional tightening disk, again a contradiction of Lemma 9.1. Lastly, in case 2(b), S could not have been almost normal.

So deduce that the transverse orientation which F_n gives α must point toward A . Thus A and the support of α are disjoint. Let γ be an arc which runs along P from α to β . Let α' be a push-off of α along A , towards β . This push-off bounds a disk in one of the components of $\tau \setminus \mathcal{F}$, by Lemma 9.4. This disk does not intersect $P \subset F_n \subset \mathcal{F}$ and hence fails to intersect γ . This is a contradiction. \square

Remark 9.8. By Lemma 9.1 all simple curves of F_i are innermost. It follows that the “tubes” analyzed in Lemma 9.7 do not run through each other. In addition, analysis similar to that needed for Lemma 9.5 implies that these tubes are unknotted, but this last fact is not needed in the sequel.

10. CAPPING OFF

Here we construct our candidate for C_S , the compression body associated to S .

Let $\mathcal{F} \subset |T|$ be the image of the map constructed above. Recall that $\partial\mathcal{F} = S \cup F_n$ where S is the almost normal surface we started with and F_n is the surface obtained by “tightening” S . Note that, since \mathcal{F} is the embedded image of $S \times [0, n]$, in fact F_n is isotopic to S in $|T|$. (It cannot be normally isotopic as it has lower weight.)

Definition 10.1. A two-sphere which is embedded in $|T|$ but disjoint from T^2 is called a *bubble*.

From Lemma 9.7, Corollary 9.6, and Remark 9.8 we deduce:

Corollary 10.2. *Let F'_n be the surface obtained by surgering all simple curves of F_n . Then F'_n is a disjoint collection of bubbles and normal*

surfaces. Each bubble bounds a ball with interior disjoint from $T^2 \cap F'_n$. \square

Construct C_S as follows: For every simple curve α of F_n attach a two-handle to \mathcal{F} along α . Attach so that the core of the two-handle is the subdisk of T^2 cut out by α . Call this \mathcal{F}' . As noted in Remark 9.8 all simple curves of F_n are innermost. So \mathcal{F}' is an embedded compression body. At this point there may be components of $\partial_- \mathcal{F}'$ which are not normal. By Corollary 10.2 all of these are bubbles bounding a ball disjoint from all of the other bubbles. So cap off each bubble to obtain C_S . Set $\tilde{S} = \partial_- C_S$. The next section proves that $v(\tilde{S})$ does not depend on the choices made in the construction of \mathcal{F} .

Remark 10.3. The reason why two-spheres are allowed in $\partial_- C_S$ should now be clear: we cannot prevent normal two-spheres from appearing in the normalization process. In particular, if S is an almost normal two-sphere then, for one of the two possible transverse orientations, there will always be a normal two-sphere appearing in \tilde{S} .

11. PROOF OF THE NORMALIZATION THEOREM

Suppose that S is almost normal and equipped with a transverse orientation. Before proving Theorem 11.1 recall from Definition 7.1 that C_S , a compression body in $|T|$, is *associated* to S if $\partial_+ C_S = S$, $\partial_- C_S$ is normal, the transverse orientation on S points into C_S , and any normal surface $S' \subset |T|$ disjoint from S may be normally isotoped to one disjoint from C_S .

We now have:

Theorem 11.1. *Given a transversely oriented almost normal surface S there exists a compression body C_S associated to S and C_S is unique (up to normal isotopy). Furthermore there is an algorithm that, given the triangulation T and the surface vector $v(S)$, computes the surface vector of $\partial_- C_S = \tilde{S}$.*

Proof. We proceed in several steps.

Claim. There exists a compression body C_S associated to S .

There are two cases. Either the transverse orientation for S points at the exceptional surgery disk (implying that S contained an almost normal annulus) or the transverse orientation points at an exceptional tightening disk.

In the first case, C_S is obtained by thickening S slightly and adding a regular neighborhood of the exceptional surgery disk. It is clear that C_S is a compression body, $\partial_+ C_S = S$, and $\partial_- C_S$ is normal. Suppose

that S' is any normal surface in T which is disjoint from S . Then, perhaps after a normal isotopy of S' (rel S), we have that S' is disjoint from the exceptional surgery disk for S . It follows that S' may be isotoped out of C_S .

In the second case the transverse orientation of S points at an exceptional tightening disk of S . As in Section 8 form \mathcal{F} with $\partial\mathcal{F} = S \cup F_n$. As in Section 10 attach two-handles to \mathcal{F} along the simple curves of F_n to obtain \mathcal{F}' . Cap off the bubbles with their three-balls to obtain C_S . Again, C_S is a compression body with $\partial_+C_S = S$.

