

# Boundary-twisted normal form and the number of elementary moves to unknot

Chan-Ho Suh

ABSTRACT. Suppose  $K$  is an unknot lying in the 1-skeleton of a triangulated 3-manifold with  $t$  tetrahedra. Hass and Lagarias showed there is an upper bound, depending only on  $t$ , for the minimal number of elementary moves to untangle  $K$ . We give a simpler proof, utilizing a normal form for surfaces whose boundary is contained in the 1-skeleton of a triangulated 3-manifold. We also obtain a significantly better upper bound of  $2^{120t+14}$  and improve the Hass–Lagarias upper bound on the number of Reidemeister moves needed to unknot to  $2^{10^5n}$ , where  $n$  is the crossing number.

## CONTENTS

1. Introduction	390
2. Boundary-twisted and boundary-restricted normal forms	391
2.1. Marked triangulations	391
2.2. Boundary-twisted normal form	391
2.3. Boundary-restricted normal form	393
3. Enumeration of twisted normal discs	394
3.1. Marked tetrahedron with only marked vertices	394
3.2. Tetrahedron with one marked edge	395
3.3. Tetrahedron with one marked edge and one marked vertex	395
3.4. Tetrahedron with one marked edge and two marked vertices	395
3.5. Tetrahedron with two marked edges	396
3.6. Obtaining normal discs from twisted normal discs by truncating	397
4. Isotoping to boundary-twisted normal form	398
5. Existence of a fundamental unknotting disc	402
6. A bound on the number of elementary moves to unknot	404
6.1. Bounding the number of Reidemeister moves	406

Received May 17, 2011.

2010 *Mathematics Subject Classification*. Primary 57M, 57N10; Secondary 68Q25.

*Key words and phrases*. Reidemeister move, unknotting, normal surface.

Research was partially funded by the National Science Foundation (VIGRE DMS-0135345 and DMS-0636297).

## 1. Introduction

Suppose  $M$  is a triangulated, compact 3-manifold with  $t$  tetrahedra;  $K$  is a simple, closed curve in the 1-skeleton of the triangulation; and  $K$  is an *unknot*, i.e., bounds a disc in  $M$ . Recall that  $K$  can be isotoped in  $M$  using polygonal moves across triangles called *elementary moves*. J. Hass and J. Lagarias obtained an upper bound of  $2^{10^7 t}$  on the minimum number of elementary moves to take  $K$  to a triangle contained within one tetrahedron [3].

The central idea of their proof is to use normal surface theory. Take a double barycentric subdivision of the triangulation, and let  $N(K)$  denote the simplicial neighborhood of  $K$ . Since  $K$  is an unknot, a homotopically nontrivial curve  $l$  on  $\partial N(K)$  bounds a normal disc  $D$  in  $M - \text{int}(N(K))$ . Normal surface theory gives an exponential upper bound in  $t$  on the number of triangles in a minimal such  $D$  [4]. Naively we might think to move  $K$  across  $D$  and obtain the bound on elementary moves that way, but this overlooks that  $K$  must first be moved from the interior of  $N(K)$  to  $l$  which lies on  $\partial N(K)$ . Thus Hass and Lagarias break up the complete isotopy of  $K$  to a triangle into three parts. First isotope  $K$  to  $\partial N(K)$ , isotope across  $\partial N(K)$  to  $l$ , and finally isotope across  $D$ . The bound on  $D$  gives a similar bound for the number of elementary moves to isotope across  $D$ . The bounds for the number of elementary moves realizing the other isotopies are obtained from an involved analysis, which takes up a good part of [3]. Independently, but also using normal surface theory, S. Galatolo worked out a bigger bound for the specialized case of a polyhedral  $M$  in  $\mathbb{R}^3$  [2].

By considering a normal form for surfaces whose boundary is contained in the 1-skeleton of a triangulated 3-manifold, we avoid the retriangulation and subsequent isotopies related to  $N(K)$ . This not only substantially simplifies the proof but improves the upper bound on the number of moves to  $2^{120t+14}$ .

A corollary of bounding the number of elementary moves is a bound on the number of Reidemeister moves to unknot an unknot diagram with crossing number  $n$ . Our result lets us improve the bound given by Hass and Lagarias from  $2^{10^{11}n}$  to  $2^{10^5 n}$ .

In Section 2 we define the normal form and relate it to a restricted version of normal surface theory in truncated tetrahedra. Section 3 explains how to enumerate the normal discs, which is a key technical detail for this paper. In Section 4 we explain how an essential surface spanning a link in the 1-skeleton can be isotoped into boundary-twisted normal form. Section 5 proves the existence of a fundamental spanning disc for the unknot. In Section 6, using the previous results, we obtain an improved upper bound on the number of triangles needed in a spanning disc for an unknot. From this bound, we obtain bounds which improve the main results of [3].

**Acknowledgments.** The author acknowledges the inspiration and support of Professor Joel Hass, who supervised this work as part of a Ph.D. dissertation. The author also thanks Professor Chuichiro Hayashi for his encouragement and helpful comments.

## 2. Boundary-twisted and boundary-restricted normal forms

**2.1. Marked triangulations.** Given a pair  $(M, L)$  consisting of a compact 3-manifold  $M$  and link  $L$  in  $M$ , we say  $\mathcal{T}$  is a *marked triangulation* for  $(M, L)$  if  $\mathcal{T}$  is a triangulation of  $M$  with  $L$  contained in the 1-skeleton. The vertices and edges of  $\mathcal{T}$  which belong to  $L$  are called *marked vertices and edges*.

If two edges of the same face of a tetrahedron are marked and not part of a triangle component of  $L$ , there is an isotopy of these edges to the third edge of the face reducing the number of marked edges. It will suffice to consider that  $L$  has no triangle component and the number of edges is minimal.

**Convention 2.1.** No component of  $L$  is a triangle, and in any tetrahedron of a marked triangulation, every face has at most one marked edge.

**2.2. Boundary-twisted normal form.** A tetrahedron,  $T$ , of a marked triangulation is of 9 possible types and is called a *marked tetrahedron*. Normal disc types in a marked tetrahedron are more complicated than the usual triangles and quads.

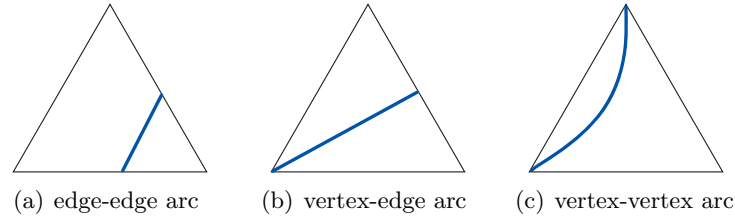


FIGURE 1. Normal arcs in a face

**Definition 2.2.** We define a *normal arc* in a face of  $T$  to be a properly embedded arc such that one of the following holds:

- It starts in the interior of an edge and ends in the interior of another edge (Figure 1(a)).
- It starts at a marked vertex and ends in the interior of the opposite edge (Figure 1(b)).
- It starts at a marked vertex and ends at a different marked vertex (Figure 1(c)).