Suppose now that S' is some normal surface in T which is disjoint from S . Then, by Corollary 9.3, the surface S' is disjoint from \mathcal{F} (perhaps after a normal isotopy of S' rel S). Since S' is normal it cannot meet any of the disks (in T^2) bounded by simple curves of F_n . So $S' \cap \mathcal{F}' = \emptyset$ as well. Finally, suppose that A is a bubble component of $\partial_-\mathcal{F}'$. Let B be the three-ball which A bounds (such that $B \cap T^2 = \emptyset$). Then no component of S' meets B as $S' \cap A = \emptyset$ and S' is normal. Deduce that $S' \cap C_S = \emptyset$. The claim is complete.

Claim. The associated compression body C_S is unique (up to normal isotopy).

Suppose that C_S and C'_S are both associated to S . Let $A = \partial_-C_S$ and $A' = \partial_-C'_S$. Then A and A' are normal surfaces, both disjoint from S . It follows that there exists a normal isotopy \mathcal{H} which moves A' out of C_S , rel S , and conversely another normal isotopy \mathcal{H}' which moves A out of C'_S , rel S .

Consider any face $f \in T^2$ and any normal arc $\alpha \subset f \cap S$. Let $X \subset f \cap C_S$ be the component containing α . Also take X' to be the component of $f \cap C'_S$ which contains α . We must show that X and X' have the same combinatorial type. Suppose not. After possibly interchanging X and X' there are only six situations to consider:

1. X is a critical rectangle and X' is a terminal rectangle.
2. X is a critical rectangle and X' is a critical hexagon.
3. X is a critical rectangle and X' is a terminal hexagon.
4. X is a critical hexagon and X' is a terminal hexagon.

In any of these four cases let δ be the normal arc of $A = \partial_-C_S$ on the boundary of X . Note that $\partial X'$ contains α (as does ∂X) and also another normal arc $\beta \subset f \cap S$ which does not meet X (as $S = \partial_+C_S$). Now note that it is impossible for \mathcal{H}' to normally isotope δ out of X' while keeping S fixed pointwise (as δ would have to cross β).

5. X is a terminal rectangle and X' is a critical hexagon.
6. X is a terminal rectangle and X' is a terminal hexagon.

In either of these cases let β be the other normal arc of $S \cap \partial X$. Then β intersects the interior of X' , a contradiction.

This proves the claim.

Claim. There is a algorithm that, given the triangulation T and the surface vector $v(S)$, computes the surface vector of $\partial_- C_S = \tilde{S}$.

We follow the proof of Lemma 9.1: We keep track of the intersection between the image of \mathcal{F}_k and every face $f \in T^2$. These are unions of components, with all allowable kinds shown (up to symmetry) in Figures 6 and 7. There is at most one hexagon in each face and perhaps many rectangles, arranged in three families, one for each vertex of f . At stage n there are no bent arcs remaining. Now delete all simple curves of F_n and all normal arcs of S . The normal arcs left completely determine \tilde{S} and from this we may find the surface vector $v(\tilde{S})$. This proves the claim and finishes the proof of Theorem 11.1. \square

Of course, the algorithm just given is inefficient. It depends polynomially on $\text{size}(T)$ and $\text{weight}(S)$. In the next section we improve this to a algorithm which only depends polynomially on $\text{size}(T)$ and $\log(\text{weight}(S))$.

As a corollary of Theorem 11.1:

Corollary 11.2. *If $S \subset |T|$ is a transversely oriented almost normal two-sphere then C_S is a three-ball, possibly with some open three-balls removed from its interior. (These have closures disjoint from each other and from S .)*

Now an orientable surface in an orientable three-manifold may be transversely oriented in exactly two ways. By Theorem 11.1, if S is an almost normal surface, for each transverse orientation there is a associated compression body. Call these C_S^+ and C_S^- .

From Corollary 11.2 deduce:

Theorem 11.3. *If $S \subset |T|$ is an almost normal two-sphere and both $\partial C_S^+ \setminus S$ and $\partial C_S^- \setminus S$ are (possibly empty) collections of vertex-linking two-spheres, then $|T|$ is the three-sphere.*

Proof. By hypothesis $\partial C_S^+ \setminus S$ is a collection of vertex linking spheres. For each of these add to C_S^+ the corresponding vertex neighborhood. Let \mathbb{B}^+ be the resulting submanifold of $|T|$. By the Alexander trick \mathbb{B}^+ is a three-ball. Do the same to C_S^- to produce \mathbb{B}^- . Applying the Alexander trick again deduce that the manifold $|T| = \mathbb{B}^+ \cup_S \mathbb{B}^-$ is the three-sphere. \square

12. AN EXAMPLE

Here we give a brief example of the normalization procedure. Let T be the one vertex triangulation shown in Figure 11.

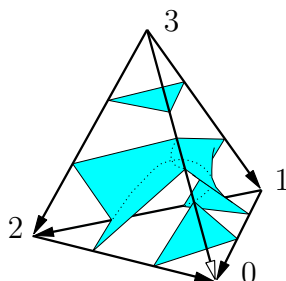


FIGURE 11. A one tetrahedron triangulation of S^3 . It is a simple exercise to list all normal and almost normal surfaces in T . It is a pleasant exercise to draw the graph T^1 as it actually sits in S^3 . It is slightly harder to draw the two-skeleton.

The front two faces (1 and 2) are glued to each other as are the back faces (0 and 3). The faces are glued to give the edge identifications shown. The surface S depicted in T is an almost normal two-sphere with two triangles and one almost normal octagon. Check this by computing $\chi(S) = 3 - 7 + 6 = 2$.

The sphere S has two exceptional tightening disks: D meeting the edge $(0, 3)$ of the model tetrahedron and D' meeting edge $(1, 2)$.

Tightening along D gives F_1 which is the vertex link. Tightening along D' gives F'_1, F'_2, F'_3 . Here F'_3 is a weightless two-sphere in T with a single simple curve and no other intersection with the two-skeleton. As a note of caution: F'_1 drawn in the model tetrahedron has four bent arcs – however $F'_1 \cap T^2$ contains only two. These are independent of each other and doing these moves in some order gives F'_2 and F'_3 . To obtain the normalization of S on the D' side, surger the simple curve of F'_3 and cap off the two resulting bubbles.

So, on the D side of S the normalization is the vertex link. On the D' side the normalization is the empty set. It follows from Theorem 11.3 that $|T|$ is the three-sphere. This finishes the example.

13. NORMALIZING QUICKLY

The normalization procedure can be made much more efficient.

Theorem 13.1. *There is a polynomial-time algorithm that, given T and the surface vector $v(S)$, produces as output $v(\tilde{S})$, the normalization of S . Here S is assumed to be a transversely oriented almost normal surface and T is assumed to be a triangulation of a three-manifold.*

As in Remark 7.2 we will restrict to the case where S is separating. Recall that N_S is the closure of the component of $|T| \setminus S$ which the transverse orientation points into. Then \widehat{N}_P is the union of all product blocks in N_S and \widehat{N}_C is the union of all the core blocks. Also N_P is a regular neighborhood of \widehat{N}_P , taken in N_S . Finally $N_C = \overline{N_S \setminus N_P}$. We will prove Theorem 13.1 by altering our original normalization procedure three times. First we will show that the order of the tightening moves is irrelevant. Then we will show that surgeries on simple curves and capping off of bubbles may happen during the normalization process, instead of being held until the end. Finally we show that tightening inside of N_P can be done very quickly. These three modifications combine to give an efficient algorithm.

13.1. Changing the order of the tightening moves. As stated in Remark 8.1 the isotopy $\mathcal{F}: S \times [0, n] \rightarrow M$ need not be unique. But we do have:

Lemma 13.2. *Any order for the tightening moves (performed in the construction of \mathcal{F}) gives the same surface \tilde{S} once the simple curves of F_n have been surgered. \square*

This follows immediately from the first sentence of Theorem 11.1.

13.2. Surgery on simple curves and deleting bubbles. We now alter the tightening procedure in a more substantial fashion:

Recall that $S \subset |T|$ is a transversely orientable separating almost normal surface. Recall that D is the exceptional tightening disk for S . Transversely orient S to point into the component of $|T| \setminus S$ which meets D . Here is the G -tightening procedure:

1. Let $G_0 = S$. Let $D_0 = D$.
2. Now do a small normal isotopy of G_0 in the transverse direction while also tightening G_0 along D_0 . Call the surface so obtained G'_0 . Now surger all simple curves of $f \cap G'_0$ for every $f \subset T^2$ to obtain G''_0 . Then delete any bubble components of G''_0 (two-sphere components which are contained in the interior of tetrahedra). Call the resulting surface G_1 . Note that G_1 inherits a transverse orientation from G_0 .
3. At stage $k \geq 1$ there are two possibilities. Suppose first that G_k has an outermost bent arc α with the transverse orientation of G_k

pointing into the tightening disk D_k , which is cut out of T^2 by α . Then perform a small normal isotopy of G_k in the transverse direction while tightening G_k across D_k . Call the surface so obtained G'_k . Now surger all simple curves of $f \cap G'_k$ for every $f \in T^2$ to obtain G''_k . Then delete any bubble components of G''_k . Call the resulting surface G_{k+1} . Note that G_{k+1} inherits a transverse orientation from G_k .