**Definition 2.3.** We define a *twisted normal disc*  $D$  in a marked tetrahedron  $T$  to be a properly embedded disc such that its boundary  $\partial D$  satisfies the following conditions ( $e$  is an edge and  $\Delta$  is a face):

- (1)  $\partial D \cap \text{int}(\Delta)$  consists of normal arcs for every face  $\Delta$ .
- (2)  $\partial D \cap e$  is one of the following:
  - (a) empty,
  - (b) one endpoint of  $e$ ,
  - (c) an arc and  $e$  is marked,
  - (d) an interior point of  $e$  and  $e$  is unmarked,
  - (e) both endpoints of  $e$  and  $e$  is unmarked.
- (3) A pair of normal arcs of  $\partial D$  abutting the same point or arc of  $\partial D \cap e$  must be in different faces.
- (4)  $\partial D \cap \Delta$  does not have two normal arcs with endpoints at the same marked vertex.

**Definition 2.4** (Boundary-twisted normal form). Let  $(M, L)$  be a 3-manifold with link  $L$  and marked triangulation  $\mathcal{T}$ . Suppose also  $S$  is a surface with boundary contained in  $L$ . We say  $S$  is in *boundary-twisted normal form* if its intersection with every tetrahedron consists of twisted normal discs.

In attempting normal surface theory with boundary-twisted normal surfaces, a technical obstacle arises: it is necessary to consider an additional kind of surface punctured by marked edges. Rather than work with that setup, we have chosen to work in a more familiar setting. We will truncate each marked tetrahedron at its marked vertices and edges. Then we do normal surface theory with respect to these truncated tetrahedra while imposing some additional conditions on the surface's boundary behavior in the truncated regions.

**Definition 2.5.** Let  $T$  be a marked tetrahedron. Then the truncation of  $T$ , denoted  $T_{tr}$ , is a polyhedron obtained by the process indicated in Figure 2. Given a marked triangulation  $\mathcal{T}$ , the *truncated triangulation*  $\mathcal{T}_{tr}$  is the polyhedral decomposition given by truncating every marked tetrahedron of  $\mathcal{T}$ .

**Definition 2.6.** A properly-embedded disc  $D$  in a truncated tetrahedron  $T_{tr}$  is *normal* if it is the restriction to  $T_{tr}$  of a twisted normal disc in  $T$ . A surface in a truncated triangulation is normal if it intersects each triangulation in normal discs.

**Remark 2.7.** This is not the same as the usual generalization of the concept of normal disc from a tetrahedron to a polyhedron, e.g., [1] or [6]. Our definition follows naturally from the view of boundary-twisted normal surfaces and has the advantage of reducing the number of disc types.

Since the truncation  $T_{tr}$  can be considered a subset of  $T$ , a normal disc in  $T_{tr}$  naturally sits inside  $T$ . It can be extended to a twisted normal disc of  $T$  in a natural way. In general, this extension is far from unique, since one can choose different directions to twist and in some cases twist either in the interior of an edge or at an end (see Figure 3).

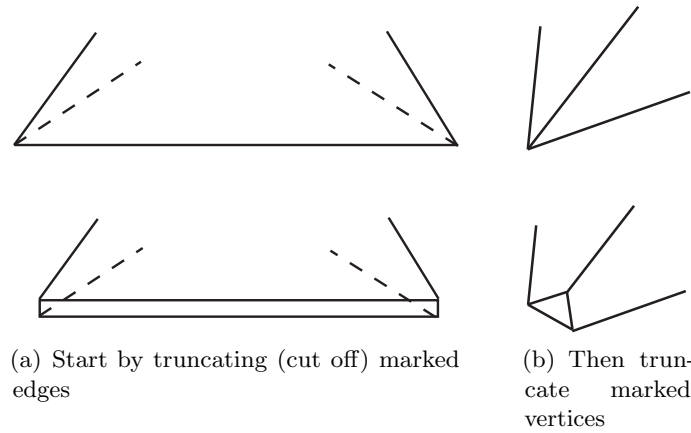


FIGURE 2. Obtaining truncated tetrahedra from marked tetrahedra

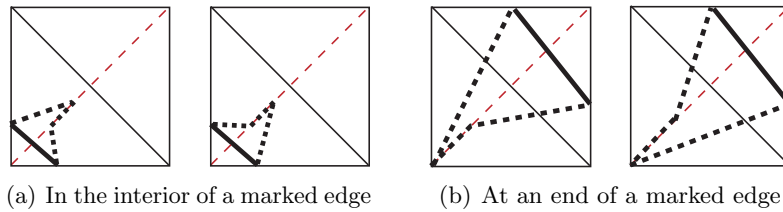


FIGURE 3. Discs differing by an opposite twist along a marked edge

**2.3. Boundary-restricted normal form.** A boundary-twisted normal surface is a normal surface in the truncated triangulation  $\mathcal{T}_{tr}$ , whose boundary has been extended through the truncated region to  $L$ . In order for this extension to happen in a simple way, the normal surface must satisfy additional conditions on its boundary. We call this type of normal surface in  $\mathcal{T}_{tr}$  a *boundary-restricted normal surface*.

There are two kinds of boundary regions given by the truncation, triangles and rectangles. Rectangles should be visualized as “long” with the long sides corresponding to the longitudinal direction and the short sides corresponding to the meridional direction.

We define a boundary-restricted normal surface to be a normal surface  $S$  in  $\mathcal{T}_{tr}$  such that for any rectangle  $R$  we require  $\partial S \cap R$  look as in Figure 4. We either have one longitudinal normal arc,  $n$  meridional arcs, or  $n$  meridional arcs with one or two corner-cutting arcs ( $n \geq 0$ ). If there are two corner-cutting arcs, they must be at opposite corners.

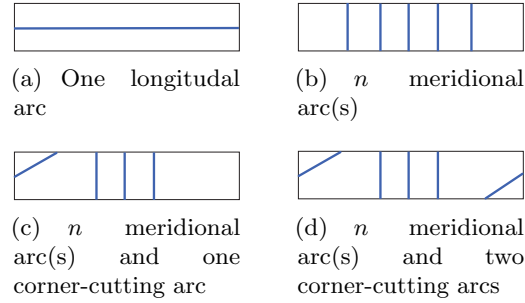


FIGURE 4. Pictures of the restrictions on the boundary of a boundary-restricted normal surface

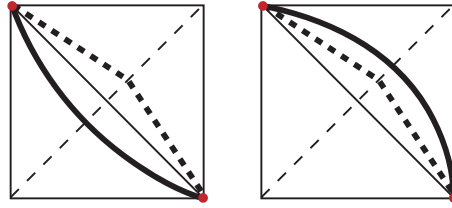


FIGURE 5. Discs that differ from flipping a vertex-vertex arc to an adjacent face

### 3. Enumeration of twisted normal discs

Figures 6–10 show pictures of some twisted normal discs in the different marked tetrahedra. To obtain further disc types from a particular figure, apply a symmetry of the tetrahedron (preserving markings), change the direction of twisting along a marked edge (Figure 3), and/or flip a vertex-vertex arc to an adjacent face (Figure 5). Each figure's caption includes a number denoting the total number of disc types obtainable from a figure in this manner. Also note that some of the discs in one kind of marked tetrahedron also show up in other kinds, but the figures illustrate only the nonredundant ones. For example, all of the disc types for a tetrahedron with one marked vertex and one marked edge show up in the tetrahedron with two marked vertices and one marked edge.