If there is no such outermost bent arc $\alpha \subset G_k$ then set $n = k$ and the procedure halts.

Lemma 13.3. *The surface G_n is normally isotopic to \tilde{S} , the normalization of S .*

Proof. Recall that Lemma 9.1 gives a complete classification of the possible components of intersection of $\text{image}(\mathcal{F}_k)$ with the faces of T^2 . Again, see Figures 6 and 7. The only components containing a simple curve are the terminal rectangle with hole and terminal hexagon with hole. Hence their names.

Since the terminal with hole rectangles and hexagons do not contain normal or bent arcs of F_k they remain unchanged in the F -tightening procedure until F_n is reached. Then all simple curves are surgered and bubbles capped off. Thus it makes no difference to the resulting surface \tilde{S} if we surger simple curves and delete bubbles as soon as they appear. \square

13.3. Tightening in I-bundle regions. We now give the final modification of the tightening procedure. Suppose that $v(S)$ is an almost normal surface vector. Suppose also that S has a transverse orientation pointing at an exceptional tightening disk.

Recall that N_S is the blocked submanifold cut from $|T|$ by the surface S (so that the transverse orientation points into N_S). Also, N_P is the I -bundle part of N_S while $N_C = \overline{N_S \setminus N_P}$ is the core of N_S .

We require slightly more sophisticated data structures. First define $\text{product}(S)$ to be the list $\{v_j\}_{j=1}^m$ where the j^{th} element is the vector $2 \cdot v(N_P^j)$ – here $v(N_P^j)$ is the block vector for the j^{th} component of \widehat{N}_P , found by Theorem 6.2. That is, $\sum v_j$ counts the normal disks of S which make up the horizontal boundary of the product blocks in N_S .

Put a copy of the horizontal boundary of N_C in $\text{core}(S)$. That is, record in $\text{core}(S)$ all of the gluing information between edges of disks which are on the horizontal boundary of core blocks. Also record, for each edge in $\partial \partial_v N_C$, which disk of $\text{core}(S)$ contains it and which component $N_P^j \subset N_P$ it is glued to. Build a model of N_C . That is,

deduce what core blocks occur in which tetrahedra and how they are glued across faces of T^2 .

We now turn to constructing a sequence of surfaces H_k . Each H_k will be represented by the two pieces of data: $\text{core}(H_k)$ and $\text{product}(H_k)$. Here is the *H-tightening procedure*:

1. Let $\text{core}(H_0) = \text{core}(S)$. Let $\text{product}(H_0) = \text{product}(S)$. Let $D_0 = D$. At stage k there is a tightening disk D_k used to alter H_k .
2. If the D_k has empty intersection with N_P then perform the tightening move as in the G -sequence. This effects only the pieces in $\text{core}(H_k)$ and we use the tightening move to compute $\text{core}(H_{k+1})$. Set $\text{product}(H_{k+1}) = \text{product}(H_k)$ and go to stage $k + 1$.
3. Suppose D_k intersects a component of N_P , say N_P^j . Then set $\text{product}(H_{k+1}) = \text{product}(H_k) \setminus \{v_j\}$; that is, remove v_j from the product part. We also alter the disks in the core as follows: Let $\text{core}(H'_k) = \text{core}(H_k) \cup \partial_v N_P^j$. Let $D'_k = D_k \setminus N_P^j$ (that is, remove a small neighborhood of T^1 from D_k). See Figure 12. Then D'_k is a surgery disk for $\text{core}(H'_k)$. So surger along D'_k , surger along all simple curves of $\text{core}(H'_k)$, and delete all bubbles in $\text{core}(H'_k)$. This finally yields $\text{core}(H_{k+1})$. Go on to stage $k + 1$.
4. At stage $k + 1$ there are two possibilities: either there is a bent arc in $\text{core}(H_{k+1})$ or there is not. If there is then we have a tightening disk D_{k+1} and proceed as above. If there is no bent arc then sum the vectors in $\text{product}(H_{k+1})$ and add to this vector the number of normal disks of each type in $\text{core}(H_{k+1})$. Output the final sum $v(H_n)$.