Except for these operations and the nonillustration of redundant disc types, these figures are complete. The organization of the figures is meant to suggest the method of enumeration. We now explain the enumeration in further detail.

**3.1. Marked tetrahedron with only marked vertices.** For an unmarked tetrahedron, the only twisted normal discs are the standard normal discs: triangles and quads (Figure 6(a)). This gives the usual 7 disc types.

For a marked tetrahedron with one or more marked vertices but no marked edges, in addition to the previous triangles and quads, we have vertex touching triangles (Figure 6(b)) and possibilities utilizing vertex-vertex arcs (Figures 6(c), 6(d), 6(e), and 6(f)).

For a tetrahedron with one marked vertex, we have 7 discs from the unmarked case and the three vertex touching triangles, making 10 disc types.

For a tetrahedron with two marked vertices, we have 4 quads, 3 triangles, 6 vertex touching triangles, 2 triangles with a vertex-vertex arc, and 1 bigon. This gives 16 total.

For a tetrahedron with three marked vertices, 4 quads, 3 triangles, 9 vertex touching triangles, 6 triangles with one vertex-vertex arc, 4 triangles with all vertex-vertex arcs, and 3 bigons. This gives 29 total.

For a tetrahedron with four marked vertices, we have 4 triangles, 3 quads, 6 quads with all vertex-vertex arcs, 12 vertex-touching triangles, 12 triangles with one vertex-vertex edge, 16 triangles with all vertex-vertex arcs, 6 bigons, giving 59 total.

**3.2. Tetrahedron with one marked edge.** For the one marked edge case, we have 30 total disc types.

There are two triangles and one quad that don't use the marked edge at all. Moving on to discs that utilize an endpoint of the marked edge, there are 6 vertex-touching triangles, as in a previous case.

The remaining discs utilize the entire marked edge or a subarc of it (see Figure 7). The first row illustrates those using the entire marked edge, while the last two rows show disc types using an interior or exterior subarc, resp.

**3.3. Tetrahedron with one marked edge and one marked vertex.**

There are 47 total disc types, with 30 types from the 1 marked edge case, and 17 new types, which we will enumerate and illustrate. The 17 new disc types must utilize the marked vertex (see Figure 8). The first row of the figure are 2 discs not utilizing a subarc of the marked edge: a vertex-touching triangle and a bigon. Each subsequent row of the figure, as before, illustrates the disc types utilizing a particular kind of subarc of the marked edge.

**3.4. Tetrahedron with one marked edge and two marked vertices.**

There are a total of 93 disc types. From the 1 marked edge case, we have 30. From the 1 marked edge and 1 marked vertex case, we get double contributions (since we now have two marked vertices), giving  $2 \cdot 17 = 34$ . We enumerate the 29 new disc types below. Note that they must utilize both of the marked vertices.

There are 9 disc types not using any arc of the marked edge: 1 bigon (Figure 9(a)) and 8 triangles with all vertex-vertex arcs (Figure 9(b)).

The remaining disc types are all quads. There are 4 quads with three vertex-vertex arcs which utilize the full marked edge (Figure 9(c)). There are 12 quads utilizing an exterior subarc of the marked edge (Figure 9(d)). There are 4 quads utilizing an interior subarc of the marked edge (Figure 9(e)).

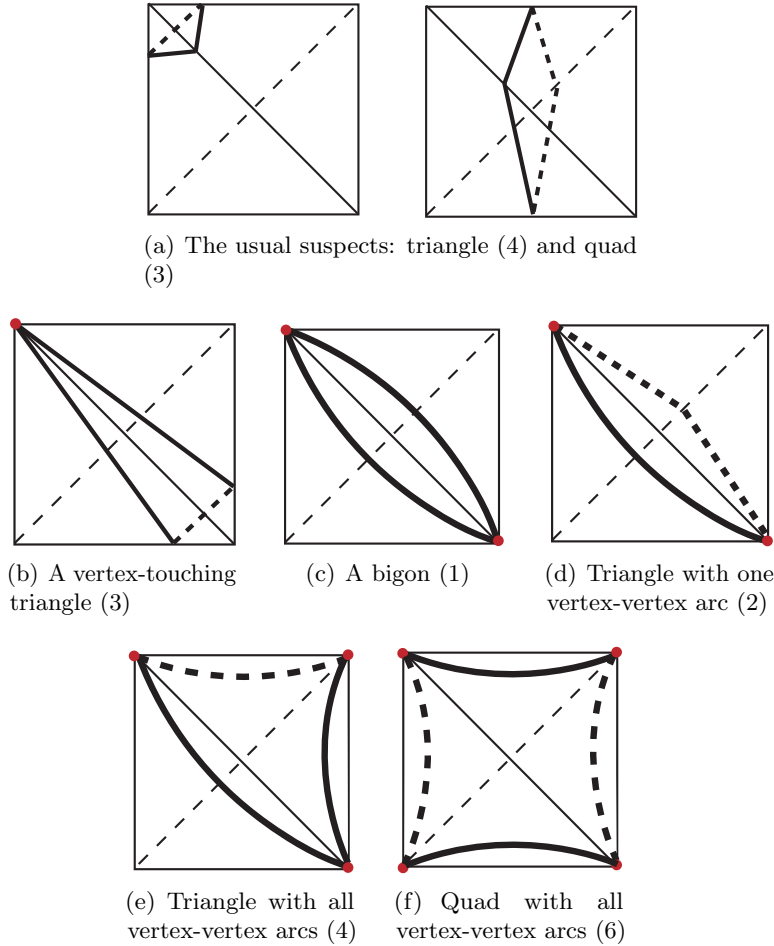


FIGURE 6. Twisted normal discs in a tetrahedron with  $n$  marked vertices ( $n \geq 0$ )

**3.5. Tetrahedron with two marked edges.** There are a total of 148 disc types. There are 66 new disc types, which are illustrated in Figure 10.

As before, the discs from the previous cases should be counted, but not all of them are compatible with this type of marked tetrahedron. From the 1 marked edge case, we have 1 normal quad, 4 vertex-touching triangles, 8 quads with an interior subarc, and 8 quads with an exterior subarc. The total is 21.

From the 1 marked edge and 1 marked vertex case, we have 2 bigons, and a quadruple contribution of the rest of the disc types, except for the vertex-touching triangles which are considered in the previous paragraph. This gives  $2 + 4 \cdot 15 = 62$  total.



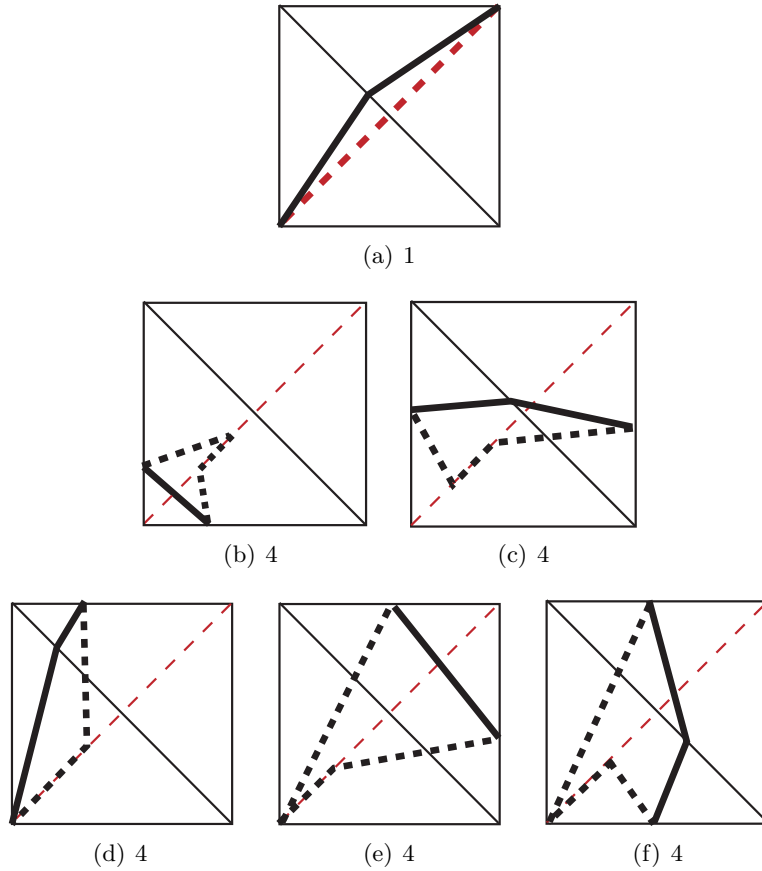


FIGURE 7. Twisted normal discs in a tetrahedron with one marked edge

The nonredundant discs from the 1 marked edge and 2 marked vertex case are not allowable here.

**3.6. Obtaining normal discs from twisted normal discs by truncating.** For normal surface theory, we need to know the maximal number of normal discs in any truncated tetrahedron. Table 2 shows the total number of normal discs in a truncated tetrahedron. These are obtained by truncating the tetrahedron and observing carefully how different twisted normal disc types become the same normal disc type. For example, after truncation of the edges, the two discs illustrated in Figure 10(d) are amongst the discs obtained from Figure 10(a). The last column of the table will be used later in Section 6. It is the maximum number of disc types whose boundary has a common normal arc in a face.

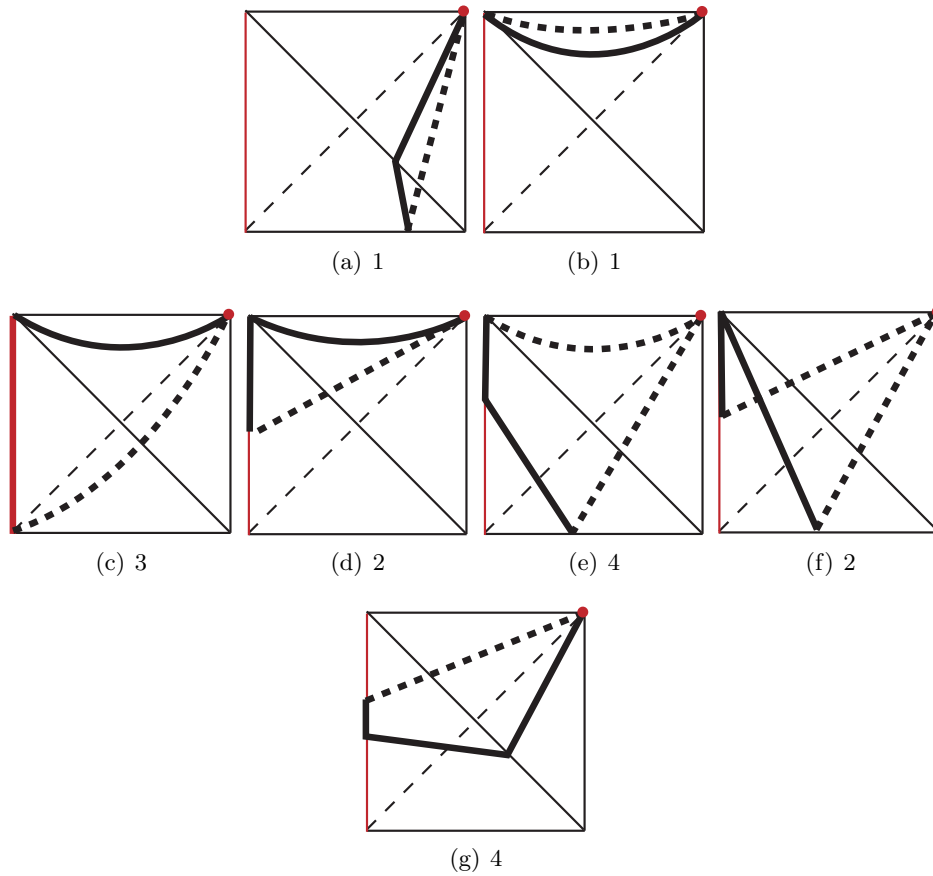


FIGURE 8. Twisted normal discs in a tetrahedron with one marked edge and one marked vertex

#### 4. Isotoping to boundary-twisted normal form

**Theorem 4.1.** *Let  $M$  be a compact irreducible 3-manifold with triangulation  $\mathcal{T}$  and  $L$  be a polygonal link contained in the 1-skeleton of  $\mathcal{T}$ . Suppose  $L$  has at most one edge in each triangle of  $\mathcal{T}$  and bounds an incompressible surface  $S$ . Then  $S$  can be isotoped (rel  $\partial$ ) into boundary-twisted normal form.*

**Proof.** Isotope  $S$  to be in general position with respect to the 2-skeleton of  $\mathcal{T}$ . Then  $\text{Int}(S) \cap \partial\Delta^3$  for any tetrahedron  $\Delta^3$  consists of simple closed curves and/or embedded (open) arcs with endpoints on marked edges or vertices.