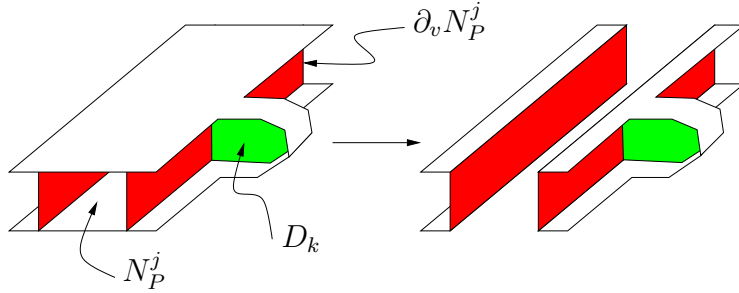


FIGURE 12. Removing the horizontal boundary of N_P^j and adding the vertical.

This is our final modification of the tightening procedure.

13.4. Correctness and efficiency.

Proof of Theorem 13.1. Note that if the transverse orientation on S points towards an exceptional surgery disk of S then the theorem is trivial. So suppose instead that a tightening disk is pointed at.

Claim. The H -tightening procedure outputs $v(\tilde{S})$.

That is, we claim that H_n is normally isotopic to \tilde{S} . Since the H procedure is identical to the G procedure in N_C we need only consider the situation where a tightening disk meets N_P . Consider the smallest k such that $D_k \cap N_P \neq \emptyset$. Recall that $\partial D_k = \alpha \cup \beta$ where $\alpha \subset H_k$ and $\beta \subset T^1$. Also by our hypothesis on k the arc β is contained in $T^1 \cap \partial_v \widehat{N}_P$ while only a small neighborhood of $\partial\alpha$ (taken in α) is contained in N_P . Suppose that N_P^j is the component of N_P containing β .

We have assumed that $H_k = G_k$. We will show that we can reorder tightening moves in the G -procedure to obtain $G_{k+k'}$ normally isotopic to H_{k+1} . Then it will follow from Lemma 13.3 that the H procedure produces correct output.

Recall that \widehat{N}_P^j and N_P^j are I -bundles. Let π be the natural quotient map which squashes I -fibres to a point. Let $E = \pi(N_P^j)$. Let $\widehat{E} = \pi(\widehat{N}_P^j)$. Note that \widehat{E} is not necessarily a surface. However E is a surface with boundary, \widehat{E} naturally embeds in E , and there is a small deformation retraction of E to \widehat{E} . Note that \widehat{E} and E inherit cell structures from \widehat{N}_P^j and N_P^j . Choose a spanning tree U for the one-skeleton \widehat{E}^1 of \widehat{E} rooted at $b = \pi(\beta)$. Choose an ordering of the vertices of U , $\sigma: U^0 \rightarrow (\mathbb{N} \cap [1, k'])$, so that for any vertex d with parent c we have $\sigma(c) < \sigma(d)$. Here $k' = |U^0|$ is the number of vertices in U^0 .

We now have a sequence of tightening moves to perform in the G procedure. At step one do the tightening move along the disk D_k , surger all simple curves, and delete bubbles. At step $i, i > 1$, examine the edge e between c and d (where $\sigma(d) = i$ and c is the parent of d). Then, by induction and the fact that $\sigma(c) < \sigma(d) = i$ there is a bent arc of G_{k+i-1} in the rectangle $\pi^{-1}(e)$ with endpoints on the segment $\pi^{-1}(d) \subset T^1$. Do this tightening move, surger simple curves, delete bubbles, and go to stage $i + 1$.

After stage $k' = |U^0|$ we obtain the surface $G_{k+k'}$ which is normally isotopic to the following: $(G_k \setminus \partial_h N_P^j) \cup \partial_v N_P^j$ surgered along the disk D'_k , surgered along simple curves, with bubbles deleted. Here $D'_k = \overline{D_k \setminus N_P^j}$. So $G_{k+k'}$ agrees with H_{k+1} and the claim is proved.

Claim. Precomputation for the H procedure takes time at most polynomial in $\text{size}(T)$ and $\log(\text{weight}(S))$.

Theorem 6.2 computes the vectors $\{v_j\}_{j=1}^m$ in the required amount of time. This gives $\text{product}(N_S)$. Then, since there are only a linear number (in $\text{size}(T)$) of core blocks in N_S (Remark 6.1) we can also compute their gluings and so compute $\text{core}(S)$ in the allotted time.

Claim. The number of steps in the modified normalization procedure is polynomial in $\text{size}(T)$.

Each step either reduces the weight of $\text{core}(H_k)$ by two or removes a vector from $\text{product}(H_k)$. Since the weight of $\text{core}(H_k)$ is at most linear (again, Remark 6.1), and since there are at most a linear number of components of N_P (see Remark 6.3), the claim follows.

Claim. Performing each step of the modified normalization procedure takes time at most polynomial in $\text{size}(T)$ and $\log(\text{weight}(S))$.

If the tightening disk is disjoint from N_P then we only have to alter $\text{core}(H_k)$ in the tetrahedra adjacent to the disk. There are only a linear number of these.

If the tightening disk meets a component of N_P , say N_P^j , then delete v_j from $\text{product}(H_k)$ in polynomial time (in $\text{size}(T)$). Alter $\text{core}(H_k)$ by gluing on a copy of $\partial_v N_P^j$, surgering along the remnants of the tightening disk D_k , surgering all simple curves, and deleting bubbles. As $\partial_v N_P^j$ is a subset of $\partial_v N_C$ it is at most linear in size (in terms of $\text{size}(T)$). Thus we can make the desired changes in the required time.

To sum up: we can compute the desired result, $v(\tilde{S})$, in time which is at most a product of polynomials in $\text{size}(T)$ and $\log(\text{weight}(S))$. This completes the proof of Theorem 13.1. \square

14. CRUSHING, OR: “NEW TRIANGULATIONS FOR OLD”

Crushing triangulations along normal surfaces is an important step in Casson’s algorithm [4]. As usual, we refer the reader to Jaco and Rubinstein’s work [16]. The notion of crushing is also explained in detail in Chapter 3 of Barchechar’s thesis [3].

Let T be a triangulation of a closed three-manifold. Suppose we are given a choice of quad type in a single tetrahedron, say the a^{th} type of quad in τ_i . Here $a \in \{1, 2, 3\}$ and the other two elements of $\{1, 2, 3\}$ are b and c . Recall that the a^{th} quad type separates the vertices 0 and a from the vertices b and c .

Let θ be the permutation $(0a)(bc)$. Let $\{(i, j_s, \sigma_s)\}_{s=0}^3$ be the four face pairings with i as the first element. Here σ_s glues the s^{th} face of τ_i to some face of τ_{j_s} . Note that $\{(j_s, i, \sigma_s^{-1})\}_{s=0}^3$ are also face pairings in T .

Define a new triangulation T' by *crushing* the tetrahedron τ_i along the a^{th} quad, as follows: Delete τ_i from T . Delete all of the face pairings $\{(i, j_s, \sigma_s)\}_{s=0}^3$. Replace the face pairing (j_s, i, σ_s^{-1}) (if $i \neq j_s$) with

$$(j_s, j_{\theta(s)}, \sigma_{\theta(s)} \cdot R_{(s, \theta(s))} \cdot \sigma_s^{-1}),$$

for $s \in \{0, 1, 2, 3\}$. Here $R_{(s, \theta(s))}$ is the rotation of the model tetrahedron, about the edge with vertices $\{0, 1, 2, 3\} \setminus \{s, \theta(s)\}$, which takes face s to face $\theta(s)$. Finally, no face of any model tetrahedron in T' is glued to itself – thus T' is a triangulation.

To keep track of this operation it may help to refer to the picture of a quad of type 3 shown on the right hand side of Figure 2.

Now suppose that p is a *polarization* of the triangulation T ; that is, p is a map from the set of tetrahedra to the set $\{0, 1, 2, 3\}$. Produce a new triangulation T' by *crushing* T along p : To begin with let T' be an exact copy of T . Now, for each $i = 1, 2, \dots, \text{size}(T)$ do one of two things; If $p(\tau_i) = 0$ simply go on to $i + 1$. If $p(\tau_i) \neq 0$ then remove τ_i by crushing along the $p(\tau_i)$ quad, as above, and go on to $i + 1$.

We now have:

Theorem 14.1. *There is a polynomial-time algorithm that, given a triangulation T and a polarization p , produces T' , the triangulation of T crushed along p . \square*

Of more interest is:

Theorem 14.2. *Suppose T is a triangulation so that the connect sum $\#|T|$ is a homology three-sphere. Suppose p is a polarization coming from S , a non-vertex linking normal two-sphere. Then the triangulation T' , the result of crushing T along p , satisfies $\#|T'| \cong \#|T|$.*

Proof. Theorem 5.9 of Jaco and Rubinstein's paper [16] essentially claims this result for any closed, orientable three-manifold $|T|$ with the caveat that some connect summands of $|T|$ homeomorphic to lens spaces may be omitted from the crushed $|T'|$. (See also [3, Theorem 3.1].)