Consider the portions of the marked edges of  $\Delta^3$  which abut pieces of  $S$  inside of  $\Delta^3$ . By a small isotopy of  $S$ , we can assume that they are arcs or endpoints.

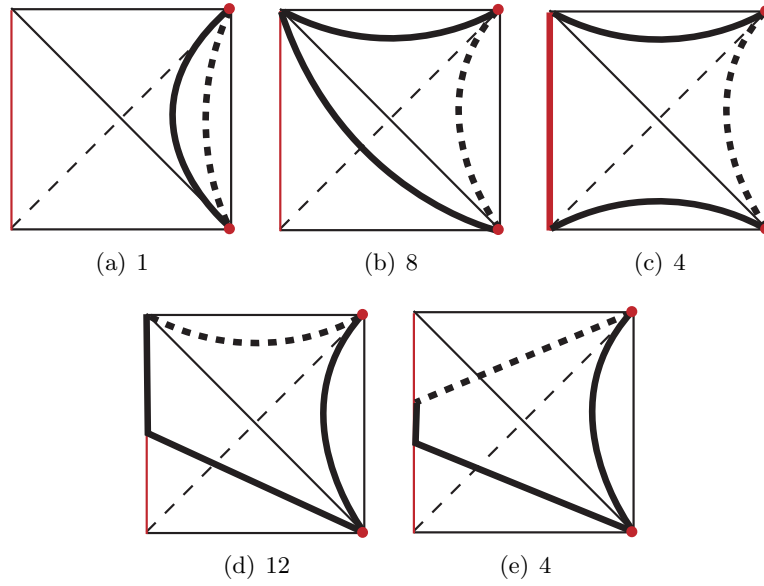


FIGURE 9. Twisted normal discs in a tetrahedron with marked edge and two marked vertices

Tetrahedron type	Total
No marked vertices or edges	7
one marked vertex	10
two marked vertices	16
three marked vertices	29
four marked vertices	59
1 marked edge	30
1 marked edge, 1 marked vertex	47
1 marked edge, 2 marked vertices	93
2 marked edges	148

TABLE 1. Number of twisted normal disc types in each type of marked tetrahedron

We call a simple closed curve in  $S \cap \partial\Delta^3$  which is a boundary component of an annulus contained in  $S \cap \Delta^3$  a *circle of intersection* of  $S$  and  $\partial\Delta^3$ . Note that strictly speaking, this is an abuse of terminology as a circle of intersection should refer to a component of the intersection which is a circle.

Define the *weight* of  $S$  to be the sum of the number of components of  $S \cap \text{int}(\Delta)$  over all faces  $\Delta$  of  $\mathcal{T}$ . We will now perform a series of weight-reducing isotopies until  $S$  is in boundary-twisted normal form. Each type

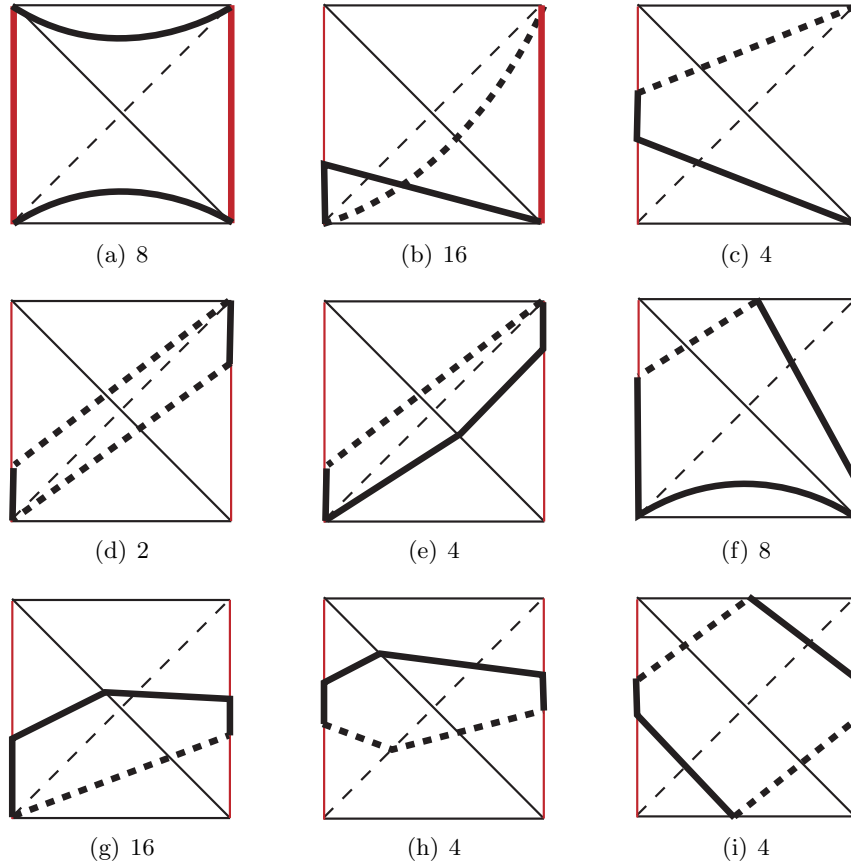


FIGURE 10. Twisted normal discs in a tetrahedron with two marked edges

Tetrahedron type	Total	Max arc #
No truncation	7	2
One truncated vertex	10	3
Two truncated vertices	16	3
Three truncated vertices	29	3
Four truncated vertices	59	3
One truncated edge	15	3
Two truncated edges	17	6
One truncated edge, one truncated vertex	22	3
One truncated edge, two truncated vertices	40	6

TABLE 2. Number of normal disc types in each type of truncated tetrahedron

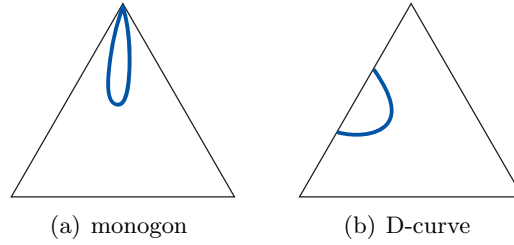


FIGURE 11. Some nonnormal curves

of isotopy will eliminate an unwanted situation and bring  $S$  closer to being in boundary-twisted normal form. After an isotopy, we may have disturbed our work in previous stages, but we can repeat the entire simplification procedure up to that point, which drives down the weight, to ensure all the previous conditions still hold. We will assume this is done in the following descriptions.

Consider a particular  $\Delta^3$ . For each circle of intersection, starting with innermost ones, we can take the disc bound by it in  $\partial\Delta^3$  and push it in slightly. By doing so we obtain a compression disc for  $S$ . Since  $S$  is incompressible and  $M$  is irreducible, we can isotope  $S$  until its intersection with  $\Delta^3$  coincides with these compression discs. Therefore we can arrange that each circle of intersection on  $\partial\Delta^3$  bounds a disc inside it.

Consider a circle of intersection which is in the interior of a face. We can isotope the disc of  $S$  bound by the circle through the face, removing one or more circles of intersection. We repeat this to remove any further circles of intersection inside a face.

Now consider any circle of intersection whose restriction to a face has a nonnormal curve. The arc either has both its endpoints on a marked vertex, a *monogon* (Figure 11(a)), or at least one endpoint is in the interior of an edge, a *D-curve* (Figure 11(b)). A monogon can be eliminated by pushing the disc it bounds into the next tetrahedron.

In the *D-curve* case, we can suppose it is innermost. Such an innermost curve and a segment of the marked edge bounds a disc in the face. This is a compression disc, so we can isotope  $S$  to the disc and then push through the face. This isotopy must reduce the weight.

Because we are decreasing the weight, by repeating this procedure, we can ensure that the surface  $S$  in  $\Delta^3$  satisfies:

- Its intersection with each tetrahedron consists of discs.
- The boundary of each disc consists of normal arcs and arcs of marked edges (Definition 2.3 (1) and (2c)).

We can eliminate a disc with two normal arcs in a face touching the same vertex by pushing the disc through the face near the vertex (Definition 2.3 (4)). This reduces the weight by combining the two arcs into one.

It may be the case that the boundary of a disc  $D$  intersects an edge  $e$  more than once, including at least once in its interior. We can suppose  $D$  is innermost. We now have two cases: 1)  $e$  is not marked. 2)  $e$  is marked.

Case 1): There is a disc  $E$ , whose interior is disjoint from  $S$ , such that  $\partial E = \alpha \cup \beta$ , where  $\alpha$  is an arc on  $S$  connecting two points of  $\partial D \cap e$  and  $\beta$  is a subarc of  $e$ . We can push  $S$  across  $E$  so that  $\alpha$  is taken to  $\beta$ . A further small isotopy of  $S$  through  $e$  will split  $D$  into two discs in the tetrahedron. If the two points of  $\partial D \cap e$  were in the interior of  $e$ , these two discs have combined weight less than that of  $D$ . Otherwise, one point was at an endpoint of  $e$  and it is possible the two discs have the same weight as that of  $D$ . Nonetheless, the intersection of  $S$  with the 1-skeleton is simplified. Since none of our weight-reducing isotopies increase intersection with the 1-skeleton, we can eliminate this situation also.

Case 2): If  $D$  intersects  $e$  only at its endpoints, we need to avoid violating Definition 2.3 (2e). In this case, there must be an arc  $\alpha$  on  $D$  joining the endpoints of  $e$ . Then  $\alpha$  and  $e$  bound a compression disc which gives a weight-reducing compression. Otherwise  $D$  intersects  $e$  in its interior in at least one arc. By taking an arc  $\alpha$  of the marked edge adjacent to two innermost components of  $\partial D \cap e$  and an arc  $\beta$  in the interior of  $S$  connecting the endpoints of  $\alpha$ , we obtain a simple closed curve  $\gamma = \alpha \cup \beta$ .  $\gamma$  bounds a compression disc in  $\Delta^3$  for  $S$ . Thus we can isotope  $S$  to the disc, reducing the weight.

Now consider a circle of intersection which touches a marked edge but does not pass through from one face to another (this violates Definition 2.3 (3)). Suppose the curve bounds the disc  $D$ . We can reduce the weight by pushing the part of  $D$  near the marked edge through a face.

Eventually we arrive at a situation where the pieces of  $S$  inside  $\Delta^3$  are exactly what we defined previously as twisted normal discs (see Definition 2.3). In particular, if Definition 2.3 (2) is not satisfied, clearly we can do one of the above isotopies.

Now we move onto another tetrahedron and repeat the entire process until  $S$  is cleaned up to the requisite form inside the tetrahedron. Note this may ruin our work in the previous tetrahedron. We move onto yet another tetrahedron and do the process. By cycling through the tetrahedra and noting that the weight is strictly decreasing, eventually the weight is at a minimum and  $S$  is in boundary-twisted normal form.  $\square$

## 5. Existence of a fundamental unknotting disc

Normal surface theory in truncated triangulations is analogous to that in standard triangulations. Each normal disc type in a truncated tetrahedron is represented by a variable. There are integer linear equations in these

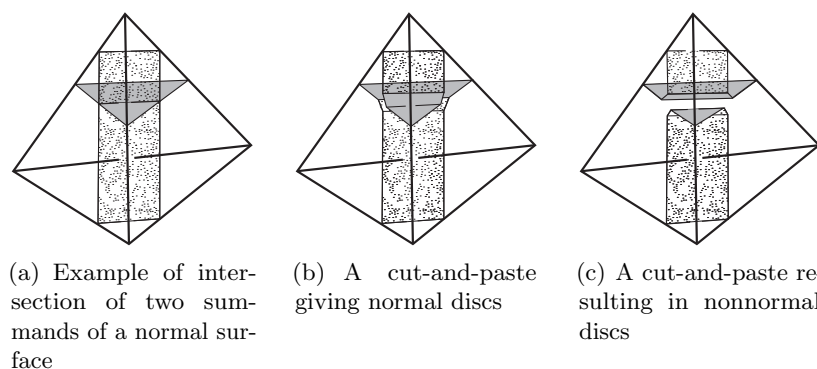


FIGURE 12. The Haken sum

variables, *matching equations*, given by each pair of truncated tetrahedra sharing a face. A coordinate vector is obtained from a normal surface by simply counting the number of normal discs of each type it contains. A normal surface’s vector must be a solution to the equations, but nonnegative integral solutions are not necessarily normal surfaces. Nonnegative integer solutions which satisfy a no-intersection condition are *uniquely* realized as a normal surface. The condition is that certain pairs of coordinates cannot both be nonzero; the pairs correspond to normal disc types that cannot be realized at the same time as disjoint discs.

A normal surface with vector  $v$  such that  $v \neq v_1 + v_2$  where each  $v_i \neq 0$  is a nonnegative integer solution to the matching equations is called *fundamental*. All such  $v$  constitute the minimal Hilbert basis for the system of equations. This basis is a finite, generating set for all nonnegative integer solutions and can be found using methods of integer linear programming.

Let  $S$  be a normal surface with solution vector  $v(S) = v_1 + v_2$  and each  $v_i$  is a nonnegative integer solution vector. The no-intersection condition on  $v(S)$  passes to each  $v_i$  so that  $v_i = v(S_i)$  for a normal surface  $S_i$ . There is a *Haken sum*, a cut-and-paste addition of normal surfaces, such that the Haken sum of  $S_1$  and  $S_2$  is  $S$  (Figure 12). Two important properties are that Euler characteristic is additive under Haken sum, and there is a way of cutting and pasting the “wrong way”, the so-called *irregular switch*, that creates a nonnormal surface.