However, by Lemma 4.3 no non-trivial lens space appears as a connect summand of the homology three-sphere $|T|$. Finally, omitting S^3 summands does not change the connect sum. The result follows. \square

15. THOMPSON'S THEOREM

We will need to use Casson's version [4] of the proof of Thompson's Theorem [26] (which in turn draws extensively from work of Rubinstein [23]). Chapter 6 of [3] gives a more detailed exposition of Casson's algorithm.

Theorem 15.1 (Casson [4]). *The three-sphere recognition problem lies in PSPACE; there is an exponential-time but polynomial-space algorithm that, given a triangulation T , decides whether or not $|T|$ is homeomorphic to the three-sphere.*

Proof. We give only a sketch of Casson's version of Thompson's algorithm. Begin with a triangulation $T_0 = T$. Check, using Theorems 4.1 and 4.2, that T_0 is a homology three-sphere. Inductively we have a triangulation T_i .

If T_i is not zero-efficient then apply Lemma 5.11 to find $S_i \subset |T_i|$, a fundamental non-vertex-linking normal two-sphere. Let T_{i+1} be the triangulation obtained by crushing along S_i . This requires Theorem 14.1.

If T_i is zero-efficient use Lemma 5.11 to search for almost normal two-spheres. If some component of T_i does not contain an almost normal two-sphere then by Theorem 14.2 and Theorem 5.10 the manifold $|T|$ was not the three-sphere. If S_i is an almost normal two-sphere inside a component T' of T_i then let $T_{i+1} = T_i \setminus T'$.

This completes the description of Casson's algorithm. If T_n is non-empty, then $|T|$ was not the three-sphere. If T_n is empty then $|T|$ was homeomorphic to the three-sphere. Both of these again use Theorem 14.2.

Note that $\text{size}(T_i) + i \leq \text{size}(T)$ as either crushing along a polarization or deleting a component always reduces the number of tetrahedra by at least one. This completes the sketch. \square

16. SHOWING THE PROBLEM LIES IN NP

We are now in a position to prove:

Theorem 16.1. *The three-sphere recognition problem lies in NP.*

Proof. Suppose that T is a triangulation of the three-sphere. The certificate is a sequence of pairs $(T_i, v(S_i))$ with the following properties.

- $T = T_0$.
- S_i is a normal or almost normal two-sphere, contained in $|T_i|$, with $\text{weight}(S_i) \leq \exp(\text{size}(T_i))$.
- If S_i is normal then S_i is not vertex linking and T_{i+1} is obtained from T_i by crushing along S_i .
- if S_i is almost normal then S_i normalizes to vertex linking two-spheres, in both directions. Also, T_{i+1} is obtained from T_i by deleting the component T' of T_i which contains S_i .
- Finally, the last triangulation T_n is empty, as is S_n .

Note that existence of the certificate is given by our proof of Theorem 15.1. So the only task remaining is to check the certificate. Here

we find two subtle points – we will not attempt to verify that the S_i are fundamental nor will we try to check that the T_i containing almost normal two-spheres are zero-efficient.

Note instead, since the S_i are fundamental, they obey the weight bounds given in Lemma 5.7; that is, $\text{weight}(S_i) \leq \exp(\text{size}(T_i))$.

So suppose a certificate $(T_i, v(S_i))$ as above, for the triangulation T , is given to us. First check, using Theorems 4.1 and 4.2, that T is a triangulation of a homology three-sphere.

By Lemma 4.4 check that $T = T_0$. Using Theorem 5.5 verify that S_i is a connected normal or almost normal surface. Using Lemma 5.4 compute the Euler characteristic of S_i . (Here we are using the fact that $\text{weight}(S_i) \leq \exp(\text{size}(T_i))$ in order to compute Euler characteristic in time only polynomial in $\text{size}(T_i)$.) This verifies that S_i is a two-sphere.

If S_i is normal, by Theorem 14.1, crush T_i along S_i in time at most polynomial in $\text{size}(T_i)$. Then check, using Lemma 4.4, that T_{i+1} agrees with the triangulation obtained by crushing T_i .