In our setup, we work with boundary-restricted normal surfaces. In order that our arguments work, we need only ensure that passing to a summand lets us continue working with boundary-restricted normal surfaces. That is the content of the following simple, but crucial, lemma.

**Lemma 5.1.** *Let  $S$  be a boundary-restricted normal surface and suppose  $v(S) = v_1 + v_2$ , where each  $v_i$  is a nonzero solution to the matching equations. Then  $v_i = v(S_i)$  for a unique boundary-restricted normal surface  $S_i$  and  $S$  is the Haken sum  $S_1 + S_2$ .*

**Proof.** From the remarks about normal surface theory preceding the lemma, clearly  $v_i = v(S_i)$  for a unique normal surface  $S_i$  and  $S = S_1 + S_2$ . It remains only to check that  $S_i$  is boundary-restricted. But the conditions on the boundary (Figure 4) are inherited from  $v$  by each  $v_i$ , so this follows.  $\square$

Now recall that the *weight* of a normal surface is the number of points of intersection with the 1-skeleton.

**Theorem 5.2.** *Let  $(M, K)$  have a marked triangulation  $\mathcal{T}$  with  $K$  a knot in  $\mathcal{T}^{(1)}$ . Suppose  $\mathcal{T}_{tr}$  is the truncation of  $\mathcal{T}$  and  $D$  is a boundary-restricted normal disc in  $\mathcal{T}_{tr}$ . If  $D$  is of least weight over all such discs, then it is fundamental.*

**Proof.** Standard techniques as in [5] are applicable here. So we are brief on some points and refer the reader to [5] for more details.

Suppose  $D$  is not fundamental. Then  $D = D_1 + D_2$ . By Schubert's lemma ([5], 1.9, p. 199), we can suppose the  $D_i$ 's are connected. The condition on  $\partial D$  means that one summand, say  $D_2$ , is a surface with only meridional boundary components (if any), while the other spans  $K$ . Since Euler characteristic is additive under Haken sum, we must have two cases: 1)  $D_1$  a disc and  $D_2$  is a torus, Klein bottle, annulus, or Möbius band. 2)  $D_1$  is a punctured torus or Klein bottle and  $D_2$  is a sphere.

In the first case,  $D_1$  is a normal disc spanning  $K$  of smaller weight than  $D$ . In the second case, Schubert's lemma also says we can suppose that no curve of intersection is separating on both  $D_i$ 's. Thus since every curve separates on a sphere, no curve of intersection separates  $D_1$ . Pick such an innermost curve on  $D_2$  that bounds a disc  $E$  of least weight over all innermost curves. This curve must be 2-sided on  $D_1$ . By compressing along  $E$ , we obtain a disc  $D'$  with at most weight of  $D$ . If the disc is not normal, then normalization will reduce the weight, contradicting the definition of  $D$ .

Suppose  $D'$  is normal. Since  $D$  is of minimal weight,  $D'$  must have the same weight. This implies that the weight of  $D_2 - E$  equals the weight of  $E$ . If there is no other curve of intersection, compressing along the disc  $D_2 - \text{Int}(E)$  will result in a nonnormal disc and we obtain a contradiction as before. If there is another curve of intersection, examination shows that there is another innermost curve of intersection on  $D_2$  which has less weight than  $E$ , a contradiction.  $\square$

## 6. A bound on the number of elementary moves to unknot

J. Hass and J. Lagarias obtained an upper bound of  $2^{10^7 t}$  on the minimum number of elementary moves to take  $K$  to a triangle in one tetrahedron [3]. Their key insight was to use normal surface theory to use a normal spanning disc  $D$  for the unknot  $K$ , which was of exponential size (in  $t$ ). But this disc  $D$  was normal with respect to a doubly barycentric subdivision of  $M$  with a simplicial neighborhood of  $K$  removed. Before we can move  $K$  across the disc  $D$ ,  $K$  must first be moved to  $\partial D$ . Recall that the bulk of their efforts



was on working out how to do this while bounding the number of elementary moves.

The Hass–Lagarias bound is obtained by isotoping  $K$  by elementary moves across an annulus connecting  $K$  to a curve on the boundary of the removed neighborhood of  $K$ , isotoping across the torus boundary of the regular neighborhood to the boundary of a normal disc, and finally isotoping across the normal disc to a single triangle. Since the number of triangles in the normal disc is at most  $2^{8t+6}$ , the chief culprit for their large final bound of  $2^{10^7t}$  is because of their large bound for the first two isotopies.

We will now show how to improve the upper bound on the number of moves to  $2^{120t+14}$ , and subsequently improve the upper bound on the number of Reidemeister moves to unknot an unknot diagram with crossing number  $n$  from  $2^{10^{11}n}$  to  $2^{10^5n}$ . The idea is to use boundary-twisted normal form to implement normal surface theory in the most direct way possible: Theorem 5.2 implies there is a boundary-twisted normal disc  $D$  of bounded size. We straighten out the discs to be piecewise-linear, and move the unknot along  $D$  using elementary moves until it becomes a triangle in a tetrahedron.

First, we need to understand how to bound the size of the fundamental normal disc given by Theorem 5.2. In [4] an upper bound was given for the maximal coordinate of any fundamental solution. This bound depends only on two particular features of the system of integer linear equations: the number of variables,  $n$ , and the maximum,  $m$ , over the sum of the absolute values of the coefficients of an equation. The upper bound is  $n \cdot m^{\frac{n-1}{2}}$ .

In the last column of Table 2, for each type of truncated tetrahedron, we give the maximum number of normal disc types that share a particular normal arc type on a face. Since 6 is the max over all tetrahedra types, we see that any matching equation will have at most 12 as the sum of the absolute values of its coefficients.

Thus plugging these numbers into the bound from [4], we obtain  $59t \cdot 12^{\frac{59t-1}{2}}$ . We get a slightly nicer form by relaxing the bound to  $59t \cdot 2^{118t-2}$ .

So the total number of discs of a fundamental surface is at most  $59t \cdot 59t \cdot 2^{118t-2} \leq 2^{120t+10}$ .

This bound with the next basic result will give a bound on the number of elementary moves to deform  $K$  to a triangle. Recall that an *elementary move* of a polygonal link in a piecewise-linear 3-manifold with specified triangulation consists of two kinds of moves (and their inverses) in a tetrahedron (Figure 13):

- 1) A segment of the link is divided into two by inserting a vertex.
- 2) Two segments which are edges of a triangle otherwise disjoint from the link are moved to the third edge of the triangle.

**Lemma 6.1** ([3]). *Let  $M$  be a triangulated 3-manifold with  $S$  a normal disc in  $M$  with  $w$  triangles. Then  $\partial S$  can be isotoped to a triangle by a series of*

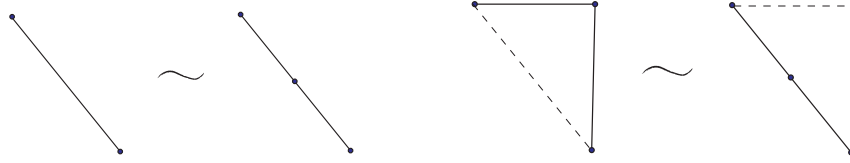


FIGURE 13. Elementary moves taking place in a tetrahedron

at most  $2w$  elementary moves in  $M$ , each of which takes place in a triangle or edge in  $S$ .