If S_i is almost normal, we need to check that T' , the component of T_i containing S_i , has $|T'| \cong S^3$. Using Theorem 13.1 normalize S_i in both directions in time at most polynomial in $\text{size}(T_i)$ (again, because $\log(\text{weight}(S_i)) \leq \log(\exp(\text{size}(T_i))) = \text{size}(T_i)$). If all components of the two normalizations S_i^+ and S_i^- are vertex linking two-spheres then T' is a triangulation of the three-sphere, by Theorem 11.3. Finally, use Lemma 4.4 to check that the triangulation $T_i \setminus T'$ is identical to T_{i+1} . \square

REFERENCES

- [1] Ian Agol. Personal communication.
- [2] Ian Agol, Joel Hass, and William Thurston. 3-manifold knot genus is NP-complete. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 761–766 (electronic), New York, 2002. ACM. arXiv:math/0205057.
- [3] Alexandre Barchecheat. *Minimal triangulations of 3-manifolds*. PhD thesis, U.C. Davis, 2003. arXiv:math/0307302.
- [4] Andrew J. Casson. The three-sphere recognition algorithm. 1997. Talk given at MSRI.
- [5] Andrew J. Casson and Cameron McA. Gordon. Reducing Heegaard splittings. *Topology Appl.*, 27(3):275–283, 1987.
- [6] Bruce Randall Donald and David Renpan Chang. On the complexity of computing the homology type of a triangulation. In *32nd Annual Symposium on Foundations of Computer Science (San Juan, PR, 1991)*, pages 650–661. IEEE Comput. Soc. Press, Los Alamitos, CA, 1991. <http://citeseer.ist.psu.edu/chang94computing.html>.
- [7] David Gabai. Foliations and the topology of 3-manifolds. III. *J. Differential Geom.*, 26(3):479–536, 1987.

- [8] Michael R. Garey and David S. Johnson. *Computers and intractability*. W. H. Freeman and Co., San Francisco, Calif., 1979. A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences.
- [9] Wolfgang Haken. Some results on surfaces in 3-manifolds. In *Studies in Modern Topology*, pages 39–98. Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.), 1968.
- [10] Joel Hass, Jeffrey C. Lagarias, and Nicholas Pippenger. The computational complexity of knot and link problems. *J. ACM*, 46(2):185–211, 1999. arXiv:math/9807016.
- [11] Alan Hatcher. Notes on basic 3-manifold topology. 2001. <http://www.math.cornell.edu/~hatcher/3M/3Mdownloads.html>.
- [12] Costas S. Iliopoulos. Worst-case complexity bounds on algorithms for computing the canonical structure of finite abelian groups and the Hermite and Smith normal forms of an integer matrix. *SIAM J. Comput.*, 18(4):658–669, 1989.
- [13] S. V. Ivanov. Recognizing the 3-sphere. *Illinois J. Math.*, 45(4):1073–1117, 2001. <http://www.math.uiuc.edu/~hildebr/ijm/winter01/final/ivanov.html>.
- [14] S. V. Ivanov. The computational complexity of basic decision problems in 3-dimensional topology. *Geom. Dedicata*, (131):1–26, 2008.
- [15] William Jaco and Ulrich Oertel. An algorithm to decide if a 3-manifold is a Haken manifold. *Topology*, 23(2):195–209, 1984.
- [16] William Jaco and J. Hyam Rubinstein. 0-efficient triangulations of 3-manifolds. *J. Differential Geom.*, 65(1):61–168, 2003. arXiv:math/0207158.
- [17] William Jaco and Jeffrey L. Tollefson. Algorithms for the complete decomposition of a closed 3-manifold. *Illinois J. Math.*, 39(3):358–406, 1995.
- [18] Simon A. King. The size of triangulations supporting a given link. *Geom. Topol.*, 5:369–398 (electronic), 2001. arXiv:math/0007032.
- [19] S. V. Matveev. Algorithms for the recognition of the three-dimensional sphere (after A. Thompson). *Mat. Sb.*, 186(5):69–84, 1995.
- [20] Sergei Matveev. *Algorithmic topology and classification of 3-manifolds*, volume 9 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2003.
- [21] Christos H. Papadimitriou. *Computational complexity*. Addison-Wesley Publishing Company, Reading, MA, 1994.
- [22] J. Hyam Rubinstein. Polyhedral minimal surfaces, Heegaard splittings and decision problems for 3-dimensional manifolds. In *Geometric topology (Athens, GA, 1993)*, pages 1–20. Amer. Math. Soc., Providence, RI, 1997.
- [23] Joachim H. Rubinstein. The solution to the recognition problem for S^3 , 1992. Lectures in Haifa, Israel.
- [24] Joachim H. Rubinstein. An algorithm to recognize the 3-sphere. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 601–611, Basel, 1995. Birkhäuser.
- [25] Saul Schleimer. *Almost normal Heegaard splittings*. PhD thesis, U.C. Berkeley, 2001. <http://warwick.ac.uk/~masgar/Maths/thesis.pdf>.
- [26] Abigail Thompson. Thin position and the recognition problem for S^3 . *Math. Res. Lett.*, 1(5):613–630, 1994.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UK

E-mail address: s.schleimer@warwick.ac.uk