**Theorem 6.2** (Bounding elementary moves to unknot). *Let  $M$  be a triangulated 3-manifold with  $t$  tetrahedra and  $K$  is a knot in the 1-skeleton with at most one edge in each face. Suppose  $K$  is an unknot, i.e., bounds a disc in  $M$ . Then there is a series of at most  $2^{120t+14}$  elementary moves taking  $K$  to a triangle lying in a tetrahedron.*

**Proof.** By Theorem 4.1,  $K$  bounds a boundary-twisted normal disc  $D$ , and thus there is a boundary-restricted normal disc  $D'$  in the truncated triangulation. We will suppose  $D'$  is of least weight so that Theorem 5.2 applies to  $D'$  and obtain a bound of  $2^{120t+10}$  on the number of normal discs in  $D'$ . After extending each normal disc to a twisted normal disc, the number of twisted normal discs of  $D$  is also bounded by the same number. Before we can isotope  $K$  across  $D$  by elementary moves, we need to straighten each twisted normal disc to a piecewise-linear disc.

Almost every twisted normal disc can be straightened to a piecewise-linear disc by straightening any vertex-vertex arcs to become an edge of the tetrahedron. The only exception is a “bigon”, which has boundary composed of exactly two vertex-vertex normal arcs.

After straightening twisted normal discs and collapsing bigons to edges, a priori, the result may not be embedded. This can only happen when two of the tetrahedra around an unmarked edge contain twisted normal discs which have vertex-vertex arcs with endpoints on that edge. Since the disc is not embedded, we can pick two such arcs so that they do not bound a chain of bigons. The arcs bound a compression disc, and after the compression we have a disc with fewer points of intersection with the 1-skeleton. Normalizing the disc to boundary-twisted normal form and then truncation will give a boundary-restricted normal disc of lesser weight than  $D'$ , which is a contradiction.

By examining twisted normal discs, we see that there are at most 6 sides. So after straightening, every disc can be divided up into at most 6 piecewise-linear triangles. Using Lemma 6.1 with our bound, we obtain  $2(6) \cdot 2^{120t+10} = 2^{120t+14}$  as an upper bound on elementary moves.  $\square$

**6.1. Bounding the number of Reidemeister moves.** We recall some results from [3]:

**Lemma 6.3** (Triangulating a knot diagram). *Given a knot diagram  $D$  of crossing number  $n$ , there is a triangulated convex polyhedron  $P$  in  $\mathbb{R}^3$  with at most  $140(n+1)$  tetrahedra so that it contains a knot in the 1-skeleton which orthogonally projects to  $D$  on a plane. Furthermore, each face of a tetrahedron contains at most one edge of the knot.*

**Remark 6.4.** Hass and Lagarias actually get a larger number because they need to assume the knot is in the interior of  $P$ .

**Lemma 6.5** (Reidemeister bound for a projected elementary move). *Let  $L$  and  $L'$  be polygonal links in  $\mathbb{R}^3$ . Suppose that  $L$  (resp.  $L'$ ) has at most  $n$  edges and has a link diagram  $D$  (resp.  $D'$ ) under orthogonal projection to the plane  $z = 0$ .*

*If  $k$  elementary moves take  $L$  to  $L'$ , then at most  $2k(n + \frac{1}{2}k + 1)^2$  Reidemeister moves take  $D$  to  $D'$ .*

Now we can prove:

**Theorem 6.6.** *Let  $D$  be an unknot diagram with  $n$  crossings. Then there is a sequence of at most  $2^{10^5 n}$  Reidemeister moves taking  $D$  to the standard unknot.*

**Proof.** We use Lemma 6.3 to obtain a triangulated polyhedron  $P$  of at most  $140(n+1)$  tetrahedra such that a knot in its 1-skeleton orthogonally projects to  $D$ . Let  $t$  be the number of tetrahedra.

By Theorem 6.2 there is a sequence of at most  $2^{120t+14}$  elementary moves taking  $K$  to a triangle in a tetrahedra. Using the bound on the projection of an elementary move (Lemma 6.5) and noting  $K$  contains at most  $2t$  edges, we obtain the following bound on Reidemeister moves:

$$2^{120t+14} \left( 2 \cdot t + \frac{2^{120t+14}}{2} + 1 \right)^2$$

This quantity is less than  $2^{360t+43} \leq 2^{360 \cdot 140(n+1)+43}$ . Except for  $n = 1$ , for which the Reidemeister bound obviously holds, the last is less than  $2^{10^5 n}$ , which was the desired bound.  $\square$

## References

- [1] BRITTENHAM, MARK. Essential laminations and Haken normal form II: regular cell decompositions. Preprint, 1991. [www.math.unl.edu/~mbrittenham2/papers/pdf/hnf2r2.pdf](http://www.math.unl.edu/~mbrittenham2/papers/pdf/hnf2r2.pdf)
- [2] GALATOLO, STEPHEN. On a problem in effective knot theory. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **9** (1998), no. 4, 299–306. [MR1722788](#) (2000j:57010), [Zbl 1001.57007](#)
- [3] HASS, JOEL; LAGARIAS, JEFFREY C. The number of Reidemeister moves needed for unknotting. *J. Amer. Math. Soc.* **14** (2001), no. 2, 399–428. [MR1815217](#) (2001m:57012), [Zbl 0964.57005](#), [arXiv:math/9807012v1](#).
- [4] HASS, JOEL; LAGARIAS, JEFFREY; PIPPENGER, NICHOLAS. The computational complexity of knot and link problems. *J. ACM* **46** (1999), no. 2, 185–211. [MR1693203](#) (2000g:68056), [Zbl 1065.68667](#), [arXiv:math/9807016v1](#).

- [5] JACO, WILLIAM; OERTEL, ULRICH. An algorithm to decide if a 3-manifold is a Haken manifold. *Topology* **23** (1984), no. 2, 195–209. [MR0744850](#) (85j:57014), [Zbl 0545.57003](#).
- [6] LACKENBY, MARC. Classification of alternating knots with tunnel number one. *Comm. Anal. Geom.* **13** (2005), no. 1, 151–186. [MR2154670](#) (2006e:57007), [Zbl 1084.57007](#), [arXiv:math/0103217v2](#).

UNIVERSITY OF CALIFORNIA, ONE SHIELDS AVENUE, DAVIS, CA 95616  
[suh@math.ucdavis.edu](mailto:suh@math.ucdavis.edu)

This paper is available via <http://nyjm.albany.edu/j/2012/18-22.html>.